Mathematical Modeling of Real World Problems

Interdisciplinary Studies in Applied Mathematics

MATHEMATICS RESEARCH DEVELOPMENTS

MATHEMATICAL MODELING OF REAL WORLD PROBLEMS

INTERDISCIPLINARY STUDIES IN APPLIED MATHEMATICS

MATHEMATICS RESEARCH DEVELOPMENTS

Additional books and e-books in this series can be found on Nova's website under the Series tab.

MATHEMATICS RESEARCH DEVELOPMENTS

MATHEMATICAL MODELING OF REAL WORLD PROBLEMS

INTERDISCIPLINARY STUDIES IN APPLIED MATHEMATICS

ZAFER ASLAN
FUNDA DÖKMEN
ENRICO FEOLI
AND
ABUL H. SIDDIQI
EDITORS

nova
science publishers
New York

NOTICE TO THE READER

Library of Congress Cataloging-in-Publication Data

Names: Aslan, Zafer, editor. | Dökmen, Funda, editor. | Feoli, E., editor. | Siddiqi, A. H., editor.
Title: Mathematical modeling of real world problems : interdisciplinary studies in applied mathematics / Zafer Aslan, Funda Dökmen, Enrico Feoli and Abul H. Siddiqi, editors.
Description: New York : Nova Science Publishers, [2019] | Series: Mathematics research developments | Includes bibliographical references and index. |
Identifiers: LCCN 2019041486 (print) | LCCN 2019041487 (ebook) | ISBN 9781536162677 (hardcover) | ISBN 9781536162691 (adobe pdf)
Subjects: LCSH: Mathematical models. | Data mining. | Engineering--Mathematical models. | Environmental sciences--Mathematical models. | Medicine--Mathematical models.
Classification: LCC TA342 .M354 2019 (print) | LCC TA342 (ebook) | DDC 511/.8--dc23
LC record available at https://lccn.loc.gov/2019041486
LC ebook record available at https://lccn.loc.gov/2019041487

Published by Nova Science Publishers, Inc. † New York

CONTENTS

PREFACE

This book covers applications of data processes in multidisciplinary topics. We provide a decomposition of spatio-temporal series into signal and noise components. From the overall spectrum, independent spectra can be extracted allowing the separate detection and monitoring of the world problems. The authors discuss major issues of importance for integrated mathematical implementation, including development experiences worldwide engineering applications.

Data mining provide avenues for proper understanding of real world problems. These scientific papers are intended to provide opportunity for interaction of scientists, engineers, planners, social scientists, environmental scientists, agricultural scientists, medical scientists who are applying wavelets in their research. Mathematical studies are often aimed to analyze and visualize real world problems. Discussion on this theme during IWW Conferences mooted the idea of writing the book.

The book comprises three parts and 16 chapters. The first part is related with engineering and environmental problems like drought survey, precipitation and erosivity, cloud clarification applications, estimation of convection scheme and non-linear time series analyses of air pollution. The second part is related with water management. This part covers two chapters on water quality and river pollution. The third part is on medical sciences like ECG analyses, neurosurgery, computational neuroscience, brain disasters, Parkinson diseases, support vector machine, logic and mathematics.

Keeping in view importance of mathematical tools, it was provided that all scientists interacted with scientists from different countries by gathering specialist researchers in this field and it created opportunity. Joint study subjects were developed and young scientist reunited with internationally famous scientists by providing opportunity to get in contact with scientists from foreign countries. We take this opportunity to thank to the International Center for Theoretical Physics (ICTP, Trieste, Italy) for their encouragement. We also thank to Prof. Dr. Osman UÇAN (Kemerburgaz University),

Prof. Dr. Pammy MANCHANDA (Guru Nanak Dev University) and Dr. Gökhan ERDEMİR (Istanbul Zaim University) for their comments on it.

Editorial Board

Z. Aslan, F. Dökmen, E. Feoli, A. H. Siddiqi
January, 2019

PART I. ENGINEERING AND ENVIRONMENT

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 1

DROUGHT SURVEY OF BASIN OF BÜYÜK MENDERES BY SPI

A. Tokgözlü[1], E. Özkan[1], Funda Dökmen[2,*] and Zafer Aslan[3,4]
[1]Süleyman Demirel University, Science and Literature Faculty,
Isparta, Turkey
[2]Kocaeli University, Food and Agricultural Vocational School,
Campus of Arslanbey, Kocaeli, Turkey
[3]Istanbul Aydın University, Faculty of Engineering, Istanbul, Turkey
[4]Abdus Salam International Centre for Theoretical Physics,
Trieste, Italy

ABSTRACT

Drought is a natural phenomenon composed of several components that is closely related to human life, with a significant impact on the quality of life and means of livelihood of the humankind. It is caused by the decrease in the groundwater level and the water deficit in the soil, which in turn are caused mainly by climatic changes. In this study which was carried out in the Basin of Buyuk Menderes, long term data taken from the stations located in the center of the cities Afyonkarahisar, Usak, Denizli, Aydın and Mugla which are part of the basin in terms of city borders were used as well as the general climatological properties of the basin. The arid periods of the basin have been determined through Standardized Precipitation Index (SPI) and central tendency (linear) analyses.

Keywords: basin, drought, River of Buyuk Menderes, SPI

* Corresponding Author's E-mails: f_dokmen@hotmail.com & funda.dokmen@ kocaeli.edu.tr.

INTRODUCTION

This is how drought is defined in the International Convention for Combating Desertification: a natural phenomenon, which affects the field and resource production systems and causes severe hydrological imbalances as a result of a significant decrease in the usual recorded levels of precipitation (WMO, 1997). Although there are several definitions in literature regarding drought types, three significant drought type definitions are commonly used. In addition, these are:

- Meteorological Drought
- Agricultural Drought
- Hydrological Drought

Meteorological drought is defined as the decreasing of precipitation below the normal levels in a long period. It is usually regional and constitutes the base for an exact understanding regarding regional climatology. It is among the main indications in expressing drought. Agricultural drought, defined as the lack of a sufficient amount of water in the soil as to meet the needs of the plant, is seen when the humidity loss increases and water sources are scarce. During the growing period, if at a certain critical phase when the plant needs water and the soil is not humid enough, agricultural drought occurs.

Agricultural drought is a typical occurrence that comes forth following meteorological drought and preceding hydrological drought. Hydrological drought means the decreases that occur in surface and groundwater as a result of a prolonged term with insufficient precipitation. It can be tracked by river flow measurements, lake reservoirs and groundwater level measurements. Even after the meteorological drought has ended, hydrological drought can continue (Wilhite and Glantz, 1987).

In short, drought is defined as the “decrease in the amount of water as to reach a level that is much lower than the usual level for a certain amount of time in any region.” We could say that the main factors that cause drought are meteorological and climatological events. Changes in the type and intensity of precipitation, the constantly increasing temperatures and over-evaporation in certain regions reveal the drought that is caused by climatological changes and the meteorological events behind this occurrence, in many respects.

According to the conclusions of a survey, which was carried out in 87 countries by the World Meteorological Organization, Turkey is among the 74 countries that receive the greatest impact from the drought (WMO, 1992). IPCC (Intergovernmental Panel on Climate Change), predicts that at the beginning of the year 2030, the amount of CO_2 at the beginning of this century will double and in the area that covers South Europe, including Turkey, the temperatures will rise by 2 ºC in the wintertime and by 2-3 ºC in

the summertime in the future. The rise in the temperatures may cause our dry areas to desertificate as well. As the drought becomes greater, the sharing and management of many international, national and local water sources will become more challenging, including the use of rivers that transcend the city or country borders (Ozgurel et al., 2003). Therefore, governments need to work out several action plans. The making of action plans requires the calculation of drought analyses. Various approaches and methods have been suggested for examining drought events. Among these, Palmer Drought Severity Index (PDSI, Palmer, 1965) and Standardized Precipitation Index (SPI), (Mckee et al., 1993) are accepted as methods that are effective in determining, evaluating and tracking droughts, (Patel et al., 2017, Santos et al., 2017). The conclusions that are arrived at through the use of these drought indexes, help in gaining information about drought tendencies (increases or decreases) that occur in certain periods, as well as providing information regarding the climate of a certain country or a region. Such information can be utilized in improving drought management, skills, and means for combating drought.

Figure 1. Location map of Büyük Menderes Basin.

To summarize the leading studies that are conducted in our country; Sırdas S. has written a doctoral thesis titled “Meteorological Drought Modeling and Its Application to Turkey.” He made use of SPI in order to determine meteorological drought in this study (Sırdas, 2002). Sırdas S. and Sen Z. revealed the dry periods in GAP (South-East Anatolia Project) region in their study of 2003, titled “Investigation of Dry Period Properties in GAP Region,” by making use of precipitation data taken from 9 different

stations within the GAP region in SPI (Sırdaş and Sen, 2003). Caldag B. and colleagues have also made use of SPI in a study they conducted in the year 2004 for determining the drought situation in the region of Thrace.

Yilmaz B. and colleagues have utilized the SPI method for drought analyses in the study they conducted on dry periods in the city of Manisa as well. In 2007 Keskin M.E. and his colleagues, in their study "Meteorological Drought Analysis of Isparta Region" have used the SPI method drought analyses (Keskin et al., 2007). Again, in 2007, drought probabilities for Turkey were calculated by the use of SPI method in a study conducted by Turkes M. and Tatli H. (Turkes and Tatli, 2007). Ilgar's study of 2010, titled "Determining Drought Situation and Tendencies in the City of Canakkale through the Use of Standardized Precipitation Index," is another example to studies that draw attention to drought events taking place in our country, by the use of the SPI method.

Location of the Basin

Basin of Büyük Menderes is among the 25 major basins specified by the Ministry of Forestry and Water Affairs. The borders of the basin go through the city borders of Usak, Denizli, Afyonkarahisar, Aydın and Muğla. In addition, the basin neighbors the basins of Akarçay, Burdur, Antalya, Batı Akdeniz, Küçük Menderes and Gediz (Figure 1).

Properties of the Survey Area

The basin, located in the southern and southeastern parts of the Aegean Region, which is divided into two parts as Aegean, and Central West Anatolian regions, is 23,937 km2 wide. This counts for almost one third of the Aegean Region, which covers an area of 79,000 km2 in total. The reason why this study was conducted on a basin-basis is that basins have climatic properties that are distinct from those of the environment that surrounds them. Behaviors of air masses affecting a certain region can show differences even within the basin itself (Figure 1). The nautical effect of approximately 150 km that comes in through the Aegean horst-graben system manifests itself in the basin. With the addition of the altitude factor, the difference is unmistakably felt. While the yearly average temperature in the center of the city Usak, which is located in the upper northeastern line of the basin is around 11 °C, the yearly average temperature in the city of Aydin, which is located in the lower western line of the basin goes up to 20 °C.

Data

For the drought index calculations in this study, the monthly and annual total precipitation totals and again monthly and annual average temperature data of the years 1960-2015 taken from the General Directorate of Meteorology were used. These data is collected in the stations located in the city centers. The main reasons for the use of data from the city centers are the fact that these data are more reliable and that there is not much lack of data in these stations.

METHOD

The graphs of temperature and precipitation data taken from the relevant stations - within our study area have been prepared by calculating the average values of the monthly average temperature data and monthly total precipitation data (between the years 1960-2015)* in Microsoft Excel 2010.

Drought is a multi-dimensional natural phenomenon with many variables. Therefore, an effort was made for putting together concrete indications regarding drought by various mathematical formulas. It was determined that the SPI (Standardized Precipitation Index) be used for the study. In addition, through central tendency (linear) analyses, information about the extent and severity of dry periods was set forth.

Standardized Precipitation Index

SPI value is derived by subtracting the precipitation rate belonging to a certain duration from the average and dividing it by the Standard deviation rate. “Conceptually, SPI means the oscillating of the Z-score or the standard deviation rate under or above the average. Of the values derived, the ones under the average (negative SPI values) correspond to dry periods while the ones above the average (positive SPI values) correspond to humid periods” (Ilgar, 2010).

Bringing down all precipitation parameters into a single numerical value with the purpose of defining the drought situation in regions with different climates has importance in the tracking of drought. Efe and Ozgur (2014), Mckee et al. (1993) have developed Standardized Precipitation Index in order to express it numerically. In this index, precipitation (Xi) at a given period is subtracted from the long-term average (Xiave) and divided by the standard deviation (σ), (Eq. 1);

$$SPI = \frac{X_i - X_i^{ave}}{\sigma} \quad (1)$$

"In SPI, gamma probability density function is used" (Sirdas, 2002). "The fact that values and probabilities in SPI are computable for 1, 3, 6, 12, 24 and 48-month or longer time, slots allow the tracking and evaluation of droughts in different time slots. For example, 6-month SPI values are calculated based on the 6-month sliding averages of monthly precipitation values", (Turkes and Tatli, 2007).

Table 1. SPI values

SPI	Drought Intensity
≥ 2	Extremely wet
1.50 ~ 1.99	Very wet
1.00 ~ 1.49	Moderately wet
0.99 ~ 0	Normal
0.99 ~ -0.99	Near normal
-1.00 ~ -1.49	Moderately dry
-1.50~-1.99	Severely dry
≤ - 2	Extremely dry

Source: McKee et al., 1993, VMO 2012

In this study, in order to examine drought on a yearly basis so it provides a general opinion, 12-month SPI values were calculated based on 12-month sliding averages. The SPI values derived were first tabulated and then arranged into a graph. In the graph, central tendency line was used in order to express the tendency of the drought situation. In addition, in the next step, graphs were interpreted (Table 1).

RESULTS

General Climatic Characteristics of the Basin of Büyük Menderes

Turkey is a country that is located in the variable zone moderate climatic area. The main air masses that influence our land move from the north to south depending on the season. To put it in a more detailed way; Turkey is located on a transitional area in terms of genetic-dynamic factors and is under the influence of microclimatic, Mediterranean climate conditions with subtropical rains. Seasonal movements of tropical and polar air masses and the area of influence characterize this type of climate; intensity and frequency of frontal activities that take place along the polar front between them are seasonal (Erinc, 1984).

Temperature and Precipitation Characteristics of the Basin

Among the air masses which influence Turkey and can be evaluated within the category of planetary factors we can name the Siberian HP (cPk) which is generally prevalent in the winter causing days of frost in the central and eastern parts of our country. The air masses which are influential throughout the year, induced by subtropical seas (Mediterranean moving cyclones and Azor HP) (mTw) and again, the Subtropical area lands (North Africa, South Balkans and the Middle East in general) that are influential throughout the year.

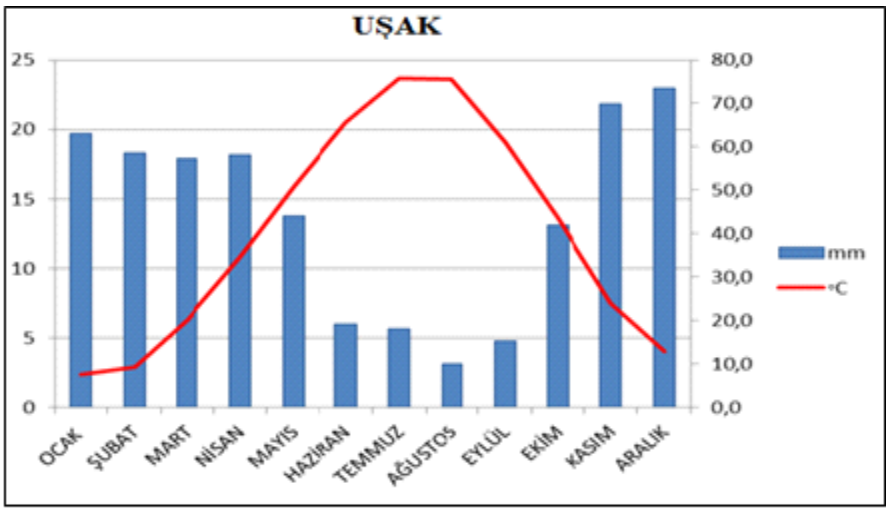


Figure 2. Monthly total precipitation and average air temperature, Uşak, (1960-2015).

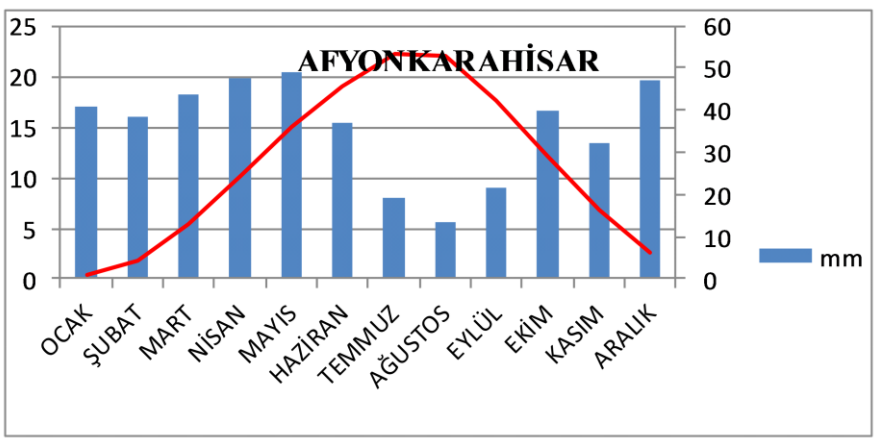


Figure 3. Monthly total precipitation and average air temperature, Afyonkarahisar, (1960-2015).

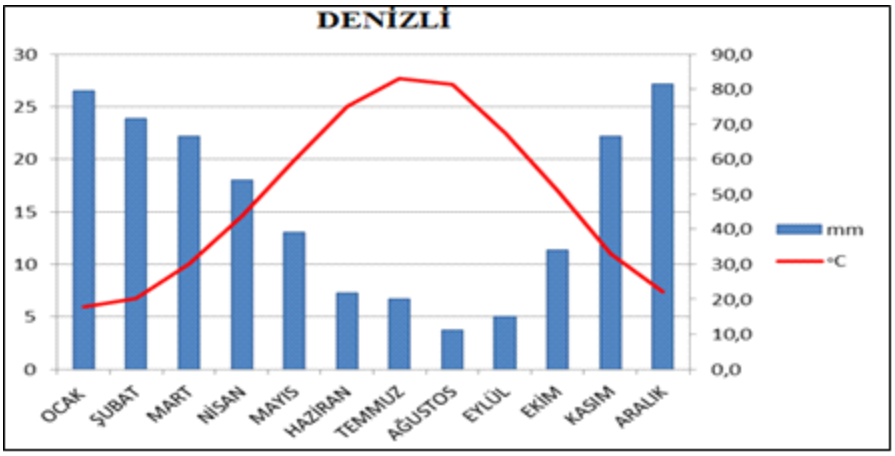


Figure 4. Monthly total precipitation and average air temperature, Denizli, (1960-2015).

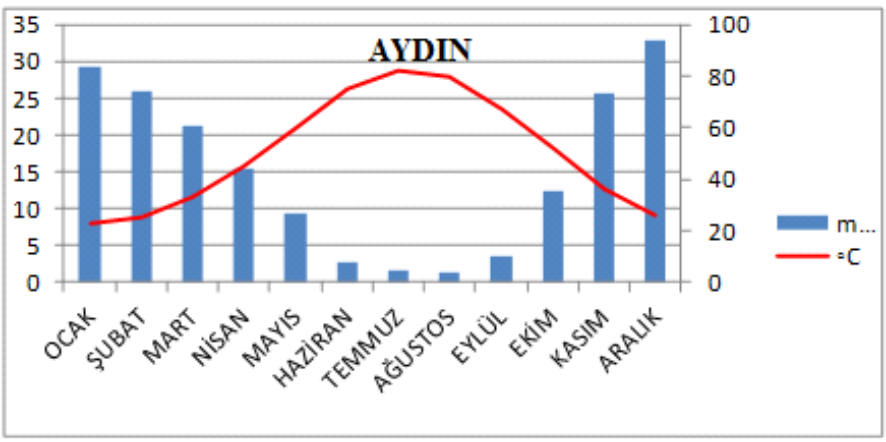


Figure 5. Monthly total precipitation and average air temperature, Aydın, (1960-2015).

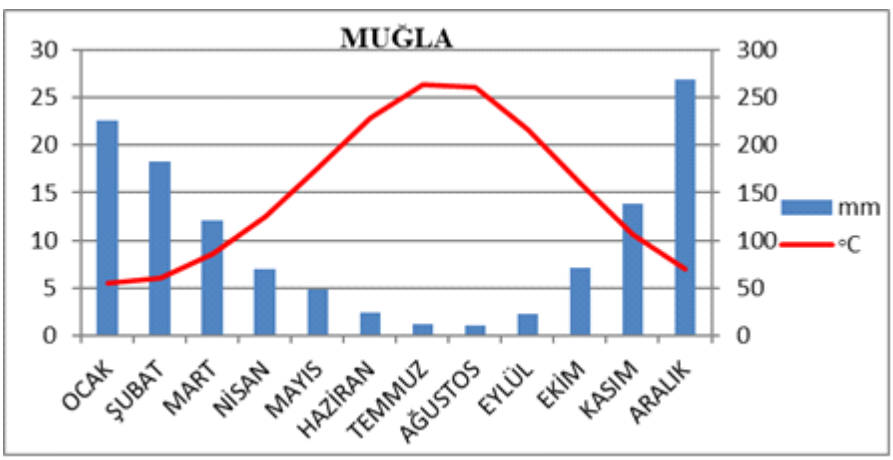


Figure 6. Monthly total precipitation and average air temperature, Muğla, (1960-2015).

Basin of Buyuk Menderes is an important area located in the western part of our country and reaching towards the coast in the form of a channel. Whereas the temperatures can get high during the summer, it is seen that in the wintertime the weather in the central and western parts of the basin is warm but on the contrary, high-altitudes are cold and with days of frost. In the wintertime where the Siberian HP is influential in Turkey, especially in stations such as Dinar, Sandikli (Afyonkarahisar) and Banaz (Usak) the temperatures get lower than 0 ˚C. During this period, this polar air mass, which mostly faces the moving Mediterranean low pressure, causes much rainfall in the basin. However, in some years, when the Siberian HP can enter all the way into Aydin through the Saraykoy (Denizli) channel valley, it can be seen that in the central parts of the basin the temperature can fall below 0˚C and cold and dry days can occur.

In the monthly total precipitation, graph of the cities Usak and Afyonkarahisar, the months March and April call for attention. Precipitation is distinctively greater in comparison to other stations in March and April in these two cities, which stand out as locations where continental influences are the greatest within the basin. In addition, the fact that certain frontal activities occur as the Siberian HP starts leaving the country and the impact of the moving Mediterranean low pressure is felt more strongly can be taken as other factors causing this occurrence (Figures 3 and 4).

As we move towards the central parts of the basin, with the prominence of nauticality and decreasing altitudes due to grabens, we see that average wintertime and summertime temperatures increase. Aydin and Denizli are good and clear examples of this. When the values derived in these two stations that are 120 km apart from each other, are examined together significant differences are seen. Denizli receives less rainfall per year than Aydin (by 60.5 mm). In addition, the average temperature is 3-4 ◦C lower than Aydin as well (Figures 4 and 5).

In the lower parts of the basin, especially the streams that join into the Brook of Cine are in most part within the city borders of Mugla. Yatagan and Kavaklidere counties of the city Mugla are located within the borders of the Basin of Buyuk Menderes anyway. Due to geographical influences, the city of Mugla comes forth as the station that receives the greatest amount of rainfall within the basin (~1250 mm). This value is much greater even than the national averages. Located in the Southwestern part of the country and differing from the rest of the Aegean Region due to the direction of mountain chains in relation to the sea, Mugla demonstrates the impact of relief and altitude as well as the differences in precipitation and temperature values (Figure 6).

Examination of the Relevant Station Data by SPI

When the central tendency of the SPI values of the cities Usak and Afyonkarahisar are examined, it is seen that while until 1980’s the drought level was “near-normal,”

starting with 1990's the values began to rise even if just a little bit and the drought level reached "normal" conditions. However, when it is considered as the transition to the normal period, especially after 2000's, it has been observed that the rise in the humidity levels in Afyonkarahisar has continued with an accelerating pace. In addition to this, it was established that in the years 1961 and 2008 "severe drought" occurred in Usak and in the years 1974 and 2007 "very severe drought" occurred in Afyonkarahisar (Figures 7 and 8).

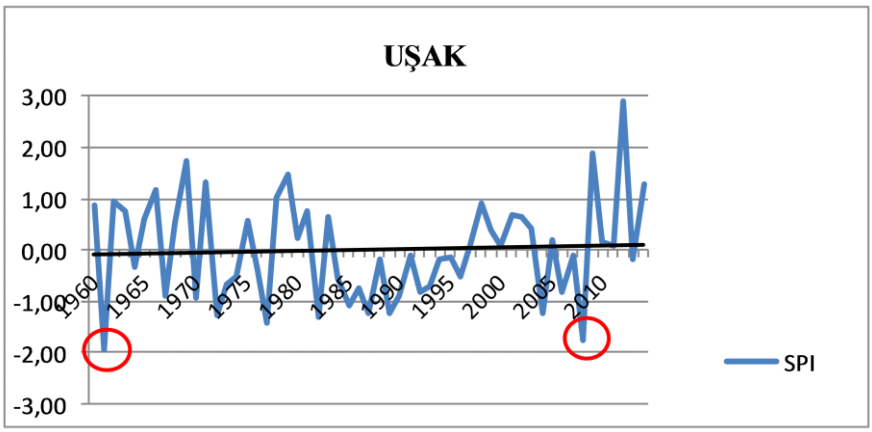


Figure 7. Anunal variation of SPI in Uşak, (1960-2015).

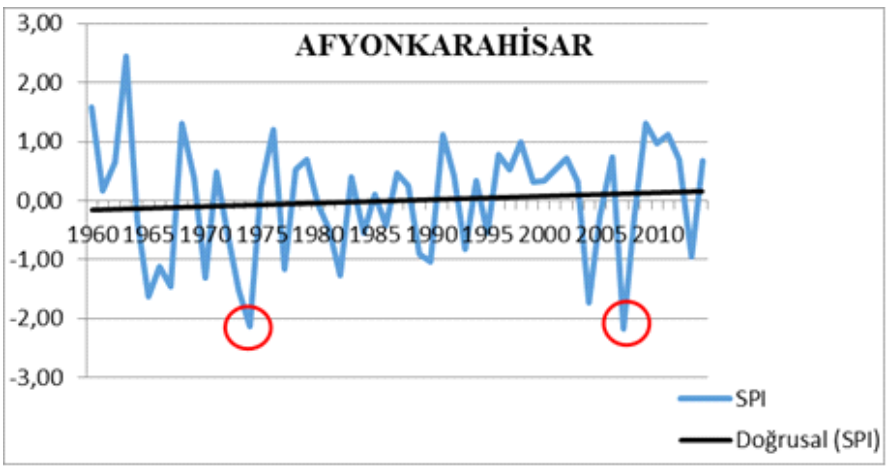


Figure 8. Annual variation of SPI in Afyonkarahisar, (1960-2015).

SPI values of Denizli, have kept their "normal" course during this 55-year period, as illustrated by the linear tendency line. However, in the years 1974, 1977 and 1989 "severe drought," and in the year 2008 "very severe drought" was observed (Figure 9).

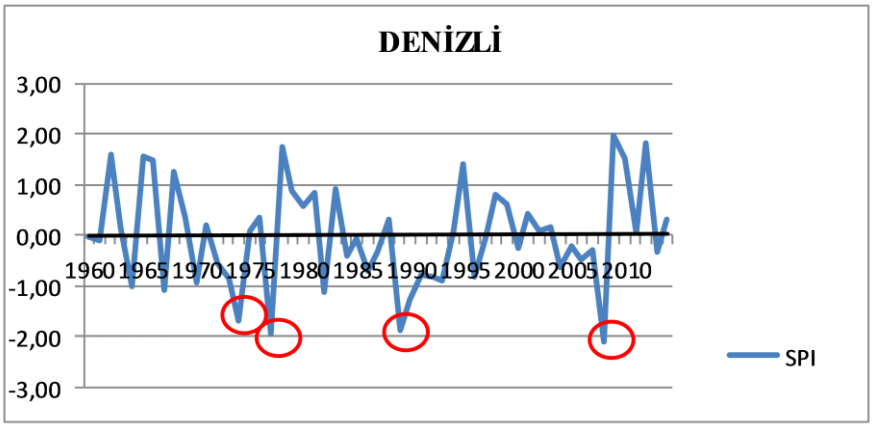


Figure 9. Annual variation of SPI in Denizli, (1960-2015).

Based on the data from the Aydin station it was concluded that the values remained within "near-normal drought" limits until 2000's and that after 2000's a transition towards the "normal" was seen and this was the continuing course. "Severe drought" occurred in the years 1972, 1989, 1992 and 2008 (Figure 10).

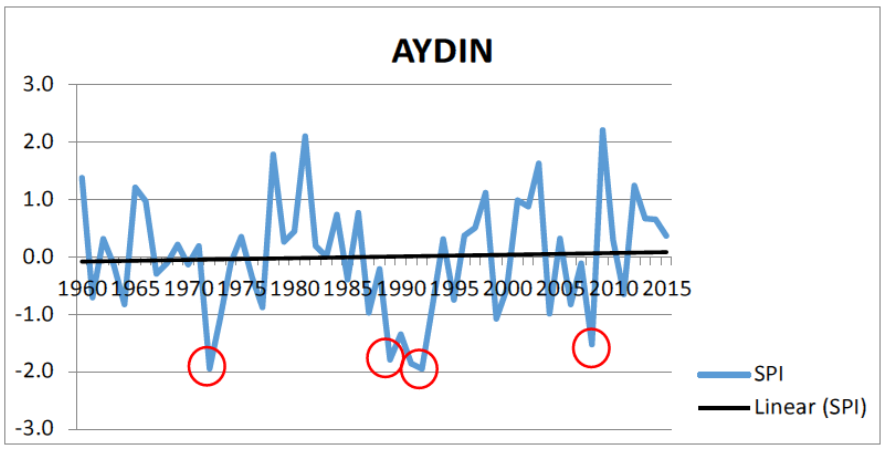


Figure 10. Annual variation of SPI in Aydın, (1960-2015).

Mugla comes to the fore with a central tendency line whose course deviates from the other stations in our study. As the location with the greatest levels of precipitation within the basin, Mugla is the only center where drought took a positive course. In Mugla, "severe drought" occurred in the years 1972, 1977, 1990 and 1992. The year 2008 which was a dry one throughout the Basin of Buyuk Menderes (Usak: severe drought; Afyonkarahisar and Denizli: very severe drought; Aydın: moderately severe drought) was the year a "very severe drought" occurred in Mugla as well (Figure 11).

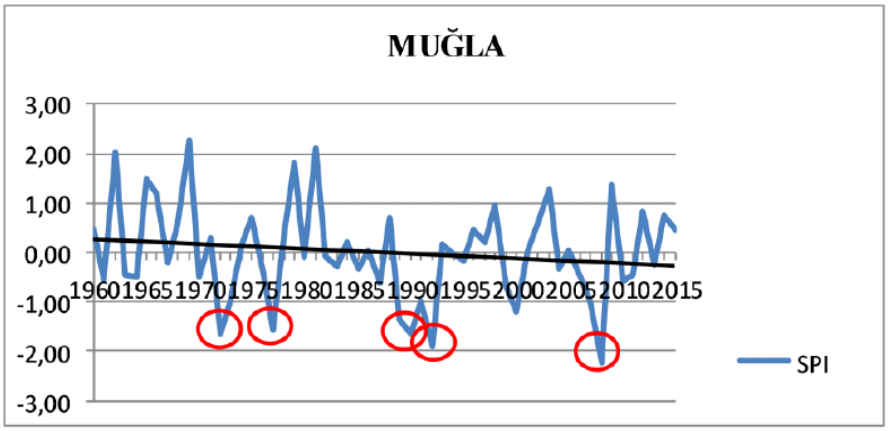


Figure 11. Annual variation of SPI in Mugla, (1960-2015).

CONCLUSION

Specific results of the study are listed below:

- Within the Basin of Buyuk Menderes, Mugla is the only center with a drought tendency.
- Aydin is the city where the drought severity is the lowest within the basin.
- The year 2008 stands out as a year that was dry all across the basin.
- It is clear that in the 2000's the central tendency followed a course that moved from drought to normal conditions.
- It is obvious that there have been climatic oscillations within that 55-year period. In addition, just like the rest of the world, the Basin of Buyuk Menderes saw many dry periods during this time. Another considerable point is that the number of these short drought periods was higher until the 1990's. The years where drought was at peak were mostly the years before the 1990's.
- When the graphs are taken into account, a detail that needs attention is the fact that the year 2009, right after the year 2008 that was a dry one all across the Basin was a year with a lot of precipitation.
- Drought is a natural disaster that needs to be tracked in a multi-dimensional way.
- Defining drought and making action plans to combat it is of great importance both globally and locally. In the city of Afyon, the action plans made by the General Directorate of State Hydraulic Works have constituted a great example. A decrease in the droughts seen in the city of Afyon has been observed.
- In the Basin, there are other drought-prone cities as well. When action plans are being made, an effort should be made so that they allow for basin-wide

evaluations and drought situations of various stations must be determined by scientific methods.

- The fact that the agricultural and population potential of both our country and the Basin of Buyuk Menderes is high is an indicator of how far the impact of drought can extend. Therefore, it is of outmost importance to take the necessary precautions in order to minimize the damage to agricultural produce and to predict the course of climatic events accurately.
- In identifying and tracking drought, the years in which droughts are on a positive course must be paid specific attention to and it must be examined which global system in those years has had a greater impact on our country and the reasons behind it should be studied.

REFERENCES

Da Cunha, L. V. (1983). Drought, environment and society. "Coping with droughts", 3-11. *Ankara Üniversitesi Çevrebilimleri Dergisi* 4(2), 1-32, Ankara.

Efe, B., and Özgür, E. (2014). "Standart Yağış İndeksi (SPI) ve Normalin Yüzdesi Metodu (PNI) İle Konya ve Çevresinin Kuraklık Analizi" ["Analyses of drought in and near vicinity of Konya by standart percipitation index (SPI) and normal percentage method"] *II. Uluslararası Katılımlı Kuraklık ve Çölleşme Sempozyumu, Konya*.

Erinç, S. (1984). *Klimatoloji ve Metotları* [*Climatology and Methods*], İstanbul Üniversitesi Deniz Bilimleri ve Coğrafya Enstitüsü Yayınları, No 2, İstanbul.

Ilgar, R. (2010). "Çanakkale'de Kuraklık Durumu ve Eğilimlerinin Standartlaştırılmış Yağış İndisi İle Belirlenmesi" ["Determination of drought by standart percipitation index in Çanakkale"] *Marmara Coğrafya Dergisi*, 22, 183-204.

McKee, T. B., Doesken, N. J., and Kleist, J. (1993). "The relationship of drought frequency and duration of time scales. Presented at the Eighth Conference on Applied Climatology," *American Meteorological Society*, (Jan 17-23), Anaheim CA, pp. 179-186.

NIDIS Drought and Water Assessment. *NIDIS Intermountain West Drought Early Warning System May 9, 2017*. Retrieved from: http://climate.colostate.edu/~drought/archive_summaries/2017_May09_assessment.pdf, (December, 2017).

Özdemir, Y. (2009). *Büyük Menderes Havzasının Su Yönetimi ve Arazi Kullanımı Açısından İncelenmesi* [*Investigation of basin of Büyük Menderes in terms of water management and land using*]. Doktora tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

Pamuk, G., Özgürel, M., and Topçuoğlu, K. (2004). "Standart Yağış İndisi (SPI) ile Ege Bölgesinde Kuraklık Analizi" ["Analyses of drought by

Standard Precipitation Index (SPI) in Aegean Region, Ege Üniv. Ziraat Fak. Derg"] *Ege University Journal of Agricultural Faculty*, 41 (1):99-106 ISSN 1018-8851, İzmir.

Patel, N. R., P. Chopra, and V. K. Dadhwal. (2007). Analyzing spatial patterns of meteorological drought using standardized precipitation index, *Meteorological Applications,* 14: 329–336, DOI**:** 10.1002/ met.33

Santos, C. A. G., N. R. M. Brasil, J. S. A., and Passos, R. M. da Silva. (2017). Drought assessment using a TRMM-derived standardized precipitation index for the upper São Francisco River basin, Brazil, *Environmental Monitoring and Assessment* Jun;189(6):250. Doi: 10.1007/s10661-017-5948-9. Epub *(2017* May 3).

Sırdaş, S. (2002). *Meteorolojik Kuraklık Modellemesi ve Türkiye Uygulaması* [*Meteorological drought modelling and applications in Turkey*], Doktora tezi, İTÜ Fen Bilimleri Enstitüsü, İstanbul.

Standardized Precipitation Index User Guide, (2012). WMO-No. 1090, pp, 16, ISBN 978-92-63-11091-6, Switzerland.

Türkeş, M. (2007). "Türkiye'nin kuraklığa, çölleşmeye eğilimi ve iklim değişikliği açısından değerlendirilmesi" "Evaluated of Turkey in terms of drought, desertification and climate chancing," *Pankobirlik* 91, 38-47.

Türkeş, M., and Tatlı, H. (2008). "Aşırı kurak ve nemli koşulların belirlenmesi için yeni bir Standartlaştırılmış Yağış İndisi (yeni-SPI): Türkiye'ye Uygulanması" "New standard precipitation index (New-SPI) for determination of humidity conditions and excessive droughts: Application of Turkey," *IV. Uluslararası Atmosfer Bilimleri Sempozyumu Bildiriler Kitabı*, (25-28 Mart), İstanbul.

Türkeş, M. (2012). "Türkiye'de Gözlenen ve Öngörülen İklim Değişikliği, Kuraklık ve Çölleşme" "Observation and prediction of climate changing, drought and desertification in Turkey," *Ankara Üniversitesi Çevrebilimleri Dergisi*, 4(2), 1-32, Ankara.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 2

AN OBSERVATIONAL AND NUMERICAL STUDY OF PRECIPITATION CONCENTRATION AND EROSIVITY INDICES

Z. N. Çağlar[1], B. Oğuzhan[1], D. O. Demirci[1], S. Söğüt[1], N. Yeniçeri[1], A. Tokgözlü[2] and Zafer Aslan[3,4,*]

[1]Kandilli Obs. and Earthquake Res. Inst., Istanbul, Turkey
Boğaziçi University, Istanbul, Turkey
[2]Faculty of Science and Literature, Süleyman Demirel University, Isparta, Turkey
[3,4]Faculty of Engineering, Istanbul Aydın University, Florya, Istanbul, Turkey
[4]Abdus Salam International Centre for Theoretical Physics Trieste, Italy

ABSTRACT

Erosivity risk due to precipitation was studied using observations and numerical model estimates. Four case studies were performed using observed precipitation data collected between 1975 and 2006 at Isparta, Kandilli, Sakarya and Eskişehir meteorological stations. The future precipitation estimates for years 2025, 2050, and 2075 were also analyzed using the MM5 mesoscale model with a three nested grids approach after downscaling from the commit scenario of CCSM3 model. Precipitation, its trend, and associated erosivity risk from current conditions to future years were studied using Precipitation Concentration Index (PCI), Modified Fournier Index (MFI), and Bagnouls–Gaussian Aridity Index (BGI). Erosivity Index (EI) was also studied using CORINE method.

* Corresponding Author's E-mail: zaferaslan@aydin.du.tr.

Keywords: rainfall intensity, precipitation concentration index, modified fournier index, climate change, corine

INTRODUCTION

The observed surface temperature analyzed over the last hundred years shows an average increase of about 0.90C globally, and the future estimates obtained from various climate change models suggest an additional increase between 1.1 to 6.40C (IPCC, 2007, 2017). Since the warming will not be evenly distributed over the earth' surface, it is not known exactly how much the temperature would increase in an area and how it will affect other atmospheric conditions. For example, the temperature increase could result in a decrease in vegetative cover and its type, as well as an increase or decrease in the amount of precipitation in a given area.

Hence, it would be beneficial to estimate future precipitation distribution and soil erosion risks in an area for especially agricultural purposes.

The risk of erosion in Mediterranean Iberian Peninsula was studied by Luis et al., (2010) using the total annual precipitation, precipitation concentration index (PCI), and Modified Fournier Index (MFI) trends. Their results showed a spatial variability and complex temporal trends. Most recently, Luis et al., (2011) further studied annual and seasonal changes as well as wet and dry periods of PCI in the conterminous Spain for periods 1946-1975 and 1976-2005. Zhang et al., (2009), also analyzed changing properties of PCI in China. They studied the precipitation concentrations across the Pearl River basin and the associated spatial patterns using daily precipitation data collected from 42 rain-gauging stations during the period 1960–2005. A significant increase in PCI after 1990 was detected in the West River, lower North River, and upper Beipan River. These changes of PCI in the Pearl River basin are likely to be associated with the consequences of the well-evidenced global warming, Zhang et al., (2009).

The risk of soil erosion in Turkey was discussed by using CORINE method by applying various indices gradually including MFI and Bagnouls– Gaussian Aridity Index (BGI) (CEC, 1992; Dengiz and Akgul, 2005). Recently, Aslan et al., (2010 a, b) also calculated the Erosivity Index (EI) using ground and remote sensing data collected in Turkey. They generated a detailed an actual water erosivitiy and wind erosivity risk map of Turkey.

Precipitation trends were assessed over the period of 1954–2003 using parametric ordinary least square fits and non-parametric Mann–Kendall technique (Pal and Al-Tabbaa, 2010). They applied this model on the study area in India considered whole Keralastate and Indian Peninsula. A decreasing trend was recorded in the spring and during monsoon rainfall season and an increasing trend in the autumn and winter were observed. Temporal variability of precipitation over 96 years was also studied by the

application of the wavelet transform to five precipitation series in Northern California (Kim, 2004). The increasing precipitation trend seen in Northern California can be interpreted as the coupled effect of the extremely long-period component and multi-decadal period component.

Gabriels (2003) showed that the impact of raindrops on erosivity might be evaluated by considering the energy of a rainstorm. Soil losses from erosion were predicted by using Universal Soil Loss Equation (USLE). The coupled moisture flux was proved to capture continuous property of climate system and providing extra information to determine rainfall probability and rainfall amount (Yang at al, 2010). The application was made to simultaneously downscale daily precipitation at multiple sites within the Rhine River basin. The results show that the model can reproduce statistical properties of daily precipitation time series.

In conjunction with climate change, extreme low or high precipitation may cause floods or drought (Yeşilırmak, 2015). Frequency analysis, precipitation homogeneity indices are used for examining extreme rainfalls. In this study, temporal and spatial changes in the precipitation concentration index (CI), which is a rainfall homogeneity index, were investigated using long-cycle daily precipitation measurements recorded at seven stations in Büyük Menderes Basin. The results show that there is no statistically significant temporal change in CI values across the watershed.

The supply of cultivable land is limited, and the threat posed by climate change (especially reduced seasonal rainfall) is acute in the Mediterranean region. The importance of the water-for-food nexus is underlined by the fact that agriculture is estimated to account for over 80% of total water usage in many SEMCs (compared with a global average of 70%). Intensive tillage has also led to the depletion of organic matter and—together with an over-use of chemical fertilizers—resulted in a reduction in fertility.

Our analysis indicates that the Mediterranean region faces multiple challenges in terms of food sustainability. Threats common to most of the countries (albeit to varying degrees) include climate change; water scarcities; mounting environmental pressures (particularly along the coast); soil degradation; unsustainable farming practices; and poor nutrition. This latter is marked by rising rates of obesity, and the “multiple burden” of obesity, under-nutrition and micronutrient deficiencies in many of the SEMCs.

The agricultural lands are decreasing worldwide, especially in the Mediterranean region (Barilla, 2017). Climate change, especially the decrease in seasonal rainfall, is a significant threat to the Mediterranean region. This research study shows that the Mediterranean region faces many challenges in terms of food sustainability. In most of the countries, there are common threats in the following areas, at varying degrees: climate change; water scarcity; environmental constraints on coastal areas; soil degradation; unsustainable agricultural practices.

Another paper deals with soil erosion and loss of physical properties of the soil (Bardhi and Ramollari, 2017). Natural characteristics such as soil characteristics, precipitation amount and precipitation density, slope and length of the slopes are taken into consideration. Here, maps are created according to the CORINE program; areas with low and high erosion risk are clearly shown. Soil erosion analysis based on precipitation is very important in terms of erosion risk detection, (García-Marín et al., 2017). Seasonal change and rainfall intensity should be considered in evaluating the potential of erosion. Based on the modified Fournier Index (MFI) values, the risk of soil erosion can be determined. SI, PCI, MFI and rainfall values were analyzed using nonparametric Mann-Kendall test for different seasons (autumn, winter, spring, summer, autumn-winter and spring-summer). The results show that precipitation has a distinct seasonality characteristic in a total 225 rain-gauge stations spread all over region for the Western region of Andalusia. In the stations, there was a tendency to decrease by 72% in spring-summer rainfall.

This study analyzes the changes in observed precipitation and examines its effect on soil erosion in Turkey. The changes in PCI, MFI, BGI, and EI were analyzed to study the effect of climate change on precipitation using the observed data collected between 1975 and 2006 at four surface stations. In addition, a numerical modeling study was performed to estimate the precipitation levels within the next 75 years to estimate the impact of climate change on precipitation using the same indices.

STUDY AREA, OBSERVATIONS AND METHODOLOGY

Precipitation data were collected at Kandilli, Isparta, Sakarya and Eskişehir stations between January 1975 and October 2006. Their geographical coordinates and heights above mean sea levels are given in the Table 1. The rainfall rate was measured by an electronic pluviograph at 10 minutes intervals and Hellmann Precipitation Recorder measured continuous precipitation.

PCI (Oliver, 1980), MFI (Arnoldus, 1980) and BGI (Michiels and Gabriels, 1996) were calculated on an annual basis for each study region using:

$$PCI = 100 * \sum p_i^2 / P^2 \quad (1)$$

$$MFI = \sum p_i^2 / P \quad (2)$$

$$BGI = \Sigma (2t_i - p_i) k_i \quad (3)$$

Here, p_i is monthly total precipitation (mm), P annual total precipitation (mm), t_i is monthly mean temperature (^oC), and k_i is the proportion of the month during which $2t_i-p_i > 0$. Erosivity Index (EI) is then defined using MFI and BGI values:

EI = [MFI Class No] *[BGI Class No]

PCI, MFI, BGI, and EI classes and their descriptions are given in Tables 2 through five respectively.

Table 1. The geographical information of the stations used in case studies

Station Name	**Longitude (deg)**	**Latitude (deg)**	**Height above sea level (m)**
Kandilli (Istanbul) (Marmara Region)	41.06°N	29.06°E	114.5 m
Sakarya (Marmara Region)	40.77oN	30.42oE	31 m
Isparta (Mediterranean Region)	37.77°N	30.55°E	1050 m
Eskişehir (Central Anatolia)	39.82°N	30.52°E	787 m

Table 2. PCI classes

PCI	**Temporal Rainfall Distribution (PCI)**	**Class**
< 10	Uniform	1
11 – 15	Moderate Concentrated (Moderate Seasonal)	2
16 – 20	Concentrated (Seasonal)	3
21 – 50	Strong Concentrated (Strong Seasonal)	4
> 50	Isolated (Irregular)	5

Table 3. MFI classes

MFI (range)	**Description**	**Class**
<60	Very low	1
60-90	Low	2
90-120	Moderate	3
120-160	High	4
>160	Very high	5

The CORINE Soil Erosion Assessment Methodology (CEC, 1992) is described in Figure 1. In this scheme, soil texture, soil depth, stoniness, MFI, BGI, soil erodibility (K), EI, slope angles, land cover (V), Potential Soil Erosion Risk (E_p), and Actual Soil Erosion

Risk (E_A) have been taken into account to compare actual water erosivitiy and potential erosivity risk maps in study areas.

Table 4. BGI classes

BGI (range)	Description	Class
<50	Very low	1
50-75	Low	2
75-100	Moderate	3
100-125	High	4
>125	Very high	5

Table 5. EI classes

EI	Class	Class No
<4	Low	1
4-8	Moderate	2
>8	High	3

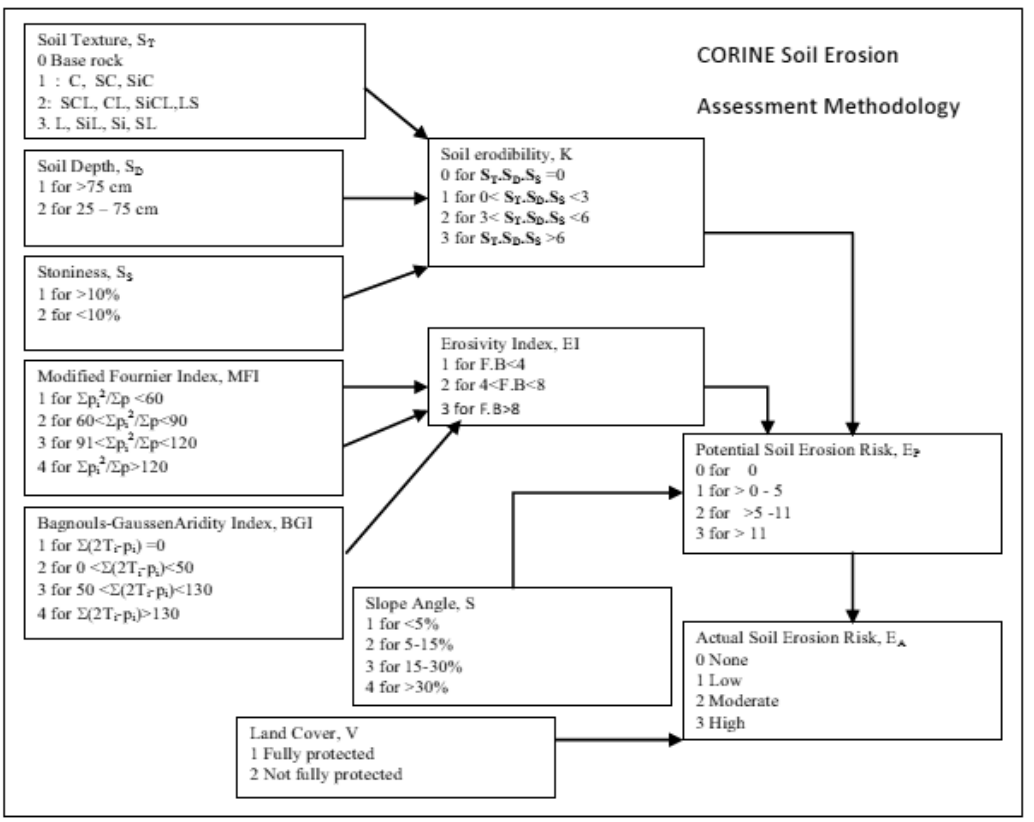


Figure 1. Algorithm of CORINE soil erosion assessment methodology (Erhard et al., 2003).

Numerical Experiments

A numerical model was setup using three nested grids using (45x55), (70 x 91), and (79 x 142) grid points in (north-south x west-east) directions using 108, 36, and 12 km horizontal grid spacing, respectively (Figure 2). The outmost grid was centered at 40^0N and 33^0E and inner grids were placed within their respective coarser grids. 32 sigma levels were used in the vertical with the first layer depth of 20m near the surface with 15 vertical layers located within the first 1-km above the surface and model domain extending to 15-km at the top.

MM5 was configured by using simple ice microphysics is for the treatment of excess moisture cumulus parameterization of convective motions boundary layer parameterization to resolve the atmospheric flow. Rapid Radiative Transfer Model (RRTM) is for calculation of radiation. 5-layer soil model is for the calculation of surface fluxes (Dudhia, 1989), (Grell et al., 1995), (Blackadar, 1976 and 1979) and (Mlawer et al., 1997). The model was initialized using the commit scenario of the CCSM3 data available at Earth System Grid after downscaling to the modeling domain. The model runs were conducted on a dual-core Pentium based desktop Linux System, and lasted more than a year to complete due to limited computing resources.

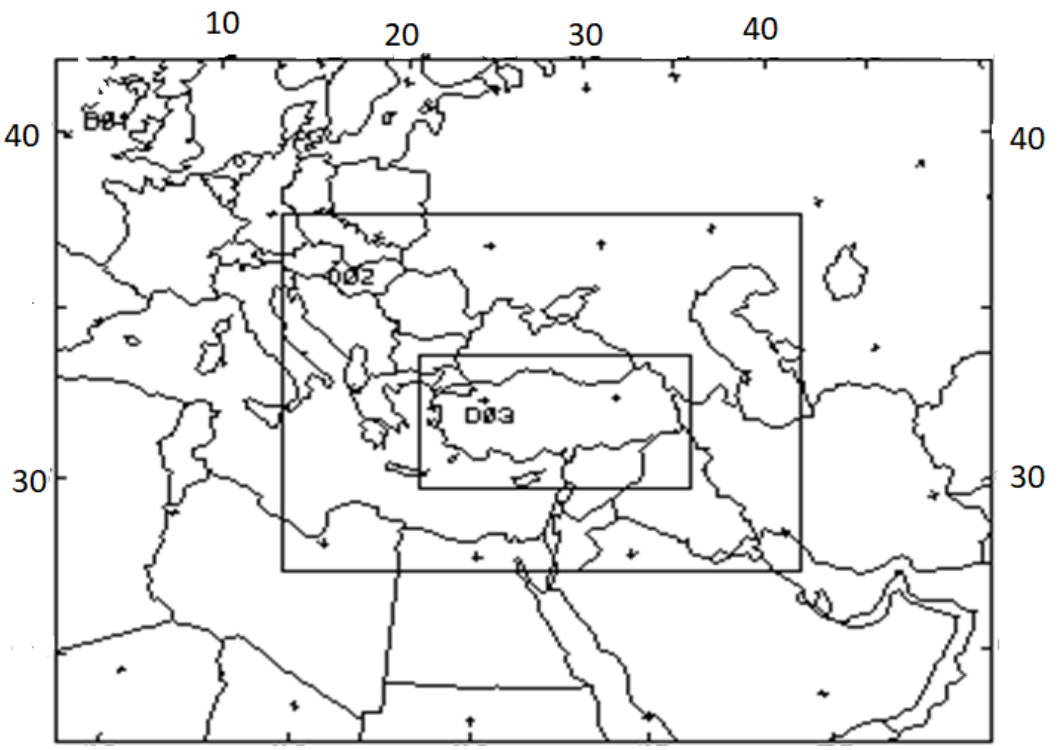
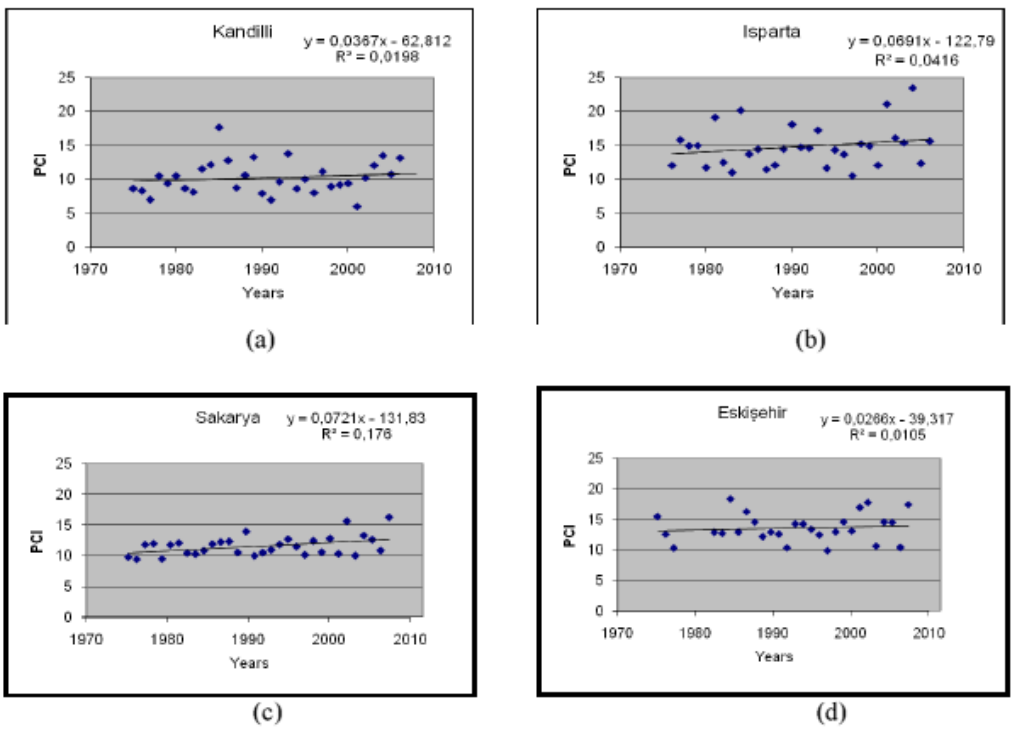


Figure 2. Three nested grids for a numerical model.

Analyses of the Observed and Simulated Precipitation

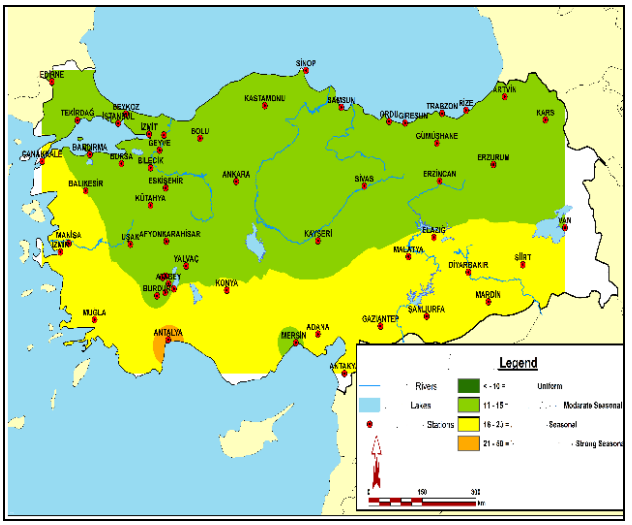
Station based analyses of PCI given in Figures 3a-d show an increasing trend throughout the analysis period, particularly after year 2000 at all four stations. The distribution at Kandilli station was mostly uniform (PCI < 10) before year 2000. Although PCI values point out an increasing trend in the last part of study period at all four stations, the type of distribution changes from uniform to moderate concentration (moderate seasonal). There is sufficient evidence of this type of relation at Sakarya (Figure 3c) with $\alpha = 0.10$ and at other locations with $\alpha \cong 0.25$ significance levels. PCI values at Isparta are influenced by strong concentrated (strong seasonal) conditions (PCI > 21) in recent years and small increase in tendency is seen especially within the last decade at this location (Figure 3b, and Table 2).



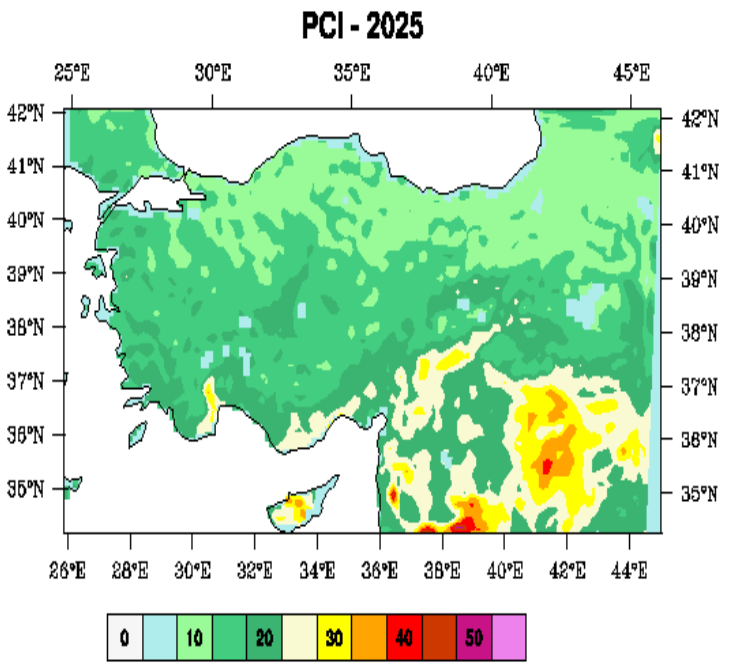
Figures 3a-d. PCI values in (a) Kandilli, (b) Isparta, (c) Sakarya and (d) Eskişehir (1975-2006).

The analysis of observed PCI distribution performed throughout Turkey (Figure 4a) shows a temporally uniform rainfall distribution in Northern Turkey, moderate concentrated (moderate seasonal) and concentrated (seasonal) distribution in Central and Southern Turkey, respectively (Aslan et al., 2010 a, b). A localized strong concentrated (strong seasonal) distribution near Antalya is also seen. These results were found to be like other analyses performed for the period between 1976 and 2004 in Turkey (for example, Türkeş et al., 1999; and Apaydin et al., 2006). The comparison of the current

PCI distribution (Figure 4a) against the future estimates (Figures 4c-d) shows two features: First, both the current and future estimates are in general agreement with the distribution of uniform, moderate seasonal and seasonal distributions. Second, strong seasonal distribution seen in current observed conditions is expected to be more widespread than observations in the future, but would be concentrated along especially Southern coastal areas and southeastern Turkey.

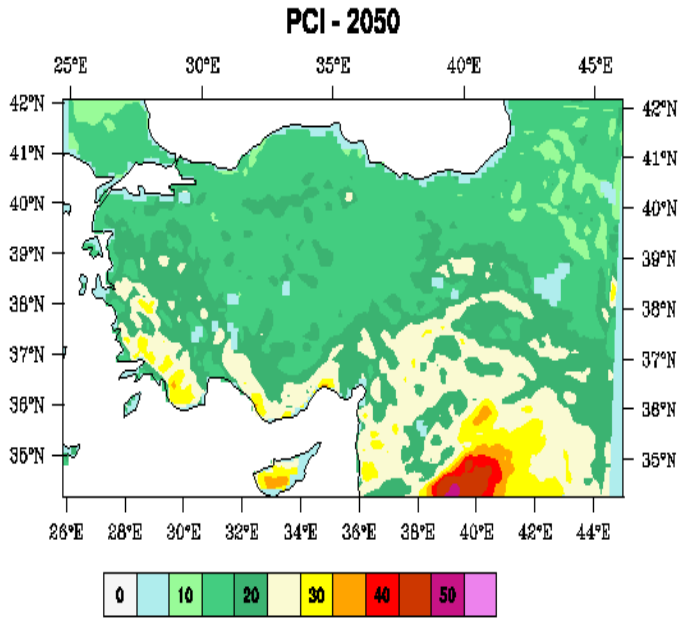


(a)

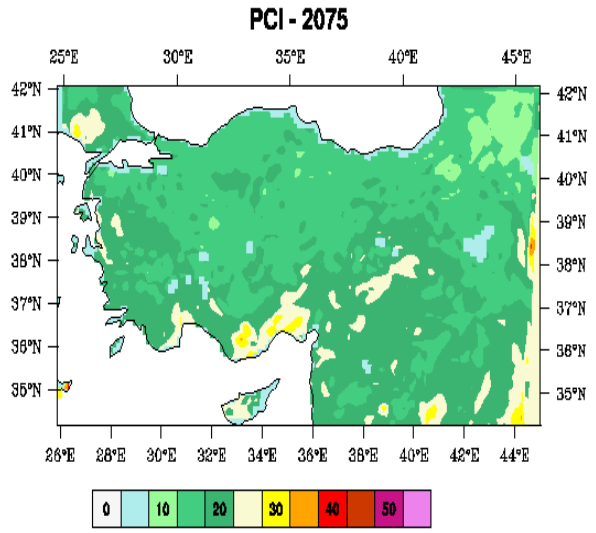


(b)

Figure 4. (Continued).



(c)



(d)

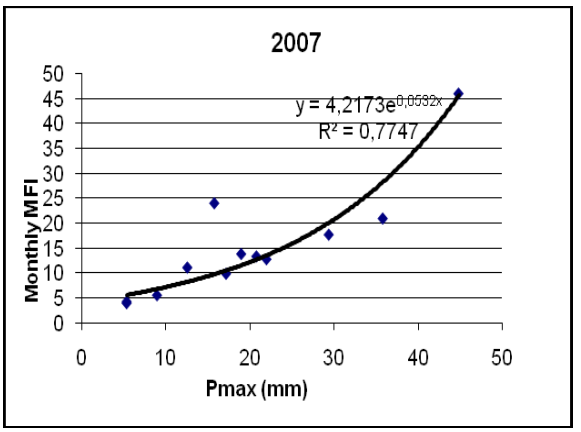
Figures 4a-d. Spatial variation of PCI for (a) observed conditions between1975-2006, and future estimates for years (b) 2025, (c) 2050, and (d) 2075.

Numerical model results suggest that seasonal effects would extend into Southern Aegean Sea region in 2050. However, the model also suggests that these seasonal effects would show a decline by 2075, and the effects of climate change on precipitation would be seen more dramatically by 2050 compared to 2025 and 2075. The reason for this model behavior is not known and it requires further analyses to find its cause. Since the model's microphysics has one of the dominant factors affecting the formation of

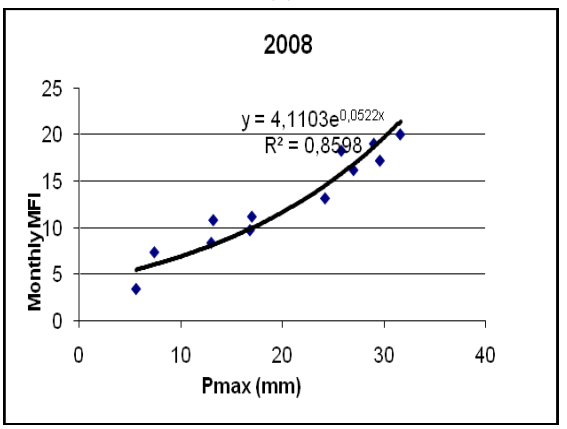
precipitation, a sensitivity analysis was conducted using the Reisner microphysics scheme. The sensitivity analyses didn't show any major deviations from the original estimates of PCI distribution. Another possibility is that, since the CCSM3 climate model, any meteorological under, drove the MM5 model or overestimation generated by the CCSM3 model would also affect the outcome of our future meso-scale model estimates.

Analyses of Modified Fournier Index

"Empirical relations" between monthly maximum precipitation and, water erosivity risk values based on observations in Sakarya have been presented in Figure5 (a) and (b). There is sufficient evidence of this empiric relation between monthly MFI and maximum precipitation values with the confidence level; $\alpha = 0.01$.

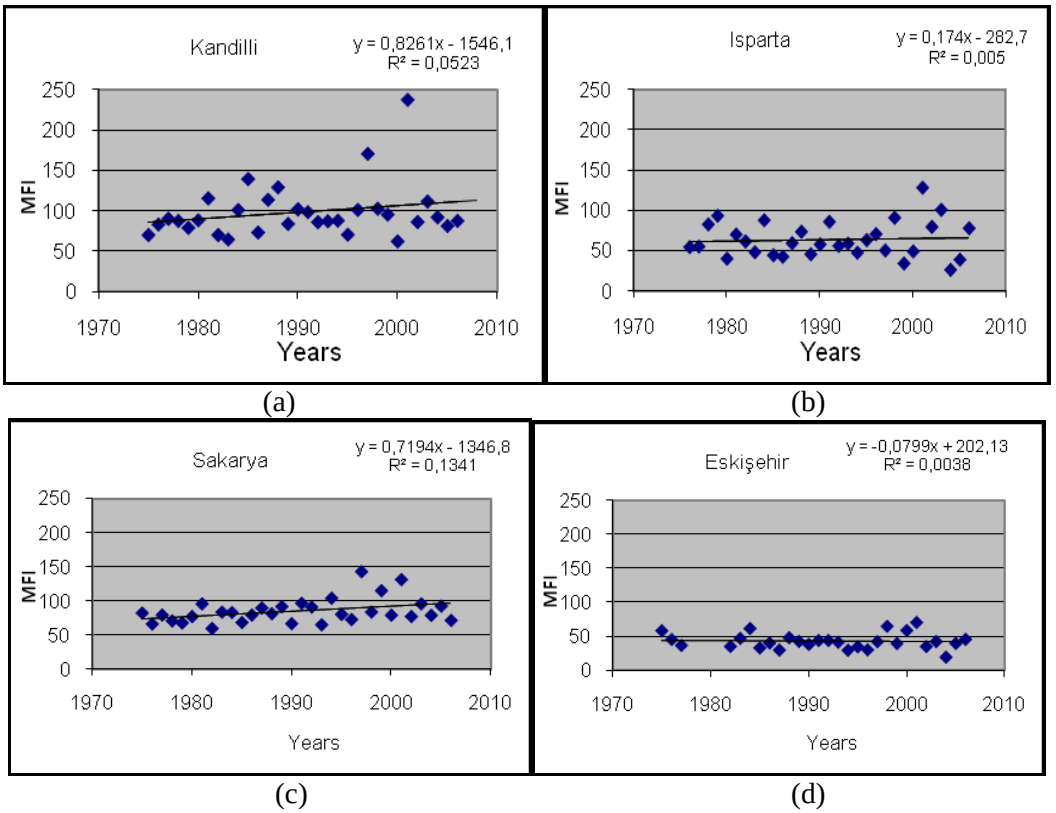


(a)



(b)

Figures 5a-b. Relation between maximum precipitation and monthly MFI values in Sakarya, (a) in 2007, (b) in 2008.

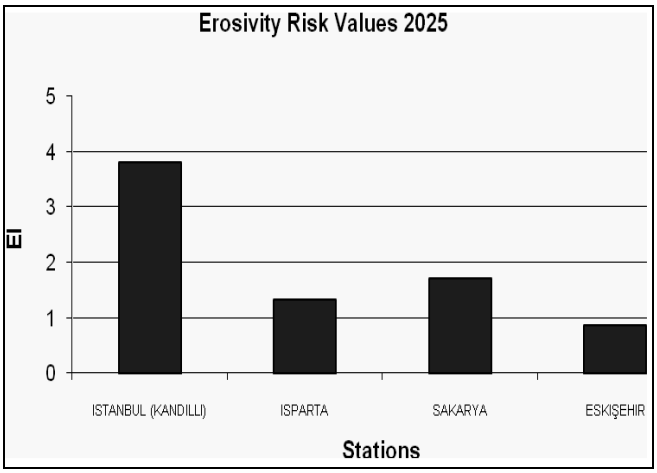


Figures 6 a-d. Annual variations of MFI values in (a) Kandilli, (b) Isparta, (c) Sakarya and (d) Eskişehir.

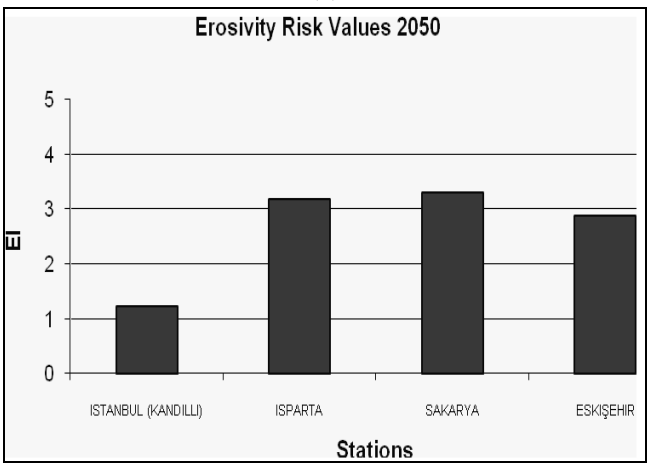
Figures 6 a-d show temporal variations of MFI calculated from 1975 to 2006 at four stations. In general, Kandilli is under low to moderate erosivity risks. However, high-risk characteristics were also recorded in 1985, 1988, 1997, and 2002. Similar to Kandilli, erosivity risk is generally low (less than 90) in Isparta while moderate to high water erosion risk values were recorded in recent years. Sakarya and near vicinity had low erosivity risk, as well, except it was also under the high erosivity risk within the last ten years. Erosivity risk using MFI variations shows very low characteristics in Eskişehir during the entire study period compared to other stations. There is a significant increasing trend of MFI values in Sakarya with alpha = 0,05.

Analyses of Observed and Simulated Erosivity Index

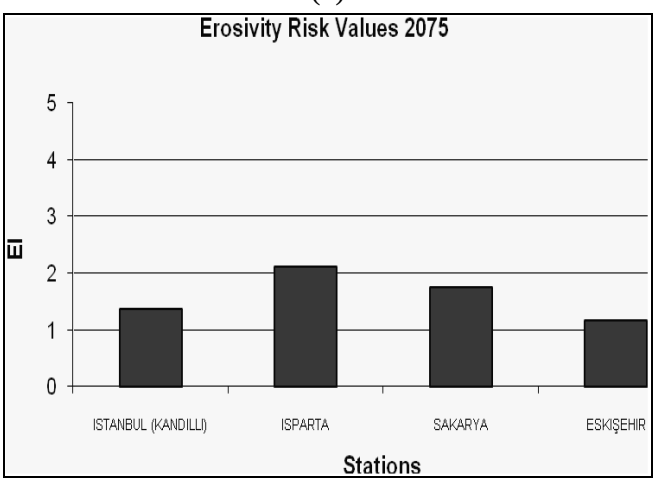
There are low and moderate erosivity risk (EI) characteristics in Eskişehir and Sakarya (Table 6). High erosivity risk (Maximum EI class No = 3) values have been recorded in Kandilli and Isparta in the last decade.



(a)



(b)



(c)

Figures 7a-c. Erosivity risk scenarios at Kandilli, Isparta, Sakarya and Eskişehir in (a) 2025, (b) 2050 and (c) 2075 (cont.).

Erosivity risk values based on a single climate change scenario show some increasing trend until 2050 in Isparta, Sakarya and Eskişehir (Figs 7a-c). In the following 25 years, a slightly decreasing risk trend is being estimated. On the other hand, an opposite or decreasing trend is calculated in Kandilli compared to Isparta, Sakarya, and Eskisehir.

Table 6. Statistical Descriptive of EI values in Kandilli, Isparta, Sakarya and Eskişehir for ten years interval

Station	1975-1984		1985-1995		1995-2006	
Kandilli	Min.	1 (L)	Min.	1 (L)	Min.	1 (L)
	Max.	1 (L)	Max.	2 (M)	Max.	3 (H)
	Mean	1 (L)	Mean	2 (M)	Mean	2 (M)
Sakarya	Min.	1 (L)	Min.	1 (L)	Min.	1 (L)
	Max.	1 (L)	Max.	2 (M)	Max.	2 (M)
	Mean	1 (L)	Mean	1 (L)	Mean	1 (L)
Eskişehir	Min.	1 (L)	Min.	1 (L)	Min.	1 (L)
	Max.	2 (M)	Max.	2 (M)	Max.	2 (M)
	Mean	1 (L)	Mean	1 (L)	Mean	1 (L)
Isparta	Min.	1 (L)	Min.	1 (L)	Min.	1 (L)
	Max.	2 (M)	Max.	2 (M)	Max.	3 (H)
	Mean	1 (L)	Mean	1 (L)	Mean	2 (M)

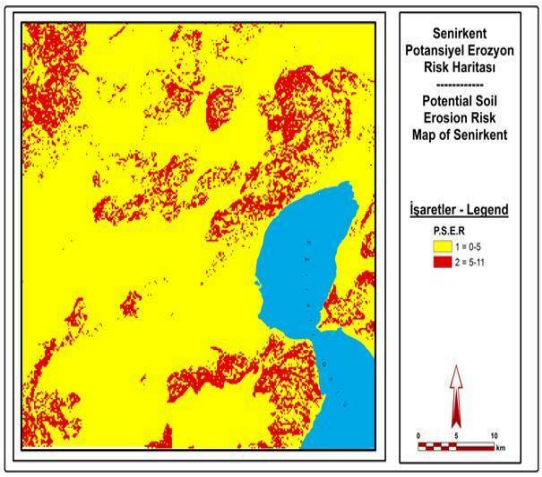
L (1): Low erosivitiy risk, M (2): Moderate erosivitiy risk), H (3): High erosivitiy risk

Spatial Variation of Erosivity

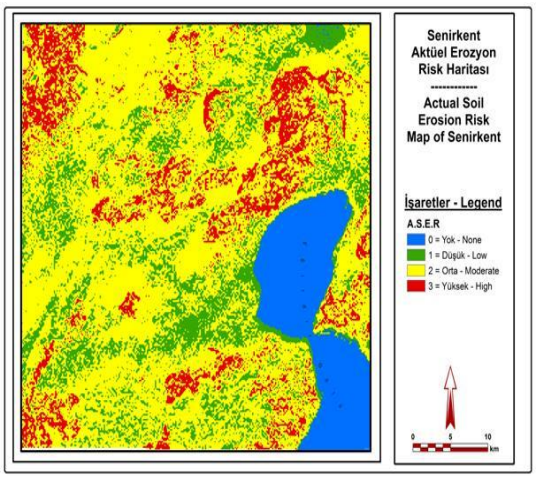
Figures 8a, b show potential, and actual water erosion risk maps for Isparta (Senirkent). Potential erosion risk map shows there two class values: Class number - 1 represents low risk areas and class number - 2 represents high-risk areas. However, in detailed analysis of actual water erosion risk map evaluates four different classes form no erosivity risk to high erosivity risk areas.

To estimate potential erosivitiy risk for a given study area, there are sufficient data, but for predicting actual water erosivity risk, actual land use maps and more data on agricultural fields are needed. By using CORINE method, land use planning programs, statistical and remote sensing analyses in time and space domain, the more reliable results have been evaluated (Erhard et al., 2003).

When potential and actual soil erosion risk maps are compared with each other in three study areas, roles of surface characteristics, slope and altitude, the differences between two kinds of maps are better illustrated. Similarly, potential and actual water erosivity risk variations are compared with two other study areas; Eskişehir and Sakarya. Actual erosion and potential erosion risk maps of whole Turkey have also some differences as it is defined in these three pilot areas, (Aslan et al., 2010, a, b).

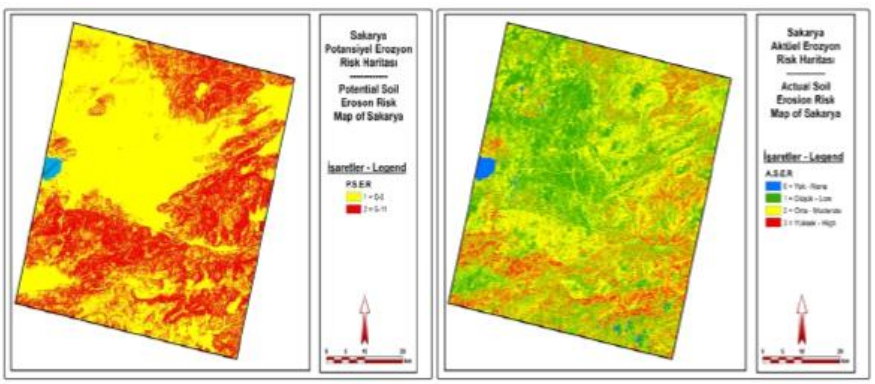


(a)



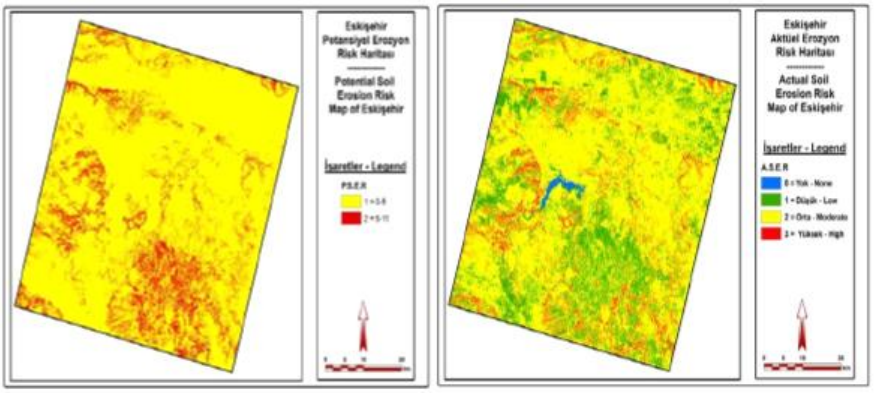
(b)

Figures 8a-b. (a) Potential and (b) actual soil erosion risk map at Isparta.



(a) (b)

Figures 9a-b. (a) Potential and (b) actual soil erosion risk map at Sakarya.



Figures 10a. (a) Potential and (b) actual soil erosion risk map at Eskişehir.

RESULTS AND CONCLUSION

Spatial and temporal variations of PCI, MFI, and BGI were analyzed using observed precipitation at four locations, and precipitation estimates obtained from MM5 numerical model for years 2025, 2050, and 2075. EI was also studied based on the CORINE method.

PCI analyses show an increasing trend especially beginning from year 2000 in Kandilli. In the last part of the study period, its type changes from uniform to moderate seasonal although an increasing trend is seen. However, strong seasonal conditions were recorded in recent years in Isparta.

Kandilli is generally under low to moderate erosivity risks. However, intermittent and very high water erosivity risk values were also recorded in 1985, 1988, 1997, and 2002. Erosivity risk values in Eskişehir were less than values observed in other regions. Sakarya is generally under the low erosivity risk except last ten years in which it is under the high erosivity risk. When the estimated future precipitation values were examined, erosivity risks show an increasing trend in Isparta, Sakarya and Eskişehir while it shows a decrease in Kandilli.

In the second part of the paper, CORINE Soil Erosivity Assessment Methodology is applied to define potential and actual water erosion and soil loss in study areas. As a first step, soil erodibility (K; soil texture, depth and stoniness) was taken into account by using remote sensing data and topographic maps. To define EI, MFI and BGI were computed. Potential soil erosion risk (EP) was defined by considering K, EI and slope angle (S). With the last additional factor (V; land cover), actual water erosivity risk (EA) was evaluated for each case.

Based on CORINE Soil Erosion Assessment Methodology, more detailed analyses of actual water and potential water erosion risks for Isparta, Eskişehir and Sakarya were presented in last part of his paper. Potential erosion risk maps of study areas show only

two classes (low and high-risk areas) while a detailed analysis indicates four different classes of soil erosion risk in pilot areas (from no erosivity risk to high erosivity risk). There are also great differences between actual and potential soil erosion risk maps in Sakarya and Eskişehir, (Aslan et al., 2010 a, b). Because of these differences, other case studies and detailed analyses of spatial soil erosion risk variations are essential for definition of actual water erosivity risk in Turkey.

ACKNOWLEDGMENTS

The authors thank to Prof. Dr. Donald GABRIELS, Ghent University, Belgium, TUJJB (Turkey National Geodesy and Geophysical Union), Boğaziçi University, Kandilli Observatory and Earthquake Research Institute and Computer Technician Mr. Şenol SOLUM. This paper is related with the research project supported by Turkish HGK (General Command of Mapping) with the project number: TUJJB-TUMEHAP–01–06.

REFERENCES

Apaydin, H., G. Erpul, I. Bayramin, and D. Gabriels, (2006). Evaluation of indices for characterizing the distribution and concentration of precipitation: A case for the region of Southeastern Anatolia Project, Turkey. *J. of Hydrology*, 328, 726-732.

Arnoldus, H. M., (1980). An Approximation of the Rainfall factor in the Universal Soil Loss Equation. In Assessments of Erosion, de Boodts, M, and D. Gabriels (Eds). John Wiley and Sons Ltd, Chichester 127–132.

Aslan, Z, D. Gabriel, C. Ayday, D. N. Yeniçeri, D. Okçu, A. S. Söğüt, B. Oğuzhan, Z. N. Çağlar, A. Tok-gözlü, M. Gümrükçüoğlu, K. Gürer, and G. Erpul, (2010). *Yüzey Gözlemlerine Ve Uzaktan Algılama Tekniklerine Dayalı Aktüel Erozyon Riski Analizi* [*Actual Water Erosivity Risk Analysis Based on Ground Measurements and Remote Sensing Data*], Project Number: TUJJB-TUMAHAP-01-06.

Aslan, Z, D. Gabriels, C. Ayday, D. N. Yeniçeri, D. Okçu, A.S. Söğüt, B. Oğuzhan, Z. N. Çağlar, A. Tokgözlü, M. Gümrükçüoğlu, K. Gürer, and D. Lobo, (2010). *Actual Erosivity Risk Analysis Based on Ground Measurements and Remote Sensing Data*: Workshop Report, Boğaziçi University, pp.42.

Barhdi A., and G. Ramollari, (2017). Risk assessment of potential and actual soil erosion at the territory of Albania via System Geographical Information (GIS) and its combination of technology assessment according to Corine Land Cover, *Albanian j. Agric. Sci.* (2017) p. 121-127., Special edition.

Blackadar, A. K., (1976). Modeling the Nocturnal Boundary Layer. Preprints of Third Symposium on Atmospheric Turbulence and Air Quality, Raleigh, NC. 19-22, (October 1976), *Amer. Meteor. Soc.*, Boston, 46-49.

Blackadar, A. K., (1979). *Advances in Environmental Science and Engineering*, 1. No. 1. Pfaflin and Ziegler, Eds. Gordon and Breach Publishers, 50-85.

CEC (Commission of the European Communities), (1992). *CORINE Soil Erosion Risk and Important Land Resources*. EUR 13233, Luxembourg.

Dengiz O, and Akgül S., (2005). Soil Erosion Risk Assessment of the Golbasi Environmental Protection Area and Its Vicinity Using the CORINE Model. *Turk J. Agric. For*., 29, 439-448. TUBITAK.

Dudhia, J., (1989). Numerical Study of Convection Observed During the Winter Monsoon Experiment Using a Two-Dimensional Model. *J. Atmos. Sci.*, 46, 3077-3107.

Erhard, M., Boken, H., Glante, F. (2003). The Assessment of the Actual Soil Erosion Risk in Germany, based on CORINE Land-cover and Statistical Data from the Main Representative Survey of Land Use, *OECD Expert Meeting on Soil Erosion and Soil Biodiversity Indicators*, Rome, Italy.

Gabriels, D., (2003). The USLE for Predicting Rainfall Erosion Losses. *ICTP-SMR*.705-3. Trieste.

García-Marín, A. P.; Ayuso-Muñoz, J. L.; Cantero, F. N.; Ayuso-Ruiz, J. L., (2017). *Spatial and Trend Analyses of Rainfall Seasonality and Erosivity in the West of Andalusia* (Period 1945–2005), Soil Science: (April 2017) - Volume 182 - Issue 4 - p 146–158, doi: 10.1097/SS.0000000000000020.

Grell, G. A., J. Dudhia and D. R. Stauffer, (1995). *A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5)*. NCAR Technical Note, NCAR/TN-398+STR.

IPCC, (2007). Climate Change (2007).The Physical Science Basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Solomon, S., D. Qin, M. Manning (Eds).

Kim S., (2004). Wavelet Analysis of Precipitation Variability in Northern California, U.S.A, *KSCE, Journal of Civil Engineering*, Vol. 8, No. 4, pp. 471~477.

Luis M. De, González- Hidalgo J.C., and Longares L.A., (2010). Is Rainfall Erosivity Increasing in The Mediterranean Iberian Peninsula. *Land Degrade. Develop*. 21: 139–144.

Luis M. De, González-Hidalgo J. C., Brunetti M., and Longares L. A., (2011). Precipitation concentration changes in Spain (1946–2005), *Nat. Hazards Earth Syst. Sci.*, 11, 1259-1265.

Michiels, P., and Gabriels, D., (1996). Rain Variability Indices for the Assessment of Rainfall Erosivity in the Mediterranean Region. In: Rubio, J.L., Calva, A. (Eds.), *Soil*

Degradation and Desertification in Mediterranean Environments. Geoforma Editions, Logrono, Spain, pp. 49-70.

Mlawer, E. J., Taubman S. J., Brown P. D., Iacono M. J., and Clough S. A., (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the long wave. *J. Geophys. Res.*, 102(D14), 16,663-16,682.

Oliver, J. E., (1980). Monthly precipitation distribution: A comparative index. *Professional Geographer*. 32: 300–309.

Pal I. and Al-Tabbaa A., (2010). Assessing seasonal precipitation trends in India using parametric and non-parametric statistical techniques, *Theoretical Applied Climatology*, DOI 10. 1007/s00704–010–0277–8.

Turkes, M., (1999). Vulnerability of Turkey to Desertification with Respect to Precipitation and Aridity Conditions, *Turkish Journal of Engineering and Environmental Science,* Vol. 23, p. 363-380.

Yang W, Bárdossy A., Caspary H.J., (2010). Downscaling daily precipitation time series using a combined circulation- and regression-based approach, *Theoretical Applied Climatology*, DOI 10. 1007/s00704–010–0272-0.

Yeşilırmak, E., (2015). " Büyük Menderes Havzasi'nda günlük yağış konsantrasyonunun analizi: Analysis of Daily Precipitation Concentration at Buyuk Menderes Watershed", *Journal of Adnan Menderes University Agricultural Faculty,* 12(2): 55 – 71.

Zhang, G. Q., C. Xu, Gemmer M., Chen Y.D., and Liu C., (2009). Changing properties of precipitation concentration in the Pearl River basin, China, *Stoch Environ Res Risk Assess*, 23: 377–385, DOI 10. 1007/s00477-008-0225-7.

Websites

http://www.ipcc.ch/report/ar5/ (December 31, 2017).

http://foodsustainability.eiu.com, (August 1, 2018).

www.barillacfn.com, Barilla Center for Food & Nutrition, fixing food: towards a more sustainable food system, written by The Economist Intelligence Unit, (December 28, 2017).

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 3

SHORELINE EXTRACTION FROM CLOUD REMOVED LANDSAT 8 IMAGE: CASE STUDY LAKE ERÇEK, TURKEY

K. Kalkan[1,*], D. Maktav[1,†] and B. Bayram[2]
[1]Istanbul Technical University, Geomatics Engineering Department, Istanbul, Turkey
[2]Yildiz Technical University, Geomatics Engineering Department, Istanbul, Turkey

ABSTRACT

Cloudless satellite images may not be found for each desired time for shoreline change monitoring tasks. On the other hand, working with cloudy satellite images does not allow to obtain accurate information due to cloud and shadow effect. They can cause disruptive effects, especially for shoreline extraction issues. In this study, difficulties of shoreline extraction from cloudy satellite image are coped with using information cloning technique from cloudless images with similar spectro-temporal characteristics. Landsat 8 images are used to test cloud removal approach. First, the cloud removal is realized by a developed method based on object-based image analysis. Then, Simple Linear Iterative Clustering (SLIC) method and Normalized Difference Water Index (NDWI) are examined to extract shorelines. The efficiency of proposed cloud removal method is investigated on shoreline extraction process. For this purpose, a lake is selected as the study area to implement the proposed algorithm. Spectral and structural features are used to extract cloud and shadow features to use in information cloning for cloudless image generation. In this study, Digital Shoreline Analysis System (DSAS) is used for comparison of differences between manually digitized and automatically extracted shorelines.

* Corresponding Author's E-mail: kaankalkan@gmail.com.
† Corresponding Author's E-mail: maktavd@itu.edu.tr.

Keywords: cloud removal, SLIC, NDWI, image processing, Landsat 8

INTRODUCTION

Global warming and its effects are one of the main concerns of scientists. Global average surface temperature has increased by about 0.74 °C in the last century according to the global surface temperature measurement results (Deng et al., 2014). Therefore, prediction of response of biodiversity to environmental changes became one of the most important research fields. Lakes consist of large variety of habitats, which are vital for survival of biodiversity (Gammal et al., 2017). The biological and physical characteristics of lakes can change unprecedentedly as a response of global warming and anthropogenic processes (Pareeth et al., 2017). According to International Union of Geological Sciences (IUGS), shorelines are one of the most important environmental heritages (Li et al., 2001). Many of the lake shorelines are in danger because of human-based, economic developments and ecological reasons. Therefore, monitoring of shoreline plays a key role to protect environmental heritage areas (Addo 2013). Up-to-date data is obligatory to take efficient decisions for lakes (Jawak et al., 2015).

Different techniques are proposed for shoreline extraction and monitoring which are based on mainly ground survey, photogrammetry, satellite imagery with various resolution, RADAR (Radio Detection and Ranging), and LIDAR (Light Detection and Ranging) systems (Gens, 2010). Up-to-date, accurate, temporal, and reliable information on lakes can be derived by using optical remote sensing data (Kutser et al., 2012). Monitoring of water bodies with satellite remote sensing provides instantaneous and temporal data collection, which may not be available by field measurements (Trochta et al., 2015). The 30 m (visible, NIR (Near Infrared), SWIR (Short-wave Infrared)) spatial resolution Landsat series give a great opportunity for mapping coastal areas (Vincent et al., 2004).

Satellite imagery can be applied to detect changes of lakes by adapting various segmentation algorithms such as pixel-based and object-based, and remote sensing indices (Song et al., 2014; Machado et al; 2014). Iterative Self Organized Data Analysis (ISODATA) (Guariglia et al., 2006), NDWI (Zheng et al., 2011), thresholding and morphological filtering (Pardo-Pascual et al., 2012), Otsu thresholding (Bouchahma and Yan, 2012), non-separable wavelet transform, active contour models (Zhang et al., 2013), genetic algorithm based methods (Yousef and Iftekharuddin, 2014), particle swarm optimization method (PSO) (Bayram et al., 2016), mean-shift segmentation, object oriented/fuzzy classification approaches (Purkis et al., 2016; Bayram et al., 2013; Bayram et al., 2008), normalized cut approach (Ding and Li, 2014), region segmentation and edge detection based methods (Paravolidakis et al., 2016) can be counted as mainly used methods for shoreline extraction.

Traditional cloud removal methods could be categorized into three methods: image filtering method, mathematical morphologic method, and multi-temporal cloud free areas composition (Zhengke et al., 2011).

Image filtering method aims to remove thin clouds (Feng et al., 2004). The limitation of this filtering method is that it can remove thin clouds from images but cannot recover information under cloud cover effectively; meanwhile, the filtering process can cause radiometric loss. The mathematical morphologic strategy uses statistical test to predict pixel values under clouds and cloud shadows in reference scenes from multi-temporal dataset (Helmer and Ruefenacht, 2005). Cloning from cloud-free areas method uses multi-temporal data which you can check radiometric consistency and then can yield better results in case of radiometry (Gabarda and Cristobal 2007; Tseng et al., 2008).

Lin et al., (2013) categorize image reconstruction methods into three categories such as in-painting, multispectral and multi-temporal-based methods. In in-painting-based methods, information of cloud covered regions is synthesized based on using the techniques of image synthesis and in-painting (Maalouf et al., 2009; Lorenzi et al., 2011). In multispectral-based methods, multispectral data are used in cloud detection and information reconstruction steps (Rakwatin et al., 2009; Roy et al., 2008; Zhang et al., 2009). Multi-temporal-based methods (Melgani 2006; Benabdelkader and Melgani 2008; USGS 2004; Gabarda and Cristobal 2007; Helmer and Ruefenacht 2005; Jiao et al., 2007; Tseng et al., 2008) compared with the in-painting and multispectral-based methods, which rely on both temporal and spatial relationships, show better results with reconstruction of opaque cloud covered pixels. Li et al., (2003) also developed a threshold-based approach to identify the best cloud-free and non-shadow pixels of a given region. A cloud-free image is then generated by mosaicking the selected cloud-free pixels. While methods above can yield good results for homogenous regions, these methods based on data fusion can show difficulties with clouds over heterogeneous land cover. Lin et al., (2013) proposed an information-cloning algorithm that selects cloud-free patches using a quality assessment index and reconstruct patches by using a global optimization process. Thus, this method can yield good cloud-free results for opaque clouds (Lin et al., 2013).

In this study, cloud patches are cloned with cloudless satellite images from multi-temporal dataset by using most correlated image for any cloud area. This study is an application of information reconstruction technique to cloudless images generation. The primary aim of this study was to test the efficiency of shoreline extraction from cloudy satellite image by using a cloud removal method. For this purpose, cloudless and cloud removed Landsat 8 images are processed by implementing SLIC and NDWI+SLIC methods for shoreline extraction. MATLAB environment is used for examining the methods. Accuracy assessment is performed by measuring perpendicular distances between shorelines, which digitized manually and obtained by using the DSAS an extension for ArcGIS v.9.0.

STUDY AREA AND DATA USED

Lake Erçek is selected as the study area. This lake with geographic coordinates 38.67º N and 43.57º E is a salt lake in Van Province in Eastern Turkey, about 30 km east of Lake Van. It sits on an elevation of about 1800 meters, and has an area of 106 km^2 and an average depth of 18 meters (maximum depth 40 m). Its main inflow is from the Memedik Stream. Lake Erçek is an important site for breeding and migrating water birds (Van İl Kültür Turizm Müdürlüğü, 2016).

The cloudless and cloudy Landsat 8 images of two different dates, September 13 and 29, 2014 were used for analysis. Cloud and shadow areas are classified and cloudless image is created by using cloud removal method proposed by Kalkan and Maktav, 2017. Landsat 8 data are provided as raw Digital Numbers (DN). Data are converted to the Top of Atmosphere (ToA) reflectance and brightness temperature using radiometric rescaling coefficients stored in the product metadata file (MTL file). By this conversion, image data is converted to physically meaningful units.

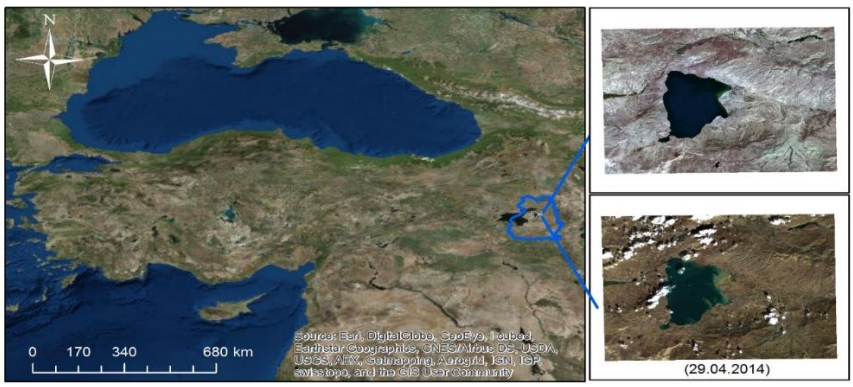


Figure 1. The study area.

METHODS

Cloud and cloud shadow detection procedures are still an ongoing research topic in remote sensing. Our proposed method detects clouds and cloud shadows from Landsat 8 images by using both spectral and geometrical properties after a super-pixel segmentation process. Beside these properties, discrimination of cold surfaces (snow, ice) and discrimination of cloud/shadow are strengthened by using thermal infrared channels of Landsat 8. Neighborhood relations are used to improve detection accuracy of cloud shadow regions around cloud areas (Kalkan, 2017).

After accurate and reliable detection of cloud and cloud shadow features, information reconstruction concept is applied to cloudy satellite images for creating cloudless images

from multi-temporal image dataset. Cloning cloudy patches from cloudless multi-temporal dataset is the information reconstruction concept of this study. Super-pixel segmentation detects cloudy patches accurately to start Flood Fill process to reconstruct cloudless information for cloud patches. Information reconstruction from multi-temporal cloudless images to cloud image is implemented by using Flood Fill algorithm by pixel-by-pixel which is one of the most known graphical filling algorithm. Choosing best image for cloning process is an important step to conserve spectral consistency. Patch by patch correlation calculations between multi-temporal dataset are used to find best image for cloning in terms of spectral consistency. Vector intersection calculations are also used to find best cloudless dataset for cloning process to achieve full cloudless patch cloning. Spectral and structural consistency is the main goal of our proposed cloning algorithm.

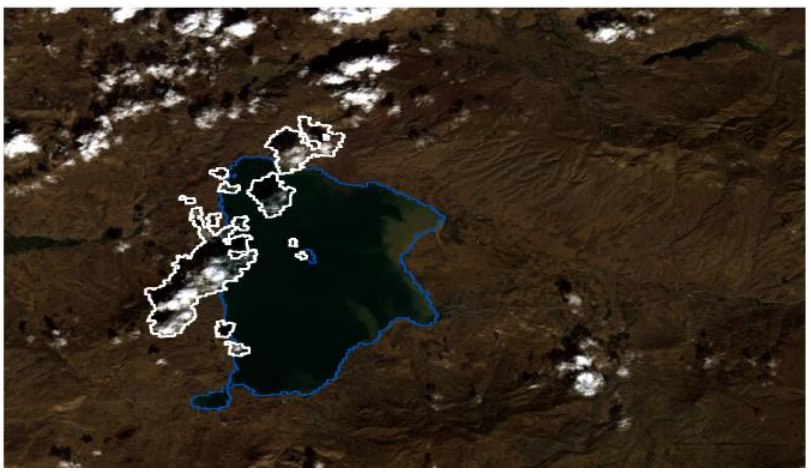


Figure 2. Cloud and cloud shadow classification results intersected with Erçek Lake water body.

Structural Similarity (SSIM) index is used to evaluate the quality of cloning in terms of spectral and structural consistency with cloudless reference image. Information reconstruction for cloudy images by our proposed method gives sufficient results for passive remote sensing applications, which are negatively affected, by clouds.

This study consists of four main steps: (i) generation of cloudless image, (ii) shoreline extraction from cloud removed Landsat 8 NIR band by SLIC, (iii) shoreline extraction from cloud removed Landsat 8 image by NDWI+SLIC, (iv) accuracy assessment.

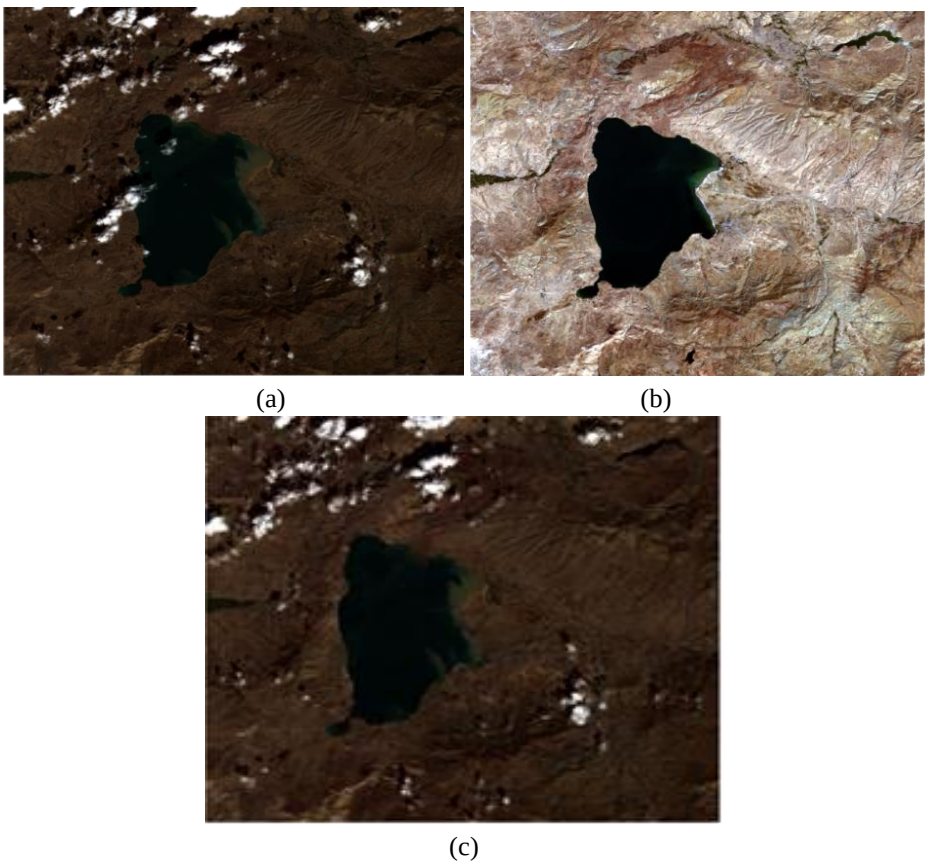
(i) Generation of Cloudless Image from Cloudy Landsat 8 Image

Different procedures are used to classify cloud and cloud shadow areas such as; detection of cold patches using thermal bands, projection of cloud features to classify

cloud shadows, water and shadow discrimination using NDWI values and different band ratios and indices to classify cloud patches (Figure 2). Spectral similarities are calculated to find best cloudless image for cloning process and image dated September 13, 2014 is chosen for cloning process, which shows higher spectral correlation with cloudy image (Figure 3).

Superpixels are used to combine pixels into meaningful groups to create pixel groups. Merging pixels having similar information is speeding up image processing tasks. SLIC algorithm is an efficient method for segmentation of image, which is based on spatially localized version of K-means clustering method. Fundamental specifications and advantages of SLIC method are evaluated in Achanta et al., 2012. This clustering method divides image into M×N regular grids. M and N values are given as an input, where (1)

$$M = \frac{imageWidth}{regionSize}, N = \frac{imageHeight}{regionSize} \quad (1)$$



(a) (b)

(c)

Figure 3. (a) Cloudy image (September 19, 2014), (b) cloudless image (September 13, 2014), (c) cloned image.

A super-pixel is processed by initializing from each grid center (2)

$$x_i = round_i \frac{imageWidth}{regionSize}, y_i = round_j \frac{imageWidth}{regionSize} \quad (2)$$

Following this step, regions are obtained by running K-means clustering, started from the centres (3).

$$C = \begin{Bmatrix} \Psi(xi,yj) \\ i = 0,1,\dots,M-1 \\ j = 0,1,\dots,N-1 \end{Bmatrix} \quad (3)$$

K-means clustering uses the standard Lloyd algorithm alternating by assigning pixels to the closest centers (Lloyd, 1982). Differences of SLIC compared to standard K-means is that each pixel can be assigned only to the center, which comes from the neighbor tiles. After creation of super-pixels, each super-pixel is taken into account to check if area is less then minimum region size value, which is taken as an input from user (Vedaldi and Fulkerson, 2010).

Shadows and water bodies are misclassified to each other because of their dark behavior. In this study, NDWI is used to overcome these two misclassification problems mentioned above and classify water bodies (Gao, 1996). The constant threshold is used to classify water bodies automatically (equation 4).

The NDWI for Landsat 8 is given in (4):

$$NDWI = \frac{NIR - Green}{NIR + Green}$$

(Water – NDWI < -0.2) (4)

RGB band combination and NDWI image of cloudy image are given in Figure 4.

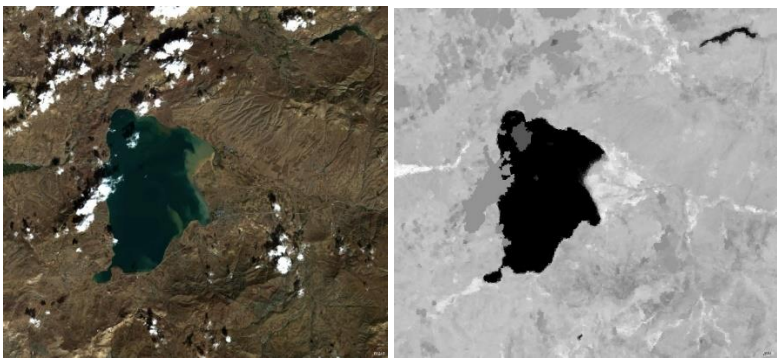
The cloudless and cloud removed NIR bands of Landsat 8 images are given in Figure 5.

(ii) Shoreline Extraction from Cloud Removed Landsat 8 NIR Band by SLIC

In the second step of the study SLIC was applied to both cloud removed and cloudless Landsat 8 NIR bands to extract shoreline.

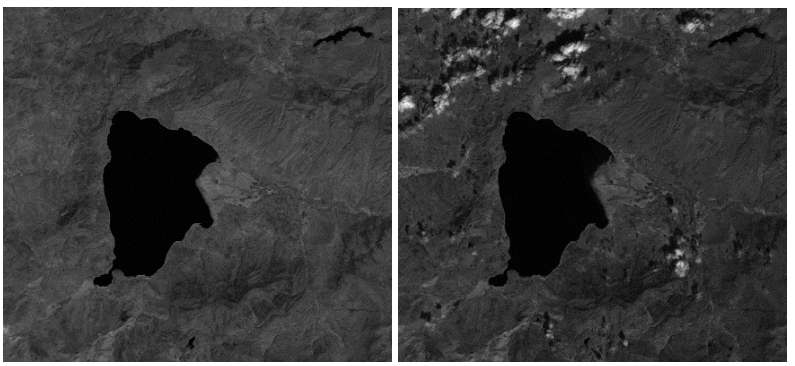
The same parameters for SLIC were used for both cloudless and cloud removed NIR bands as 120000 (number of super-pixels) and 7.2 (compactness). These parameters are defined after several tests and chosen empirically. Thresholding is applied to obtain binary images which include water body and land classes from SLIC segmentation

results. The threshold values are chosen for cloudless and cloud removed images as 33-255 and 25-255 respectively. Region analysis is implemented to remove noisy pixels by calculating the area of largest blob in binary image and deleting others (Mathworks, 2017). The obtained binary water body and land classes for cloudless, cloud removed images, and extracted shorelines are given in Figure 6 and Figure 7 respectively.



(a) (b)

Figure 4. (a) RGB combination, (b) NDWI image of cloudy image.



(a) (b)

Figure 5. Cloudless (a) and cloud removed (b) NIR bands of Landsat 8 images.

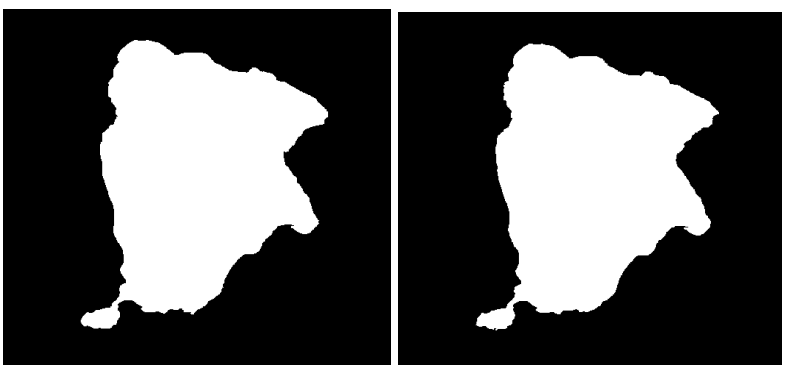


Figure 6. Noise removed images after SLIC method for cloudless (left) and cloud removed (right) images.

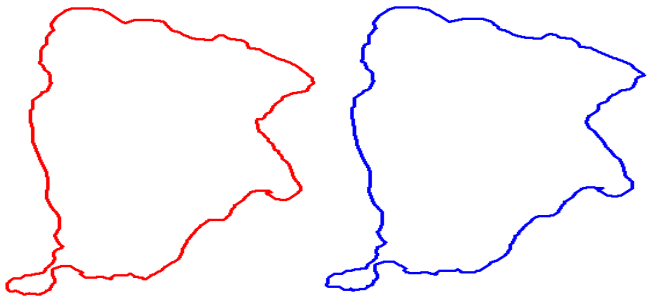


Figure 7. Extracted shorelines from cloudless (left) and cloud removed (right) NIR bands of Landsat 8 images.

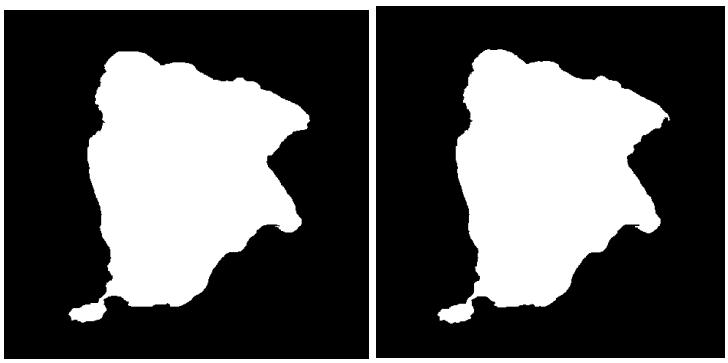


Figure 8. Noise removed images after applying NDWI+SLIC (left) cloudless.

(iii) Shoreline Extraction from Cloud Removed Landsat 8 Image by NDWI+SLIC

In the third step of the study, first, NDWI images have been created from both cloud removed and cloudless Landsat 8 images and then SLIC has been applied to both cloudless and cloud removed images. Same as the second step, the parameters for SLIC were set for both images as 120000 (number of superpixels) and 7.2 (compactness). Threshold values were set as 78-255 and 57-255 for cloudless and cloud removed images respectively. Thus, water body and land classes are created by NDWI+SLIC segmentation. Following to it, noise removal process is applied same as in the second step. The results of both cloudless and cloud removed images and extracted shorelines are given in Figure 8, 9 respectively.

(iv) Accuracy Assessment

Accuracy assessment is performed by comparing the manually digitized shorelines to obtained results. First, shorelines are manually digitized from cloudless and cloud

removed images to investigate cloud removal process, and manually digitized shoreline from cloudless image is taken as reference data. Second, SLIC and NDWI+SLIC results are compared to the reference data to observe the efficiency of cloud removal process on these methods. The differences are measured by DSAS (Thieler et al., 2009) which is based on baseline measurement for evaluating of shorelines (Jayson-Quashigah et al., 2013). In the presented study, Net Shoreline Movement (NSM) tool of DSAS is used to evaluate shorelines for first and second steps. NSM measures distances between reference and search shorelines at each transects (Oyedotun, 2014). Buffer size for baseline was 60 m. The length of transects and spacing are chosen as 150 m and 250 m respectively for entire lake shoreline. Thus, the perpendicular distances between reference and search shorelines are calculated (Figure 10).

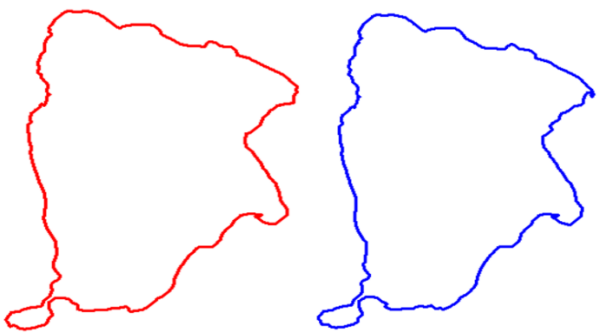


Figure 9. Extracted shorelines from Cloudless (left) and cloud removed (right) images by NDWI+SLIC.

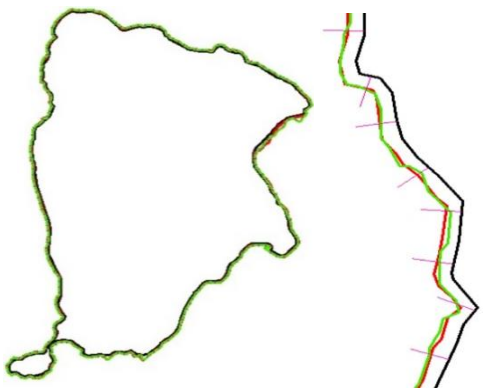


Figure 10. Evaluation of manual shoreline digitizing results.

RESULTS AND DISCUSSIONS

In this study, cloud and shadow areas near Erçek Lake have been classified with cloud and shadow detection algorithm proposed by Kalkan and Maktav, 2017. Cloud and shadow features have been cloned with cloudless satellite image to acquire cloudless

satellite image for shoreline extraction. Shorelines are extracted from both cloudless and cloud removed image by using different approaches to test efficiency.

The accuracy assessment results between manually digitized shoreline from cloudless (reference shoreline) and cloud removed images, reference shoreline and SLIC, and reference shoreline and NDWI+SLIC are given in Table 1. The results are obtained by using absolute values of calculated differences.

According to the results, the absolute average differences of shorelines between cloudless and cloud removed, cloudless and SLIC, cloudless and NDWI+SLIC are calculated around half pixel size (Table 1).

Table 1. Accuracy assessment results for entire lake

Case	**1**	**2**	**3**
Reference	Manually digitized shoreline from cloudless image		
Search	Manual digitized shoreline from cloud removed image	SLIC	NDWI+SLIC
Count of transects	209	189	
Spacing (m)	250		
Abs. Avr. (m)	14.93	19.62	14.67
Abs.Stdev (m)	18.57	14.2	14.23
Abs.max (m)	162.18	68.81	85.78
Abs.min (m)	0.01	0	0.5
RMS (m)	20.05	24.21	20.47

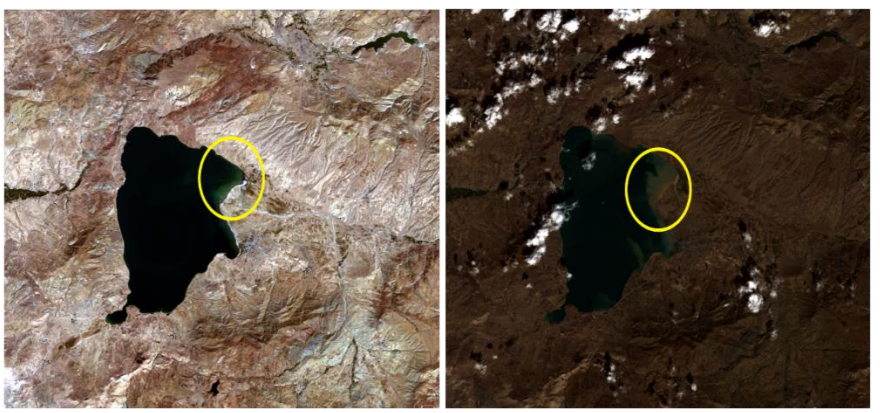


Figure 11. Water surface condition between cloudless and cloudy images (a) cloudless image, (b) cloudy image.

As it can be seen in Figure 11, the lake water surface conditions in marked areas are different. Water surface in cloud removed image is very fuzzy compared to cloudless image, although time period between both images is very short (16 days). This affects

results directly, even manual digitizing. Due to the difficulties in visual interpretation, the maximum difference, average between two manually digitized shorelines is calculated as 162.18 m. For this reason, accuracy assessment is done by ignoring the marked area and calculations are made with 189 transects. The standard deviation calculations also show that the distributions of differences are in half pixel size.

The ratio of calculated distances for 0-10 m, 10-15m, 15-20m, 20-25m, and >30 m intervals is given in Figure 12. The 66%, 45% and 64% of the ratios for Case-1, 2, 3 in Figure 12 have been calculated in half pixel size. Similar to this, 92%, 80% and 93% of the ratios for all cases are calculated as maximum in one-pixel size. The differences from comparison of manual digitizing results and comparison of reference data and NDWI results are very similar. From this point, for presented study, according to selected parameters for SLIC and NDWI, NDWI method creates more accurate results compared to SLIC method with NDWI. But, in general, the obtained results from both segmentation methods are satisfactory.

Water resources are irrevocable environmental heritages for survival of human beings. A big portion of a year can be cloudy in many countries as well in Turkey. Therefore, acquiring of cloudless satellite images may not be possible in desired time. Accurate detection and cloning of cloud and shadow areas from cloudless images is an important step for shoreline extraction from cloudy images. The proposed method has been tested for shoreline extraction and satisfactory results are obtained.

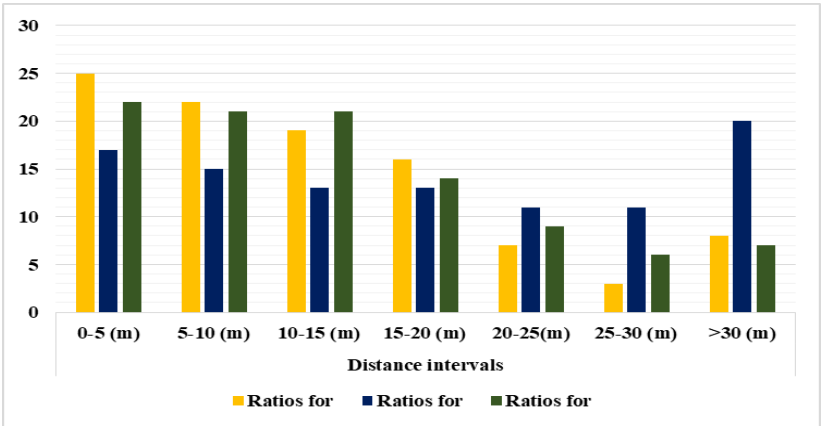


Figure 12. The ratio of calculated distances.

CONCLUSION

Since the monitoring of shorelines is very important for environmental protection, decision makers and coastal managers always need reliable data. Both natural and anthropogenic factors and their results should be multi-temporally observed. The main

contribution of presented study is to enable up-to-date and accurate data from cloudy satellite imagery by implementation of a developed cloud removal algorithm. Therefore, it will provide a big opportunity to the authorities for many issues, such as rehabilitation, planning and management of the shorelines, especially water resources. By processing of cloud removed image, almost the same results are obtained, even two different automatic segmentation algorithms are examined on the same image. According to our results, the proposed cloud removal algorithm can be implemented efficiently on cloudy satellite images to extract shoreline of lakes. Thus, difficulties of using cloudy images for shoreline extraction processes will be mostly eliminated or minimized.

As a future study, the proposed cloud removal algorithm will be applied to high-resolution multispectral satellite images. To improve accuracy of SLIC method, this algorithm will be applied to both NDWI image and multispectral image, and results will be evaluated.

ACKNOWLEDGMENTS

The authors are thankful to the students Fırat Erdem and Abdulkadir Ince for their contributions to this study.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC super-pixels compared to state-of-the-art super-pixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(11), 2274–2282. https://doi.org/10.1109/TPAMI.2012.120.

Addo, K. (2013). Shoreline morphological changes and the human factor. Case study of Accra Ghana. *Journal of Coastal Conservation*, *17*(1), 85–91. https://doi.org/10.1007/s11852-012-0220-5.

Bayram, B., Acar, U., Seker, D., & Ari, A. (2008). A Novel algorithm for coastline fitting through a case study over the Bosphorus. *Journal of Coastal Research*, *244*, 983–991. https://doi.org/10.2112/07-0825.1.

Bayram, B., Seker, D. Z., Acar, U., Yuksel, Y., Guner, H. A. A., & Cetin, I. (2013). An Integrated approach to temporal monitoring of the shoreline and Basin of Terkos Lake. *Journal of Coastal Research*, *292*, 1427–1435. https://doi.org/10.2112/JCOASTRES-D-12-00084.1.

Bayram B., Demir N., Ogurlu M., Catal R. H., and Seker D. Z., 2016. 3D shoreline extraction using orthopoto-maps and LIDAR, *37th Asian Conference on Remote Sensing*, 17-21 October 2016, Sri Lanka, Colombo.

Benabdelkader, S., & Melgani, F. (2008). Contextual spatio-spectral post reconstruction of cloud-contaminated images. *IEEE Geoscience and Remote Sensing Letters*, *5*(2), 204–208. https://doi.org/10.1109/LGRS.2008.915596.

Bouchahma, M., & Yan, W. (2012). Automatic measurement of shoreline change on Djerba Island of Tunisia. *Computer and Information Science*, *5*(5). https://doi.org/10.5539/cis.v5n5p17.

Deng, H., Chen, Y., Shi, X., Li, W., Wang, H., Zhang, S., & Fang, G. (2014). Dynamics of temperature and precipitation extremes and their spatial variation in the arid region of northwest China. *Atmospheric Research*, *138*, 346–355. https://doi.org/10.1016/j.atmosres.2013.12.001.

Ding, X., & Li, X. (2014). Coastline detection in SAR images using multiscale normalized cut segmentation. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 4447–4449). IEEE. https://doi.org/10.1109/IGARSS.2014.6947478.

Feng, C., Ma, J., Dai, Q., & Chen, X. (2004). An improved method for cloud removal in ASTER data change detection. In *IEEE International IEEE International IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004* (Vol. 5, pp. 3387–3389). IEEE. https://doi.org/10.1109/IGARSS.2004.1370431.

Gabarda, S., & Cristóbal, G., (2007). Cloud covering de-noising through image fusion. *Image and Vision Computing*, *25*(5), 523–530. https://doi.org/10.1016/j.imavis.2006.03.007.

Gammal, J., Norkko, J., Pilditch, C. A., & Norkko, A., (2017). Coastal Hypoxia and the importance of Benthic Macrofauna communities for ecosystem functioning. *Estuaries and Coasts*, *40*(2), 457–468. https://doi.org/10.1007/s12237-016-0152-7.

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.

Gens, R. (2010). Remote sensing of coastlines: detection, extraction and monitoring. *International Journal of Remote Sensing*, *31*(7), 1819–1836. https://doi.org/10.1080/01431160902926673.

Guariglia, A., Buonamassa, A., Losurdo, A., Saladino, R., Trivigno, M. L., Zaccagnino, A., & Colangelo, A., (2006). A multisource approach for coastline mapping and identification of shoreline changes. *Annals of Geophysics*, 49(1).

Helmer, E., & Ruefenacht, B., (2005). Cloud-Free satellite image mosaics with regression trees and histogram matching. *Photogrammetric Engineering and Remote Sensing*, *71*(9), 1079–1089.

Jawak, S. D., Kulkarni, K., & Luis, A. J., (2015). A Review on extraction of lakes from remotely sensed optical satellite data with a special focus on Cryospheric Lakes. *Advances in Remote Sensing*, *4*(3), 196–213. https://doi.org/10.4236/ars.2015.43016.

Jayson-Quashigah, P.-N., Addo, K. A., & Kodzo, K. S. (2013). Medium resolution satellite imagery as a tool for monitoring shoreline change. Case study of the Eastern coast of Ghana. *Journal of Coastal Research*, *65*, 511–516. https://doi.org/10.2112/SI65-087.1.

Jiao, Q., Luo, W., Liu, X., & Zhang, B. (2007). Information reconstruction in the cloud removing area based on multi-temporal CHRIS images. In Y. Wang, J. Li, B. Lei, & J. Yang (Eds.) (p. 679029). https://doi.org/10.1117/12.750462.

Kalkan, K., (2017). *Cloud detection and information cloning technique for multi Temporal satellite images*. (Doctoral Dissertation). Retrieved from Istanbul Technical University. (Accession No. 455410).

Kalkan, K., & Maktav, D., (2017). Segmentation based cloud and cloud shadow detection in satellite imagery. *Journal of Aeronautics and Space Technologies*, *10*(1), 45–54.

Kutser, T., Paavel, B., Verpoorter, C., Kauer, T., & Vahtmäe, E., (2012). Remote sensing of water quality in optically complex lakes. *In Proceedings of the XXII Congress of the International Society for Photogrammetry and Remote Sensing*, Melbourne, Australia (Vol. 25).

Li, R., Di, K., & Ma, R., (2001). A comparative study of shoreline mapping techniques. *GIS for coastal zone management*, 53-60.

Li, M., Liew, S., & Kwoh, L., (2003). Producing cloud free and cloud-shadow free mosaic from cloudy IKONOS images. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)* (Vol. 6, pp. 3946–3948). IEEE. https://doi.org/10.1109/IGARSS.2003.1295323.

Lin, C. H., Tsai, P.H., Lai, K.H., & Chen, J.Y., (2013). Cloud removal from multitemporal satellite images using information cloning. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1), 232–241. https://doi.org/10.1109/TGRS.2012.2197682.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489.

Lorenzi, L., Melgani, F., & Mercier, G., (2011). Inpainting strategies for reconstruction of missing data in VHR images. *IEEE Geoscience and Remote Sensing Letters*, *8*(5), 914–918. https://doi.org/10.1109/LGRS.2011.2141112.

Maalouf, A., Carre, P., Augereau, B., & Fernandez-Maloigne, C., (2009). A Bandelet-based in-painting technique for clouds removal from remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), 2363–2371. https://doi.org/10.1109/TGRS.2008.2010454.

Machado, C. A. S., Beltrame, A. M. K., Shinohara, E. J., Giannotti, M. A., Durieux, L., Nóbrega, T. M. Q., & Quintanilha, J. A., (2014). Identifying concentrated areas of trip generators from high spatial resolution satellite images using object-based classification techniques. *Applied Geography*, *53*, 271–283. https://doi.org/10.1016/j.apgeog.2014.06.022.

MathWorks, I. (2017). *Regionprops*. Retrieved February 15, 2017, from https://www.mathworks.com/help/images/ref/regionprops.html.

Melgani, F., (2006). Contextual reconstruction of cloud-contaminated multitemporal multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(2), 442–455. https://doi.org/10.1109/TGRS.2005.861929.

Oyedotun, T. D., (2014). Shoreline geometry: DSAS as a tool for historical trend analysis. *British Society for Geomorphology, Geomorphological Techniques*. ISSN, 2047-0371.

Paravolidakis, V., Moirogiorgou, K., Ragia, L., Zervakis, M., & Synolakis, C., (2016). Coastline extraction from aerial images based on edge detection. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *III-8*, 153–158. https://doi.org/10.5194/isprs-annals-III-8-153-2016.

Pardo-Pascual, J. E., Almonacid-Caballer, J., Ruiz, L. A., & Palomar-Vázquez, J., (2012). Automatic extraction of shorelines from Landsat TM and ETM+ multi-temporal images with subpixel precision. *Remote Sensing of Environment*, *123*, 1–11. https://doi.org/10.1016/j.rse.2012.02.024.

Pareeth, S., Bresciani, M., Buzzi, F., Leoni, B., Lepori, F., Ludovisi, A., Morabito, G., Adrian, R., Neteler, M., Salmaso, N., (2017). Warming trends of perialpine lakes from homogenised time series of historical satellite and in-situ data. *Science of The Total Environment*, *578*, 417–426. https://doi.org/10.1016/j.scitotenv.2016.10.199.

Purkis, S. J., Gardiner, R., Johnston, M. W., & Sheppard, C. R. C. (2016). A half-century of coastline change in Diego Garcia – The largest atoll island in the Chagos. *Geomorphology*, *261*, 282–298. https://doi.org/10.1016/j.geomorph.2016.03.010.

Rakwatin, P., Takeuchi, W., & Yasuoka, Y. (2009). Restoration of Aqua MODIS band 6 using histogram matching and Local Least Squares Fitting. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(2), 613–627. https://doi.org/10.1109/TGRS.2008.2003436.

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., & Lindquist, E. (2008). Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sensing of Environment*, *112*(6), 3112–3130. https://doi.org/10.1016/j.rse.2008.03.009.

Song, C., Huang, B., Ke, L., & Richards, K. S. (2014). Remote sensing of alpine lake water environment changes on the Tibetan Plateau and surroundings: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *92*, 26–37. https://doi.org/10.1016/j.isprsjprs.2014.03.001.

Thieler, E. R., Himmelstoss, E. A., Zichichi, J. L., & Ergul, A. (2009). *The Digital Shoreline Analysis System (DSAS) Version 4.0 - An ArcGIS extension for calculating shoreline change*. *Open-File Report*. Reston. Retrieved from http://pubs.er.usgs.gov/publication/ofr20081278.

Trochta, J. T., Mouw, C. B., & Moore, T. S. (2015). Remote sensing of physical cycles in Lake Superior using a spatio-temporal analysis of optical water typologies. *Remote Sensing of Environment*, *171*, 149–161. https://doi.org/10.1016/j.rse.2015.10.008.

Tseng, D.-C., Tseng, H.-T., & Chien, C.-L. (2008). Automatic cloud removal from multi-temporal SPOT images. *Applied Mathematics and Computation*, *205*(2), 584–600. https://doi.org/10.1016/j.amc.2008.05.050.

USGS. (2004). *SLC-off Gap-Filled Products gap-fill algorithm methodology*. Retrieved from http://landsat.usgs.gov/documents/L7SLCGapFilledMethod.pdf.

Van İl Kültür Turizm Müdürlüğü. (2016). Erçek Gölü Kuş Cenneti. Retrieved February 15, 2017, from http://www.vankulturturizm.gov.tr/TR,88261/Erçek-golu-kus-cenneti.html.

Vedaldi, A., & Fulkerson, B. (2010). Vlfeat: an open and portable library of computer vision algorithms. In *Proceedings of the international conference on Multimedia - MM '10* (p. 1469). New York, New York, USA: ACM Press. https://doi.org/10.1145/1873951.1874249.

Vincent, R. K., Qin, X., McKay, R. M., Miner, J., Czajkowski, K., Savino, J., & Bridgeman, T. (2004). Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sensing of Environment*, *89*(3), 381–392. https://doi.org/10.1016/j.rse.2003.10.014.

Yousef, A., & Iftekharuddin, K. (2014). Shoreline extraction from the fusion of LiDAR DEM data and aerial images using mutual information and genetic algorithms. In (*2014) International Joint Conference on Neural Networks (IJCNN)* (pp. 1007–1014). IEEE. https://doi.org/10.1109/IJCNN.2014.6889863.

Zhang, C., Li, W., & Travis, D. J. (2009). Restoration of clouded pixels in multispectral remotely sensed imagery with co-kriging. *International Journal of Remote Sensing*, *30*(9), 2173–2195. https://doi.org/10.1080/01431160802549294.

Zhang, H., Zhang, B., Guo, H., Lu, J., & He, H. (2013). An automatic coastline extraction method based on active contour model. In *2013 21st International Conference on Geoinformatics* (pp. 1–5). IEEE. https://doi.org/10.1109/Geoinformatics.2013.6626130.

Zheng, G., Peng, L., Tao, G., & Wang, C. (2011). Remote sensing analysis of Bohai Bay West Coast shoreline changes. In *Proceedings (2011) IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services* (pp. 549–552). IEEE. https://doi.org/10.1109/ICSDM.2011.5969106.

Zhengke, G., Fun, C., Jin, Y., Xinpeng, L., Fangjun, L., & Jing, Z. (2011). Automatic cloud and cloud shadow removal method for Landsat TM images. In *IEEE (2011) 10th International Conference on Electronic Measurement & Instruments* (pp. 80–84). IEEE. https://doi.org/10.1109/ICEMI.2011.6037860.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 4

THE INFLUENCE OF TWO RADIATION SCHEMES ON THE GLOBAL DAILY SOLAR IRRADIANCE USING A REGIONAL CLIMATE MODEL (REGCM4)

Samy A. Anwar[1,*] and M. M. Abdel Wahab[2,†]
[1]Egyptian Meteorological Authority - EMA, Cairo, Egypt
[2]Meteorology Department, Faculty of Science, Cairo University, Cairo, Egypt

ABSTRACT

In the current study, four experiments are conducted to study the influence of the two radiation schemes (CCM3 and RRTM) on simulating the downward shortwave radiation flux using the regional climate model (RegCM-4.4.1) on daily basis for two domains: the first one is North African area and the second one is Egypt area as the nested domain. Both simulations are evaluated against the reanalysis dataset (NASAPOWER). The results show that there is no distinctive difference between the two schemes against the NASA on daily basis. In addition, the model shows a poor performance for simulating the shortwave radiation flux on daily basis.

Keywords: global solar radiation, regional climate model and daily variability

[*] Corresponding Author's E-mail: ratebsamy@yahoo.com.
[†] Corresponding Author's E-mail: magdy@sci.cu.edu.eg.

1. INTRODUCTION

IPCC, 2013 has stated that annual temperatures are reported to be higher than the predicted ones for the whole Earth. This appears in increasing of heat waves longer and extra ones. In addition, droughts were expected to be more extreme and longer duration.

With a special focus on the Mediterranean areas which face numerous sources of risks such as submerging and erosion. In particular "North Africa," temperature rise lead to a decrease in agricultural areas, growing seasons shortening and also crop yields reduction.

Numerous risks face the African Continent countries such as diversification, so water supply and meals security, environmental safety and conservation are critical solutions for adaptation to worldwide warming effects (Radhouane 2013; IPCC 2013). With a special focus on Egypt as it is one of the most populous international locations in Africa.

The Nile River of Egypt is occupied by population of 82.2 million within an area of approximately 40,000 km^2. The huge regions of the Sahara wasteland are moderately inhabited about half of Egypt's residents live in city regions, with most people unfolding across the densely populated centers of greater Cairo, Alexandria and different important cities in the Nile Delta. To better capture the climatic changes over Africa, regional climate models (RCMs) can be a very useful tool for this purpose. RCMs cover a specific place of the globe at a higher resolution (generally round 50 km) for which situations at its boundary which is provided by the AOGCM to offer more accurate information for temperature and precipitation changes (Jones et al. 2004; Bader et al. 2008).

While GCMs are considered to be a useful tool for long-term predictions, their horizontal resolutions of the GCMs are quite low (in order of 300km). Meanwhile, the RCMs have a higher resolution than the one considered in the GCMs (Rajib et al. 2013). The goal of the current study is to examine the capability of the regional climate model (RegCM - version 4.4.1) for evaluating the overall performance of CCM3 and RRTM schemes for simulating the downward shortwave radiation flux on daily basis as the sun radiation is considered as one of the most important parameters suffering from weather exchange and additionally solar radiation management is also one of the solutions to stand the global warming (Ming et al. 2014; Robock 2012).

The RegCM has passed through many phases of developments (RegCM1, Dickinson et al. 1989 and Giorgi, 1990), (RegCM2, Giorgi et al. 1993b, Giorgi et al. 1993c), overdue 1990s (RegCM2.5, Giorgi & Mearns, 1999) and 2000s (RegCM3, Pal S. et al., 2000) and has already been used for studying various areas in the global (as example: Giorgi and Mearns, 1999; Giorgi et al. 2004b).

2. STUDY AREA, DATA AND METHODS

2.1. Study Area

1. Africa area (coarse area - 60 km resolution) extends from 10 W to 45 E and from zero – 45 N.
2. Egypt area (nested domain - 20 km resolution) extends from 22 E to 36 E and from 22 – 36 N.

2.2. Data

2.2.1. Input Data for RegCM4 Model Run

Sea surface temperature (*SST*) is derived from the OISST (Optimally Interpolated SST) datasets and initial condition and boundary condition (*ICBC*) are derived from the ERA-Interim reanalysis 1.5° x 1.5° from ECMWF (EIN15).

The input data for RegCM4 runs was obtained from: http://clima-dods. ictp.it/data/d8/cordex.

2.2.2. Data for RegCM4 Output Verification

1. NASA Prediction of Worldwide Energy Resource (POWER) for verification of daily downward shortwave radiation flux ($MJ/m^2/day$) in agrometeorology station network in Egypt. Those data cover the globe on one degree on regular grid. The dataset were provided through the website: http://power. larc.nasa.gov/cgi-bin/cgiwrap/solar/agro.cgi?email=agroclim@larc.nasa.gov).
2. Radiation Station data for 7 stations in Egypt. These data were retrieved from http://wrdc-mgo.nrel.gov/

2.3. Experiment Design

Table 1 indicates the configuration for the two radiation schemes for the two areas, the coarse one and the nested one. For this purpose, four experiments were conducted, each of 1 year, starts from January 2008 till March 2009. The first two months were considered as a spin up for the model initiation. The center of the coarse domain is latitude 25° and longitude 10°, and for the nested domain latitude 28° and longitude 28°. The default radiation scheme is derived from NCAR CCM3 (Kiehl et al. 1996). This scheme follows the d-Eddington approximation. It consists of 18 spectral periods from

0.2 to 5 μm. The absorption parameterization and cloud scattering follow that of (Slingo 1989), where in the optical properties of the cloud droplets (extinction optical intensity, unmarried scattering albedo, and asymmetry parameter) are expressed in terms of the liquid water content of the cloud and an effective droplet radius.

Table 1. RegCM4.4.1 model configuration

Settings	**Coarse Domain**		**Nested Domain**	
	CCM3	RRTM	CCM3	RRTM
Number of points in the N/S direction	100	100	90	90
Number of points in the E/W direction	100	100	90	90
Number of vertical levels	18	18	18	18
Domain cartographic projection	ROTMER			
Grid point horizontal resolution in km	60	60	20	20
Central latitude of model domain in degrees	25	25	28	28
Central longitude of model domain in degrees	10	10	28	28
Frequency in hours	6	6	6	6
Type of Sea Surface Temperature used	OI_WK			
Type of global analysis datasets used	EIN15			
Start date for ICBC data generation	2008010100			
End data for ICBC data generation	2009030100			
Boundary layer scheme	HOLTSLAG			
Radiation scheme	0	1	0	1
land surface model	BATs model			

RRTM is a new radiation scheme inserted recently in the RegCM4 model. It is a rapid, accurate, correlated-okay, radiative switch model (Mlawer et al. 1997) developed for the Atmospheric Radiation size (ARM) application to deal with the ARM goal of improving radiation model in worldwide weather models. a number of the essential capabilities of RRTM are its connection to radiation measurements via contrast to the drastically demonstrated (LBLRTM) line-by way of-line radiative switch model (as in Clough et al. 1992; Clough and Iacono 1995), and its use of the stepped foreward and established CKD water vapor continuum version.

Four experiments have been performed, the first two used a 60 km resolution and the other two used a resolution 20 km.

The physical configuration of the conducted experiments is illustrated in Table 1. The RegCM4 model uses the BATs (Biosphere Atmosphere Transfer Systems) land surface model, which manipulates the land-atmosphere interaction using simple parameterization (Deardoff, 1978; Dickinson et al. 1993). The boundary layer scheme was used in those experiments is the *Holtslag* scheme (Holtslag et al. 1990), which is based totally on the nonlocal diffusion concept that takes under consideration counter-gradient fluxes.

3. RESULTS AND DISCUSSION

3.1. Verification with Radiation Station Network in Egypt

The downward shortwave radiation from the RegCM4 model is evaluated against NASA dataset for twelve meteorology stations located in the Egypt domain (Marsa-Matruh, Alexandria, Port-stated, El-Arish, Ismailia, Cairo, Asyout, Luxor, Aswan, Siwa, Dakhla and Kharga).

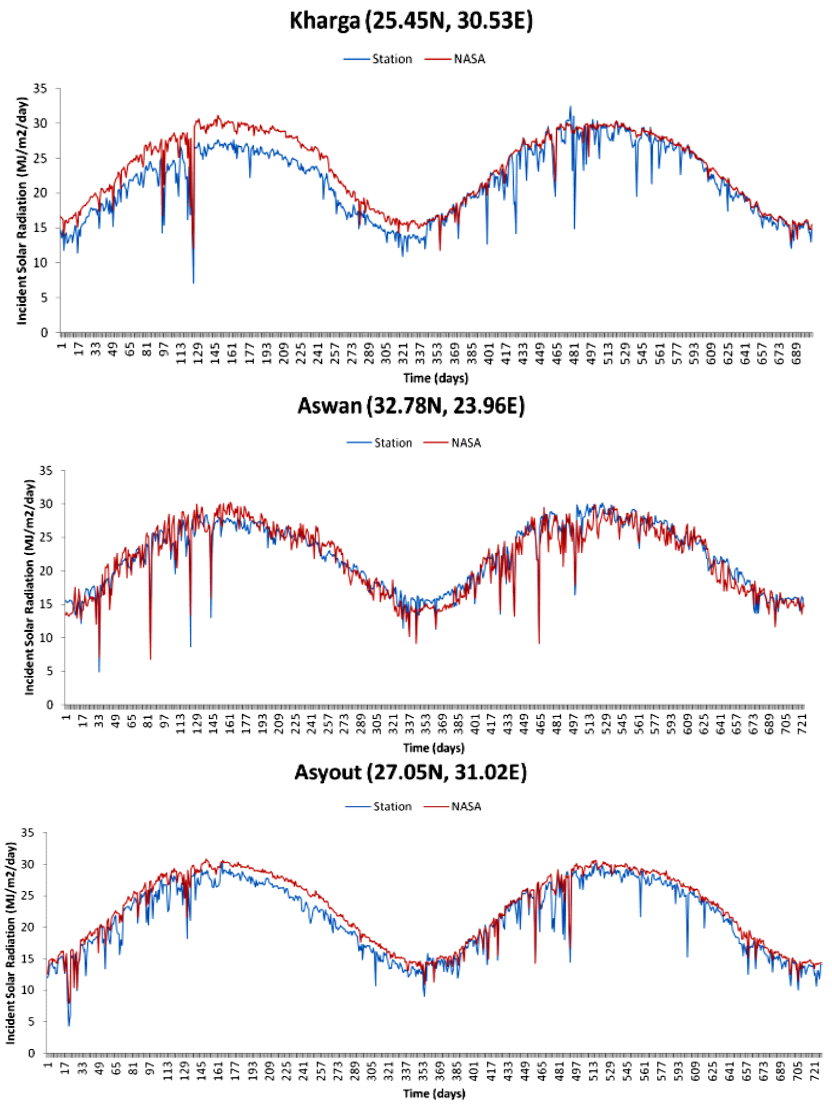


Figure 1. (Continued).

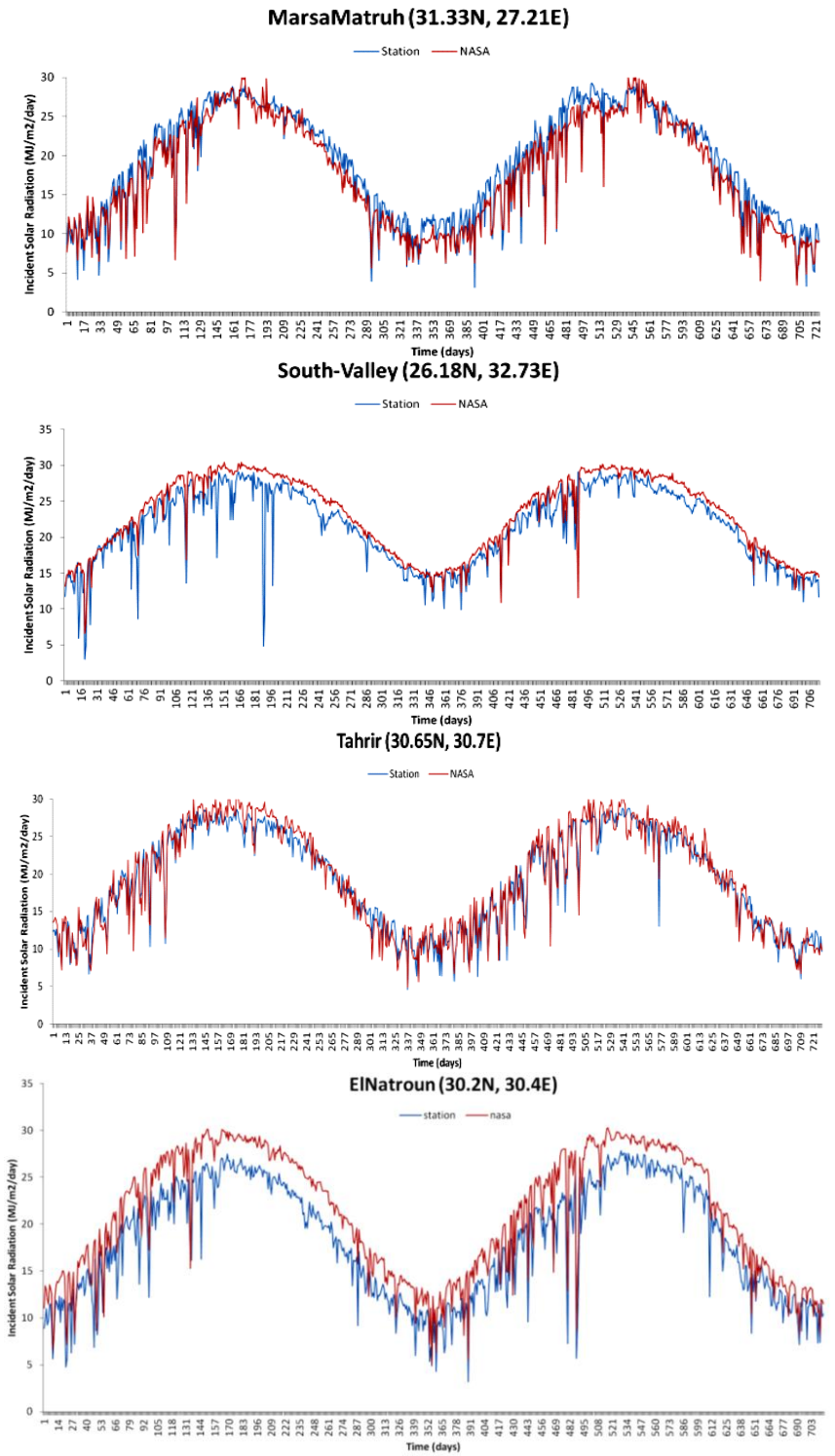


Figure 1. Evaluation of NASA POWER incident solar radiation against observations.

Before using the NASA dataset for model evaluation purpose, the NASA POWER is evaluated against 7 meteorological stations (Asyout, Tahrir, Kharga, Wadi-ElNatroun, Marsa-Matruh, Aswan and South-Valley), which are retrieved from global Radiation Centre (WRDC) (maintained for the arena Meteorological agency by way of the Russian Federal service for Hydrometeorology and Environmental tracking,St. Petersburg, Russia).

All stations except for Tahrir and Aswan, the validation period is 2008-2009 due to data availability. For Aswan and Tahrir, the time period is 1995-1996, for the same purpose, for Kharga from 2009-2010, Marsa-matruh from 1996-1997. Table 2 shows the available stations and number of days for the available observation.

Figure 1 shows the daily variation of the incident short wave radiation of NASA POWER against the observation for each station. For majority of stations, the NASA capture properly the daily variation relative to the observation. Hence NASA POWER can be used as the observation source for evaluating the RegCM4 output.

Table 2. Verification of NASA power data with station for down ward short wave radiation

Station	**No of actual days**	**Station**	**No of actual days**
Kharga	704	Aswan	724
Asyout	729	South-Valley	717
Tahrir	730	Wadi-elnatroun	715
Marsa-Matruh	725		

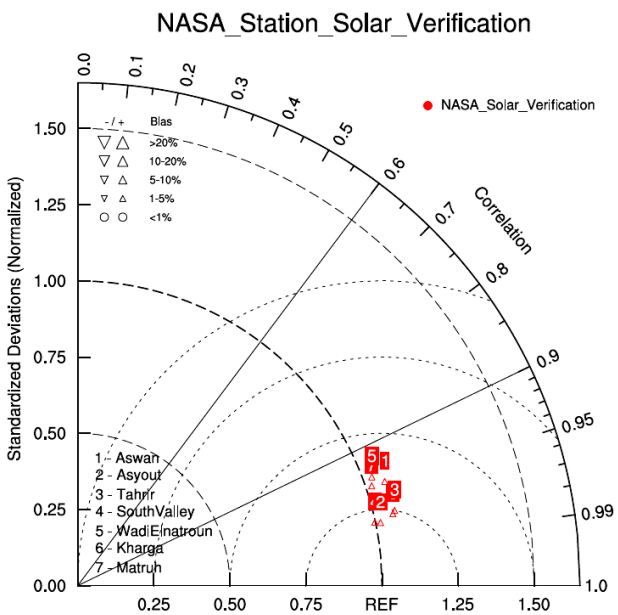


Figure 2. Verification of NASA power shortwave.

Figure 2 shows a Taylor diagram (indicates statistical measures of relative standard deviation, correlation and root imply rectangular errors (RMSE)) for evaluating the downward shortwave radiation of NASA POWER in comparison with observation. The NASA POWER (in the parent domain case) shows a high correlation in comparison with the insitu observations.

Meanwhile, for Aswan, Tahrir and Kharga stations - the NASA POWER shows a remarkable overestimation in comparison with the observations.

3.2. Verification of RegCM4 Output with NASA Power

Regarding the two radiation schemes, the incident short wave radiation from the RegCM4 model is evaluated against the NASA dataset. Figure 3 shows a quantitative evaluation of the RegCM4 model against the NASA using Taylor diagram.

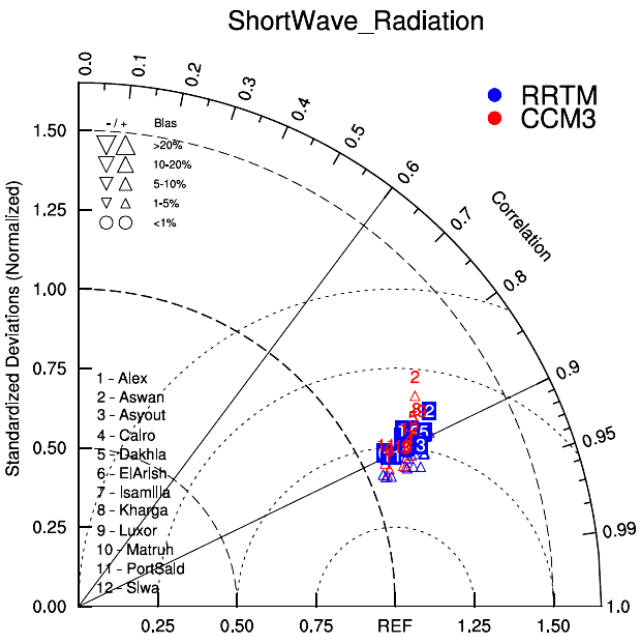


Figure 3. Verification of RegCM4 shortwave radiation.

From this figure, it can be seen that - for all stations - the RegCM4 model shows a remarkable overestimation in comparison with the NASA and a high correlation (around eighty five %). besides for Aswan (the simulated solar radiation using the CCM3 scheme has lower correlation than in the RRTM scheme), there is no big difference among the two schemes on daily basis. Despite of the high correlation among RegCM4 and NASA dataset (R = 0.72 for Aswan and R = 0.85 for Asyut), each schemes shows a poor

performance for simulating the downward short wave on daily basis as shown in Figure 4.

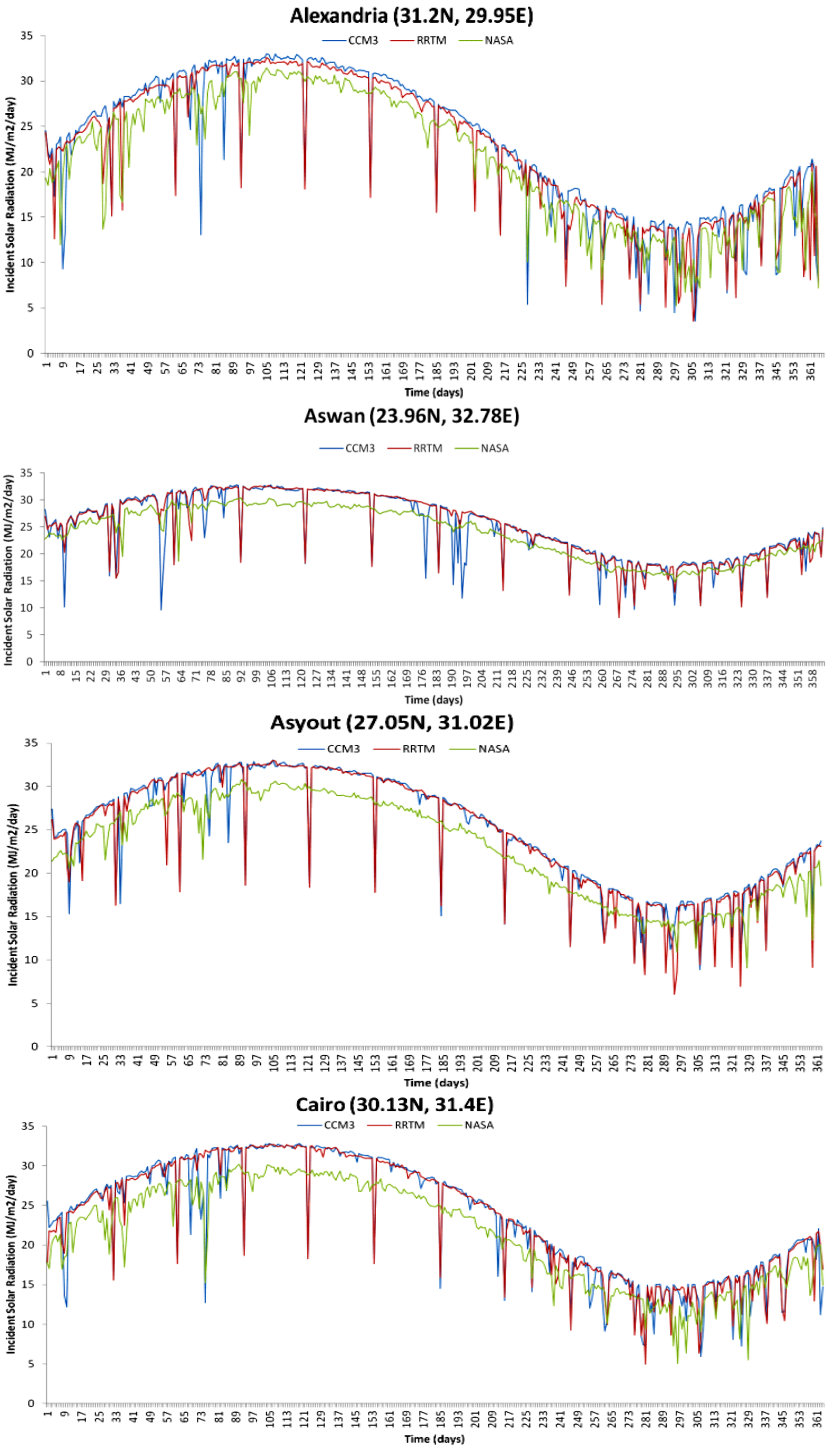


Figure 4. (Continued).

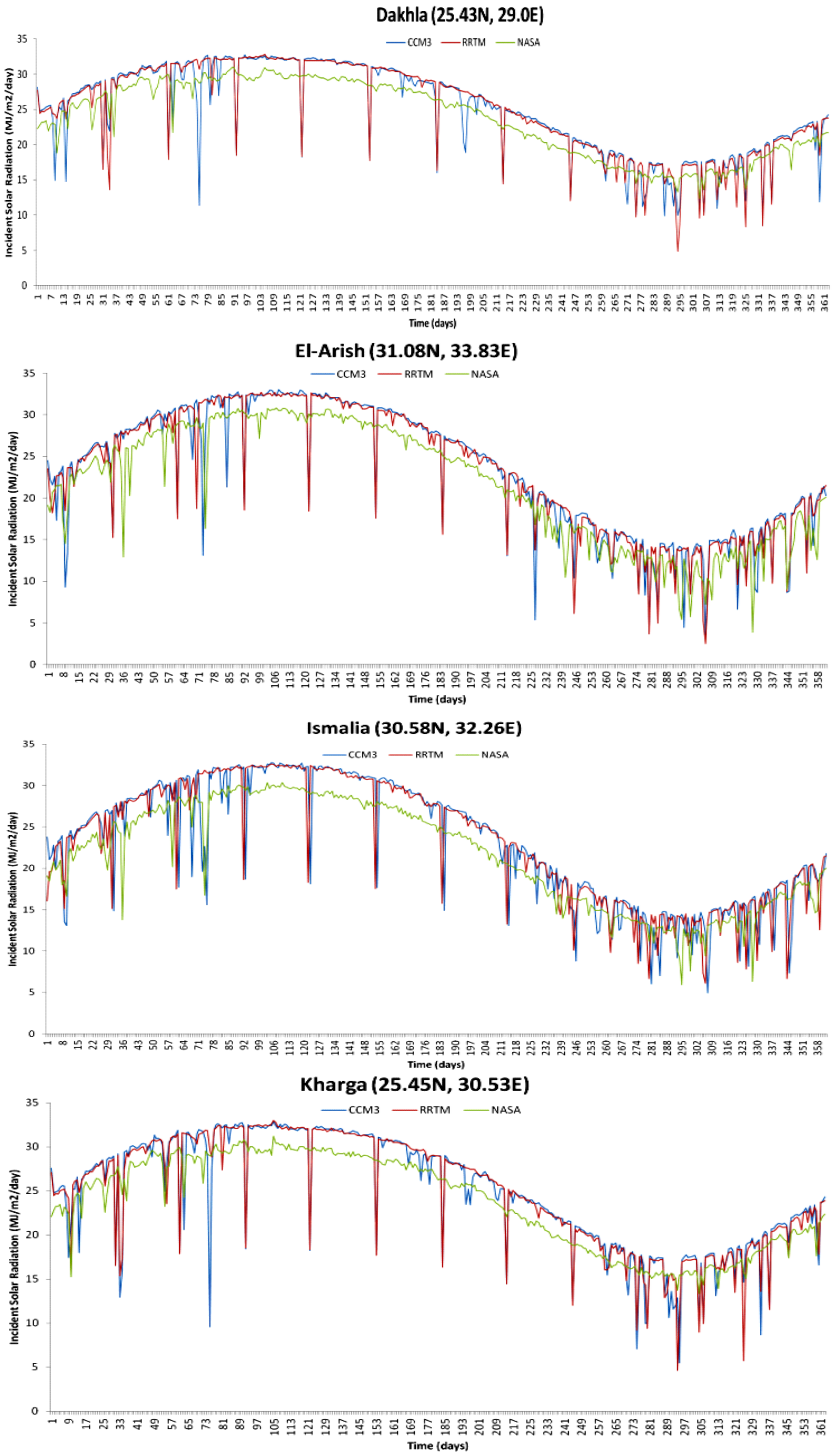


Figure 4. (Continued).

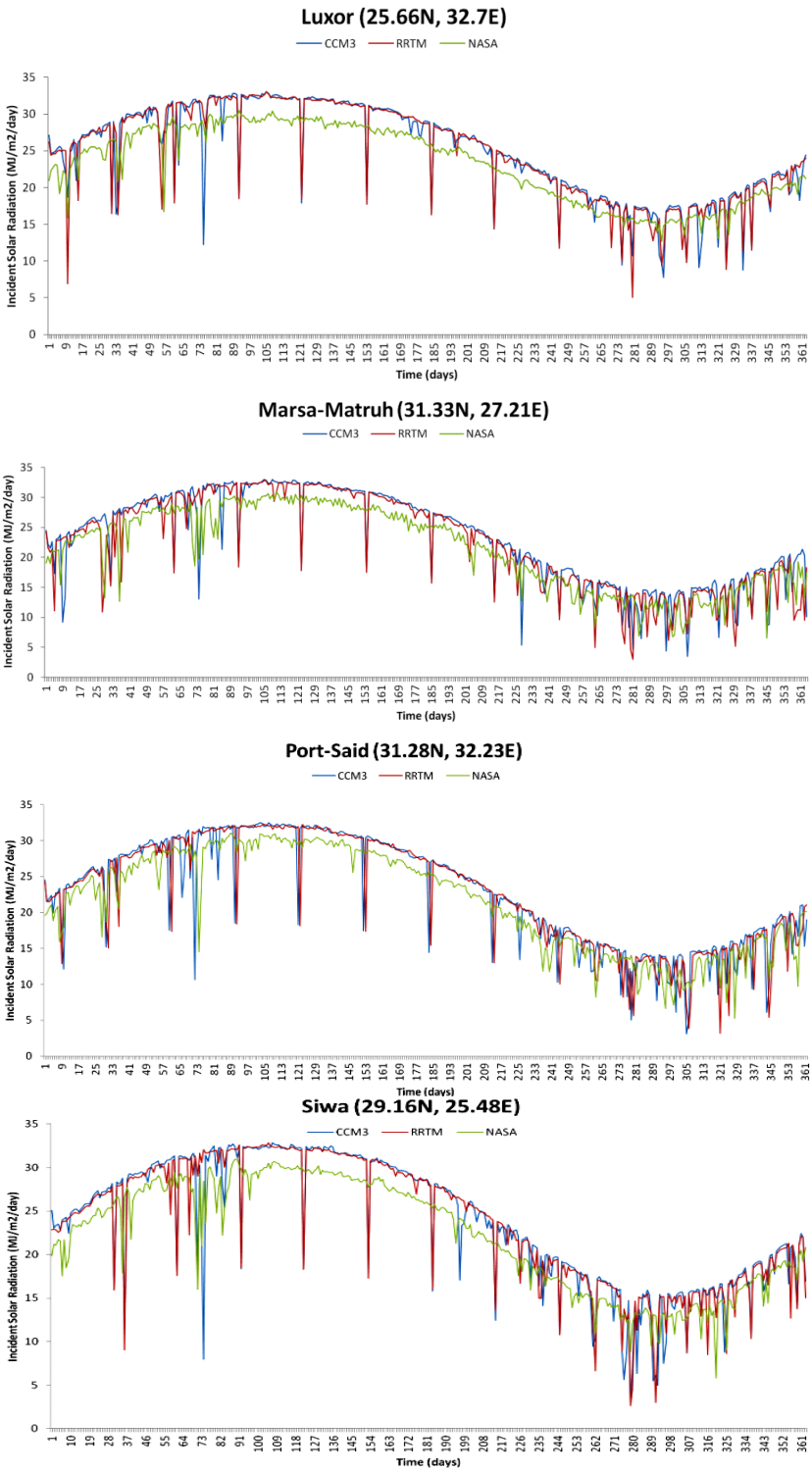


Figure 4. Daily variation of incident solar radiation from RegCM4 against NASA dataset.

This bad performance of the RegCM4 model is also observed in earlier version of RegCM model (Ohunakin et al. 2015).

CONCLUSION

The regional climate model (RegCM-4.4.1) is used for evaluating the global incident sun radiation over both Africa Continent as the coarse domain and Egypt as the nested area.

RegCM4 confirmed shows a good ability for simulating the downward shortwave radiation in comparison with the NASA power dataset. But, it showed poor performance for simulating the downward short wave with either CCM3 or RRTM radiation scheme. Additionally there is distinctive difference between the 2 schemes on daily basis.

ACKNOWLEDGMENTS

NASA Power dataset is acknowledged for providing the dataset from the website http:// power.larc.nasa.gov/ cgi-bin/ cgiwrap/ solar/ agro.cgi?email =agroclim@larc.nasa. gov. We are very grateful for global Radiation Centre (WRDC) for providing the maintaining the quality of in-situ observation for global solar radiation over Egypt.

REFERENCES

Bader D. C., Covey C., Gutowski W. J., Held I. M., Kunkel K. E., Miller R. L., Tokmakian R. T., Zhang M. H. (2008) Climate Models: An Assessment of Strengths and Limitations; *A Report by the U. S. Climate Change Science Program and the Subcommittee on Global Change Research*; Department of Energy, Office of Biological and Environmental Research: Washington, DC, USA.

Clough S. A., Kneizys F. X., Davies R. W. (1989) Line shape and the water vapor continuum. *Atmos. Res*., 23, 229-241.

Clough S. A., Iacono M. J., Moncet J. L. (1992) Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapour. *J. Geophys. Res.,* 97, 15,761-15,785.

Clough S. A., Iacono M. J. (1995) Line-by-line calculation of atmospheric fluxes and cooling rates, 2, Application to carbon dioxide, ozone, methane, nitrous oxide, and the halocarbons. *J. Geophys. Res*., 100, 16,519-16,535.

Deardoff J. W. (1978) Efficient prediction of ground surface temperature and moisture with inclusion of a layer of vegetation. *J. Geophys. Res*., 83, 1889-1903.

Dickinson R. E., Errico R. M., Giorgi F., Bates G. T. (1989) A regional climate model for the western United States. *Climatic Change*, 15, 383-422.

Giorgi F., Marinucci M. R., Bates G. T. (1993b) Development of a second generation regional climate model (regcm2) i: Boundary layer and radiative transfer processes. *Mon. Wea. Rev*., 121, 2794-2813.

Giorgi F., Marinucci M. R., Bates G. T., DeCanio G. (1993c) Development of a second generation regional climate model (regcm2) ii: Convective processes and assimilation of lateral boundary conditions. *Mon. Wea. Rev*., 121, 2814-2832.

Giorgi F., Mearns L. O. (1999) Introduction to special section: Regional climate modelling revisited. *J. Geophys. Res*., 104, 6335-6352.

Holtslag A. A. M., de Bruijn E. I. F., Pan H. L. (1990) A high resolution air mass transformation model for short-range weather forecasting. *Mon. Wea. Rev.,* 118, 1561-1575.

IPCC (2013) *Climate Change 2013: Synthesis Report*. IPCC 2013. Technical report, IPCC.

Jones R. G., Noguer M., Hassell D. C., Hudson D., Wilson S. S., Jenkins G. J., Mitchell J. F. B. (2004) Generating High Resolution Climate Change Scenarios Using PRECIS; Met Office Hadley Centre: Exeter, UK.

Kiehl J. T., Hack J. J., Bonan G. B., Boville B. A., Breigleb B. P., Williamson D., Rasch P. (1996), Description of the ncar community climate model (ccm^3), *Tech. Rep. NCAR/TN-420+STR*, National Center for Atmospheric Research.

Ming T., Richter R., Liu W., Caillol S. (2014) Fighting global warming by climate engineering: Is the Earth radiation management and the solar radiation management any option for fighting climate change? *Renewable and Sustainable Energy Reviews* 31, 792-834.

Mlawer E. J., Taubman S. J., Brown P. D., Iacono M. J., Clough S. A. (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res*., 102, 16,663-16,682.

Ohunakin O. S., Adaramola M. S., Oyewola O. K., Fagbenle R. O. (2015) Solar radiation variability in Nigeria based on multiyear RegCM3 simulations. *Renewable Energy* 74, 195-207.

Pal J. S., Small E. E., Eltahir E. A. B. (2000) Simulation of regional-scale water and energy budgets: Representation of subgrid cloud and precipitation processes within RegCM. *J. Geophys. Res.-Atmospheres*, 105(D24), 29,579-29,594.

Pal J. S., Giorgi F., BiX et al. (2007) The ICTP RegCM3 and RegCNET: Regional climate modeling for the developing world. *Bull. Amer. Meteor. Soc*., 88, 1395-1409.

Radhouane L. (2013) Climate change impacts on North African countries and on some Tunisian economic sectors. *Journal of Agriculture and Environment for International Development – JAEID*, 107 (1): 101-113.

Rajib M. A. et al. (2013) A Comprehensive Modelling Study on Regional Climate Model (RCM) Application - Regional Warming Projections in Monthly Resolutions under IPCC A1B Scenario. *Atmosphere* 2012, 3, 557-572; doi:10.3390/atmos3040557.

Robock A. (2012) Will Geoengineering With Solar Radiation Management Ever Be Used? *Ethics, Policy and Environment* Vol. 15, No. 2, June 2012, 202-205.

PART II. WATER MANAGEMENT

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 5

NONLINEAR TIME SERIES ANALYSIS OF ENVIRONMENT POLLUTANTS

R. Bhardwaj*

University School of Basic and Applied Sciences,
Non-Linear Dynamics Research Lab, Department of Mathematics,
Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India

ABSTRACT

Environmental pollution had been a reality of life for many centuries but it became a *real problem* since the start of the *industrial revolution*. Environmental pollution consists of five basic types of pollution, namely, air, water, soil, noise and light. Pollution is out and the foremost harmful kind of pollution within the surroundings mixed. Pollution is caused by the injurious smoke emitted by cars, buses, trucks, trains, and factories, specifically pollutant, oxide and N oxides. Pollution caused by industrial waste mixed into lakes, rivers, and different water bodies, has created marine life endangered Humans grime water with huge scale
disposal of garbage, flowers, ashes and another household waste. In many rural areas one can still notice people bathing and cooking at intervals in identical water of river, making it incredibly filthy.

In depth review of the assorted nonlinear techniques that are obtainable for analyzing the statistic knowledge of environmental waste matter systems is mentioned. The data gained from the investigation into the statistics analysis of advanced systems is applied to real knowledge from air and water pollutants. Within the study, daily average quantity of pollution parameters CO, NO, NO2, is employed from four observation station monitored in Delhi, India. The most stress is given on cross rippling remodel and rippling coherence for examining relationships in time frequency between twofold series of same air pollutants recorded at completely different locations. Point in time statistics is

* Corresponding Author's E-mail: rashmib@ipu.ac.in, rashmib22@gmail.com.

additionally studied to urge a lot of insight in causative relationships between the statistics. Also, for pollution, Yamuna stream water quality parameters hydrogen ion concentration (pH), COD, BOD, AMM (free ammonia), TKN (total kjeldahl nitrogen), DO (dissolved oxygen), WT (water temperature), TC (total coliform) and FC (fecal coliform) monitored at TajMahal, Agra D/S in Asian country are shapely victimization of trend and statistics analysis. Statistics and analytic thinking of water quality standards are calculable victimization and calculated by regression lines, parametric statistic, integrated moving average (ARIMA), root mean square error (RMSE), mean absolute proportion error (MAPE), mean absolute error (MAE), normalized theorem data criterion, Box- Jenkins, lag, Hurst exponent, similarity index, Prognosticative model is helpful at ninety five confidence limits.

All water parameters exhibit platykurtic curve apart from TC & FC. Molecular randomness exists for hydrogen ion concentration, DO, WT. The parameters AMM, TKN, DO, TC follows the positive trend proportion and hydrogen ion concentration, COD, BOD, WT, FC follows the negative trend proportion. All parameters crosses the prescribed limits of World Health Organization (WHO)/ US Environmental Protection Agency (EPA) and therefore it's terminated that water isn't appropriate for drinking, agriculture and industrial use.

Keywords: Hurst exponent, Brownian randomness dimension, time series, ARIMA, trend analysis, wavelet coherence, multivariate analysis

INTRODUCTION

One in all the dangerous issues that the planet is facing these days is that of environmental pollution, increasing with each passing year and inflicting grave and irreparable harm to the world. Pollution is arising too apace due to growing cities, waste production, solid waste, energy consumption, traffic and lack of surroundings rules. The most pollutants touching urban areas are stuff PM, O_3, NO_2, CO_2, SO_2. Increase in level of air pollutants higher than permissible level has effects on health of individuals inflicting respiratory issues, dizziness, having a negative impact on surroundings which ends into pathology and heart attacks. Water could be a supply of lifetime of each organism on the world and stream water is that the most vital resource for drinking water. Yamuna stream is that the largest tributary stream of the Ganga in northern Asian country. It originates from the Yamunotri ice mass at a height of 6387 meters on south western slopes of Banderpooch peaks (380 59' N 780 27' E) within the lower Himalaya Mountains in Uttarakhand. It encompasses a mixture of system of 366233 km^2 before merging with Ganga at Allahabad i.e., a complete of 42% of complete Ganga basin (CPCB, 2006). Sample location Agra D/S at West Burzi of Taj Mahal monument is concerning 310 metric linear unit downstream from Wazirabad barrage and depicts the impact of sewerage and Industrial water discharge from Delhi, Agra and Mathura.

Pollution in stream water is central to several of the environmental problems that concern society, and are studied by a large vary of specialists, engineers, hydrologists,

ecologists and geomorphologists. The various methodologies and techniques to predict pollution, foretelling with the impact of pollutants on health and therefore the dynamics of environmental and health assessment are studied by Bhardwaj and coauthors (2007, 2012, 2013, 2014, 2015, 2016). Associate degree industrial, domestic and sewerage discharge mixes with stream water and have an effect on the standard of water (river and ground). Analysis and WHO/EPA water quality standards (WHO (1971)) are used for shrewd water quality indices. Hurst exponent for long vary dependent processes rippling techniques has been calculable that provides the straight line linear relationship of the premise of constructing associate degree (Park and Park, 2009; Bhardwaj and Paramr, 2013). Fractal, Hurst exponent, Assurance index, Statistics, regression and applied math analysis are the vital tools for prediction of trend of stream water quality (Doyle and Barros, 2011; Bhardwaj and Parmar, 2013; Lupikasza, 2010). Statistic are often modelled by a theoretical account possessing long way correlation (Rangarajan and Sant, 2004, Korashey, R., 2009). Many approaches are taken to outline the connection between non-stationary signals within the time-frequency domain. (Rehman & Siddiqi 2009: Can, Aslan, Siddiqi 2005). Cross rippling remodel, Cross rippling Coherence and therefore the part distinction are used to study between air pollutants statistic recorded at completely different locations. The native prediction methodology is employed to predict O_3 concentration over Constantinople town at completely different stations and single-variable statistic knowledge to reconstruct the attractor. The dynamic model attractor is calculable and therefore the amendment in flight is foreseen by a polynomial approximation. (Koçak, Şaylan and Şen 2000).

CO has been recorded at every of the various locations at Delhi faculty of Engineering (DCE) (Industrial Area), ITO-Crossing (ITO) (Commercial Area), Siri fort (SF) (Residential Area) and by mobile Van (MV) in Delhi. Daily averaged of the above-named air pollutants is collected for an amount of 4 years from each location. Data, monitored by mobile van throughout the town, are often treated as average pollution of a mixture location. The analysis of quality of Yamuna stream water and Agra is vital to check as stream water is that the main supply of drink and conjointly individuals depend upon stream water for his or her daily usage. Quality of Yamuna stream water is that the main concern for the study because it affects immeasurable individuals across Asian country. The lag variation, residual ACF, residual PACF, multivariate analysis and shape analysis are mentioned for stream water quality parameters. The future (1999-2009) analytic thinking of every water quality parameter is additionally mentioned at Agra D/S, Yamuna stream (India). Stream map of Asian country, India is shown in Fig.1.

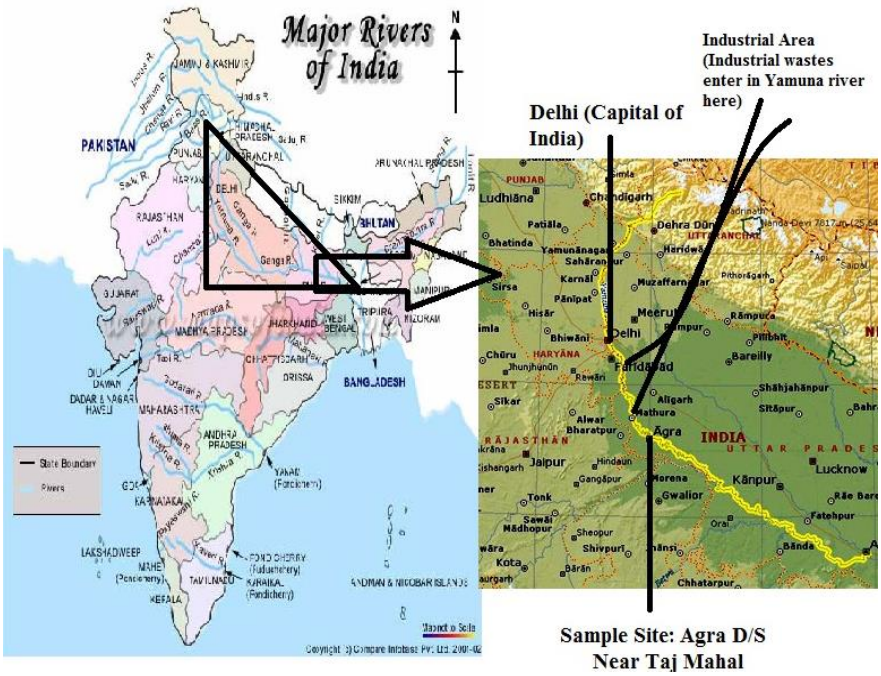


Figure 1. River Yamuna route map description of sample site.

DATA AND METHODOLOGY

Monthly average of water quality parameters pH, COD, BOD, AMM, TKN, DO, WT, TC and FC monitored by CPCB at Taj Mahal, Agra D/S of Yamuna is studied.

Hurst Exponent (H)

Hurst exponent quantifies comparative trend of time series either-to-regres-strongly-to-mean-or accumulate in a particular direction. H value equals to 0.5 indicates random-motion. In a random walk, there is no correlation between any element and a future-element. When 0.5<H<1, then persistent-behaviour exists and when 0<H<0.5-anti-persistent-behaviour persists.

Fractal Dimension (D)

Fractal dimension is the statistical magnitude which provides approximation of in which form the fractal seems to fill the space-plane, as one pans down towards refined scales. Fractal dimension simulated as:

$$D = 2 - H \tag{1}$$

Predictability Index (PI)

It refers towards behaviour of time-domain of the situation considered.

$$PI = 2\left|D - 1.5\right| \tag{2}$$

The value of 1.5 specifies Brownian motion.

Auto Correlation Functions (ACF) and Partial Auto Correlation Functions (PACF)

Prototypes that have been known through trends in their ACFs & PACFs. The autocorrelations & partial-auto-correlations have been calculated for sequential lags in the successions. First-lag shows auto-correlation in Y_{t-1}, Y_t, second-lag is the auto-correlation and-partial-autocorrelation between Y_{t-2}, Y_t similarly other lags (Fischer and Griffith, 2008). Autocorrelation can be referred to:

$$r_k = \frac{\frac{1}{N-k}\sum_{t=1}^{N-k}\left(Y_t - \overline{Y}\right)\left(Y_{t-k} - \overline{Y}\right)}{\frac{1}{N-1}\sum_{t=1}^{N}\left(Y_t - \overline{Y}\right)^2} \tag{3}$$

where N is number of observations in a whole series, k - lag. $\overline{Y}$ - mean of whole series. At lag 1, there are no previous autocorrelations, so r_0^2 is set to be 0. Standard error of an autocorrelation depends on squared autocorrelations of previous lags and referred as

$$\mathrm{SE}_{r_k} = \sqrt{\frac{1 + 2\sum_{l=0}^{k-1} r_l^2}{N}} \tag{4}$$

Standard error for a partial autocorrelation is same at all lags

$$\mathrm{SE}_{pr} = \frac{1}{\sqrt{N}} \tag{5}$$

The relation between ACF and PACF for first three lags is given as (McCleary and Hay, 1980):

$$\mathrm{PACF}(1) = \mathrm{ACF}(1)$$

$$\mathrm{PACF}(2) = \frac{\mathrm{ACF}(2) - \left(\mathrm{ACF}(1)\right)^2}{1 - \left[\mathrm{ACF}(1)\right]^2}$$

$$\mathrm{PACF}(3) = \frac{-2\left(\mathrm{ACF}(1)\right)\mathrm{ACF}(2) - \left[\mathrm{ACF}(1)\right]^2 \mathrm{ACF}(3)}{1 + 2\left[\mathrm{ACF}(1)\right]^2 \mathrm{ACF}(2) - \left[\mathrm{ACF}(2)\right]^2 - 2\left[\mathrm{ACF}(1)\right]^2} \tag{6}$$

Impact of ACF and partial autocorrelations can be evaluated with the aid of SEs. Boundary lines nearby functions - 95% confidence bounds. For-detecting-seasonality, auto-correlation function (ACF) is plotted. View for residual against a specified-time-domain is referred as lagged-auto-correlation-function. Lag – k : the difference of an event occurring at time t+k (k > 0) to another occurrence at time t. Partial autocorrelation function (PACF) is also detects trend and seasonality. PACF - amount of correlation between variable and its lag that cannot be described by correlations at all the lower-order-lags.

Trend Percentage

The percentage (%) variation of parameters is defined as

$$x\% = \frac{aN}{\bar{x}} \times 100 \tag{7}$$

where x - variable parameter; a - slope for regression line; N is number of days, months or years considered as a study-period. Statistically substantial variation at 95% confidence interval in every case.

The Morlet Wavelet

Wavelet-coefficients,- $W_X(s,\tau)$ contain joint info for both:-function $x(t)$ and analyzing-wavelet. Info of both amplitude and phase is vital for study to be performed, thus, complex-wavelet gets chosen, as yield is complex transform. Morlet-wavelet is defined as:

$$\psi_n(t) = \pi^{-\frac{1}{4}} \left(e^{i\eta t} - e^{-\frac{\eta^2}{2}} \right) e^{-\frac{t^2}{2}} \tag{8}$$

The term $e^{-\frac{\eta^2}{2}}$ being introduced to guarantee the fulfillment of the admissibility condition for $\eta \geq 5$ this term becomes negligible.

Cross Wavelet Transform and the Phase Difference

Cross wavelet transform of two-time series, $x = \{x_n\}$ and $y = \{y_n\}$, is simply defined as $W_n^{xy} = W_n^x W_n^{y^*}$ where W_n^x and W_n^y are the wavelet transform of x and y, respectively. The cross-wavelet power is given by $\left|W_n^{xy}\right|$. Wavelet-power-spectrum depicts variance of time successions, alongwith that of large variance showing large power for cross-wavelet power of two-time-series shows covariance among the time-series at every frequency. Thus, cross-wavelet-power provides computed indication for comparison of power between these-time-series.

Cross-Wavelet Coherence

It is demarcated as ratio of the cross-spectrum to the product of the spectrum of each series, and can be referred to as the local correlation between two CWTs. Wavelet coherence between two-time series $x = \{x_n\}$ and $y = \{y_n\}$ as follows:

$$R_n^2(s) = \frac{\left|S\left(s^{-1} W_n^{xy}(s)\right)\right|}{S\left(s^{-1}\left|W_n^x\right|\right)^{\frac{1}{2}} S\left(s^{-1}\left|W_n^y\right|\right)^{\frac{1}{2}}}, \tag{9}$$

where S denotes a smoothing operator in both time and scale.

$$S(W) = S_{scale}\left(S_{time}\left(W_n(s)\right)\right) \tag{10}$$

where S_{scale} denotes smoothing along the wavelet scale axis and S_{time} smoothing in time. The time convolution is performed with a Gaussian $e^{-t^2/2s^2}$, which is the absolute value of the wavelet function in each scale. The time convolution will double the edge artefact to $2 \cdot s\sqrt{2}$. The scale convolution is performed by a rectangular window with a length of 0.6 The factor 0.6 is the empirically determined scale decorrelation length for the Morlet wavelet. It is natural to design the smoothing operator so that it has a similar footprint as the wavelet used. A suitable smoothing operator for the Morlet wavelet as

$$S_{time}(W)\big|_s = \left(W_n(s) * c_1^{\frac{-t^2}{2s^2}} \right)\Bigg|_s \tag{11}$$

$$S_{time}(W)\big|_s = \left(W_n(s) * c_2 \prod(0.6s)\right)\big|_n, \tag{12}$$

having c_1, c_2 - normalization constants and $\prod$ - rectangle function. Smoothening is a necessary step, as in absence of this step, coherence is identically one at all the scales and times. In Fourier analysis, this problem is overcome by smoothening cross-spectrum before standardizing.

Phase Difference

It is essential to characterize phase relations between two series which are, $x = \{x_n\}$ and $y = \{y_n\}$. The phase of the assumed time-series, ϕ_X, is the position into pseudo-cycle of this series. Phase difference, $\phi_{x,y}$, characterizes phase relationships between the two-series. It is defined as

$$\phi_{x,y} = \tan^{-1}\left(\frac{\Im\{W_n^{xy}\}}{\Re\{W_n^{xy}\}} \right), \tag{13}$$

with $\phi_{x,y} \in [-\pi, \pi]$. Phase difference of zero directs that time-periods move together (analogous to positive covariance). When $\phi_{x,y} \in (0, \pi)$ series move in-phase, with the time series y leading x when $\phi_{x,y} \in (-\pi, 0)$ whereas, if $\phi_{x,y} \in (-\frac{\pi}{2}, 0)$ then it is x which is primary. An anti-phase relation (analogous to negative covariance) exists when $\phi_{x,y} \in \left(\frac{\pi}{2}, \pi\right] \bigcup \left(-\pi, \frac{\pi}{2}\right]$. If $\phi_{x,y} \in \left(\frac{\pi}{2}, \pi\right)$ then x is leading. Time-series y is leading if $\phi_{x,y} \in \left(-\pi, -\frac{\pi}{2}\right)$.

RESULTS AND DISCUSSION

Water Pollution

Table-1 encapsulates correlation between various water quality factors monitored at Taj Mahal, Agra D/S sample site of Yamuna River. Fig.2 shows the statistical analysis of

water-quality factors. Table-2 provides lags of ACF & PACF alongwith standard errors. The plot of residual-ACF,-residual-PACF together with lags for each factor are given in Fig.3. Table-3 tabulates outcomes:-regression-equations,-Hurst-exponent,-fractal dimension,-predictability-index and-trend-percentage. Table-4 depicts time-series analysis with values of R^2, RMSE-(Root-Mean-Square-Error), MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), Normalized BIC (Normalized Bayesian Information Criterion) etc.

Table 1. Correlation between water quality parameters at Agra D/S

	pH	**COD**	**BOD**	**AMM**	**TKN**	**DO**	**WT**	**TC**	**FC**
Ph	1	0.1784	-0.0167	-0.1565	-0.1641	0.3288	0.2415	0.1774	0.1201
COD	0.1784	1	0.8499	0.2375	0.1974	0.0739	0.2009	0.4315	0.4502
BOD	-0.0167	0.8499	1	0.2917	0.2479	-0.0121	0.0307	0.3384	0.3691
AMM	-0.1565	0.2375	0.2917	1	0.9272	-0.1885	-0.4128	-0.04693	-0.0586
TKN	-0.1641	0.1974	0.2479	0.9272	1	-0.10812	-0.3206	-0.0641	-0.0434
DO	0.3288	0.0739	-0.0121	-0.1885	-0.1081	1	0.1399	0.0189	0.0126
WT	0.2415	0.2009	0.0307	-0.4128	-0.3206	0.1399	1	0.1815	0.1770
TC	0.1774	0.4315	0.3384	-0.0469	-0.0641	0.0189	0.1815	1	0.8746
FC	0.1201	0.4502	0.3691	-0.0586	-0.0434	0.0126	0.1770	0.8746	1

Table 2(a). Lag values of ACF and PACF and respected St. Error of pH, COD, BOD

Parameters	**pH**				**COD**				**BOD**			
Lag No.	ACF	SE	PACF	SE	ACF	SE	PACF	SE	ACF	SE	PACF	SE
Lag 1	.033	0.09	.022	0.09	.022	0.091	.022	0.09	.147	0.09	.147	0.09
Lag 2	-.131	0.09	-.012	0.09	-.011	0.091	-.012	0.09	.001	0.09	-.020	0.09
Lag 3	-.073	0.09	-.084	0.09	-.085	0.091	-.084	0.09	-.101	0.09	-.100	0.09
Lag 4	-.170	0.09	-.017	0.09	-.021	0.091	-.017	0.09	.018	0.09	.049	0.09
Lag 5	.114	0.09	.113	0.09	.113	0.092	.113	0.09	.147	0.09	.141	0.09
Lag 6	.198	0.1	-.038	0.09	-.025	0.093	-.038	0.09	-.124	0.09	-.188	0.09
Lag 7	.066	0.1	-.025	0.09	-.024	0.093	-.025	0.09	-.102	0.1	-.055	0.09
Lag 8	-.187	0.1	-.083	0.09	-.101	0.093	-.083	0.09	-.129	0.1	-.072	0.09
Lag 9	-.050	0.1	-.078	0.09	-.079	0.093	-.078	0.09	-.111	0.1	-.129	0.09
Lag 10	-.057	0.1	-.005	0.09	.010	0.094	-.005	0.09	-.027	0.1	-.027	0.09
Lag 11	.093	0.1	.108	0.09	.114	0.094	.108	0.09	.054	0.1	.105	0.09
Lag 12	.149	0.1	.079	0.09	.091	0.096	.079	0.09	-.058	0.1	-.115	0.09
Lag 13	-.060	0.1	.119	0.09	.099	0.096	.119	0.09	.013	0.1	.044	0.09
Lag 14	.047	0.1	-.008	0.09	-.033	0.097	-.008	0.09	-.026	0.1	-.009	0.09
Lag 15	-.154	0.1	-.070	0.09	-.071	0.097	-.070	0.09	-.060	0.1	-.138	0.09
Lag 16	-.028	0.1	-.058	0.09	-.035	0.097	-.058	0.09	-.026	0.1	-.053	0.09
Lag 17	.065	0.1	-.091	0.09	-.058	0.097	-.091	0.09	5.911E-5	0.1	.046	0.09
Lag 18	-.087	0.1	.030	0.09	.055	0.099	.030	0.09	.057	0.1	-.020	0.09
Lag 19	.080	0.1	.140	0.09	.094	0.099	.140	0.09	.146	0.1	.168	0.09
Lag 20	-.041	0.1	.027	0.09	-.025	0.1	.027	0.09	.059	0.1	.065	0.09
Lag 21	-.092	0.1	.009	0.09	-.035	0.1	.009	0.09	.065	0.1	.015	0.09
Lag 22	-.134	0.11	-.088	0.09	-.097	0.1	-.088	0.09	-.035	0.1	-.062	0.09
Lag 23	.051	0.11	.006	0.09	.073	0.101	.006	0.09	.005	0.11	.011	0.09
Lag 24	.142	0.11	.136	0.09	.203	0.101	.136	0.09	.124	0.11	.055	0.09

Table 2(b). Lag values of ACF and PACF and respected St. Error of AMM, TKN, DO

Parameters	**AMM**				**TKN**				**DO**			
Lag No.	ACF	SE	PACF	SE	ACF	SE	PACF	SE	ACF	SE	PACF	SE
Lag 1	-.098	0.091	-.098	0.091	-.066	0.091	-.066	0.091	-.067	0.091	-.067	0.091
Lag 2	.221	0.092	.214	0.091	.151	0.091	.147	0.091	.045	0.091	.041	0.091
Lag 3	.101	0.093	.147	0.091	.056	0.092	.076	0.091	.009	0.092	.015	0.091
Lag 4	-.073	0.094	-.106	0.091	-.032	0.092	-.048	0.091	-.071	0.092	-.072	0.091
Lag 5	-.026	0.098	-.108	0.091	-.152	0.093	-.184	0.091	-.011	0.093	-.021	0.091
Lag 6	-.186	0.098	-.189	0.091	-.103	0.095	-.126	0.091	.021	0.095	.026	0.091
Lag 7	-.123	0.098	-.126	0.091	-.083	0.095	-.045	0.091	.031	0.095	.037	0.091
Lag 8	.021	0.101	.104	0.091	-.037	0.096	.016	0.091	-.088	0.096	-.092	0.091
Lag 9	-.122	0.103	-.005	0.091	-.081	0.096	-.061	0.091	.001	0.096	-.016	0.091
Lag 10	.113	0.103	.090	0.091	.047	0.097	.013	0.091	.145	0.097	.160	0.091
Lag 11	.183	0.104	.220	0.091	.202	0.098	.208	0.091	-.112	0.098	-.089	0.091
Lag 12	.065	0.109	.054	0.091	.087	0.101	.108	0.091	.140	0.098	.104	0.091
Lag 13	.173	0.111	.023	0.091	.141	0.101	.076	0.091	.049	0.101	.070	0.091
Lag 14	-.001	0.111	-.073	0.091	-.048	0.102	-.140	0.091	-.045	0.101	-.027	0.091
Lag 15	-.015	0.111	-.113	0.091	-.015	0.102	-.097	0.091	-.007	0.102	-.024	0.091
Lag 16	-.079	0.112	-.101	0.091	-.131	0.102	-.084	0.091	.106	0.102	.116	0.091
Lag 17	-.179	0.112	-.079	0.091	-.132	0.102	-.062	0.091	-.047	0.102	-.030	0.091
Lag 18	-.139	0.113	-.075	0.091	-.176	0.102	-.117	0.091	-.022	0.102	-.021	0.091
Lag 19	-.126	0.114	-.022	0.091	-.002	0.103	.038	0.091	.010	0.103	-.015	0.091
Lag 20	-.149	0.114	-.056	0.091	-.207	0.103	-.136	0.091	-.015	0.103	.003	0.091
Lag 21	-.083	0.115	-.138	0.091	-.021	0.104	-.085	0.091	-.068	0.104	-.037	0.091
Lag 22	.063	0.115	.023	0.091	.029	0.104	-.027	0.091	-.111	0.104	-.191	0.091
Lag 23	.053	0.117	.042	0.091	-.042	0.107	-.146	0.091	.038	0.107	.039	0.091
Lag 24	.119	0.117	.041	0.091	.181	0.107	.092	0.091	-.157	0.107	-.126	0.091

Table 2(c). Lag values of ACF and PACF and respected St. Error of WT, TC, FC

Parameters	**WT**				**TC**				**FC**			
Lag No.	ACF	SE	PACF	SE	ACF	SE	PACF	SE	ACF	SE	PACF	SE
Lag 1	-.108	0.09	-.108	0.09	-.058	0.091	-.058	0.09	.002	0.09	.002	0.09
Lag 2	.002	0.09	-.010	0.09	-.007	0.096	-.010	0.09	-.448	0.09	-.448	0.09
Lag 3	-.258	0.09	-.262	0.09	.137	0.096	.137	0.09	-.053	0.09	-.064	0.09
Lag 4	.014	0.09	-.047	0.09	.078	0.096	.096	0.09	.005	0.09	-.248	0.09
Lag 5	.142	0.09	.140	0.09	.096	0.096	.112	0.09	-.018	0.09	-.105	0.09
Lag 6	.220	0.09	.205	0.09	.083	0.097	.083	0.09	.022	0.09	-.130	0.09
Lag 7	.161	0.1	.244	0.09	-.060	0.097	-.073	0.09	-.019	0.09	-.130	0.09
Lag 8	-.236	0.1	-.123	0.09	.056	0.097	.011	0.09	-.043	0.09	-.148	0.09
Lag 9	-.165	0.1	-.152	0.09	.114	0.098	.081	0.09	.016	0.1	-.113	0.09
Lag10	-.164	0.1	-.202	0.09	-.087	0.101	-.084	0.09	.088	0.1	-.028	0.09
Lag11	.265	0.1	.082	0.09	.069	0.101	.047	0.09	-.023	0.1	-.113	0.09
Lag12	.200	0.1	.167	0.09	.076	0.101	.062	0.09	-.075	0.1	-.098	0.09
Lag13	.043	0.1	.056	0.09	.166	0.101	.197	0.09	.030	0.1	-.064	0.09
Lag14	-.096	0.1	.096	0.09	.061	0.101	.072	0.09	-.041	0.1	-.181	0.09
Lag15	-.297	0.1	-.148	0.09	.020	0.101	.018	0.09	.073	0.1	.030	0.09
Lag16	.107	0.11	.060	0.09	-.016	0.102	-.065	0.09	.063	0.1	-.079	0.09
Lag17	.071	0.11	-.025	0.09	.133	0.102	.042	0.09	-.059	0.1	-.024	0.09
Lag18	.185	0.11	-.101	0.09	-.001	0.103	-.051	0.09	-.003	0.1	-.005	0.09

Parameters	WT				TC				FC			
Lag No.	**ACF**	**SE**	**PACF**	**SE**	**ACF**	**SE**	**PACF**	**SE**	**ACF**	**SE**	**PACF**	**SE**
Lag19	-.052	0.11	-.046	0.09	.046	0.103	.039	0.09	-.016	0.1	-.066	0.09
Lag20	-.123	0.11	-.021	0.09	.011	0.103	-.009	0.09	-.007	0.1	-.003	0.09
Lag21	-.117	0.11	.097	0.09	.039	0.103	.043	0.09	.010	0.1	-.047	0.09
Lag22	-.083	0.11	-.052	0.09	-.039	0.104	-.079	0.09	.012	0.1	.015	0.09
Lag23	.089	0.11	-.155	0.09	-.075	0.104	-.100	0.09	-.017	0.1	-.046	0.09
Lag24	.137	0.11	.043	0.09	-.059	0.104	-.114	0.09	.000	0.1	.014	0.09

From these figures and tables, it has been observed that for pH no seasonal pattern trailed. Thus, pH follows Brownian dynamics for monthly variation and pattern follows negative lower variability. COD has 6-months period with persistent behavior. BOD follows 6-month cyclic period with negative trend. AMM follows anti-persistent positive trend. TKN trails 6-monthly cyclic patterns with positive higher variability. DO follows 6-monthly cyclic Brownian trend. WT follows six-monthly periodic pattern with random time series negative lower variability trend. For TC, Stationary-R-squared-and-R-squared-values display analogous behaviour. Thus, developed model found better than foundational model. FC follows no trend and shows negative higher variation.

Table 3. Regression equations, Fractal and Trend variation of each parameter at Agra D/S

Parameters	Regression Equation	R square	H(abs)	D(Fractal)	PI	Trend %
Ph	y = -9E-05x + 7.8276	7.00E-05	0.5	1.500035	7E-05	-0.14%
COD	y = -0.1014x + 63.40	0.0131	0.5507	1.4493	0.1014	-21.24%
BOD	y = -0.0215x + 17.90	0.0065	0.5108	1.48925	0.0215	-15.54%
AMM	y = 0.0292x + 5.166	0.0343	0.4854	1.5146	0.0292	50.54%
TKN	y = 0.0207x + 9.2232	0.0134	0.4897	1.51035	0.0207	23.72%
DO	y = 0.0008x + 4.8312	8.00E-05	0.4996	1.5004	0.0008	1.97%
WT	y = -0.0098x + 26.94	0.0037	0.5049	1.4951	0.0098	-4.46%
TC	y = 101691x + 6E+07	0.0002	-	-	-	18.72%
FC	y = -8224.8x + 6E+06	0.0001	-	-	-	-18.04%

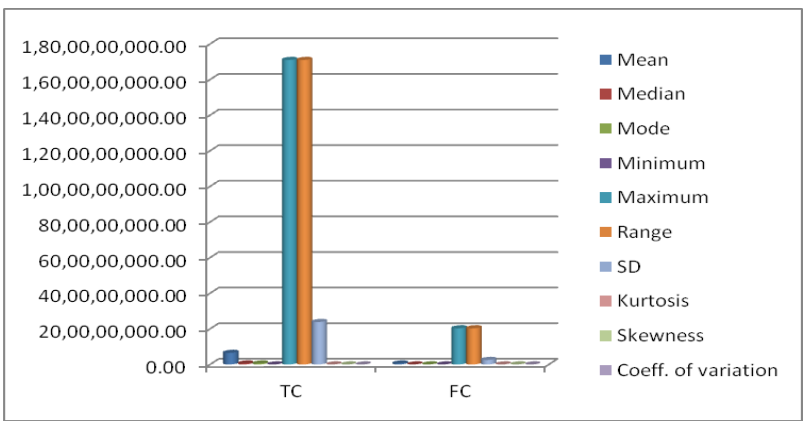


Figure 2. Graphical Representation of Statistical Analysis of Water Quality Parameters at Agra D/S Site of Yamuna River.

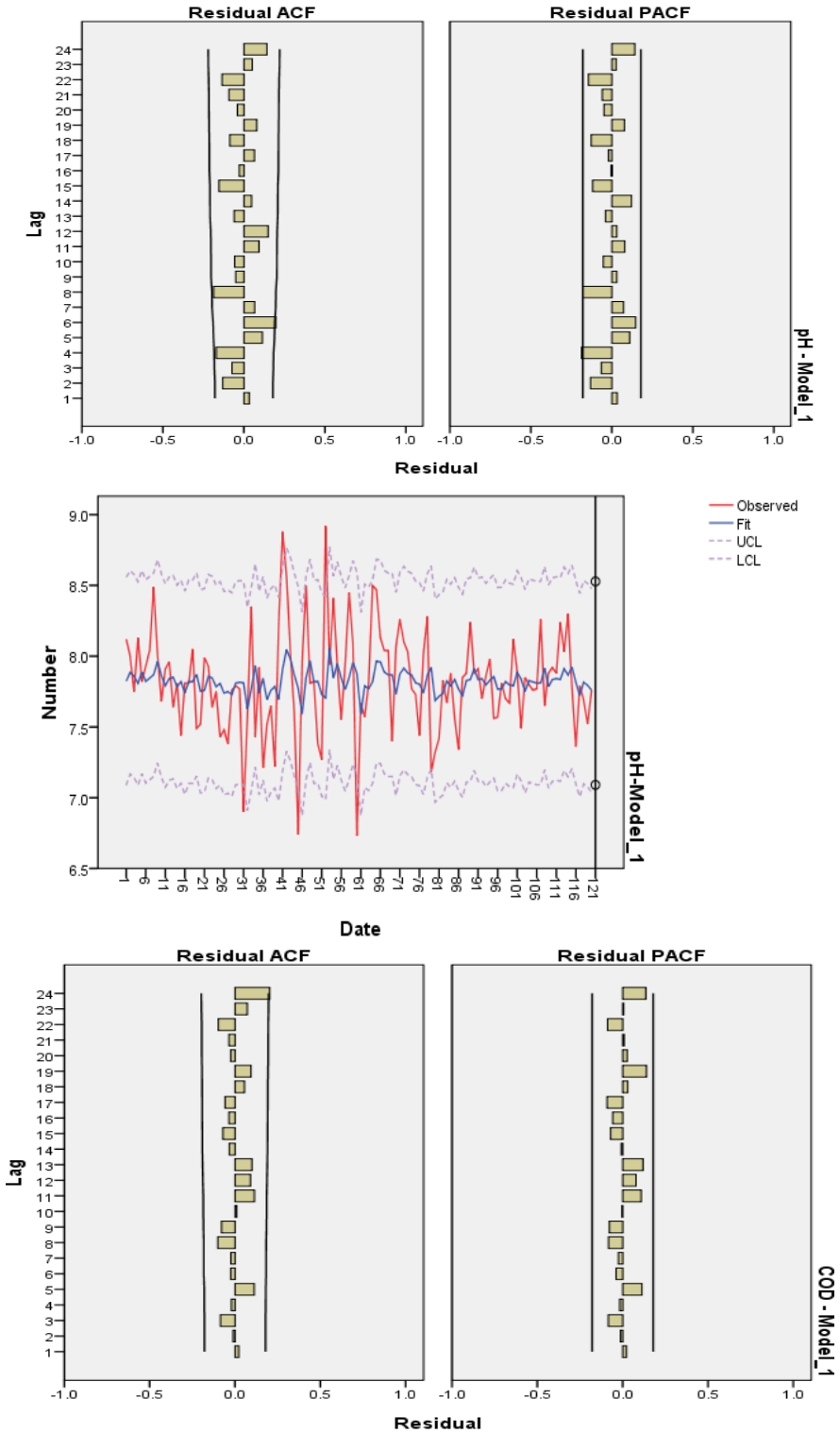


Figure 3. (Continued).

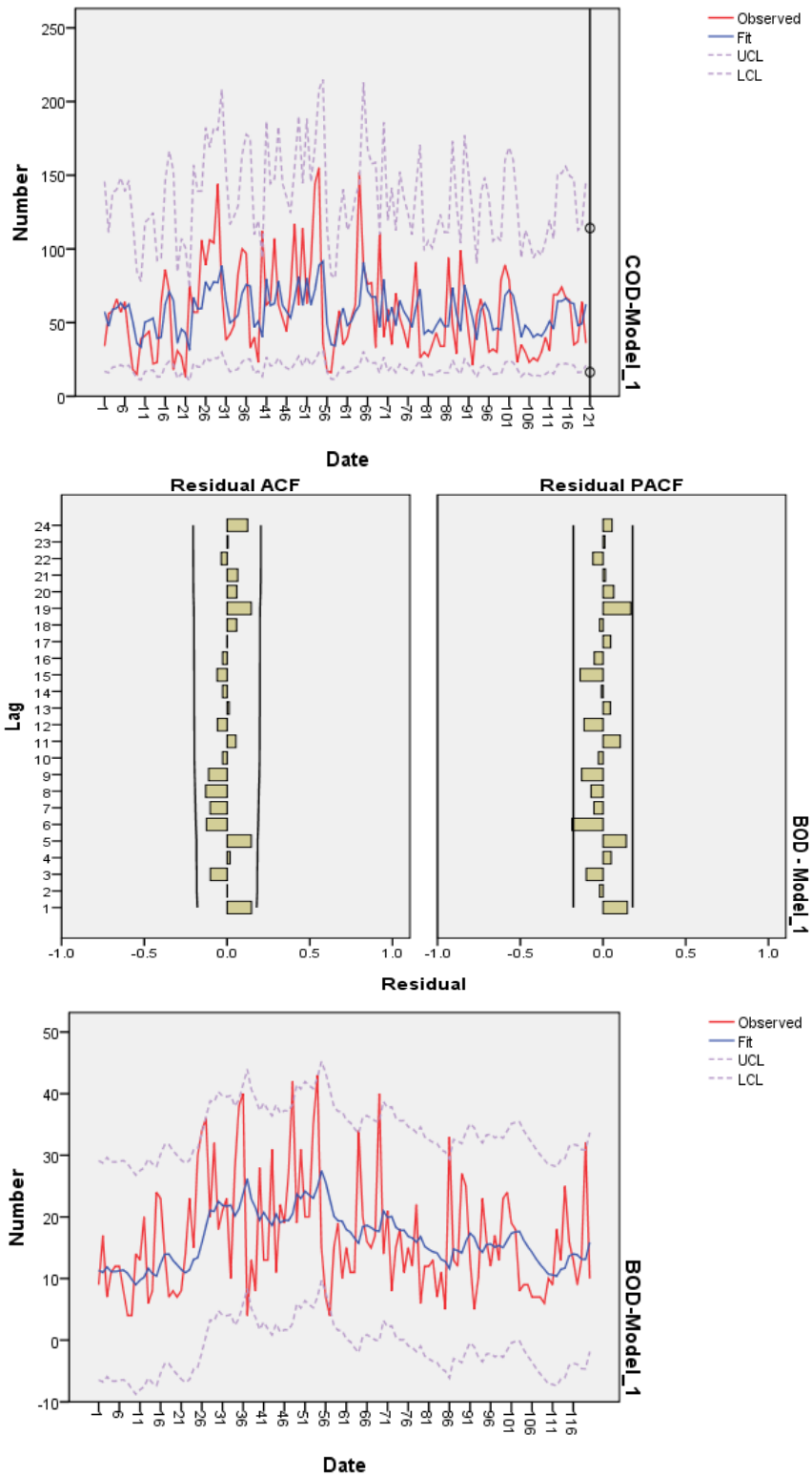


Figure 3. (Continued).

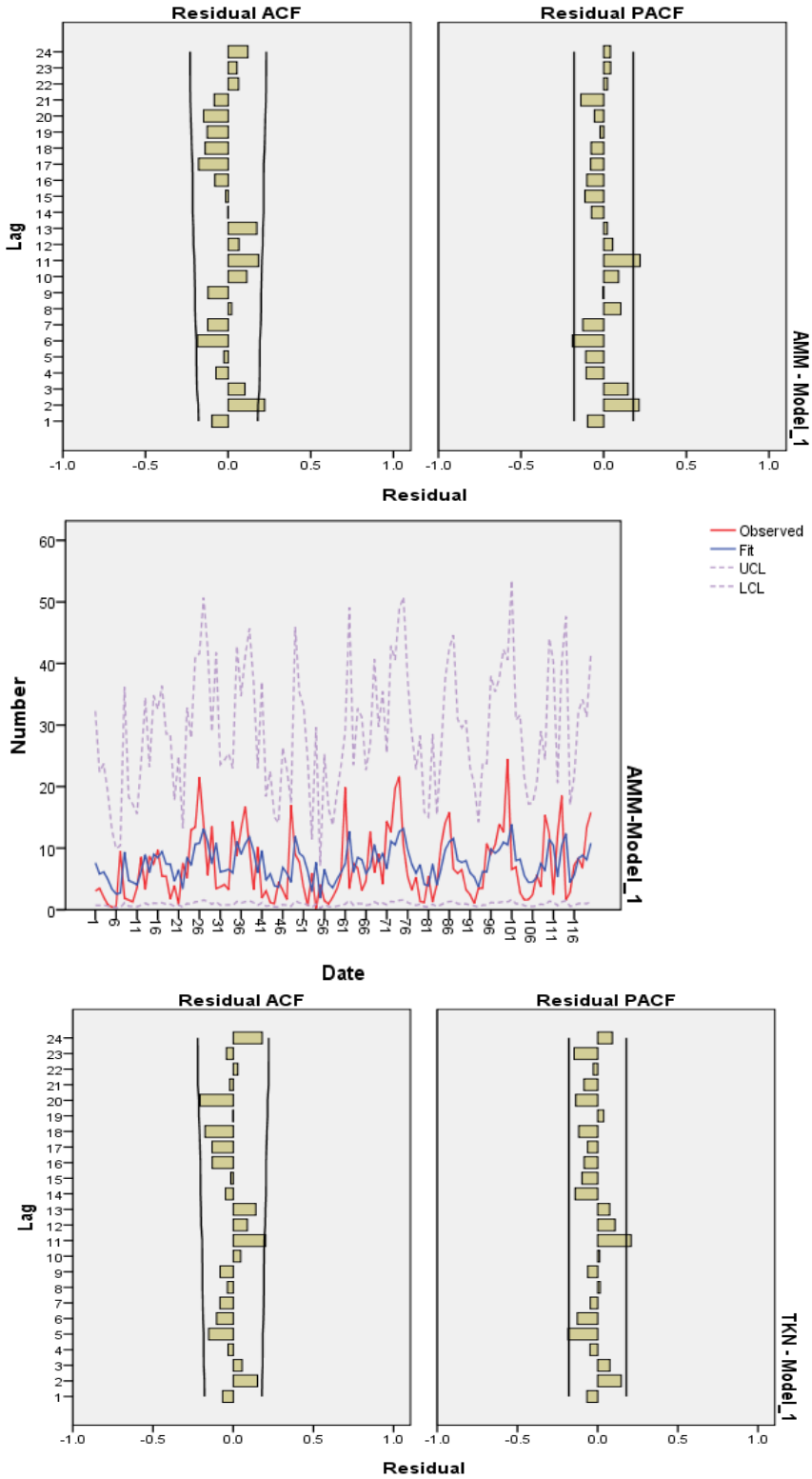


Figure 3. (Continued).

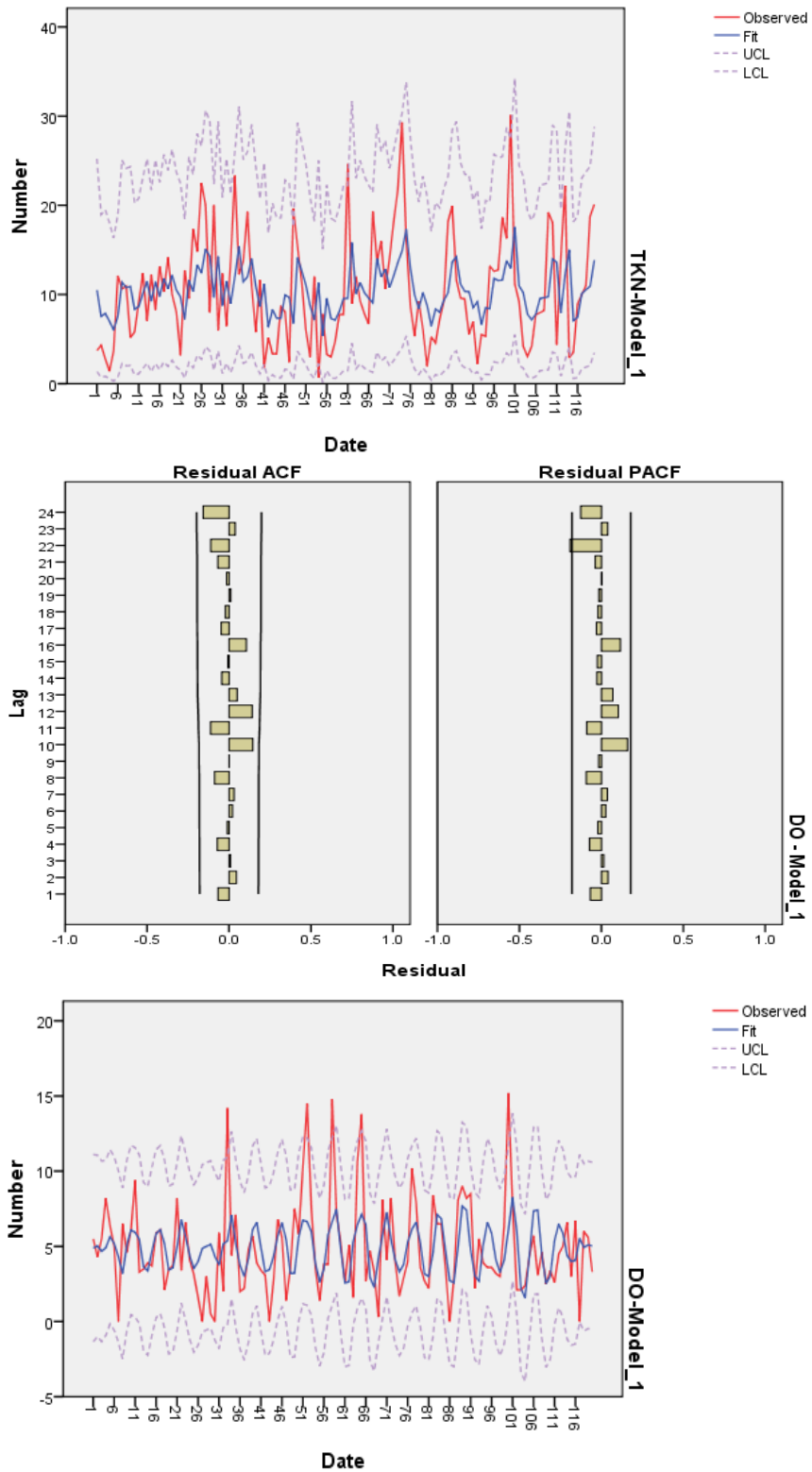


Figure 3. (Continued).

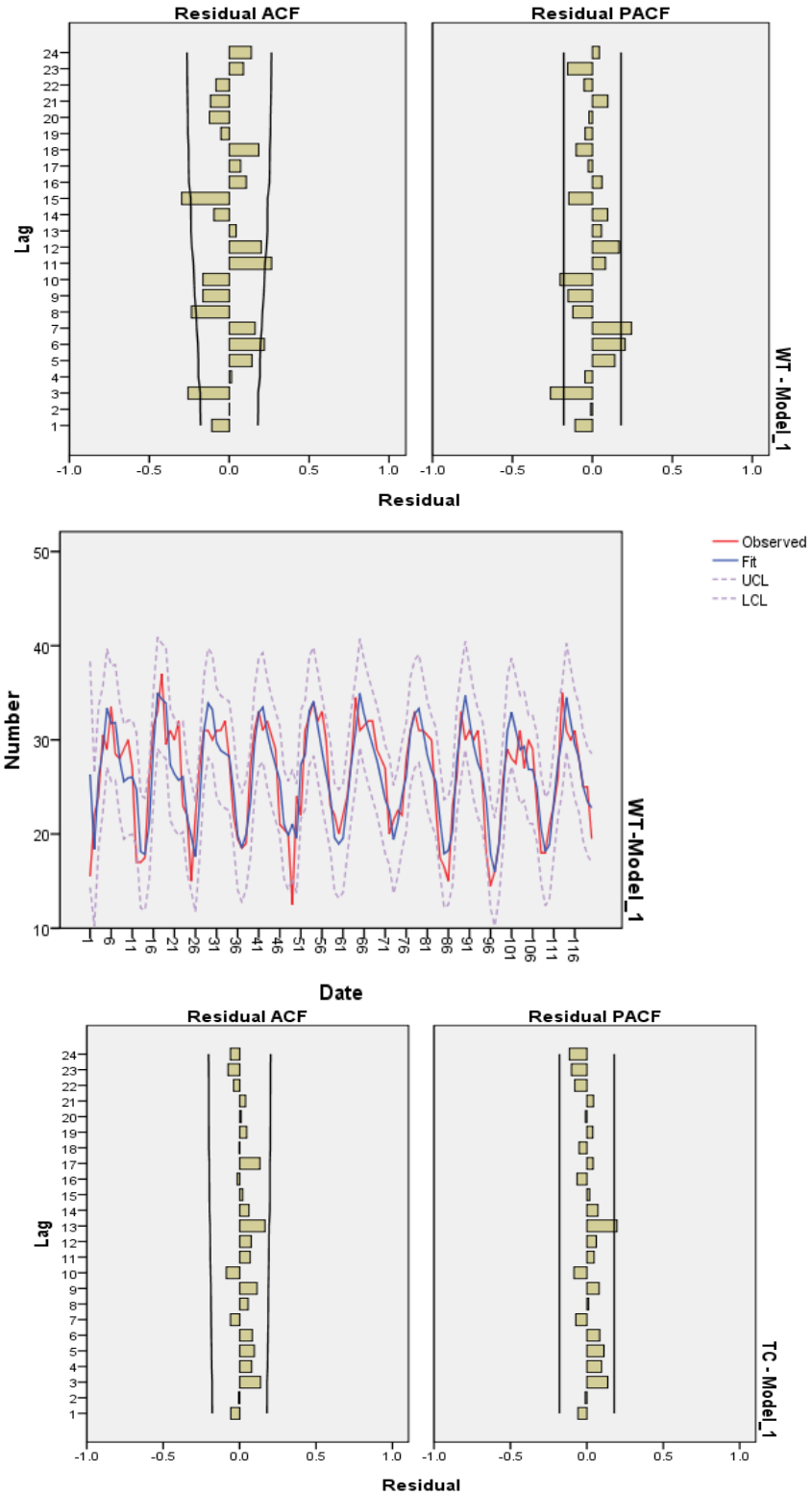


Figure 3. (Continued).

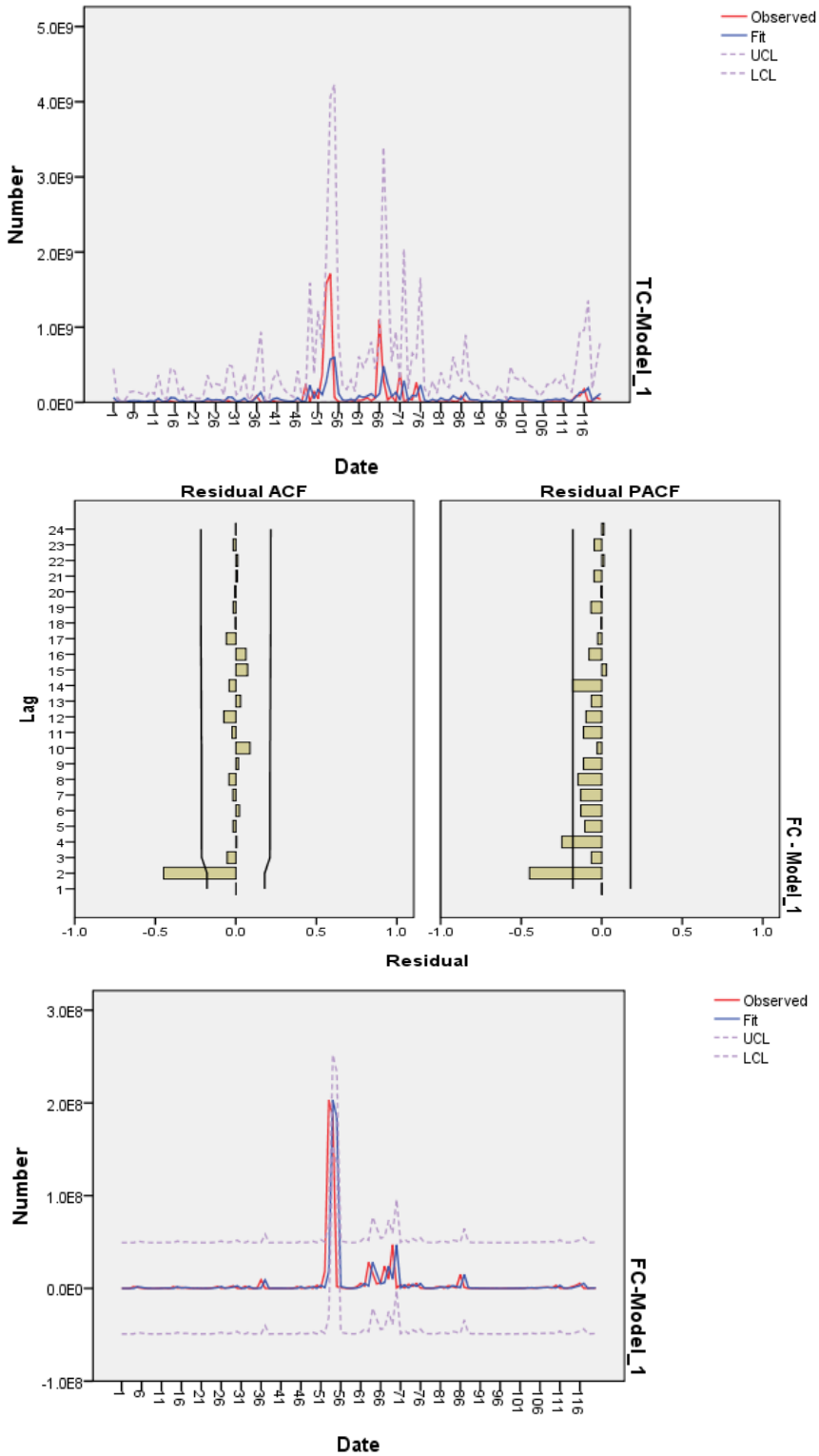


Figure 3. Residual ACF, Residual PACF plot and Time Series analysis plot with observed, fit (prediction using ARIMA), UCL, LCL of each water quality parameter at Agra D/S.

Table 4. Time series analysis of river water quality parameters

Fit Statistic		pH	COD	BOD	AMM	TKN	DO	WT	TC	FC
Stationary R2	Mean	.045	.189	.322	.174	.172	.220	.706	.243	-2.864E-7
R-squared		.045	.172	.060	.171	.169	.220	.706	.257	.037
RMSE		.363	28.175	8.976	5.013	5.680	2.816	3.084	2.063E8	2.486E7
MAPE		3.407	49.141	56.364	157.18	79.105	60.331	9.937	3.673E3	2.226E3
MaxAPE		17.046	262.85	553.77	3.742E3	1.616E3	1.296E3	69.855	1.259E5	9.645E4
MAE		.265	21.828	6.928	4.091	4.475	2.007	2.371	8.008E7	5.903E6
MaxAE		1.214	90.416	22.329	13.970	17.163	9.137	10.828	1.302E9	1.844E8
Normalized BIC		-1.947	6.757	4.429	3.304	3.554	2.270	2.492	38.370	34.097
Statistics	Ljung-Box Q(18)	29.211	11.133	15.959	36.818	29.374	12.420	72.304	17.485	29.681
DF		17	17	17	17	17	14	13	17	17
Sig.		0.0326	0.8491	0.5267	0.0036	0.0312	0.5725	3.01E-10	0.422	0.0287
ARIMA Model	Prediction	(1,0,0)	(1,0,0)	simple	(1,0,0)	(1,0,0)	(2,0,2)	(2,0,6)	(1,0,0)	Simple
Predicted Value		7.8226	57.5039	16.3898	7.521462	10.46284	4.88464	26.4121	65740506	5469920
LCL		7.1038	19.2389	-1.383	0.9075	1.9334	-0.708	20.514	272250.6	-4.4E+07
UCL		8.5415	134.95	34.162	29.005	23.576	10.478	32.3096	4.65E+08	54694594
Residual		-0.00053	0.00135	0.21019	0.00159	0.00407	-0.0038	-0.062	0.014238	1939.551

Air Pollution

The wavelet spectrum, for Carbon Monoxide at DCE and ITO are shown in figure 4(a). Figure 4(b) depicts Wavelet Cross Spectrum in both Modulus and angle of Carbon Monoxide between location Delhi College of engineering and ITO-Crossing. Magnitude of wavelet cross spectrum can be construed as absolute value of the local covariance between the two-time series. It is clear that Carbon Monoxide possesses significant input around the measure 75 over the interval [0, 180] and around the scale 110 over the time interval [350, 550]. This is consistent with the behaviour observed by visual inspection of the time-domain plot and from the individual wavelet transforms stand out as being significant. The wavelet coherence can be interpreted as the local squared correlation coefficient in the time-scale plane. Wavelet Coherence for Carbon Monoxide between location Delhi College of Engineering and ITO-Crossing is shown in Figure 4(c). Arrows in the figure represent relative phase between the two signals as a function of scale and position. Relative phase info produces a local scale of delay between two-time series. Phase info could be deduced by detecting diverse sections for time-scale plane and highlights coherent-behaviour. Certain transient minor assistances towards variability of time-series happen at smaller measures the commencement of Carbon Monoxide that

display swift fluctuations. Behaviour is non-coherent plus phase changes rapidly. Also, between positions 300 to 600 and measures more than 79, several coherent sections have been easily detected, demarcated by the stability for phase. Similarly, fig.5, 6, 7, 8 and 9 summarize discussion for wavelet cross spectrum, Coherence analysis of Carbon Monoxide between locations DCE-SF, DCE-MV, ITO-SF, ITO-MV and SF-MV respectively. Table-5 provide particulars of Hurst exponent, fractal dimension and Predictability index for every contaminant observed at Delhi College of engineering (DCE), ITO-Crossing (ITO), Siri Fort (SF) and by Mobile Van (MV) in Delhi, India.

Table 5. Hurst Exponents, Fractal Dimension, Predictability Index for each pollutant monitored at Delhi College of engineering, ITO-Crossing, Siri Fort and by Mobile Van in Delhi, India

Statistics	CO	NO	NO_2	O_3	SO_2
Station: DCE					
H	0.170	0.216	0.269	0.139	0.112
D	1.830	1.784	1.731	1.861	1.888
PI	0.33	0.284	0.231	0.361	0.388
Station: ITO					
H	0.222	0.272	0.314	0.223	0.187
D	1.778	1.728	1.686	1.777	1.813
PI	0.278	0.228	0.186	0.277	0.313
Station: SF					
H	0.190	0.284	0.033	0.206	0.165
D	1.810	1.716	1.967	1.794	1.835
PI	0.31	0.216	0.467	0.294	0.335
Station: MV					
H	0.065	0.230	0.046	0.217	0.118
D	1.935	1.770	1.954	1.783	1.882
PI	0.435	0.27	0.454	0.283	0.382

From this table, it has been observed that CO shows anti-persistent at DCE. H value is higher at ITO and lower for MV. DCE and SF trails same pattern. NO follows anti-persistent behavior. H is higher at SF and lower at DCE, thus gives the degree to which a correct prediction or forecast of a system's state can be made. H is higher at ITO and lower at SF for NO_2: It shows anti-persistent behaviour. For O3:, the value of H is higher at ITO but lower at DCE (Industrial Area). Thus, it shows anti-persistent behaviour. SO2: trails higher value of H at ITO but lower at DCE (Industrial Area). The value of Fractal dimension is higher for DCE and lesser for ITO. Thus, it shows anti-persistent behaviour.

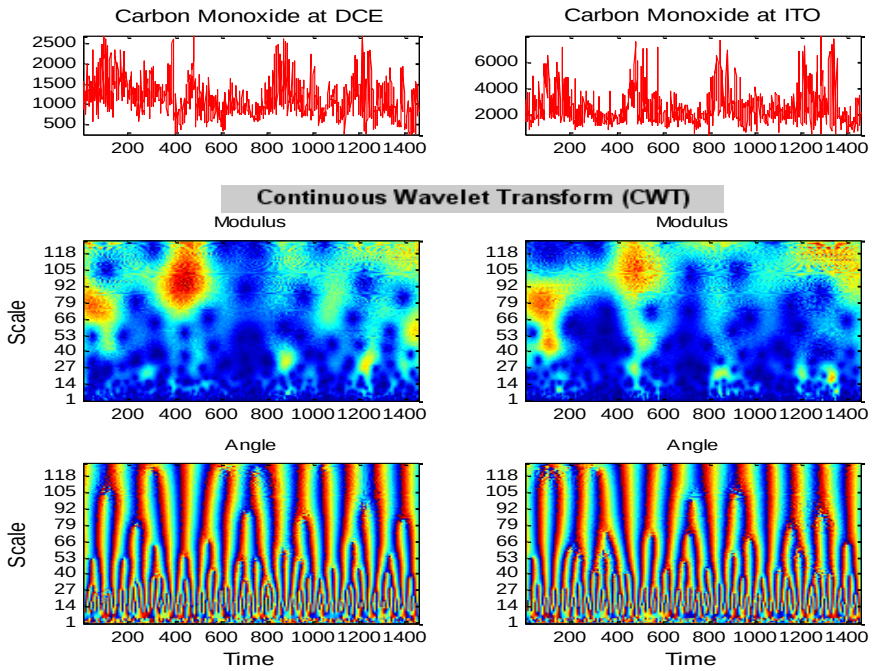


Figure 4(a). Continuous Wavelet transform using Complex Morlet Wavelet of Carbon Monoxide at locations Delhi College of Engineering and ITO-Crossing.

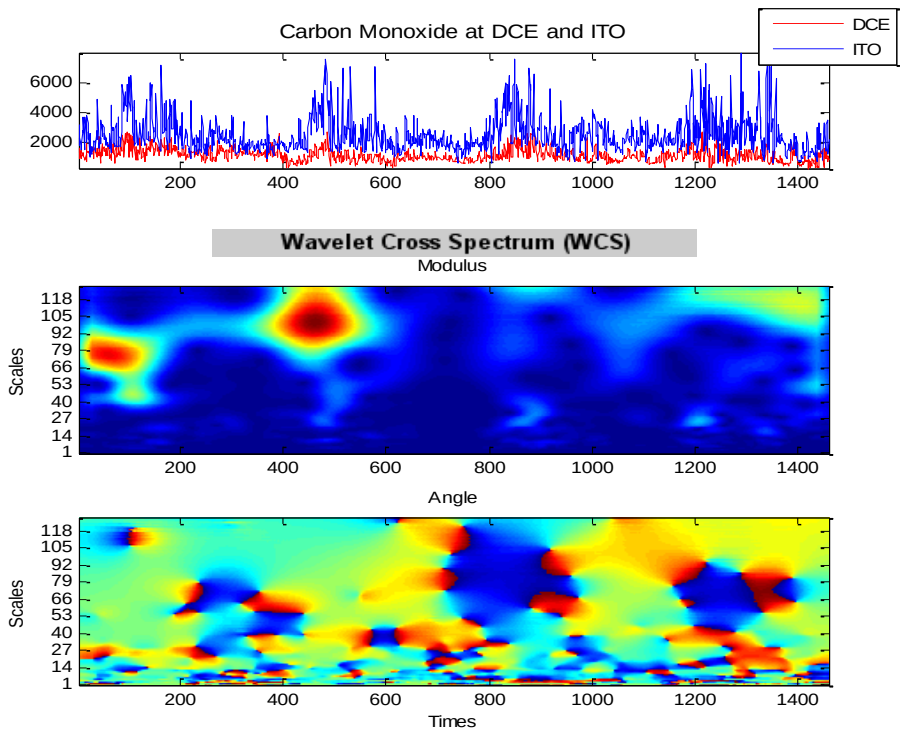


Figure 4(b). Wavelet Cross Spectrum using Complex Morlet Wavelet of Carbon Monoxide between locations Delhi College of Engineering and ITO-Crossing.

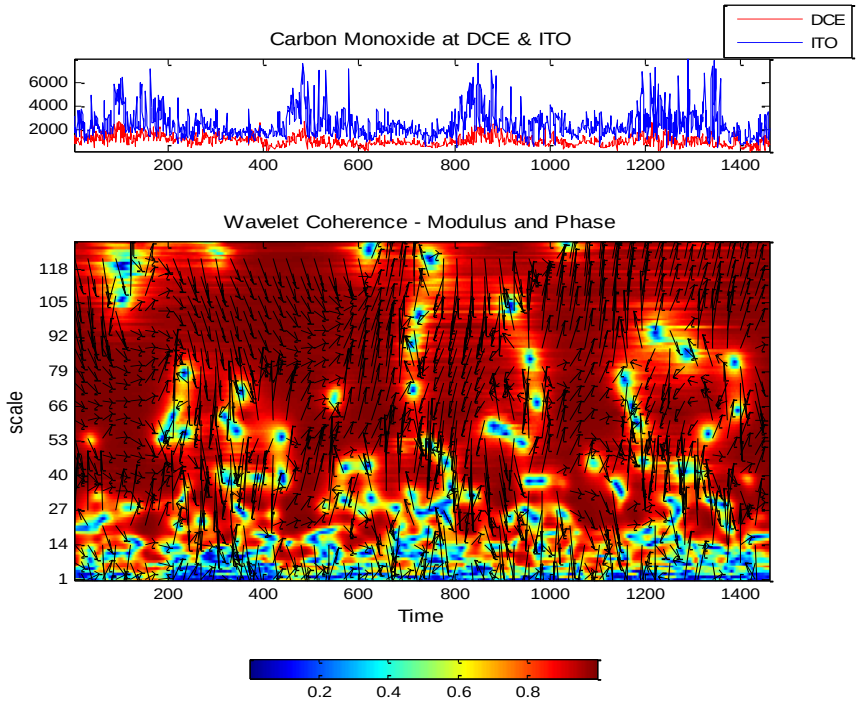


Figure 4(c). Wavelet Coherence using Complex Morlet Wavelet of Carbon Monoxide between locations Delhi College of Engineering and ITO-Crossing.

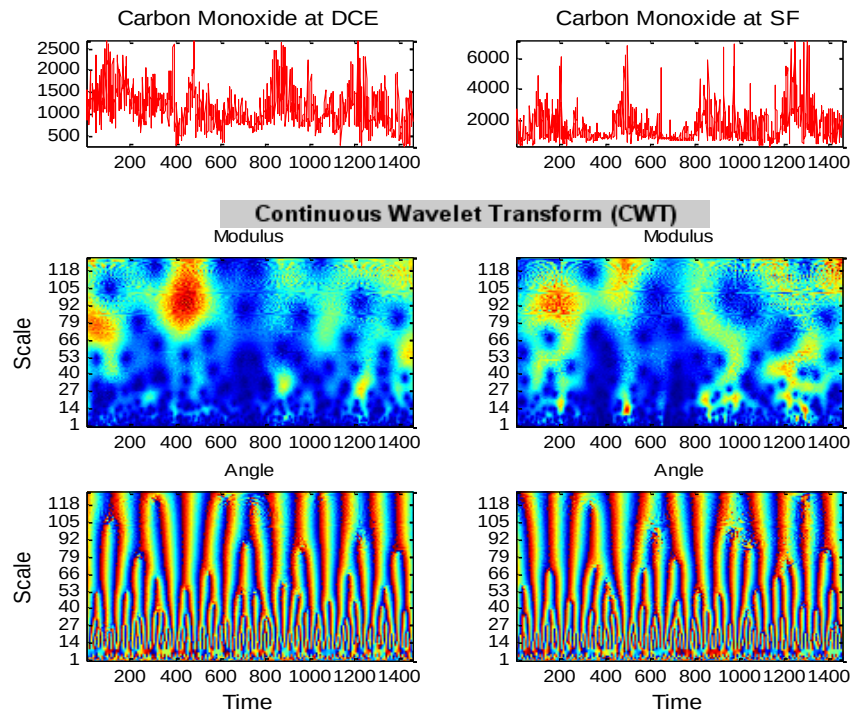


Figure 5(a). Continuous Wavelet transform using Complex Morlet Wavelet of Carbon Monoxide at locations Delhi College of Engineering and Siri Fort.

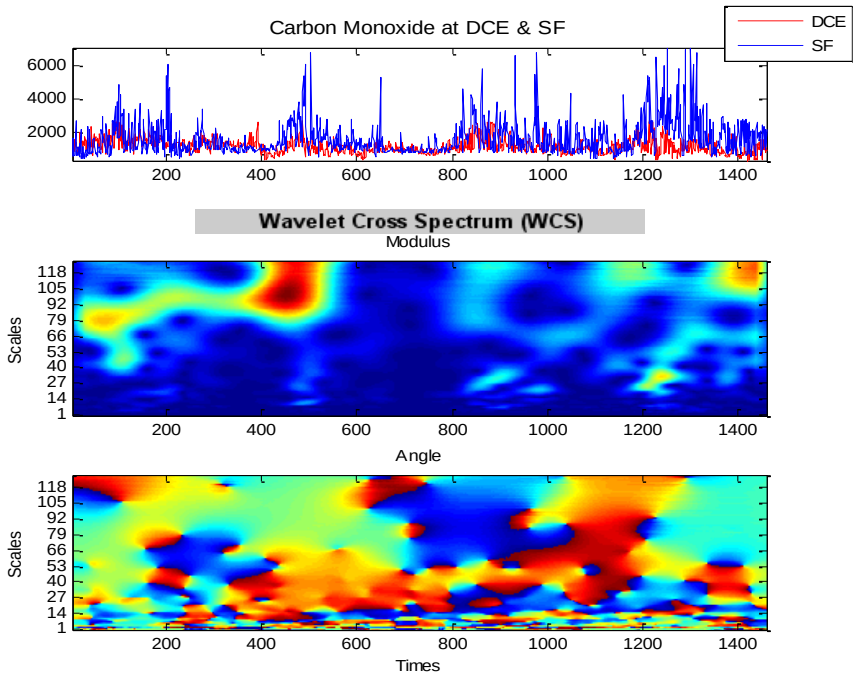


Figure 5(b). Wavelet Cross Spectrum using Complex Morlet Wavelet of Carbon Monoxide between locations Delhi College of Engineering and Siri Fort.

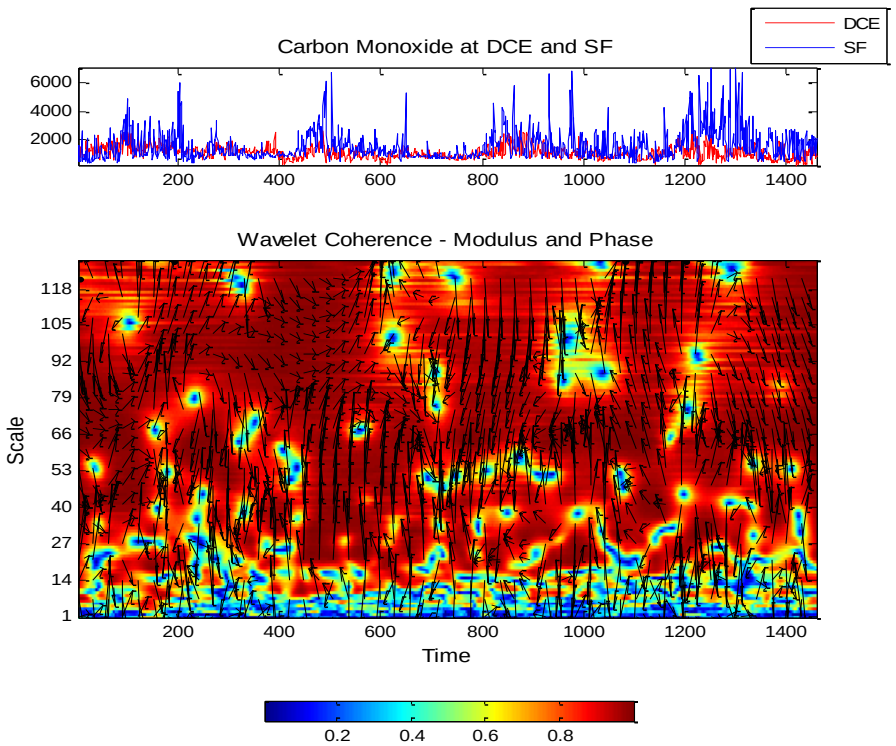


Figure 5(c). Wavelet Coherence using Complex Morlet Wavelet of Carbon Monoxide between locations Delhi College of Engineering and Siri Fort.

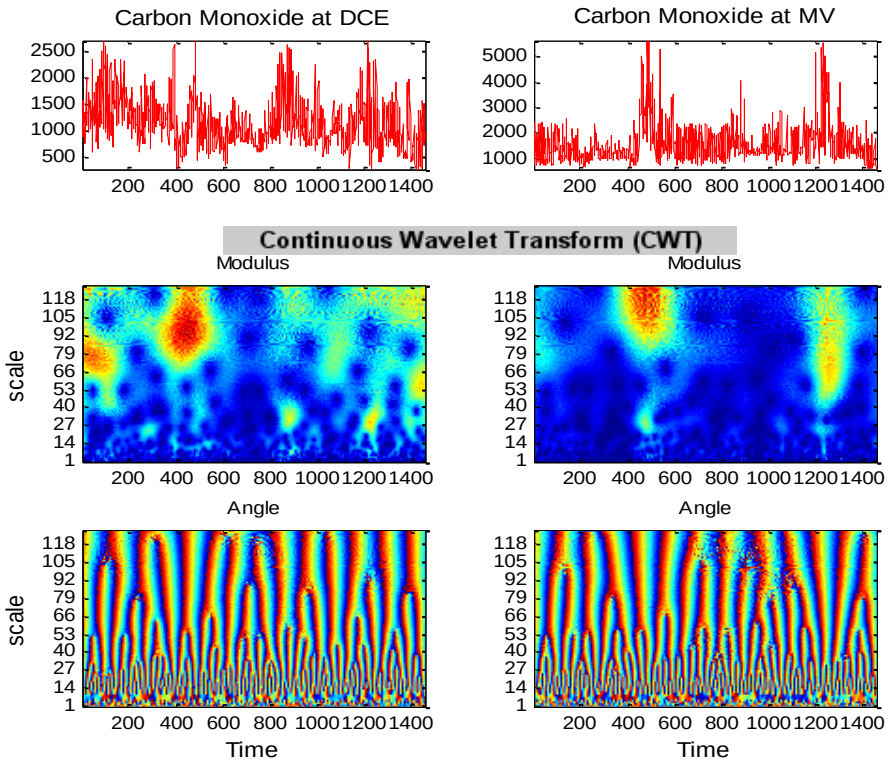


Figure 6(a). Continuous Wavelet transform using Complex Morlet Wavelet of Carbon Monoxide at locations Delhi College of Engineering and Mobile Van.

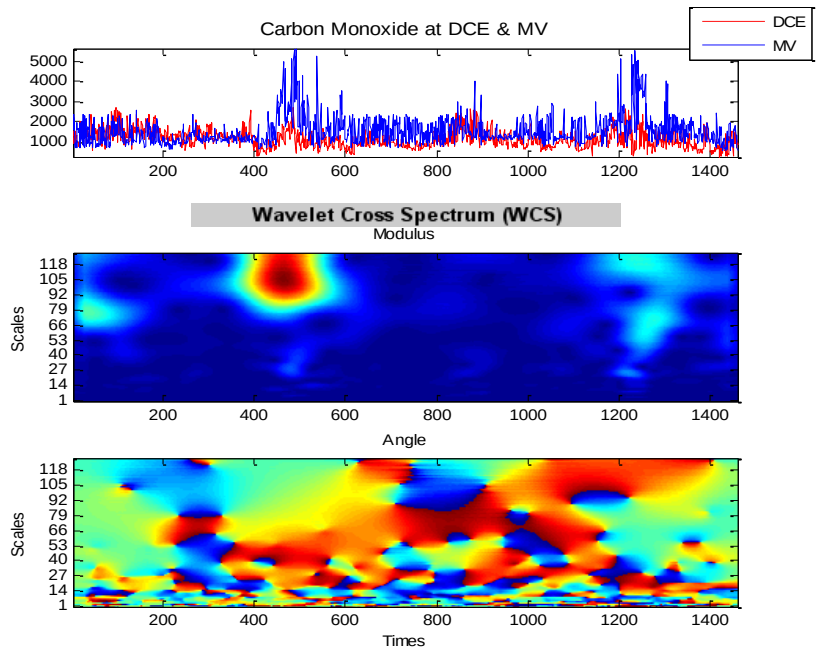


Figure 6(b). Wavelet Cross Spectrum using Complex Morlet Wavelet of Carbon Monoxide between locations Delhi College of Engineering and Mobile Van.

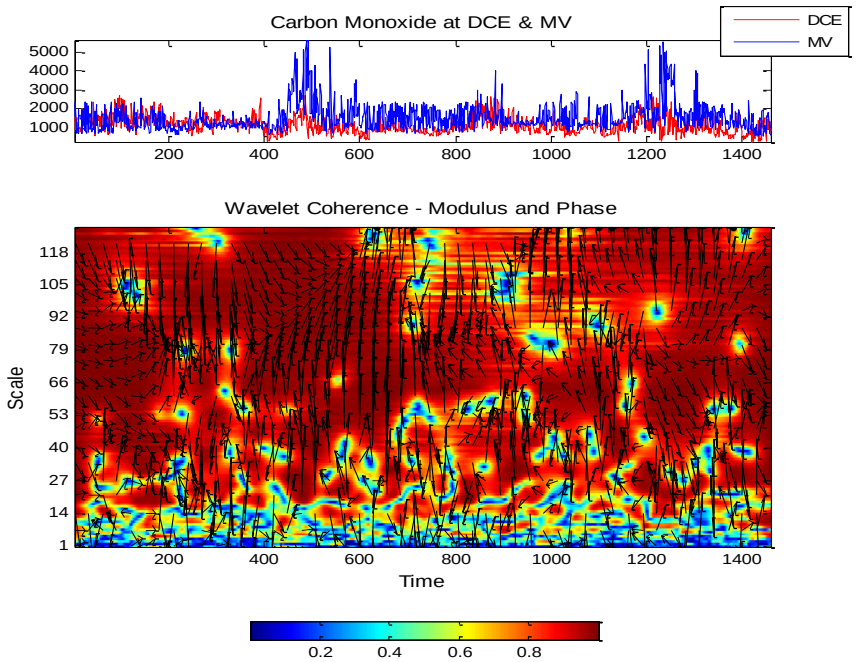


Figure 6(c). Wavelet Coherence using Complex Morlet Wavelet of Carbon Monoxide between locations Delhi College of Engineering and Mobile Van.

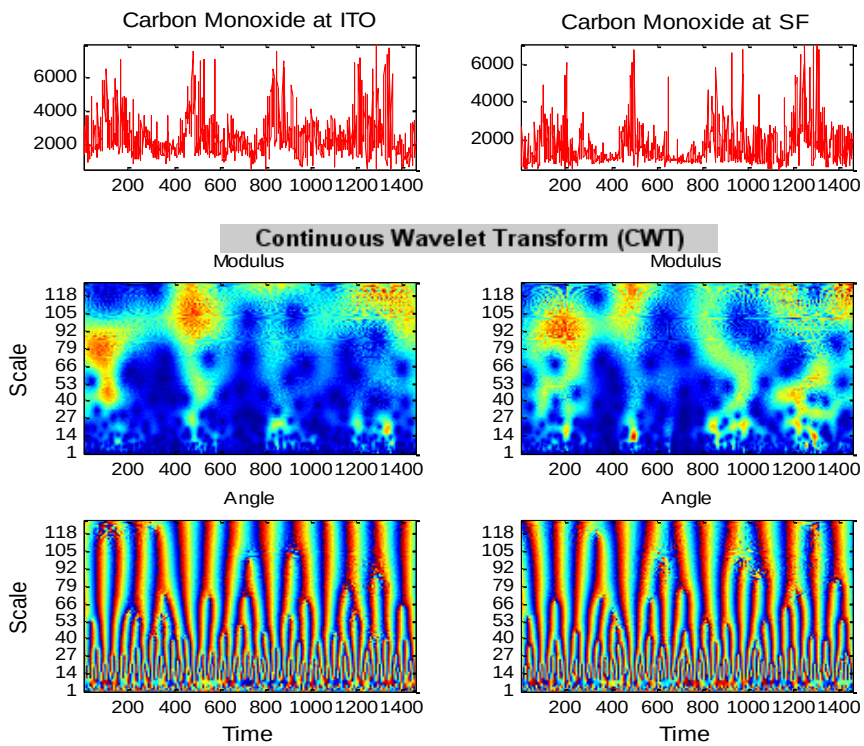


Figure 7(a). Continuous Wavelet transform using Complex Morlet Wavelet of Carbon Monoxide at locations ITO-Crossing and Siri Fort.

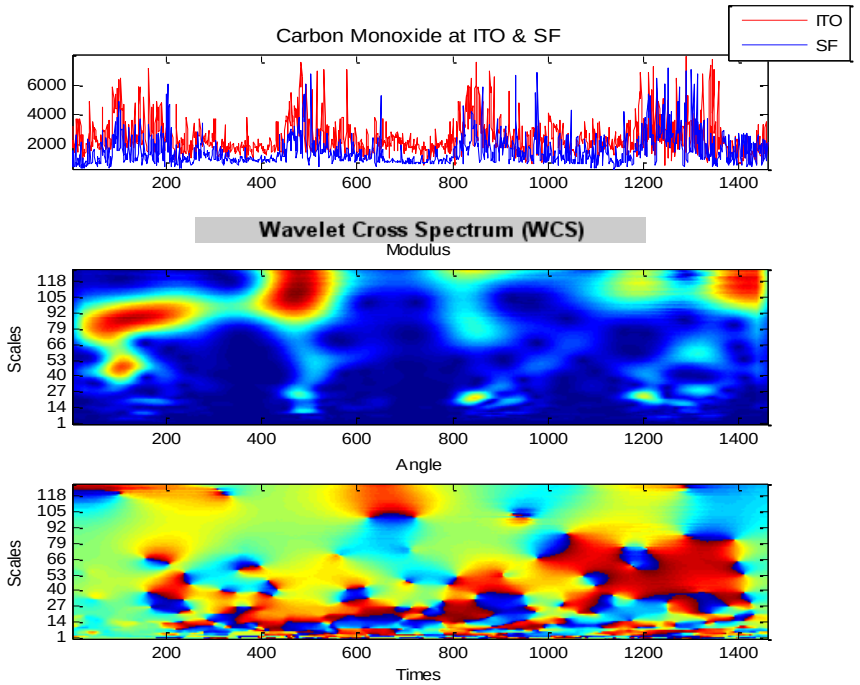


Figure 7(b). Wavelet Cross Spectrum using Complex Morlet Wavelet of Carbon Monoxide between locations ITO-Crossing and Siri Fort.

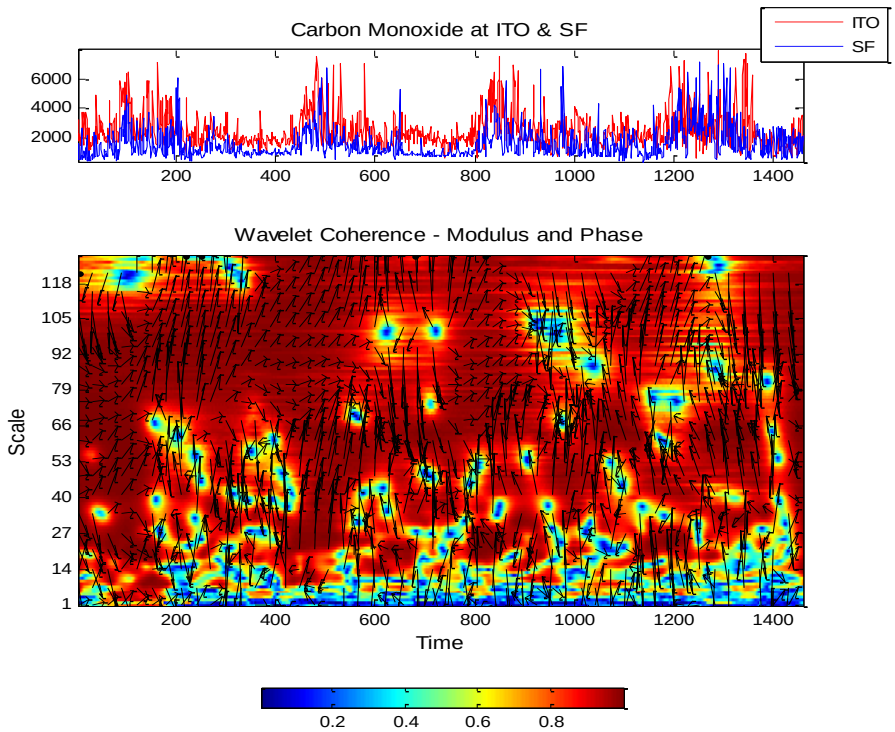


Figure 7(c). Wavelet Coherence using Complex Morlet Wavelet of Carbon Monoxide between locations ITO-Crossing and Siri Fort.

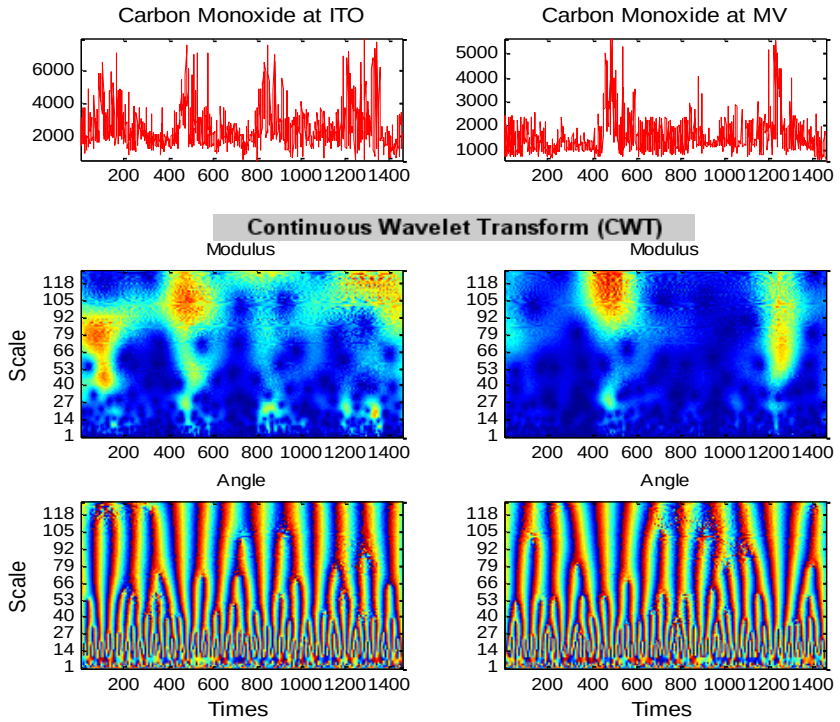


Figure 8(a). Continuous Wavelet transform using Complex Morlet Wavelet of Carbon Monoxide at locations ITO-Crossing and Mobile Van.

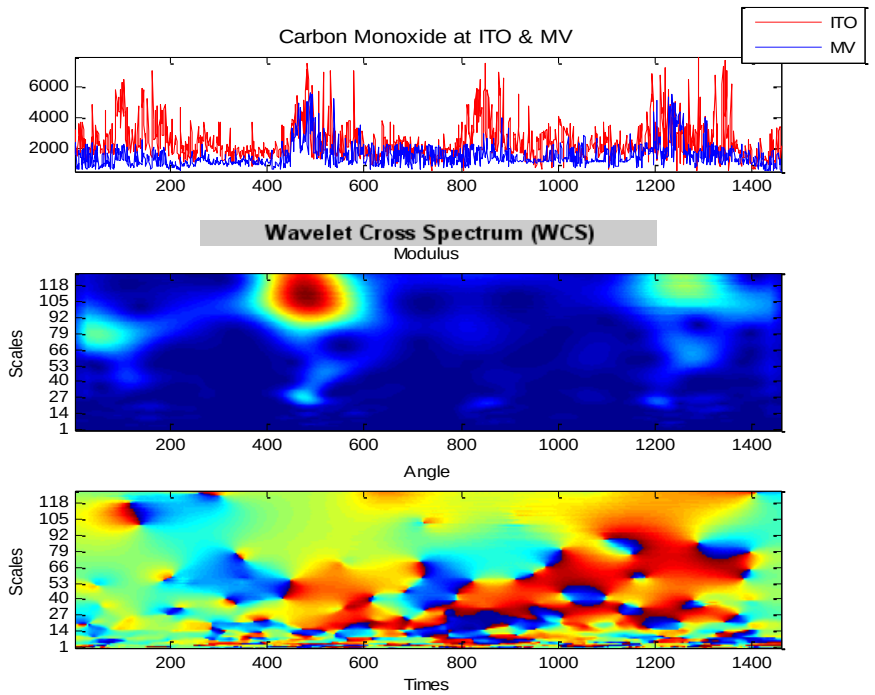


Figure 8(b). Wavelet Cross Spectrum using Complex Morlet Wavelet of Carbon Monoxide between locations ITO-Crossing and Mobile Van.

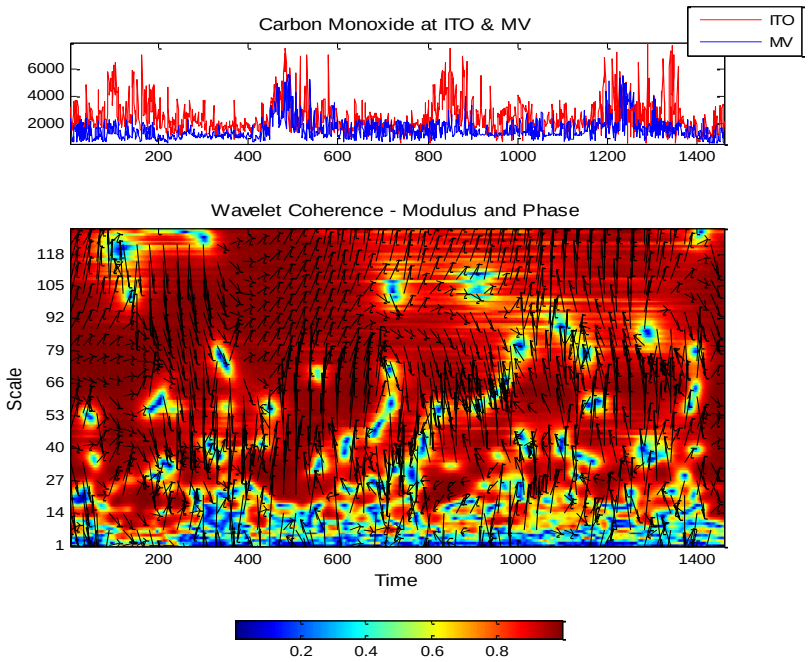


Figure 8(c). Wavelet Coherence using Complex Morlet Wavelet of Carbon Monoxide between locations ITO-Crossing and Mobile Van.

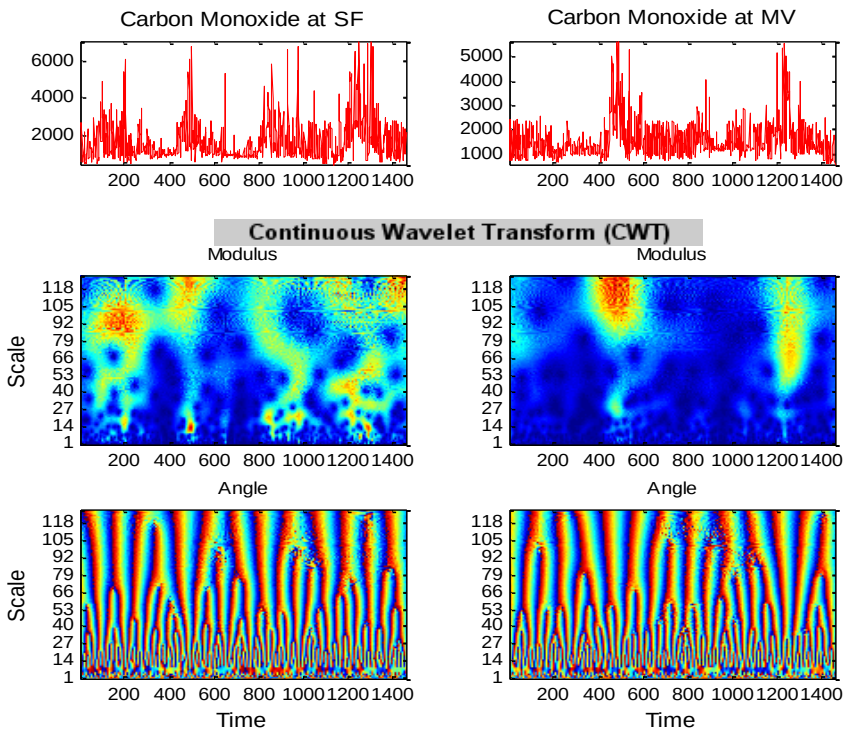


Figure 9(a). Continuous Wavelet transform using Complex Morlet Wavelet of Carbon Monoxide at locations Siri Fort and Mobile Van.

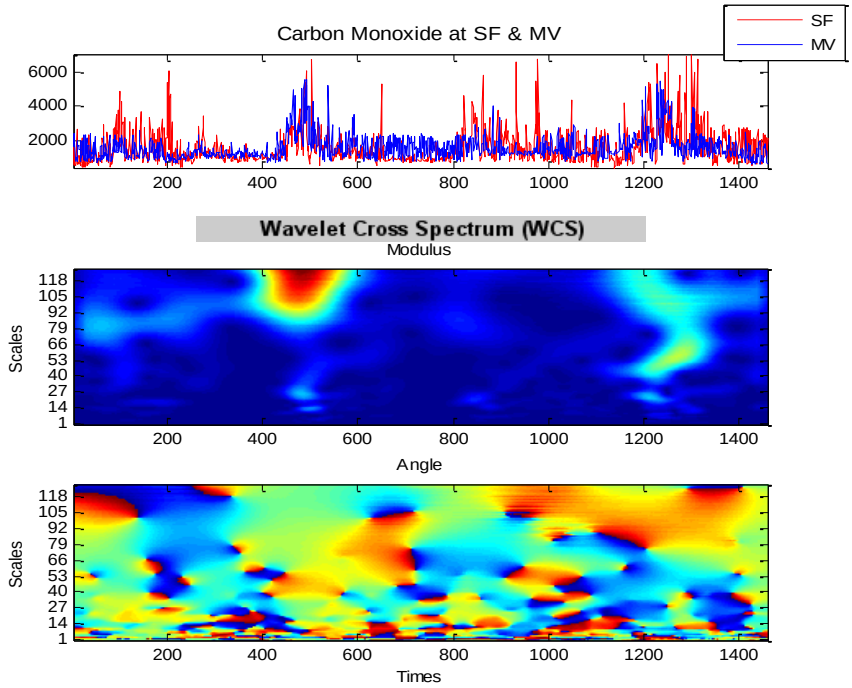


Figure 9(b). Wavelet Cross Spectrum using Complex Morlet Wavelet of Carbon Monoxide between locations Siri Fort and Mobile Van.

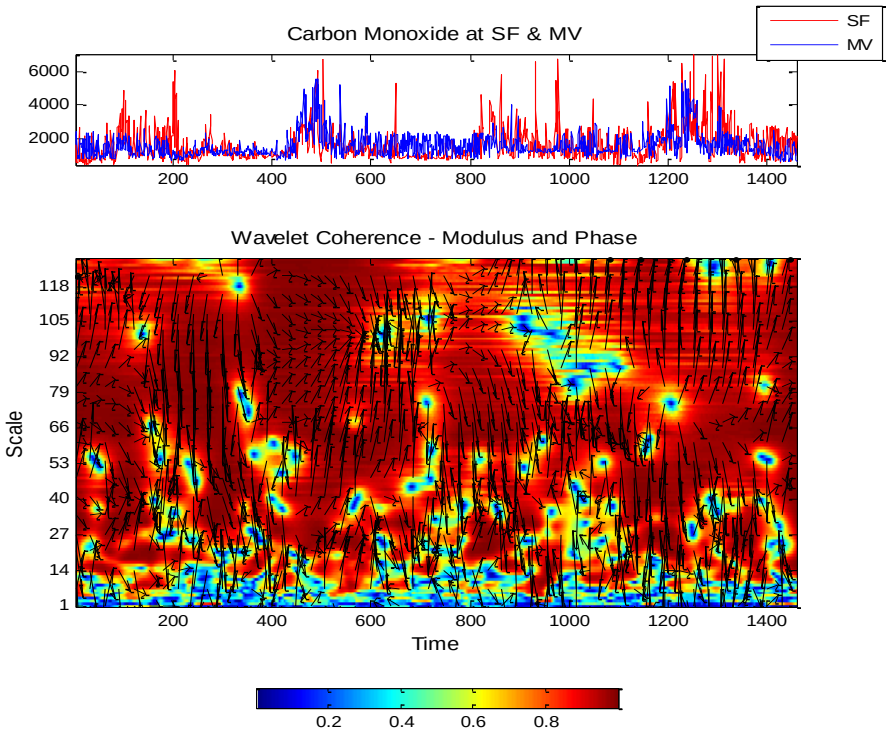


Figure 9(c). Wavelet Coherence using Complex Morlet Wavelet of Carbon Monoxide between locations Siri Fort and Mobile Van.

It is observed that all parameters show anti-persistent behaviour but pattern behaviour changes for different parameters at different locations. The value of all pollutants is higher at ITO the commercial area as all vehicles crosses and it is a hub of all commercial activities, thus it shows behaviour which is closer to Brownian approach.

From these figures, it has been observed that for CO at different locations the major values attained through AR(1) process assessed via data and enclosed via black outline. Black lines in continuous wavelet transform in fig. 4(a) to 9(a) for commencement and ending of edge indicates likelihood of edge-effects. Coherence is the "normalized" cross-spectrum with respect to the spectrum of each signal. The cross-spectrum in fig. 4(b) to 9(b) indicates how much linear information is transferred from one signal to the other (and vice-versa), i.e., the "burden" of the line transfer at each frequency. The coherence indicates how much linear information of one signal is explained by the other signal and it can be used to estimate the causality between the two signals. Wavelet Coherence in fig. 4(c)-9(c) estimate the coherence between two signals when ergodicity and stationarity assumptions are not respected and also to perform time-varying analysis of the coherence. Coherence describes all properties of the correlation between several waves. It is observed that for CO between different stations the confidence interval is obtained. The resulting intervals are effective for every point in time, as the background indications were considered as fixed. Expected value of squared-coherence at assumed scale does not alter with respect to time. It is observed that both the stations are highly affected by unforeseen movement of CO in long term.

For better improvement various leveling operators can be applied towards adjustment of fluctuating size of time-frequency spaces disguised through underlying wavelets. It is also thought-provoking to see when a more instinctive smoothening would substantiate greater computational power in comparison to the permanent-sized smoothening space. Gained outcomes are based on profile of wavelet function engaged and valid for the Morelet wavelet only.

CONCLUSION

Water quality utilizing fractal, predictability, time-series and trend-analysis is studied at Taj Mahal, Agra D/S of river Yamuna, India. It can be witnessed that for water- quality characteristics, RMSE values are relatively low that depict dependent-series are sealed alongwith model-predictive-level, so model shows 95% confidence-limits. Also, water factors depict platykurtic-curve except TC & FC which shows leptokurtic.

Through regression study, it is witnessed that water-quality-factors: COD-BOD; COD-TC; COD-FC; AMM-TKN; TC-FC are positively-correlated and AMM-WT; TKN-WT; are negatively-correlated. It can be concluded for various lags, results of Auto-correlation function (ACF) and partial auto correlation (PACF) are more than standard

error signifying the results that are distant from zero-mean. COD,-BOD,-TKN,-DO,-WT follow six-monthly cyclic pattern whereas on parameters: pH, AMM, TC, FC no such trend is detected. Water quality factors:-pH,-COD,-BOD,-WT,-FC have negative trend for monthly variation except others follows positive trend. AMM, TKN follows anti-persistent behaviour; pH, DO, WT follows chaotic behaviour while COD, BOD have persistent behaviour. AMM, TKN, TC can be observed to have the trends as positive having higher variation and DO has low positive trend. COD, BOD, FC follows negative trend with high variation and pH, WT follows the same movement having low-variation. Applying fractal, predictability index plus variability in analysis, it can be observed for all factors leaving out pH and WT cross recommended confines by WHO. It is concluded that if similar movement can be followed then characteristics of water may derogate in future.

Wavelet Cross Spectrum tends to be the more informative as it determines correspondence for local rate of recurrence performance for time series in a time-scale-plane. It is comforting that the XWT show that Carbon Monoxide between different locations are into anti-phase across all scales along with substantial common-power. Cross-Wavelet-transform reveals region with higher-common-power and info about phase relationship. Expected value for squared-coherence at a specified measure may not alter over time. It is observed that both the stations are highly affected by unforeseen movement of CO in long term. The estimated outcomes depend on profile of wavelet-function utilized and are valid only for the Morlet-wavelet. Wavelet Coherence proves to be a good time frequency method, which overcomes the limitation of conventional coherence by providing both temporal and spectral resolution incorporating continuous wavelet transform. The highest variability is observed for Carbon Monoxide between two locations.

It is observed that the air and water pollutants in India crosses the International standards and if same pattern is followed then water can never be found suitable for consumption. Also, if same standard is followed then the quality of air becomes deteriorate and it may not be fit for breathing and may cause serious health effects. Government, external agencies, public bodies should take some initiatives to curb air and water pollution for better future.

ACKNOWLEDGMENTS

Author is thankful to Guru Gobind Singh Indraprastha University, Delhi (India) for providing research facilities.

REFERENCES

Bahrdwaj R., Srivastava K., (2014). Real time Nowcast of a Cloudburst and a Thunderstorm event with assimilation of Doppler Weather Radar data. Natural Hazards. 70(2), 1357-1383.

Bhardwaj R., (2016). Wavelets and Fractal Methods with environmental applications. *Mathematical Models, Methods and Applications.* Eds: Siddiqi, A. H., Manchanda, P., Bhardwaj, R.; 173-195.

Bhardwaj R., Kumar A., Maini P., Kar S. C., Rathore L. S., (2007). Bias free rainfall forecast and temperature trend based temperature forecast based upon T-170 Model during monsoon season. *Meteorological Applications.* 14(4), 351-360.

Bhardwaj R., Siddiqi A. H. and Mittal A., (2012). Predictability Index, Fractal Dimension & Hurst Exponent Estimation of Carbon Mono-Oxide at different locations of Delhi. *Indian Journal of Industrial & Applied Mathematics.* 3(2), 91-97.

Bhardwaj, R., Parmar, K. S., (2013). Wavelet and statistical analysis of river water quality parameters. *App. Math. Comput.*, 219 (20), 10172-10182.

Box, G. E. P., Jenkins G. M., Reinsel G. C.: *Time series analysis: Forecasting and Control*; 4th Edition. John Wiley & Sons; Inc. U.K., (2008).

Can Z, Aslan Z, Oguz O, Siddiqi A. H., (2005). Wavelet transform of metrological parameter and gravity waves. *Ann Geophys* 23:659–663.

CPCB, *Water Quality Status of Yamuna River (1999–2005)*. Central Pollution Control Board, Ministry of Environment & Forests, Assessment and Development of River Basin Series: ADSORBS/41/2006-07, (2006).

Doyle, M. E. and Barros, V. R.: Attribution of the river flow growth in the Plata Basin. *Int. J. Climatol*. 31 (15), 2234–2248, (2011).

Durai V. R., Bahrdwaj R., (2014). Evaluation of Statistical Bias Correction Methods for Numerical Weather Prediction Model (NWP) Forecasts of Maximum and Minimum Temperatures. *Natural Hazards*. 73(3), 1229-1254.

Durai V. R., Bahrdwaj R., (2014). Forecasting Quantitative Rainfall over India using Multi-Model Ensemble Technique. *Meteorology and Atmospheric Physics*. 126; 31-48.

Durai V. R., Bahrdwaj R., (2014). Location Specific Forecasting of Maximum and Minimum Temperature over India by Using the Statistical Bias Corrected Output of Global Forecasting System. *Journal of Earth System Science.* 123(5), 1171-1195.

Fischer, M. M. and Griffith, D. A.: Modeling Spatial Autocorrelation in Spatial Interaction Data: An Application to Patent Citation Data In The European Union. *Journal of Regional Science*, 48(5), 969–989, (2008).

Kocak, K., Saylan, L., and Sen, O.: Nonlinear time series prediction of O_3 concentration in Istanbul. *Atmospheric Environment*, 34(8), 1267-1271, (2000).

Korashey, R.: Using Regression Analysis to Estimate Water Quality Constituents in Bahr El Baqar Drain. *J. Applied Sc. Research*, 5 (8), 1067-1076, (2009).

Lupikasza, E.: Spatial and temporal variability of extreme precipitation in Poland in the period (1951–2006). *Int. J. Climatol*. 30, 991–1007, (2010).

McCleary, R. and R. A. Hay: *Applied Time Series Analysis for the Social Sciences*, Beverly Hills, Sage, (1980).

Park, J., Park, C.: Robust estimation of the Hurst parameter and selection of an onset scaling. *Stat. Sinica*, 19 (4), 1531-1555, (2009).

Rangarajan, G., Sant, D. A.: Fractal dimensional analysis of Indian climatic Dynamics. *Chaos, Solitons and Fractals*, 19 (2), 285–291, (2004).

Rehman S., Siddiqi A. H., (2009). "Wavelet based Hurst exponent and fractal dimensional analysis of Saudi climatic dynamics." *Chaos, Solitons and Fractals*, 40, 1081-1090.

WHO: *International standards for drinking water*. World Health Organization, Geneva, (1971).

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 6

ASSESSMENT OF WATER QUALITY MODEL USING FUZZY LOGIC SYSTEM: A CASE STUDY OF SURFACE WATER RESOURCES IN YALOVA OF TURKEY

Funda Dökmen[1,*] and N. Duru[2,†]
[1]Kocaeli University, Food and Agricultural Vocational School,
Campus of Arslanbey, Kartepe-Kocaeli, Turkey
[2]Kocaeli University, Faculty of Engineering, Department of Computer,
Campus of Umuttepe, Kocaeli, Turkey

ABSTRACT

Water quality management has become an important issue in public policies and agricultural activities all over the world. Proper assessment of water quality status in surface water resources based on different observations and analyses of chemical parameters. Various classification methods have been used for estimating of usability of surface water resources. This paper presents a comparative study, a methodology based on fuzzy inference system to assess water quality is proposed. Using fuzzy logic system was compared the output generated by fuzzy with that of conventional methods. Three variables are employed for the quality assessment in terms of six different surface water resources in Yalova region. Fuzzification module will create of fuzzy set, based on fuzzy variable respectively magnesium (Mg^{+2}), calsium (Ca^{+2}) and sodium (Na^{+2}). They were taken as input variables on the model. Because, they effect the usability of agricultural activities and soil productivity. For this reason, usability was taken as an output variable in return. This research purposed to determine the quality of water resources in terms of salinity used in agricultural areas and soils selected in surface water resources of Yalova vicinity in Türkiye. According to results, find out the effects of these waters on

* Corresponding Author's E-mail: f_dokmen@hotmail.com and funda.dokmen@kocaeli.edu.tr.
† Corresponding Author's E-mail: nduru@kocaeli.edu.tr.

productivity and quality using Fuzzy Logic, which have recently been used in scientific agricultural studies relation to water management. The development of the fuzzy logic system with surface water resources are explained and discussed that better method to include water quality parameters in terms of salinity for water quality due to qualification of fuzzy logic.

Keywords: fuzzy system, monitoring, logic assessment, salinity management

INTRODUCTION

The agriculture sector is much more water user in many countries. Improving irrigation water management and increasing water productivity has become one of the research priorities in irrigation field (Barker et al. 2000).

According to the Food and Agriculture Organization, lands of the world is affected by salinity or sodicity for over 6% as totally. The term *salt-affected* refers to soils that are saline or sodic, and these are approximately 400 million hectares (Table 1). 45 million ha are salt-affected and 32 million are salt-affected to varying degrees all over the world (FAO, 2008). Salinity occurs generally natural or human and agricultural activities as biochemistry processes that result in the accumulation of dissolved salts in the soil water to an extent that prevent to plant growth. Sodicity is a secondary result of salinity in clay soils.

Table 1. Regional distribution of salt-affected soils, in million hectares

Regions	Total area	Saline soils		Sodic soils	
	Mha	Mha	%	Mha	%
Africa	1,899	39	2.0	34	1.8
Asia, the Pacific and Australia	3,107	195	6.3	249	8.0
Europe	2,011	7	0.3	73	3.6
Latin America	2,039	61	3.0	51	2.5
Near East	1,802	92	5.1	14	0.8
North America	1,924	5	0.2	15	0.8
Total	12,781	397	3.1%	434	3.4%

Source: FAO Land and Plant Nutrition Management Service.

Irrigation water adds considerable amounts of salt, even with good quality irrigation water containing only 200-500 mg L^{-1} of soluble salt. The amount of salt removed by crops is insignificant. However, more amounts of salt will accumulate in the root zone, and must be leached by supplying more water than is required by the crops (FAO, 2008). (Quirck, 2001) studied the significance of salinity and sodicity for plant cultivation. The problem of salinity is the same in Türkiye's agricultural soils. Generally, wild irrigation effected increasing of salinity on the soils.

Many researchers have been studied to water resources for agricultural crop relation to water quality modeling (Chau et al. 2005; Chen and Chau 2006; Muttil and Chau 2006; Wu et al. 2009; Lin et al. 2010; Taormina et al: 2012; Dökmen and Aslan 2013). For this reason, the method is important in terms of modern agricultural science. The water quality assessment for agricultural cultivation are fuzzy term with multiple indicators and classes. Fuzzy theory found by Zadeh (1965) was intended to interpret the fuzziness of real conditions.

Duru et al. (2010) studied soil productivity analysis by using Fuzzy Logic System. They were applied pH, salinity, lime and organic matter values of different soil types by Fuzzy Logic analysis. Raman et al. (2009) declared about fuzzy logic, if used logically, could be an effective material for some of the environmental and agricultural decision matters. Landwehr (1979) and Semiromi (2011) studied for water quality development and some physical, chemical and biological parameters of water quality range using fuzzy logic.

On this subject, some different methodologies have shown up from heuristic approach. Fuzzy logic has been experimented with real environmental issues (Chen and Chang, 2001). Fuzzy logic take into consideration by using mathematical language and their formalism. It can be translated sophisticated logical statements by mathematical equation and more rules. In addition, it means that fuzzy logic can deal with highly variable, linguistic, vague and uncertain data (Semiromi, 2011). A suitable environmental and agricultural application of inference system based on fuzzy reasoning to integrate water quality determinants has been shown on the several studies by researchers.

MATERIAL AND METHOD

Material

Yalova is located in north-west of Türkiye, and the part of southeast in Marmara region. There are Marmara Sea in the north and west of province, and Kocaeli is in the east, Bursa (Orhangazi-Gemlik) and Gemlik Gulf are in the south. Yalova is between 30-40^0 nort latitude and 29-61^0 east longitude. It is two meters high from sea level.

In this study, six different surface water resources located in the near vicinity of Yalova (at three study areas; Taşköprü, Çiftlikköy and Altınova) in Türkiye were investigated (Figure 1). The content of sodium (Na^{+2}), calcium (Ca^{+2}) and magnesium (Mg^{+2}) were examined for the relation to amount of salt. This work was accomplished analyzing six different surface water samples every month in the period of over all the 15th months (June 2010-August 2011).

In the region, annual mean temperature is 23.6^0C and precipitation is 808.4 mm $year^1$. In the long period years (1975-2016), total annual potential evaporation is 540.5

mm water year^{-1} and real evaporation of research area is 476.9 mm water (Anonymous, 2016).

Figure 1. Map of Yalova Province in Turkey.

Method

Chemical Analysis

The all samples were analyzed according to standard methods (APHA, 1985). Chemical analysis were used of sodium (Na^{+2}), calcium (Ca^{+2}) and magnesium (Mg^{+2}). FAO (Food Agricultural Organization) was taken into consideration for criteria of evaluation relation to analysis of laboratuvary in the quality of irrigation water (FAO, 1985).

Fuzyy Logic

Lotfi A. Zadeh (1965), was used Fuzzy Logic Idea as firstly that by way of expressing some uncertain expressions mathematically. It is based on the Fuzzy Set Theory by which the true and false values are determined. There is again the values of (1) and (0) here, as in classic logic (Figure 2.).

In addition, fuzzy logic goes beyond these values and by making use of intermediate values. It means that for example; not only if a certain distance is far or near, but also how far or near it is.

In this study with the use of Fuzzy Logic model, complex conclusions derived of various salinity parameters relation to water quality were evaluated all together. Fuzzy Logic and Matlab 2.2.22 modules (Fuzzy Logic Tollbox Version 2.2.22) were used in order to execute fuzzy logic applications.

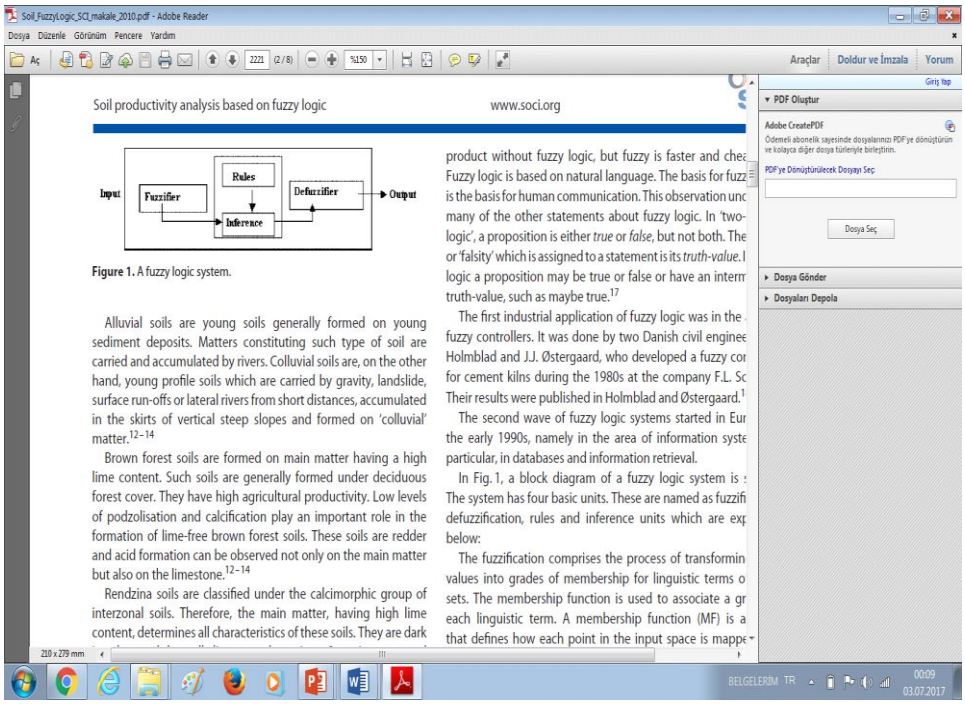


Figure 2. General Fuzzy Logic system.

The sequence of action in this stage can be explained as below:

It was defined what the input and output variables of the system will be parameters of salinity and usability. Sodium (Na^{+2}), calcium (Ca^{+2}) and magnesium (Mg^{+2}) values were taken as input variables, as usability for agricultural productivity was taken as an output variable. They are factors, which directly effect to quality of irrigation water and agricultural cultivations depend on amount of salt.

Membership functions and threshold values were determined belong to the input and output variables. Firstly, type of membership functions and the limits were determined.

In Figure 3, Membership functions for variables of Ca^{+2} is shown. Three membership functions was determined as Low, Med and High.

Like Ca^{+2}, for Mg^{+2}, Na^{+2} and Usability variables and the membership functions were recognized as seen in Figure 4-6 according to the reference values.

According to Figure 3, 4 and 5, values of parameters, which are Ca^{+2}, Mg^{+2} and Na^{+2}, were found between range of usability in water resources in terms of irrigation water quality. Figure 6 is shown that membership functions for output variables (usability).

After defining the membersip functions, the rules have beeen decided according to the experts' opinion.

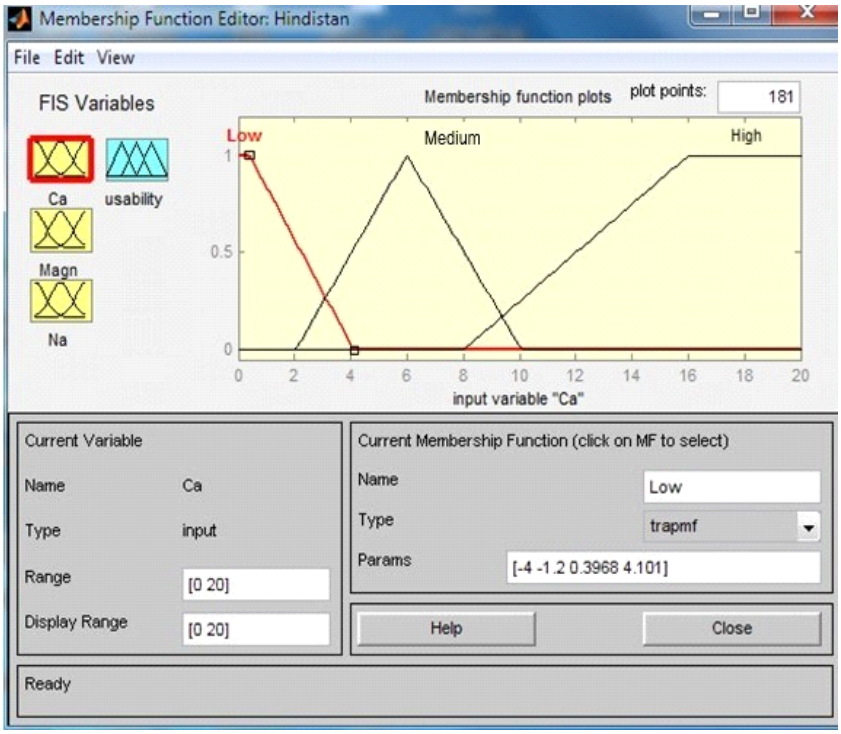


Figure 3. Membership functions for variables of Ca^{+2}.

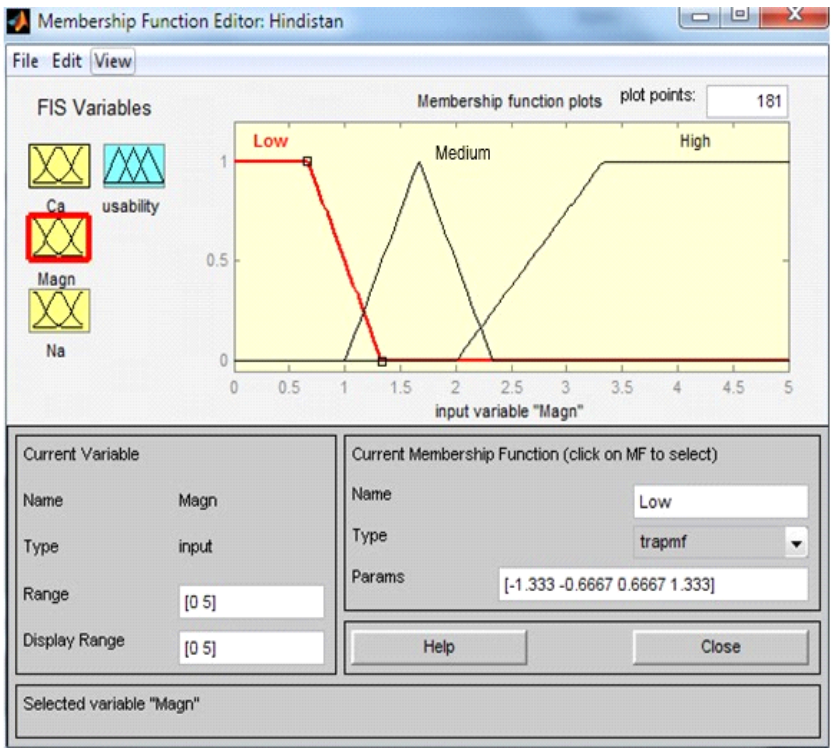


Figure 4. Membership functions for variables of Mg^{+2}.

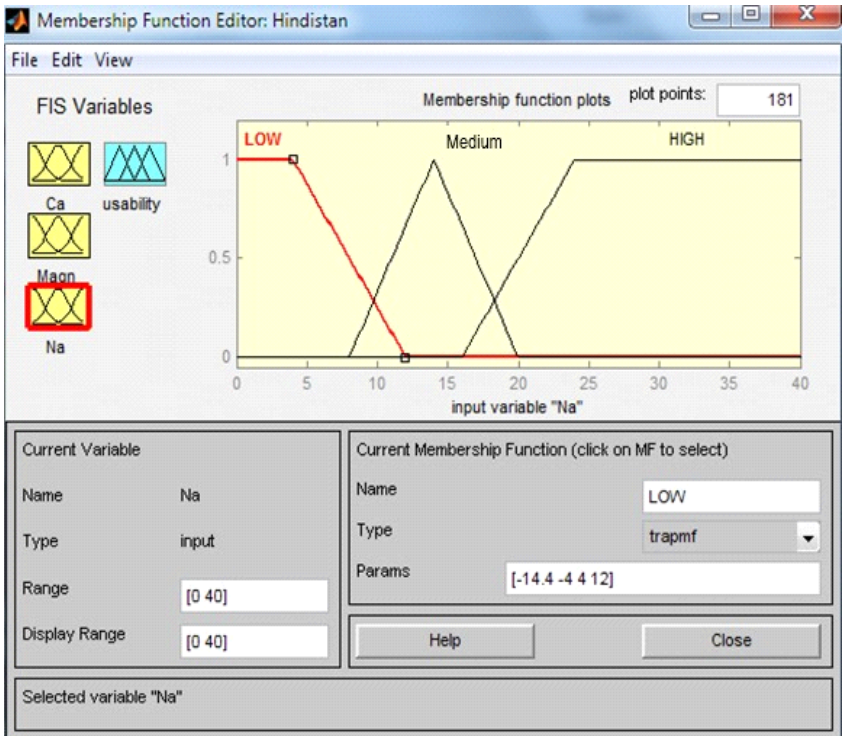


Figure 5. Membership functions for variables of Na^{+2}.

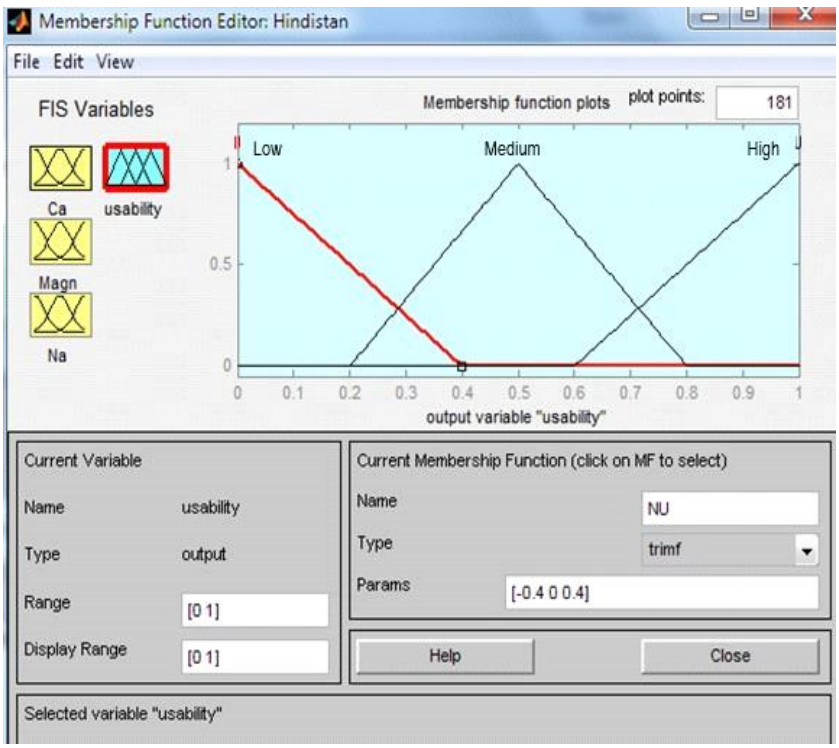


Figure 6. Membership functions for output variables (usability).

Rules have given as below:

1. IF Calcium is LOW AND Magnesium is MEDIUM AND Sodium is LOW THEN Usability is HIGH
2. IF Calcium is LOW AND Magnesium is MEDIUM AND Sodium is MEDIUM THEN Usability is MEDIUM
3. IF Calcium is LOW AND Magnesium is MEDIUM AND Sodium is HIGH THEN Usability is LOW
4. IF Calcium is LOW AND Magnesium is LOW AND Sodium is LOW THEN Usability is HIGH
5. IF Calcium is LOW AND Magnesium is LOW AND Sodium is MEDIUM THEN Usability is MEDIUM
6. IF Calcium is LOW AND Magnesium is LOW AND Sodium is HIGH THEN Usability is MEDIUM
7. IF Calcium is LOW AND Magnesium is HIGH AND Sodium is LOW THEN Usability is HIGH
8. IF Calcium is LOW AND Magnesium is HIGH AND Sodium is MEDIUM THEN Usability is LOW
9. IF Calcium is LOW AND Magnesium is HIGH AND Sodium is HIGH THEN Usability is LOW
10. IF Calcium is HIGH AND Magnesium is MEDIUM AND Sodium is LOW THEN Usability is LOW
11. IF Calcium is HIGH AND Magnesium is MEDIUM AND Sodium is MEDIUM THEN Usability isLOW
12. IF Calcium is HIGH AND Magnesium is MEDIUM AND Sodium is HIGH THEN Usability is LOW
13. IF Calcium is HIGH AND Magnesium is LOW AND Sodium is LOW THEN Usability is MEDIUM
14. IF Calcium is HIGH AND Magnesium is LOW AND Sodium is MEDIUM THEN Usability is LOW
15. IF Calcium is HIGH AND Magnesium is LOW AND Sodium is HIGH THEN Usability is LOW
16. IF Calcium is HIGH AND Magnesium is HIGH AND Sodium is LOW THEN Usability is LOW
17. IF Calcium is HIGH AND Magnesium is HIGH AND Sodium is MEDIUM THEN Usability is LOW
18. IF Calcium is HIGH AND Magnesium is HIGH AND Sodium is HIGH THEN Usability is LOW
19. IF Calcium is MEDIUM AND Magnesium is MEDIUM AND Sodium is LOW THEN Usability is LOW

20. IF Calcium is MEDIUM AND Magnesium is MEDIUM AND Sodium is MEDIUM THEN Usability is LOW
21. IF Calcium is MEDIUM AND Magnesium is MEDIUM AND Sodium is HIGH THEN Usability is LOW
22. IF Calcium is MEDIUM AND Magnesium is LOW AND Sodium is LOW THEN Usability is MEDIUM
23. IF Calcium is MEDIUM AND Magnesium is LOW AND Sodium is MEDIUM THEN Usability is LOW
24. IF Calcium is MEDIUM AND Magnesium is LOW AND Sodium is HIGH THEN Usability is LOW
25. IF Calcium is MEDIUM AND Magnesium is HIGH AND Sodium is LOW THEN Usability is LOW
26. IF Calcium is MEDIUM AND Magnesium is HIGH AND Sodium is MEDIUM THEN Usability is LOW
27. IF Calcium is MEDIUM AND Magnesium is HIGH AND Sodium is HIGH THEN Usability is LOW

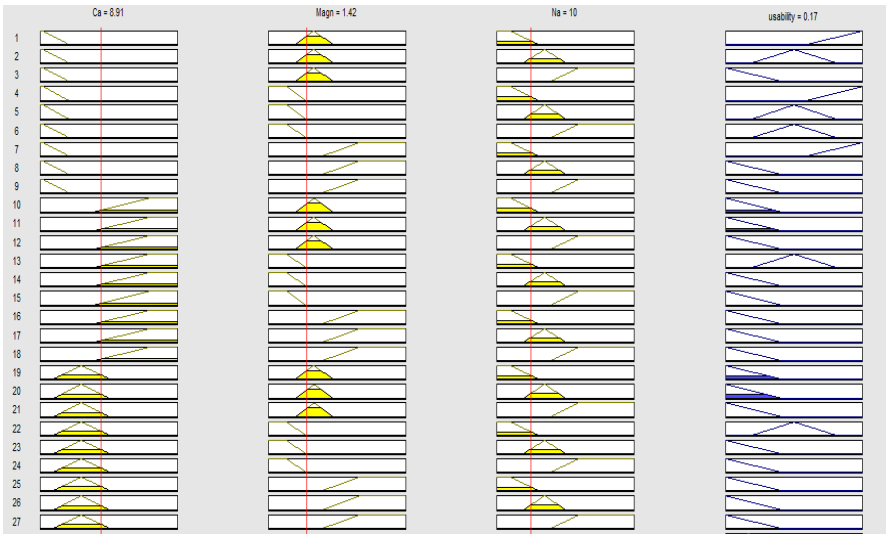


Figure 7. A case study of Ca = 8.91 and Na = 10 and Mg = 1.42.

In Figure 7, a case study of Ca=8.91 and Na=10 and Mg=1.42 are given. As seen, in this situation, usability is 0.17, that is mean 17% is obtained. It is shown that, some rules are fired while some of them not fired.

When applied all of chemical parameters' values (Ca = 8.91 and Na = 10 and Mg = 1.42), some rules are active and some rules are inactive from totaly 27 rules.

In this tread, in line with the interactions of input variables with each other, an effort was made for setting rules relation to the possible outcome. A total of 27 rules were

written for these parameters (magnesium (Mg^{+2}), calsium (Ca^{+2}) and sodium (Na^{+2})) groups.

RESULTS AND DISCUSSION

Figures 3, 4 and 5 are shown that input and output variables, membership functions and threshold values of the saltiness parameters relation to Ca^{+2}, Mg^{+2} and Na^{+2}.

In Figure 8 and 9, the 3D surface graphics is given obtained acoording to Ca^{+2}, Mg^{+2}, Na^{+2} and usabilitiy parameters from Matlab Fuzzy toolbox.

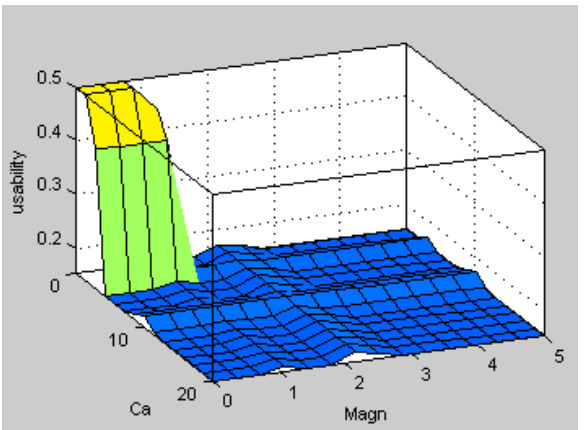


Figure 8. 3D demonstrations for the selected parameters that set up the rules are on Ca^{+2}, Mg^{+2} and usability.

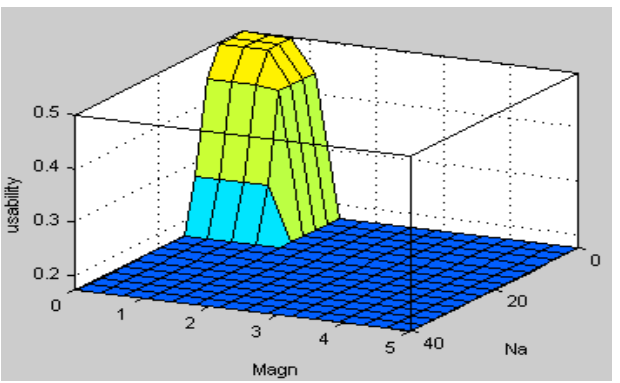


Figure 9. 3D demonstrations for relationship of Mg^{+2}, Na^{+2} and usability.

On Figure (8) are usability, the changes seen in Mg and Ca^{+2} in relation to each other; and on Figure (9) are usability and 3D indication of the rules that show the changes in Mg^{+2} and Na^{+2} in relation to each other. The system evaluates three different input and

output variables together, but only the rules concerning three selected variables can be seen on the 3D indication.

Table 2. Surface water resources in Altınova (No. 1)

Months	**Na^{+2}** **meq L-1**	**Mg^{+2}** **meq L-1**	**Ca^{+2}** **meq L-1**	**Usability range**
June	0,33	2,4	0,4	
July	1	2,08	0,72	0,838
August	0	0	0	
September	0	0	0	
October	0,95	2,4	2	
November	1,05	2,4	3,2	
December	1,2	1,2	2,48	
January	0,85	1,2	3,2	0,831
February	0	0	0	
March	0,28	0,56	2,8	
April	1,15	1	2,4	0,847
May	0,6	0,88	3,92	
June	0,7	2,48	2	
July	1,1	1,92	2,24	
August	0	0	0	

Table 3. Surface water resources in Taşköprü (No. 2)

Months	**Na^{+2}** **meq L-1**	**Mg^{+2}** **meq L-1**	**Ca^{+2}** **meq L-1**	**Usability range**
June	0,8	2,4	2	
July	1	2,08	0,72	0,838
August	0	0	0	
September	0	0	0	
October	1,4	1,6	2,8	
November	1,1	3,04	3,6	
December	1,25	1	3	
January	1,95	1,2	3,2	0,831
February	1,2	0,72	3,08	
March	0,5	0,24	2,96	
April	1,4	1,24	2,08	0,836
May	0,7	2,2	3,2	
June	0,7	2,4	2,4	
July	0,7	2,3	2,8	
August	0,7	2,35	2,6	

Table 4. Surface water resources in Altınova (No. 3)

Months	**Na^{+2} meq L-1**	**Mg^{+2} meq L-1**	**Ca^{+2} meq L-1**	**Usability range**
June	0,75	2,8	2,8	
July	0,96	2,07	4,2	0,839
August	0	0	0	
September	0	0	0	
October	1,17	1,35	5,6	
November	1	1,6	5,2	
December	1,35	1,1	6,1	
January	1	1,2	3,6	0,831
February	0,7	1,16	3,84	
March	0,4	1,12	4,08	
April	1,25	0,96	3,92	0,851
May	0,4	1,6	4,4	
June	0,8	2,6	3,6	
July	0,6	2,1	4	
August	0,7	2,35	3,8	

Table 5. Surface water resources in Çiftlikköy (No. 4)

Months	**Na^{+2} meq L-1**	**Mg^{+2} meq L-1**	**Ca^{+2} meq L-1**	**Usability range**
June	0,6	2	3,2	
July	1	2,08	0,72	0,838
August	0	0	0	
September	0	0	0	
October	1	1,8	6	
November	0,85	2	6	
December	1,15	1,6	6	
January	1	1,2	4,4	0,831
February	0,66	1,48	4	
March	0,33	1,76	3,6	
April	1,35	1,76	3,2	0.862
May	0,4	2	4	
June	0,6	2,72	2,88	
July	0,5	2,36	1,72	
August	0,5	2,54	2,3	

As seen in Tables 1, 2, 3, 4, 5, 6 and 7 for July, December and April, the measured values from different sources were applied to the fuzzy logic system and usability values were obtained. The quality of water affects soil and yield quality relationship agricultural activities. All through the research, the irrigation water to be applied to agricultural area

has been water quality examined and, salinity analyses were made both for the surface water resources.

Table 6. Surface water resources in Çiftlikköy (No. 5)

Months	**Na^{+2} meq L-1**	**Mg^{+2} meq L-1**	**Ca^{+2} meq L-1**	**Usability range**
June	1,35	2,8	2,8	
July	1,95	2,16	2,6	0,827
August	0	0	0	
September	0	0	0	
October	1,4	3,2	3,2	
November	1,25	0,4	5,6	
December	1,7	2,4	4	
January	1,35	0,8	4	0,862
February	0,8	0,44	4,4	
March	0,4	0,08	4,8	
April	1,4	1,52	3,68	0,862
May	1,1	0,8	4,8	
June	0,95	2,48	3,26	
July	1,65	1,28	4,8	
August	1,3	1,88	4,08	

Table 7. Surface water resources in Çiftlikköy (No. 6)

Months	**Na^{+2} meq L^{-1}**	**Mg^{+2} meq L^{-1}**	**Ca^{+2} meq L^{-1}**	**Usability range**
June	0,25	0,08	1,52	
July	0,28	0,44	0,88	0,87
August	0	0	0	
September	0	0	0	
October	0,55	0,6	1,6	
November	0,75	1,2	2	
December	0,6	0,7	1,7	
January	0,45	0,2	1,4	0,87
February	0,33	0,3	1,22	
March	0,22	0,4	1,04	
April	2,65	0,64	7,36	0,703
May	1,25	0,32	5,6	
June	1,7	0,8	1,2	
July	0,3	0,71	0,96	
August	1	0,75	1,08	

CONCLUSION

Fuzzy logic provides a framework to model uncertainty, the human-brain way of thinking, reasoning and apprehension process. The fuzzy logic for deciding the water quality parameters based on which, water quality rankings and its usability are given to determine the quality of water. The fuzzy analysis has been shown to be effective in avoiding the loss or non-detection of information considerable for classification and ranking of water quality in terms of salinity. As a finally, some salinity parameters were evaluated all together by using Fuzzy Logic Model relation to mixed conclusions. It was seen that Fuzzy Logic model which is made use of in water science and its quality fields for evaluation range. On the other, this model can also be used in water quality studies recommended in the agricultural water management.

REFERENCES

Anonymous, (2016). *Yalova Meteoroloji İstasyonu Kayıtları* [*Recorders of Yalova Wheather Station*], Yalova.

APHA, (1985). *Standart Methods for the Examination of Water and Wastewater American Public Health Association*, Washington.

Barker, R., Dawe, D., Tuong, T. P., Bhuiyan, S. I., and Guerrac, L. C. (2000). The outlook for water resources in the year 2020: challenges for research on water management in rice production. *Proceedings of the Nineteenth session of the International Rice commission in September 7–9*, 1998, Cairo, Egypt, pp 133–142.

Chau, K. W., Wu, C. L., and Li, Y. S. (2005). Comparison of several flood-forecasting models in Yangtze River, *J Hydrol Eng ASCE* 10(6): 485-491.

Chen, H. W., and Chang, N. B. (2001). Identification of river water quality using the Fuzzy Synthetic Evaluation approach. *J. Environ. Manage.*, 63(3): 293–305.

Chen, K. W., and Chau, K. W. (2006). Intelligent manipulation and calibration of parameters for hydrological models. *Int J Environ Pollut* 28(3-4):432-447.

Dökmen, F., and Aslan, Z. (2013). Evaluation of the parameters of water quality with wavelet techniques. *Water Resour Manage* 27(14):4977-4988.

Duru, N., Dökmen, F., Canbay, M. M., and Kurtuluş, C. (2010). Soil productivity analysis based on a Fuzzy Logic System, *Journal of the Science of Food and Agriculture*, 90(13): 2220-2227.

FAO, 2008. *FAO, Land and Plant Nutrition Management Service*, http://www.fao.org/ag/agl/agll/spush/ Access date: 08.03.2018.

Landwehr, J. M. (1979). A statistic view of a class of water quality indices. *Water Resour. Res.*, 15(2): 460-468.

Lin, C. H., Huang, T. H., and Shaw, D. (2010). Applying water quality modeling to regulating land development in a watershed. *Water Resour Manage* 24(4): 629-640.

MATLAB, 2018. *Fuzzy Logic Toolbox Version 2.2.22*, University of Kocaeli, Turkey.

Muttil, N., and Chau, K. W. (2006). Neural network and genetic programming for modelling coastal algal blooms. *Int J Environ Pollut* 28 83-4):223-238.

Quirk, JP. (2001). The significance of the treshold and turbidity concentrations in relation to sodicity and microstructure. *Aust J Soil Res,* 39: 1185-1217.

Raman, B. V., Reinier, B., and Mohan, S. (2009). Fuzzy logic Water Quality Index and importance of Water Quality Parameters. *Air, Soil Water Resources*, 2, 51-59.

Semiromi, B. F., Hassani, A. H., Torabian, A., Karbassi, A. R., and Hosseinzadeh, L. F. (2011). Water quality index development using fuzzy logic: A case study of the Karon River of Iran. *African Journal of Biotechnology*, Vol. 10 (50), pp. 10125-10133.

Taormina, R., Chau, K. W., and Sethib, R. (2012). Artificial Neural Network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng Appl Artif Intel* 25(8):1670-1676.

Wu, CL., Chau, K W., and Li, YS. (2009). Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resour Res* 45, W08432.doi:10.1029/2007WR006737.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control,* 8, 338-353.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 7

SENSITIVITY OF REGCM4.2 CONVECTION SCHEMES OVER NILE BASIN DOMAIN

Mahmoud Gaber Aly[1,*], MD, Mohamed M. Abdel Wahab[2], PhD and Asharf S. Zaky[3], PhD

[1]Nile Forecast Center, Planning Sector,
Egyptian Ministry of Water Resources and Irrigation, Giza, Egypt
[2]Astronomy, Space Science and Meteorology Department,
Faculty of Science, Cairo University, Giza, Egypt
[3]Egyptian Meteorological Authority, Cairo, Egypt

ABSTRACT

Nile River represents a major source of fresh water in Egypt. Climate change and climate variability have become the focus of attention on Nile River and its tributaries. Regional Climate Models are one of important tools for investigation of climate. The representation of Precipitation is very significant because it is one of the most important elements in climate system. We compared 4 convection schemes (Kuo – Grell with 2 closures (Arakawa Schubert and Fritsch–Chappell) – Emanuel – Tiedtke) quantitative comparison (total precipitation – convective precipitation and surface temperature) For River Nile Domain (Blue Nile region and Equatorial Lakes region) to find the best convection scheme valid for interested region. Tiedtke convection scheme was added to convection scheme sensitivity comparison. For more investigation all variables will be fixed except convection scheme using Boundary condition ECMWF - EIN15 reanalysis was validated by using CRU (surface temperature, total precipitation) and TRMM convective precipitation data. We found that Grell convection scheme as most often was more suitable in results and less bias compared with observation specifically Grell_AS was the best especially for equatorial lakes region where Grell_FC

* Corresponding Author's E-mail: mahmoudgaber@mwri.gov.eg.

give much convective especially for Blue Nile region. For Emanuel convection scheme was always overestimated for convective and total precipitation, the bias reaches to 100% or may be higher compared to observation. Kuo always underestimated for total precipitation nevertheless it was gave better result for convective precipitation which indicates that the lack may be due to that the scheme cannot handle the non-convective precipitation. For Tiedtke although it belongs to mass flux convection schemes but it was unlike Emanuel. Tiedtke as most often was under estimated compared with observation but the underestimation is not high compared to Emanuel over estimation.

Keywords: convections schemes, rainfall, Nile Basin, regional climate models, climate

INTRODUCTION

River Nile is too long to be one of the longest all over the world it is about 6,695 kilometers long starting from Kagera Basin to Delta of Egypt. It covers about 3.2 million kilometers squares area (about 10% of the area of Africa and about 2.3% of the surface area of the world) and it has significant latitudinal spreads extending from 32°N to 4°S which spread between 35 degrees of latitude. Understanding climate and hydro-meteorology of Nile River was limited due to insufficient data for climate and hydrology. Egypt's climate is generally dry so it has precipitation scarcity therefore the major source for fresh water is river Nile. It is the major supply for economy, agriculture, humanitarian activities and livestock, this explain why Egypt's water interests go beyond its borders. Nile River climate is very complicated to represent on different scales, local or large scales due to strong latitudinal gradient.

Local simulation of precipitation process is considered as one of the important topics handled by the RCMs because of its importance in controlling the climate predictions on different time scales, which is extremely important to various end user sectors. Analysis of this problem can be analyzed from two points of view: model parameterization and model evaluation. Precipitation parameterization mainly depends on the sub-grid precipitation processes and the resolved large-scale physics; however, this relationship is not totally understood. In addition, lack of high-resolution observations, both spatially and temporally is a difficult problem for the evaluation process (Vaidya 2006). In addition, modeling of precipitation in the tropics remains an unresolved challenge. Majority of the schemes are developed and tested in mid-latitude regions, where precipitation forcing differs from its corresponding in tropics. Recently (Houze 1997; Schumacher and Houze 2003) have given some insight into the forcing and structure of tropical rainfall. By utilizing these insights, parameterization can be improved to provide a good representation of tropical rainfall (Bowden and Semazzi 2009).

Neil Davis and Jared Bowden (2009) have compared between various convection schemes with observational dataset focusing on the amount and partitioning of convective and stratiform precipitation. First, the best convective scheme for the region is

determined, and then relative humidity threshold (at which clouds begin to form RHmin) and the auto-conversion scale factor Cacs in the SUBEX large-scale precipitation scheme were adjusted to provide the best configuration in this region. Both, Arakawa and Schubert closure Grell (Grell–AS) and Grell–FC showed underestimation over land, while Grell–FC (Grell–AS) showed overestimation over the ocean. Emanuel convection scheme (shortly MIT–EMAN) provides very good results from the spatial perspective (Neil Davis and Jared Bowden, 2009.

AIM OF PRESENT WORK

For our study we used RegCM4.2 with four convection schemes (Kuo – Grell with 2 closures (Arakawa Schubert, Fritsch–Chappell) – Emanuel – and the most recent scheme added to model Tiedtke), Quantitative comparison (total and convective precipitation) which have been compared with observation for our interested region.

Tiedtke convection scheme was added to convection scheme sensitivity comparison. For more investigation, all variables will be fixed except convection scheme using ECMWF - EIN15 reanalysis data from ERA interim for three-year period 2003-2005. Boundary conditions have been validated by using CRU (surface temperature, total precipitation) and TRMM convective precipitation data.

MODEL DESCRIPTION

In this study the International Centre for Theoretical Physics (ICTP) Regional Climate Model system version 4 (RegCM4) is used in this study and it is customized over East Africa (Sun et al. 1999 a, b). It adopts the following physical configuration:

1) National Center for Atmospheric Research (NCAR) Community Climate Model Version 3 (CCM3) radiative transfer as the radiation scheme.
2) The Biosphere–Atmosphere Transfer Scheme (BATS 1e; Dickinson et al. 1993) is used as the land surface scheme.
3) Holtslag et al. (1990) as the boundary layer (PBL) scheme
4) The large-scale precipitation scheme and several convective parameterization scheme options in the standard version of the model are briefly discussed below.
5) Boundary conditions and other specifications

RegCM4 requires initial and boundary conditions for wind, temperature, surface pressure, and water vapor, as well as an oceanic surface forcing. For the atmospheric

variables, we use ECMWF - EIN15 reanalysis data from ERA interim which it has spatial resolution 1.5° × 1.5°; the optimum interpolation (OI) sea surface temperature (SST) analysis is produced weekly on a one-degree grid. The analysis uses in situ and satellite SST's and SST's simulated by sea-ice cover.

Before the analysis is computed, the satellite data is adjusted for biases using the method of Reynolds (1988), Reynolds, and Marsico (1993). RegCM4 has two options for the ocean flux. This study adopts Zeng ocean flux scheme, Soil moisture and temperature were initialized using the standard RegCM4 lookup tables. The model is run using 18 vertical levels with a model top pressure of 50 hPa and 50-km grid spacing. The run, starting from the beginning of 2002 January and ending 2006 January 1 year was discarded for spin up.

6) Validation data

a). CRU TS 3.10

Monthly total precipitation is evaluated using the CRU (Climatic Research Unit) dataset (TS - version 3.10), which is available on a resolution of 0.5o x 0.5o resolution over the land surface based on upscaling of gauge stations around the world.

b). TRMM precipitation data

Satellite derived dataset for estimation the amount of surface precipitation rate on a resolution of 0.5 x 0.5 over both land and ocean. For model evaluation, the ocean has been masked.

7) Results

Three-year climatological average for JAS season and monthly mean of different convection schemes are tested over Blue Nile and Equatorial lakes regions.

a). Climatology averages

Figure 1 shows total precipitation and convective precipitation for different convection schemes, CRU total precipitation and TRMM convective precipitation. These figures show wide variation in total and convective precipitation for different convection scheme. For Emanuel convection scheme C.S creates precipitation scale ranges from 8 mm/day to 44 mm/day for Blue Nile region and up to 12 mm/day for Equatorial lakes region. While convective precipitation, which has been obtained directly from the model, was about from 4 mm/day up to 24 mm/day for Blue Nile region and up to 12 mm/day for equatorial lakes region.

Kuo C.S is simple model based on moisture convergence closure, which it well formulated for tropical region and application of low resolution (Mesoscale Meteorological Modeling -Roger A. Pielke, Sr 2002). Kuo was good comparing to convective precipitation as shown in Figure 2(i) it almost matches with observation but for total precipitation Kuo was under estimated as shown in Figure 2(d) may be return to lack of creation non convective rainfall as shown in Figure 7.

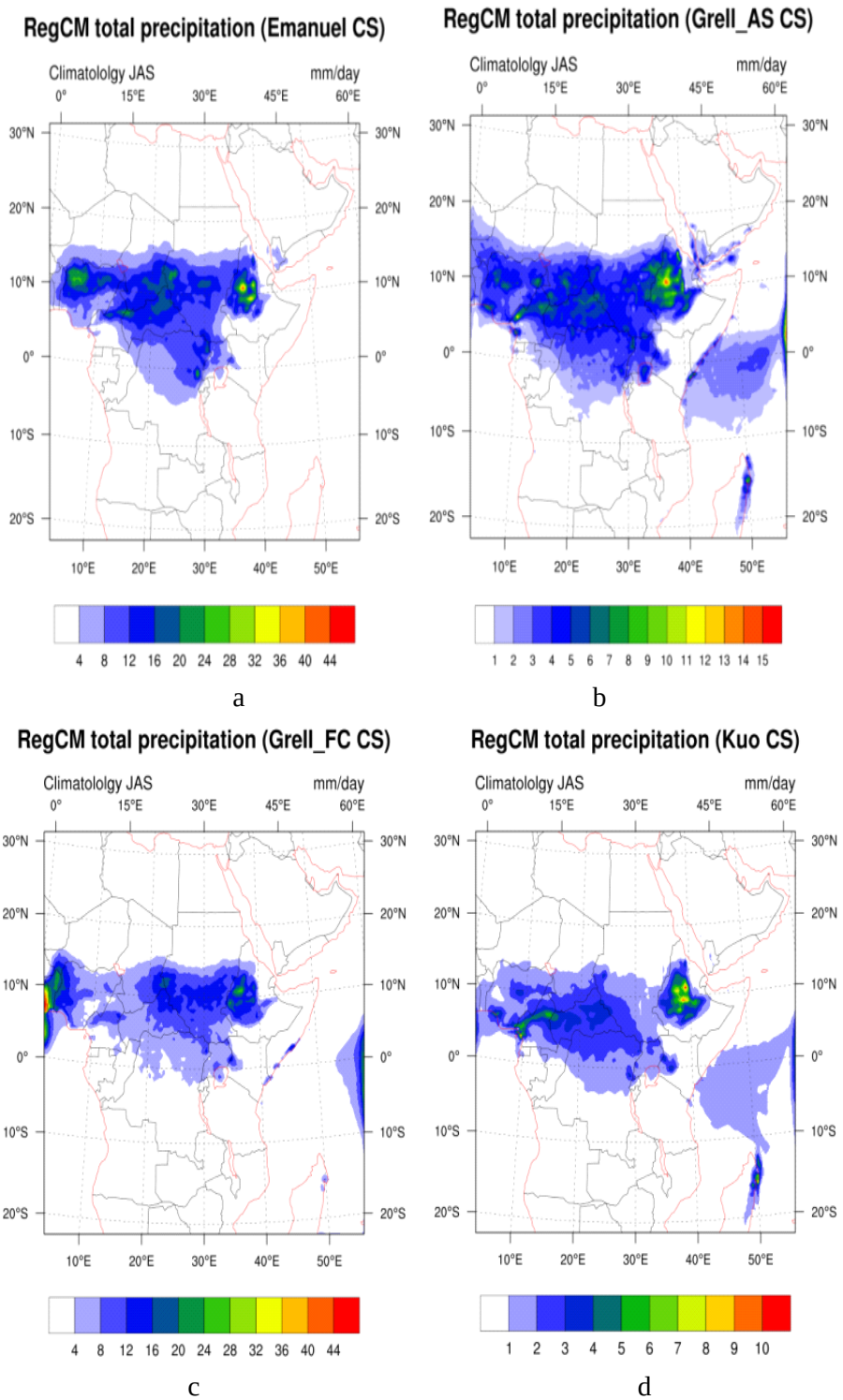


Figure 1. (a, b, c, d). Total precipitation for different convection schemes (Emanuel, Grell_AS, Grell_FC, Kuo, Tiedtke).

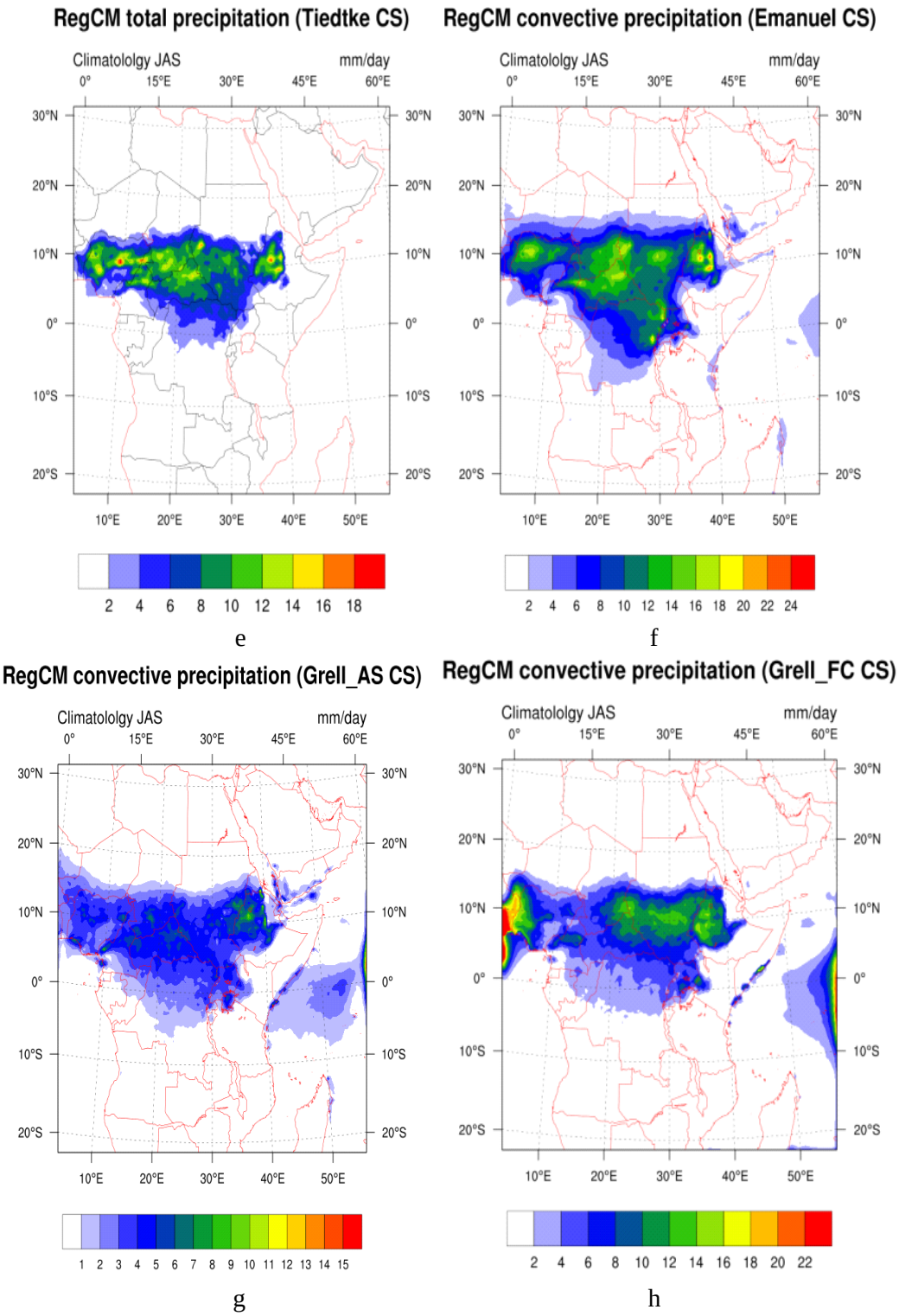


Figure 1. (e, f, g, h). Total precipitation for Tiedtke C.S and convective precipitation for different convection schemes (Emanuel, Grell_AS, Grell_FC).

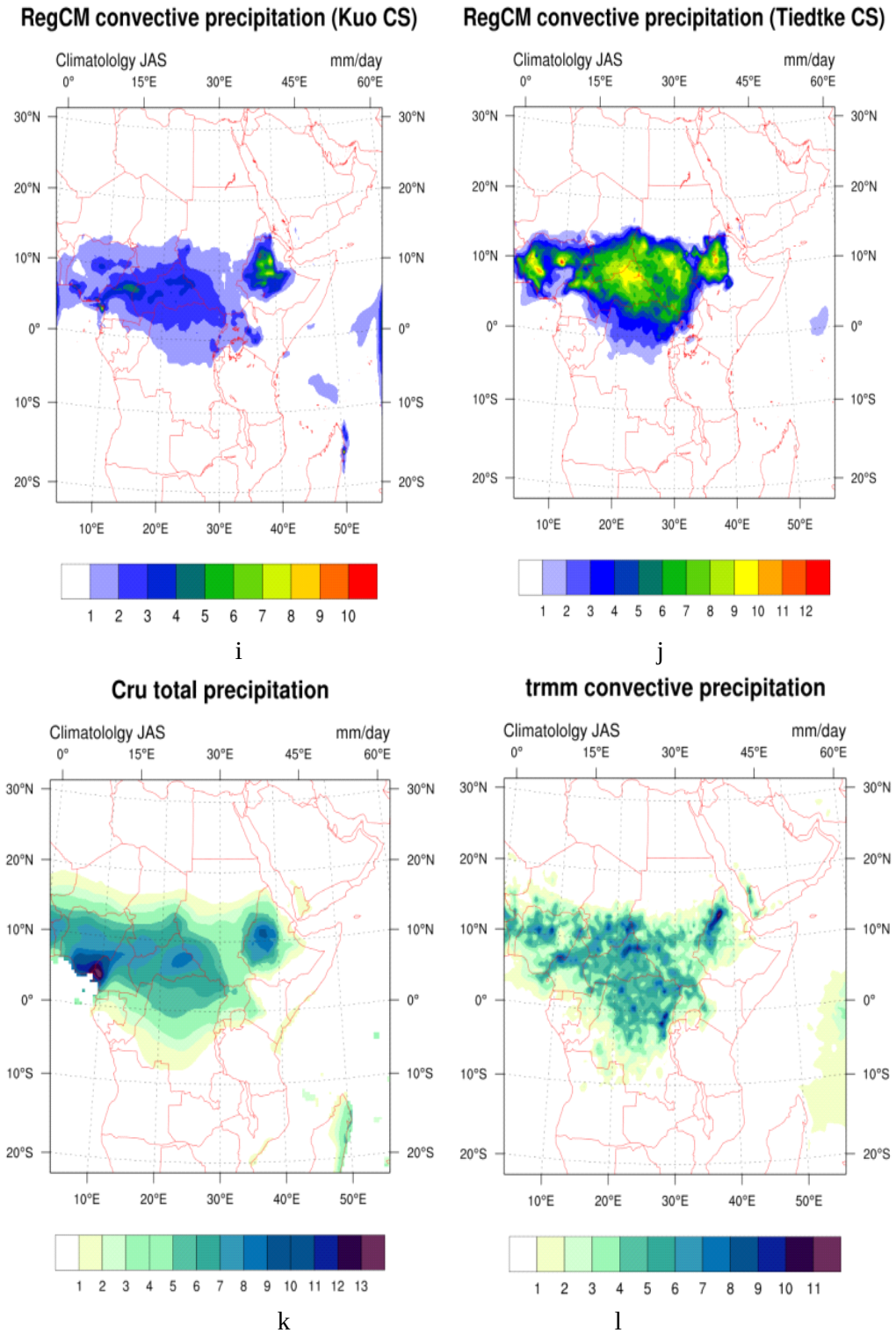


Figure 1. (i, j). Convective precipitation for different convection schemes (Kuo, Tiedtke) and (k,l) show observed total precipitation cru and observed convective precipitation trmm.

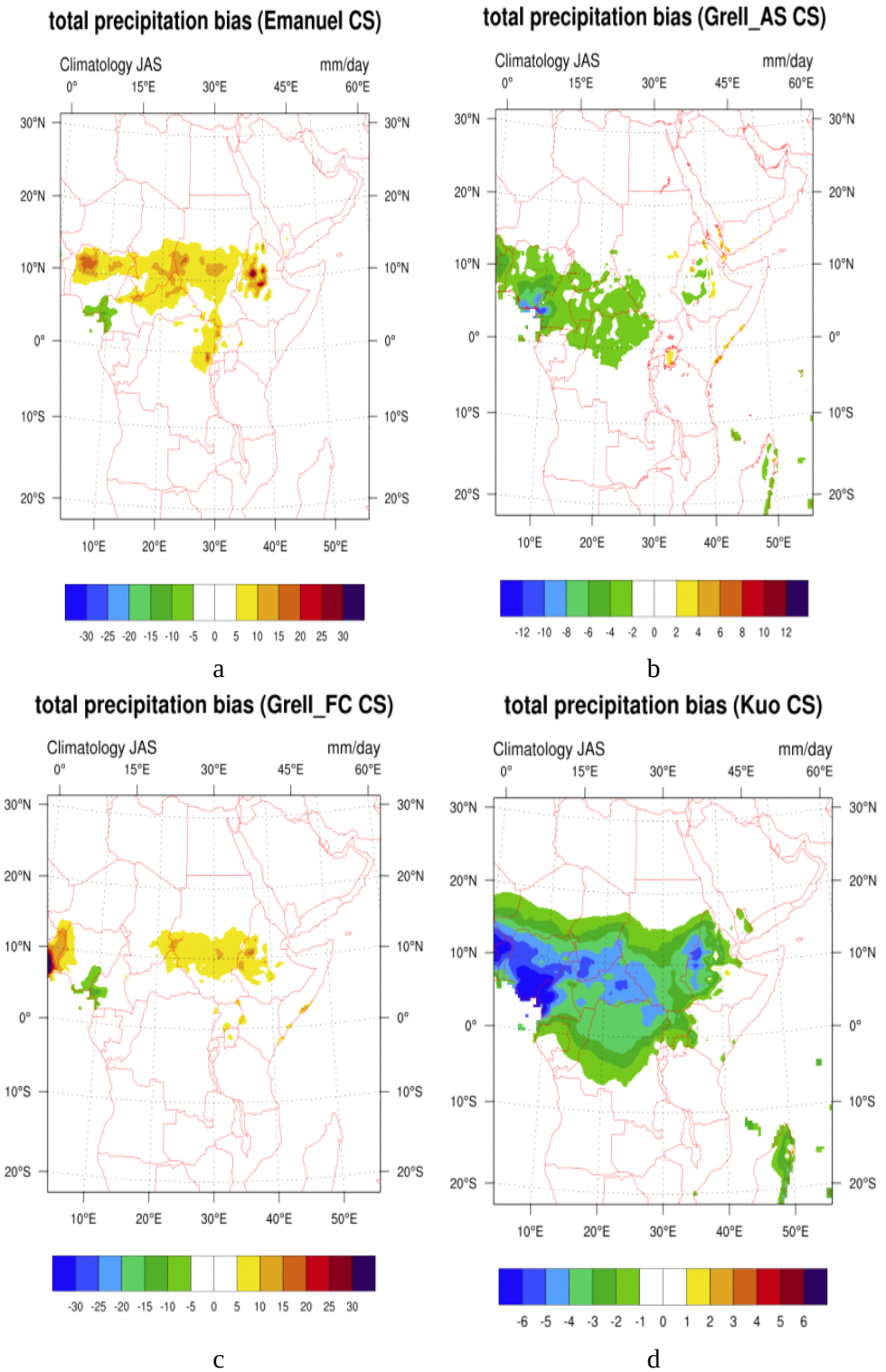
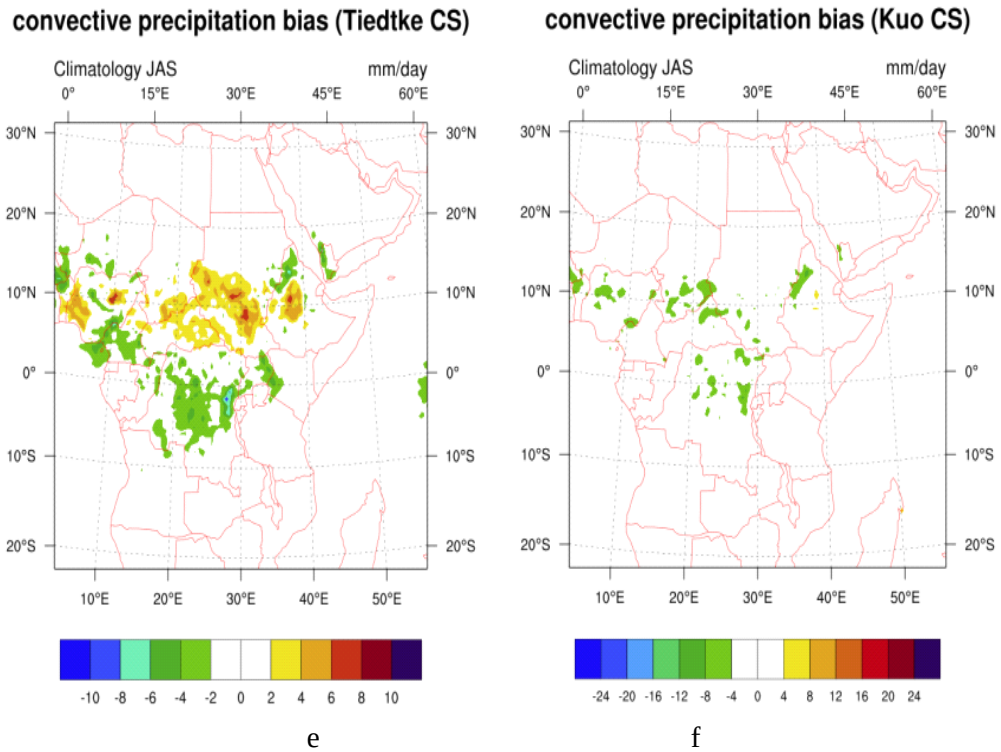


Figure 2. (a, b, c, d). Total precipitation bias for different convection schemes (Emanuel, Grell_AS, Grell_FC, Kuo).



e f

Figure 2. (e, f). Convective precipitation bias for different convection schemes (Kuo, Tiedtke).

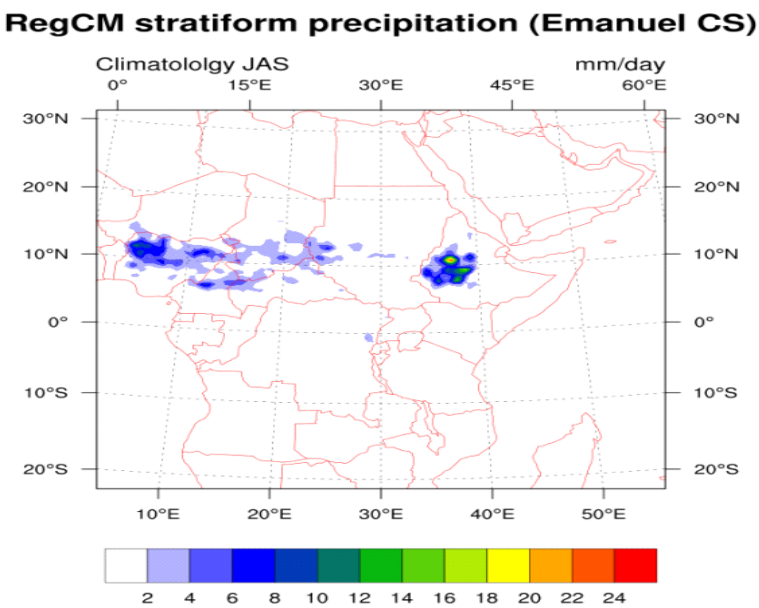


Figure 3. Non-convective rainfall created by Emanuel C.S.

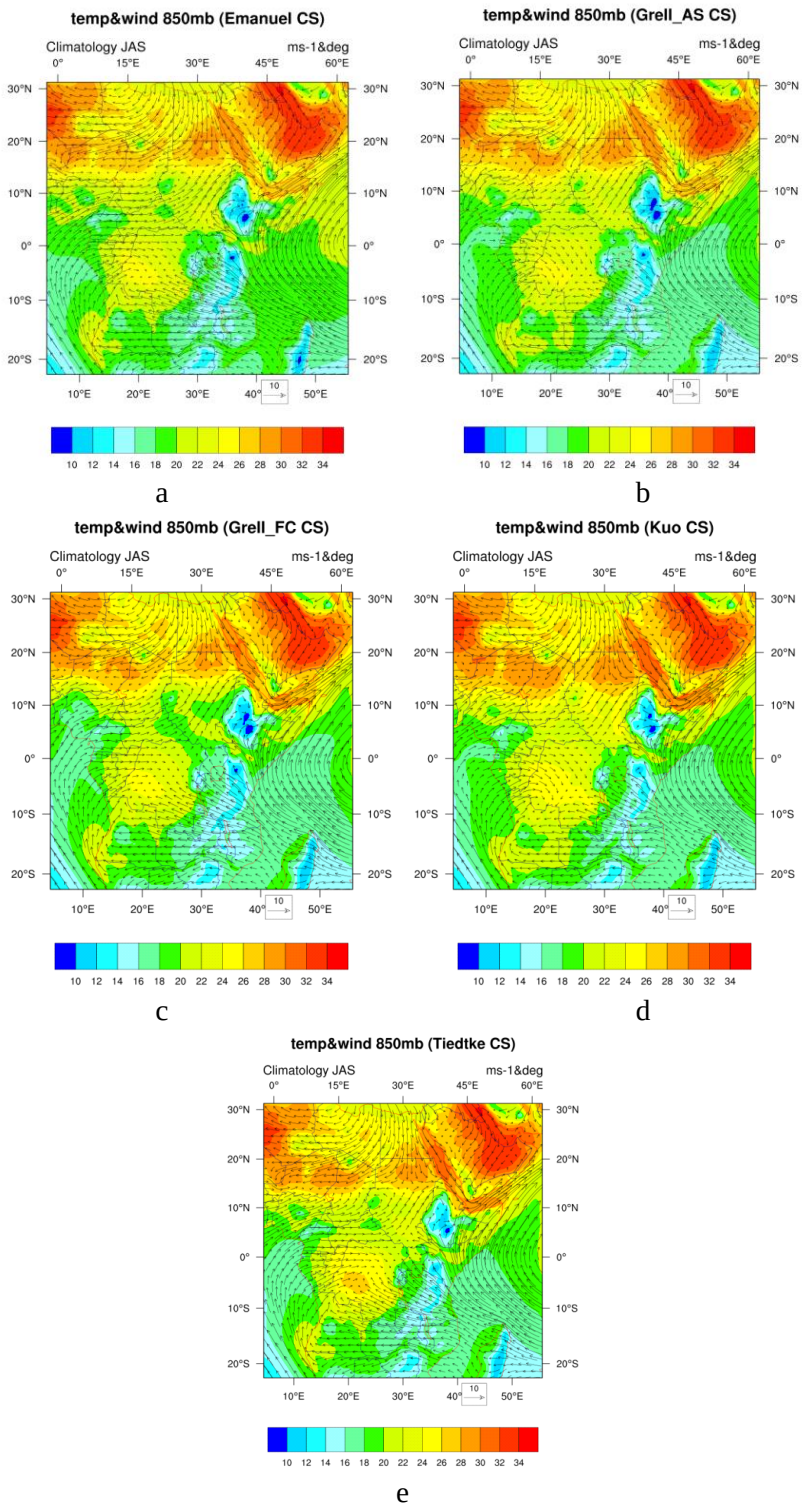


Figure 4. (a, b, c, d, e). Temperature and wind barbs at level 850mb for Emanuel, Grell_AS, Grell_FC, Kuo and Tiedtke respectively.

Grell with Arakawa–Schubert closure assumption creates precipitation scale ranges from 2 mm/day up to 15 mm/day for Blue Nile region. The total precipitation ranges from 1 mm/day up to 4 mm/day for equatorial lakes region. Convective precipitation ranges from 1 mm/day up to 11 mm/day for Blue Nile region and it ranges from 1 mm/day up to 4 mm/day for equatorial lakes. In general, closure assumption mean adjustment the amount of convection by grid scale parameters. Grell with Fritz–Chappel closure assumption creates precipitation scale ranges from 8 mm/day up to 28 mm/day for Blue Nile region. It ranges from 1 mm/day up to 8 mm/day for equatorial lakes region and convective precipitation was ranges from 6 up to 16 mm/day for Blue Nile region and it ranges from 2 mm/day to 6 mm/day for equatorial lakes region. For Kuo C.S creates precipitation scale ranges from 1 mm/day up to 10 mm/day for Blue Nile region and it ranges from 1 mm/day up to 3 mm/day for equatorial lakes region. While connective precipitation ranges from 1 up to 9 mm/day for Blue Nile region, for equatorial lakes region convective precipitation ranges from 1 mm/day up to 3 mm/day. For Tiedtke C.S creates precipitation scale ranges from 4 mm/day up to 18 mm/day for Blue Nile region and ranges from 2 mm/day up to 8 mm/day for Equatorial lakes region. While convective precipitation ranges from 2 mm/day up to 11 mm/day for Blue Nile region and it ranges from 1 up to 4 mm/day for equatorial lakes region.

Relative to Emanuel convection scheme has some advantages for the production of a good precipitation forecast over the tropical region for the following assumptions:

1) The precipitation is added to a single, hydrostatic, unsaturated downdraft of assumed constant horizontal cross section.
2) Precipitation evaporates based on a standard rate; also, it has auto conversion, threshold (which is temperature dependent). Hence, ice processes are represented (Emanuel and Zivkovic 1999). (NEIL DAVIS, JARED BOWDEN, 2009) despite that Emanuel was over estimated for total precipitation and convective precipitation also which it means that Emanuel doesn't deal properly with the instability which may be lead to create excessive of non-convective precipitation that appears in Blue Nile region resulting from increase latent heat release as shown in Figure 3. Non-convective precipitation can be getting by difference between total precipitation and convective precipitation.

As well micro physics scheme will still respond to instability resulting in an excess of latent heat this may be lead to decrease pressure at the surface accordingly low level wind respond to lower pressure with increased moisture convergence resulting more latent heat release and overestimate precipitation forecast amount. Figure 4 shows temperature and wind barbs at 850 mb this Figure (4(a)) show relatively strong northeasterly wind for Emanuel convection scheme effect on temperature advection and increase moisture convergence. For equatorial lakes region, stratiform precipitation was less than 2

mm/day. For Grell C.S with two closures AS and FC. Despite that, the dependence on degree of instability and both closures may be less valid in the tropics where there is no strong relationship between convective precipitation and ABE (Cotton and Anthes 1989). (NEIL DAVIS, JARED BOWDEN, 2009). It was almost matched with observations shown in total precipitation bias Grell_AS Figure 2(a). For convective bias Grell_AS, ranges between positive and negative bias so total precipitation may be compensated by non-convective rainfall as shown in Figure 5. For Equatorial lakes region no non-convective rainfall detected. The Arakawa Schubert – type closure used in current version of the G.S turns to minimize convective activity, since it tends dispersion of large-scale buoyant energy production within a single model time step (Filippo Giorgi and Christine Shields 1999). In addition, wind barbs and temperature at 850 mb show weak north easterlies and temperature advection minimum temperature than Emanuel C.S in Blue Nile region Figure 4(b). For Grell_FC it increase amount of total and convective precipitation as shown in Figure 2(b) & 2(h). Grell_FC has little positive bias for both total and convective precipitation. In addition, it brings not-much stratiform precipitation for Blue Nile region. For Equatorial lakes stratiform was less than 4mm/day as shown in Figure 6.

a) Tiedtke is most recent convection scheme added to RegCM model.

Tiedtke one of mass flux schemes like Emanuel. Tiedtke was over estimated over Blue Nile region but unlike Emanuel over estimation was little than Emanuel as shown in Figure 2(e), 2(j). In addition, Tiedtke is dealing well non-convective precipitation as shown in Figure 3-8. Over estimation for convective rainfall may be like Grell–FC. For equatorial Lakes Tiedtke was under estimated especially for convective rainfall. Mean region for Blue Nile and equatorial Lakes was accounted. Blue Nile 1 less bias, over equatorial lakes Grell_AS has less bias and for convective precipitation over equatorial lakes regions. Convective precipitation over Blue Nile region Kuo has less bias.

b) Monthly Averages

Monthly average (surface temperature, total precipitation, convective precipitation) for (July, August, September) (2003, 2004, 2005) for more details are detected over Blue Nile region and equatorial lakes region as shown in diagrams 2, 3, 4. Bias is equal (sim – obs) over Blue Nile region and equatorial lakes region Mean monthly July Surface temperature over Blue Nile region all convection schemes were under estimated. Kuo scheme was less bias than others, over equatorial lakes Grell_AS was the less bias. Also mean monthly August Surface temperature over Blue Nile region all convection schemes were under estimated and Kuo scheme was less bias than others, over equatorial lakes Emanuel and Grell_FC were less bias. Mean monthly September Surface temperature over Blue Nile region Tiedtke was less bias and over equatorial lakes Emanuel was less bias. For mean monthly (total rainfall bias and convective rainfall bias) competition was

most often between Grell with two closures. Kuo may give better result in-convective rainfall despite scheme simplicity.

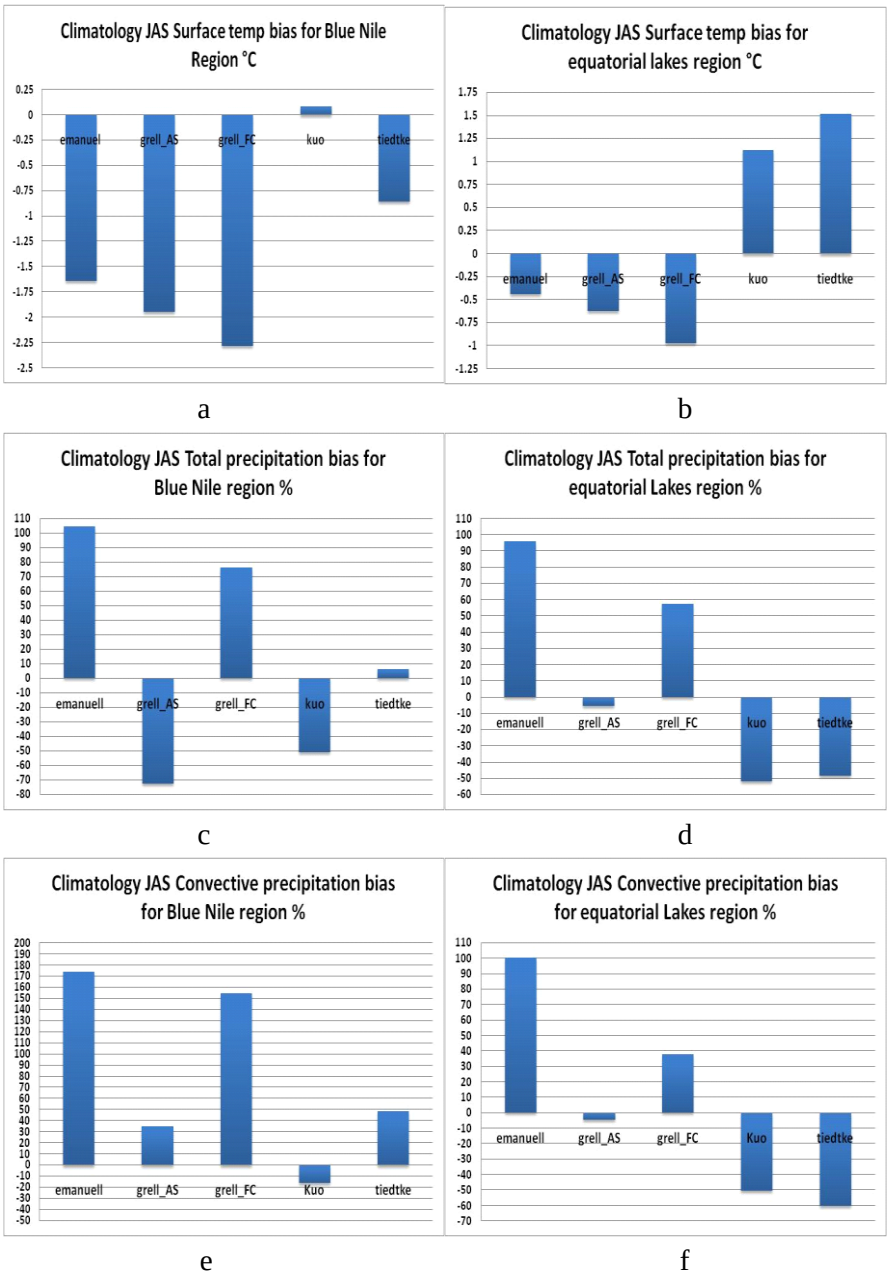


Diagram 1. (a, b, c, d, e, f). Temp, precipitation (total – convective) for Blue Nile region and equatorial lakes region respectively.

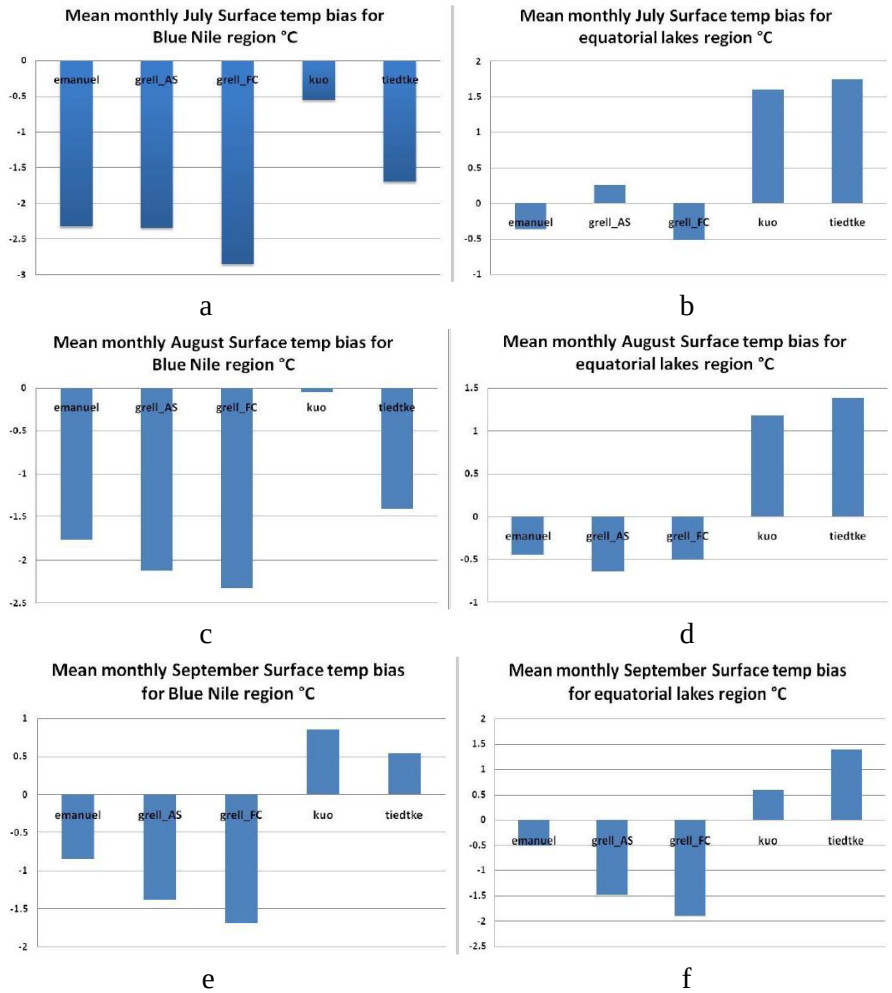


Diagram 2. (a, b, c, d, e, f). Surface temperature bias for (July, August, September) for Blue Nile region and equatorial lakes region respectively.

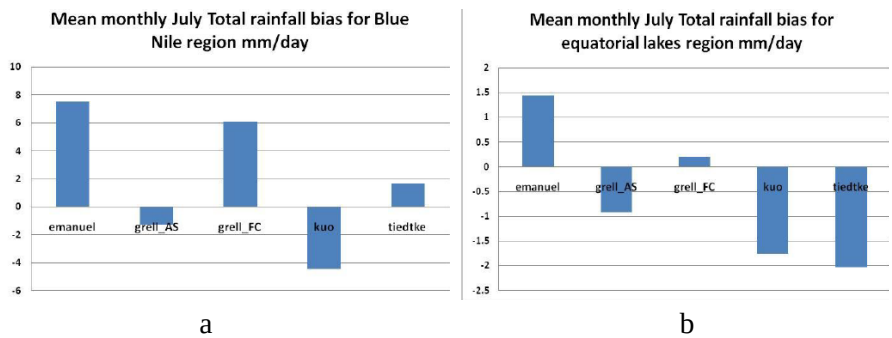


Diagram 3. (Continued).

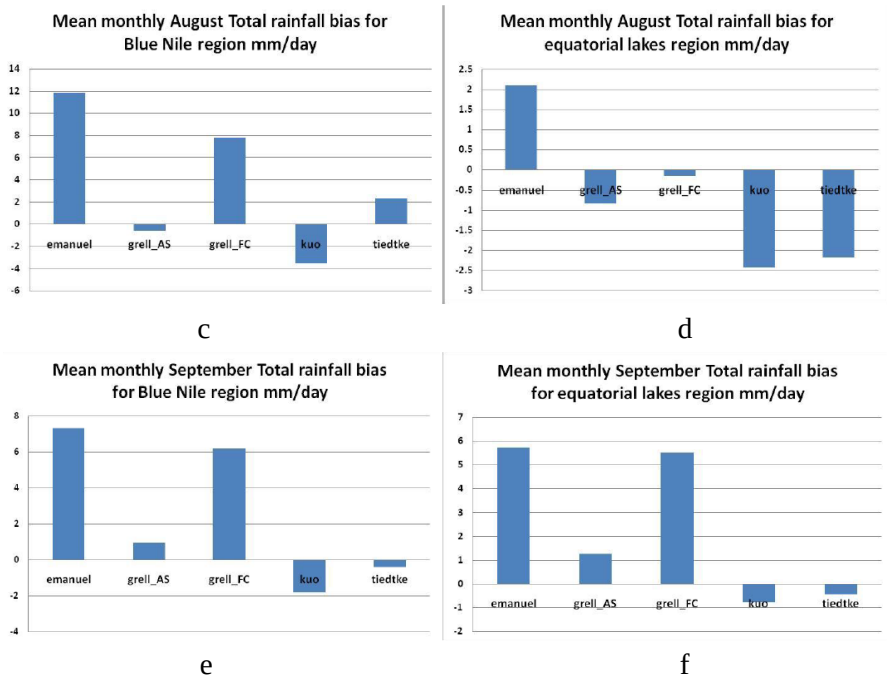


Diagram 3. (a, b, c, d, e, f). Total rainfall bias for (July, August, September) for Blue Nile region and equatorial lakes region respectively.

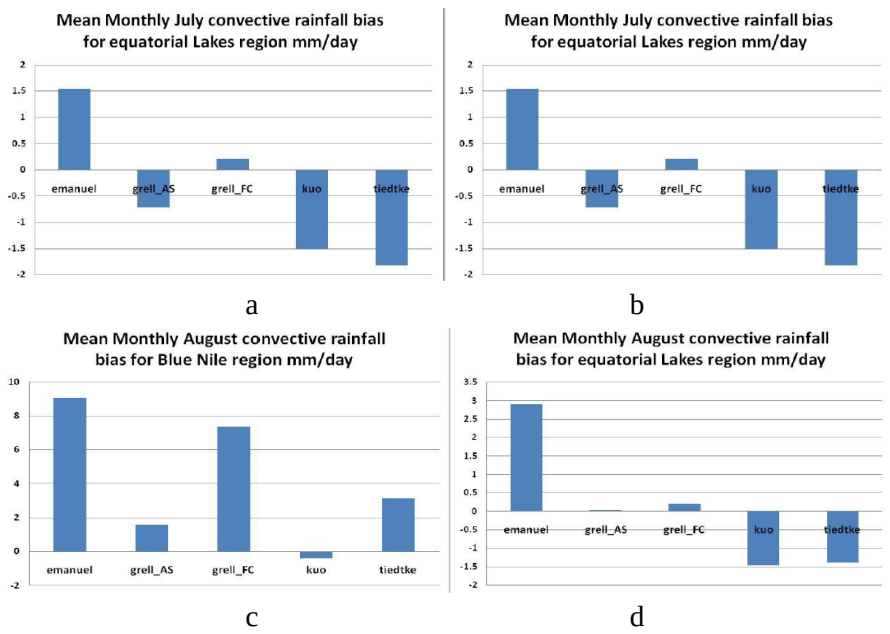
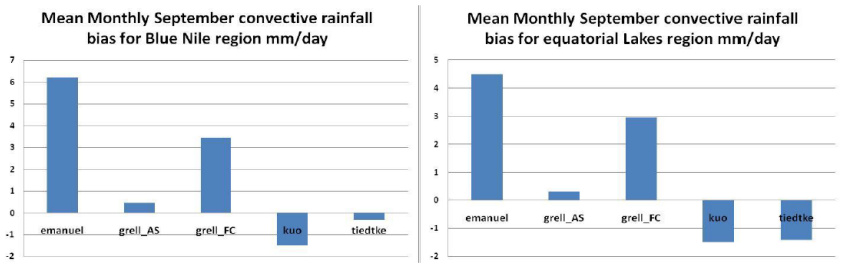
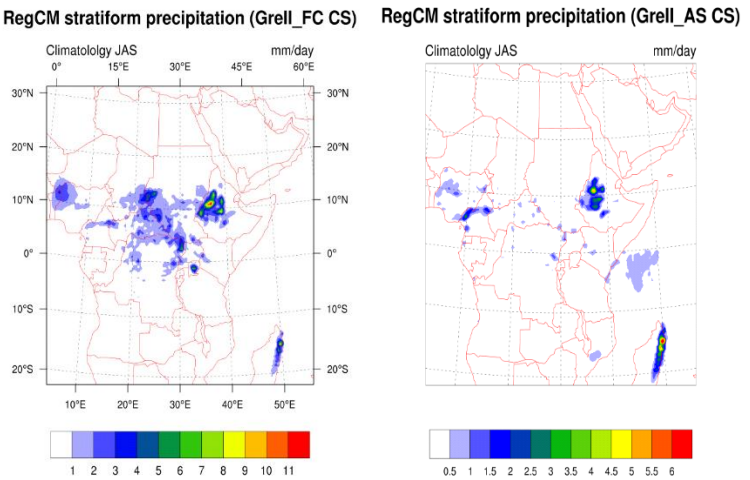


Diagram 4. (Continued).

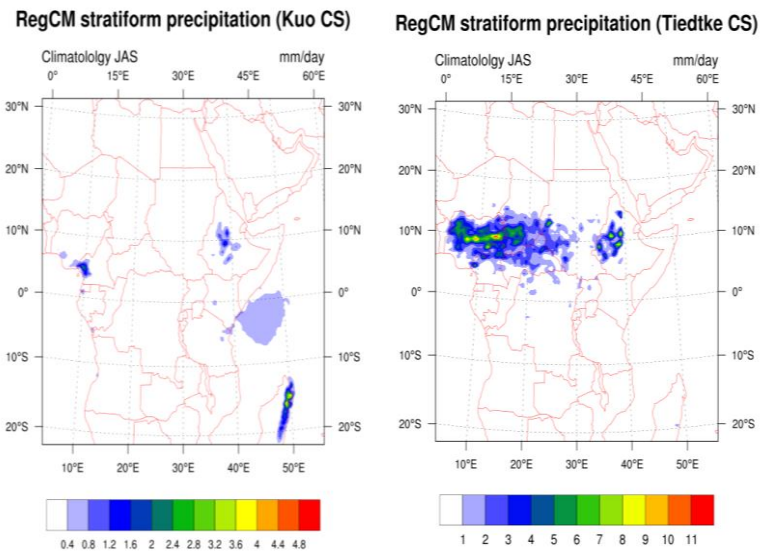


e f

Diagram 4. (a, b, c, d, e, f). Convective rainfall bias for (July, August, and September) for Blue Nile region and equatorial lakes region respectively.



5 6



7 8

Figures 5, 6, 7, 8. Non convective rainfall for different convection schemes (Grell_AS, Grell_FC, Kuo and Tiedtke) respectively.

CONCLUSION

We have tested RegCM4 convection schemes over two-interested region considered as main rainfall sources for Nile River. We compared 4 convection schemes (Emanuel – Grell with two closures (Arakawa and Schubert and Fritsch–Chappell) – Kuo – Tiedtke). We found that Grell convection scheme as most often was more stable in results and give less bias compared with observation specifically Grell_AS was the best especially for equatorial lakes region where Grell_FC give much convective especially for Blue Nile region. For Emanuel convection scheme was always overestimated for convective and total precipitation, the bias reaches to 100% or may be higher compared to observation. Kuo always underestimated for total precipitation. Nevertheless may give better result for convective precipitation, which indicates that the lack may be due to that the scheme cannot handle the non-convective precipitation. For Tiedtke although it belongs to mass flux convection schemes but it was unlike Emanuel. Tiedtke as most often was under estimated compared with observation but the underestimation is not high compared to Emanuel over estimation also in some result Tiedtke was over estimated.

RECOMMENDATIONS AND FUTURE WORK

Selection of initial and boundary conditions for models are very important because convective parameterization is considered as secondary product, which means that it is less important than meteorological forcing parameters (wind – temperature advection – moisture content). Errors in dynamical forcing variables will lead to error in precipitation forecast. Nowadays improving grid spacing for models up to 1 km especially for non-hydrostatic models is appropriate to resolve cloud. But, updraft and downdraft explicitly so convection scheme is no longer to be used but, convection scheme will remain important because sometimes research goals, and some atmospheric phenomena for interested studies, don't need high resolution, therefore we can decrease run time and use available lower computer performance. Change Grid spacing may be very important so many researches are required for our interested regions for example Kuo at grid spacing 30km or less can create extreme or unrealistic rainfall (Roger A. Pielke, Sr.2009). In addition, the Fritsch–Chappell Parameterization designed for resolution between 10–30 km (Roger A.Pielke, Sr.2009). We used SUBEX (Subgrid Explicit Moisture Scheme) RHmin as default in model (80% over land and 90% over the ocean). Many research for our interested regions are required to Vary SUBEX tuning RHmin and auto-conversion scale factor Cacs with each convection scheme separately to find best setting for each convection scheme and if it valid for our interested region or not.

ACKNOWLEDGMENTS

Primarily, enormous gratitude is due to prof. Dr. Mohamed Magdy Abdel wahab who has been there my professor and my supervisor for many years and has been unstinting in his support and constructive critique. Many thanks for my master Dr. Mohamed Abdel Ati for his advice and encouragement. Many thanks for Dr. Ashraf Saber as supervisor for his aid and foresight. Special thanks also for Dr. Zeinab Salah for her kind response and unlimited support. In addition, I cannot forget my seniors Dr. Ahmed Kamal, Mr. Mohamed Omar, Mr. Taha Kassem, Mr. Shereef Fouda, Mr. Alaa Mabrouk, my friend Mr. Mohamed Elsayed and also my deer family. We used EIN15, which one of ECMWF data sets and it is available on RegCM4 data download web site also, RegCM4 model is free and available on ICTP-ESP web site Figures were made by NCAR Command Language (NCL), which is powerful tool for data visualization, and it is available free.

REFERENCES

Arakawa A., Schubert WH (1974) Interaction of a cumulus cloud ensemble with the large-scale environment. Part I. *J Atmos Sci 31*:674–70.

Arakawa A., Wayne Howard Schuber, (1973). Interaction of a cumulus cloud ensemble with the large-scale environment, *Journal of the atmospheric science*, volume 33.

Back, L. E. and Bretherton, C. S., (2006), Geographic variability in the export of moist static energy and vertical motion profiles in the tropical Pacific, *Geophysical Research Letters*, VOL. 33.

Baik JJ., DeMaria M., Raman S., Tropical *Cyclone Simulations with the Betts Convective Adjustment Scheme. Part (III): Comparisons with the Kuo Convective Parameterization*, AMS, DOI: http://dx.doi.org/10.1175/1520-0493(1991)119 <2889:TCSWTB>2.0.CO;2.

Baldauf M., Seifert A., Förstner J., Majewski D., Raschendorfer M., (December 2011). Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: *Description and Sensitivities,* AMS, DOI. http://dx.doi.org/10.1175/MWR-D-10-05013.1.

Bowden J., Davis N., Semazzi F., Xie L., Onol B., (2007). *Customization of Moist Processes in the RegCM Regional Climate Model for Eastern Africa*, 3595.

Bowden JH., (2009), *Multi-Scale Climate change modeling study over the Greater Horn of Africa*, North Carolina State University, pp. 339.

Cotton WR., Bryan GH., Susan C. van den Heever, (2009). *Storm and Cloud Dynamics* 2nd Edition, International Geophysics science, 99.

Cotton WR., Richard A., Anthe, (1989). Storm and Cloud Dynamics, ACADEMIC PRESS, 0-12- 19253 1-5.

Daived J., Neelin (1986), *Modeling Tropical Convergence Based on the Moist Static Energy Budget*, American Meteorological Society, 115.

Davis N., Bowden J., Semazzi F., Xie L., (2009). *Customization of RegCM3 Regional Climate Model for Eastern Africa and a Tropical Indian Ocean Domain*, 3595.

Decaria A-J., (2006). Relating Static Energy to Potential Temperature: A Caution, *Journal of the Atmospheric Sciences*, 64.

Dumont, Henri J., (Ed.), (2009), The Nile: Origin, Environments, Limnology and Human Use, *Monographiae Biologicae*, Springer, 307-333.2009.

El-Khodari N., (2003). The Nile River: Challenges to sustainable development, *Canada Presentation to the River Symposium.*

Emanuel KA., (1994). *Atmospheric Convection*, Oxford University Press.

Gao., Shouting., Li., Xiaofan., (2008). *Cloud-Resolving Modeling of Convective Processes*, Springer Netherlands, 978-1-4020-8276-4.

Giorgi F., Shields C., (1999). Tests of precipitation parameterizations available in latest version of NCAR regional climate model (RegCM) over continental, *Journal of Geophysical Research*, 104.

Kain JS., Fritcsh JM., (1991). Sensitivity of numerical simulation of convective weather systems to the convective trigger function, *Meteorol. Atmos. Phys.* 49, 93-106.

Kalnay E., (2003). *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, 0 521 79629 6.

Kerry A. Emanuel David J, (1993), *The representation of cumulus Convection in numerical models*, American Meteorological Society, 24.

Krishnamurti T.N., Low-Nam S., Pasch R., (1983) AMS, DOI: http://dx.doi.org/10.1175/1520-0493(1983)111<0815:CPARRI>2.0.CO;2.

Krishnamurti TN., Lydia Stefanov AV. asubandhu Misra, (2013), *Tropical Meteorology an Introduction*, Springer Atmospheric Sciences, 978-1-4899-9813-2.

Kurihara Y., (1973). A Scheme of Moist Convective Adjustment, *Monthly Weather Review*, Vol. 101.

Mapes B. E., Warner T. T., Mei Xu, Gochis D. J., (2003). Comparison of Cumulus Parameterizations and Entrainment Using Domain-Mean Wind Divergence in a Regional Model, *Journal of the atmospheric science*, volume 61.

Mellander P-E., Gebrehiwot SG, Gärdenäs AI, Bewket W, Bishop K (2013) Summer Rains and Dry Seasons in the Upper Blue Nile Basin: The Predictability of Half a Century of Past and Future Spatiotemporal Patterns. *PLoS ONE* 8(7) e68461. doi:10.1371/journal.pone.0068461.

Nogherotto R., Tompkins A-M., Giuliani G., Coppola E., Giorgi F., (2016). Numerical framework and performance of the new multiple-phase cloud microphysics scheme

in RegCM4.5: precipitation, cloud microphysics, and cloud radiative effects, *Geosci. Model Dev.*, 9, 2533–2547, 2016.

O'Brien TA., Chuang PY., Sloan LC., Faloona IC., Rossiter DL., (2012). Coupling a new turbulence parametrization to RegCMaddsrealistic stratocumulus clouds, *Geosci. Model Dev.*, 5, 989–1008, 2012.

Odada EO., (2002). *The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity*, Kluwer Academic Publishers, 12.

Pielke RA., Sr, (2002), *Mesoscale Meteorological Modeling*- second edition, International Geophysics series, volume 78.

Program University Corporation for Atmospheric Research (UCAR), (2009). Impact of Model Structure and Dynamics - version 2. The COMET®.

Program University Corporation for Atmospheric Research (UCAR), (2009). How Models Produce Precipitation and Clouds - version 2. The COMET®

Raju PVS., a* R. Bhatla, b Almazrouia M., Assiria M., (2015). Performance of convection schemes on the simulation of summer monsoon features over the South Asia CORDEX domain using RegCM-4.3, *International Journal of Climatology*, 35.

Smith RK., (1996). The Physics and Parameterization of Moist Atmospheric Convection, *Mathematical and Physical Sciences*, 505.

Solomon S., Qin D., Manning M., Chen Z., Marquis M., Averyt KB., Tignor M., Miller HL. (Eds.). Intergovernmental Panel on Climate Change, (2007). *Climate Change 2007 the Physical Science Basis*, Cambridge University Press, 996.

Stensrud, David J., (2007), *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*, Cambridge University Press. ISBN: 9781107453258.

Sutcliffe JV., Parks YP., (1999), *The Hydrology of the Nile*, the International Association of Hydrological Sciences, 1-910502-75-9.

Tiedtke M., (1989), *A comprehensive Mass Flux scheme for cumulus parameterization in large scale model*, American Meteorological Society, (1779).

Tiedtke M., (1989). A comprehensive mass ux scheme for cumulus parameterization in largescale models. *Mon. Wea. Rev.*, Vol. 117, 1779-1800.

Tompkins A., (2010). Introduction to Atmospheric Physics (Basic Physics Diploma), ICTP, Trieste, Italy. Alan. M. Plyth, (1992), *Entrainment of cumulus cloud*, AMS, http://dx.doi.org/10.1175/1520-0450(1993)032<0626:EICC>2.0.CO;2.

University of East Anglia Climatic Research Unit; Jones, Harris P.D., I.C., (2008). *Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena*. NCAS British Atmospheric Data Centre.

Van TP., Nguyen HV., Tuan LT., Quang TN., Ngo-Duc T., Laux P., Xuan T-N., (2014). Seasonal Prediction of Surface Air Temperature Across VietnamUsing the Regional Climate Model Version 4.2 (RegCM4.2), *Advances in Meteorology*, (2014).

Wang W., Seaman NL., (1996). *A Comparison Study of Convective Parameterization Schemes in a Mesoscale Model*, American Meteorological Society, Volume 125.

WC D. R., Bechtold P., Fröhlich K., Hohenegger C., Jonker H, Mironov D., Siebesma AP., Teixeira J.,Yano JI. (2013). Entrainment and detrainment in cumulus convection: an overview. *Q. J. R. Meteorol. Soc.* 139: 1–19.DOI:10.1002/qj.1959.

WMO, (2013). *The Climate in Africa*, 1147.

Wu T., (2011). *A mass-flux cumulus parameterization scheme for large-scale models: description and test with observations*, Springer-Verlag, DOI 10.1007/s00382-011-0995-3.

Wu T., (2012). *A mass-flux cumulus parameterization scheme for large-scale models: description and test with observations, climate dynamics*, 38.

Yu X.. Lee TY., (2010). Role of convective parameterization in Simulations of a convection band at grey-zone resolutions, Tellus (2010), 62A, 617–632.

Electronic Sources

CRU TS3.10: *Climatic Research Unit (CRU) Time-Series (TS) Version 3.10 of High Resolution Gridded Data of Month-by-month Variation in Climate* (Jan. 1901 - Dec. 2009) is available Online http://catalogue.ceda.ac.uk/uuid/ac3e6be017970639a9278e64d3fd5508

Future weather - Parameterization of convective clouds available online: http://www.knmi.nl/samenw/regioklim/FW/page9.html

Huang D., Gao S., (2017) Impact of different cumulus convective parameterization schemes on the simulation of precipitation over China, Tellus A: *Dynamic Meteorology and Oceanography*, 69:1, 1406264, DOI: 10.1080/16000870.2017.1406264.

Kwon Y-V., Hong S-Y., (2016). A Mass-Flux Cumulus Parameterization Scheme across Gray-Zone Resolutions, *AMS*, https://doi.org/10.1175/MWR-D-16-0034.1

README file for TRMM Version 7 algorithm 3A-12. Is available online: http://pps.gsfc.nasa.gov/Documents/README3A12.v7

Regional Climate Model is available online: http://www.ictp.it/research/esp/models/regcm4.aspx

Stein T-H-M., Holloway C-E., (2016). Observed Relationships between Cloud Vertical Structure and Convective Aggregation over Tropical Ocean, *AMS*, https://doi.org/10.1175/JCLID-16-0125.1

The Problem of Convective Moistening, Emanuel K-A., *Program in Atmospheres, Oceans, and Climate, Massachusetts Institute of Technology, Cambridge, MA 02139* is available online: http://wind.mit.edu/~emanuel/convmoist/convmoist.html

The River Nile is available online http://www.utdallas.edu/geosciences/remsens/Nile/

Wing1 A-A., Emanuel1 K-A., (2014). Physical mechanisms controlling self-aggregation of convection in idealized numerical modeling simulations. *JAMES*, DOI: 10.1002/2013MS000269.

Working Group I: The Scientific Basis available online: http://www.ipcc.ch/ipccreports//tar/wg1/380.htm

PART III. MEDICINE

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 8

ALGORITHMIC APPROACHES FOR ECG ANALYSIS

H. Kodal Sevindir[1,*], C. Yazici[2,*] and Suleyman Çetinkaya[1]
[1]Department of Mathematics, University of Kocaeli, Kocaeli, Turkey
[2]Department of Mathematics Education, University of Kocaeli, Kocaeli, Turkey

ABSTRACT

Cardiovascular diseases constitute the main cause of mortality in many countries. The heart's mechanical activity is directly linked with its electrical activity. Therefore, electrocardiography (ECG) is an important diagnostic tool for analyzing heart function. It provides a great deal of information on the function of the heart. For almost four decades, researchers have been conducting research for automated ECG analysis algorithms. The evolution of these algorithms obviously shows the great advances in computer technology. In this study, we give an outline on some of the underlying methods to develop these software tools. Shortcomings and further research interest are discussed at the end.

Keywords: ECG waves detection, ECG de-noising stage, feature extraction, ECG database

INTRODUCTION

Heart diseases globally are the number one cause of death. The heart, a muscular pump, is formed up of four chambers: two upper and two lower chambers. The two upper

* Corresponding Author's E-mail: hkodal@kocaeli.edu.tr.
* Corresponding Author's E-mail: cuneyt.yazici@kocaeli.edu.tr.

chambers are named atria; the two lower chambers are called ventricles (See Figure 1). The heart muscle contracts due to a natural electrical system. An electrical impulse or wave travels through the heart with each beat. This wave appears due to the muscle to squeeze and pump blood from the heart (Electrocardiogram 2017).

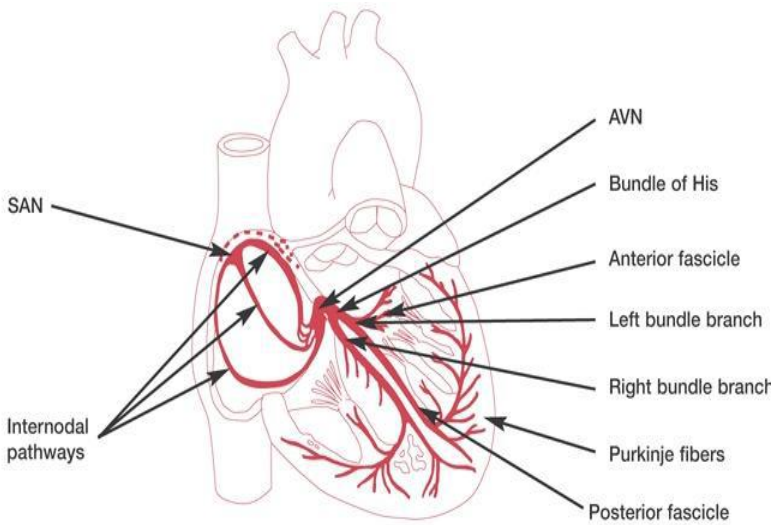


Figure 1. Cardiac depolarization root (Image adapted from (Ashley and Niebauer, 2004)).

A biomedical signal (bio signal) is any signal in living beings acquired from any organ that represents a physical variable of interest. Although the term biomedical signal may refer to both electrical and non-electrical signals, in this study, it is used to refer to bioelectrical signals. It can be continually measured and monitored. Such signals are usually a role of time and are describable regarding the amplitude, frequency, and phase. ECG signal is a biomedical signal that regulates electrical currents generated in the heart during its contraction representing muscular movements (Bio signal 2017).

An electrocardiograph is a piece of equipment measuring the electrical impulses that cause the heart to beat as they pass through the body. Electrocardiogram (ECG or EKG) shows the heart's electrical activity as line tracings on paper. The ECG signal is the key bio signal for assisting the clinical staff in heart disease diagnosis. It consists of several electrodes attached to the body of the patient and are connected by wires to the device. It is an effective, simple, noninvasive, low-cost procedure for the diagnosis of cardiovascular disorders. More precisely, ECG is a recording of tiny voltage changes as potential difference contained from the body exterior and displays the total of work potentials of the heart cells, which is transposed into a visual tracing (Kicmerova 2009). In an ECG recording, we see the graph of voltage versus moment generated by the system; the spikes and dips in the tracings, called waves, are the helpful points for detecting heart’s electrical problems.

Electrocardiography recording is done by using electrodes placed on the skin according to a prescribed arrangement. The recording usually takes 10 seconds to a few minutes for a routine recording. In general, the electrodes are placed on the patient's chest and/or limbs. With the help of these electrodes, we can detect the electrical changes occurring naturally due to the heart muscle's electrophysiologic pattern of depolarizing during each heartbeat. Then the electric voltage between two of them is measured. Each pair is called a lead, and the elements identify the same electrical heart activity but from a distinctive edge. The 12-lead ECG recording is the most widely used (Electrocardiogram 2017). Interestingly 10 electrodes are used for a conventional 12-lead ECG recording. These 10 electrodes placed on the patient's body give 12 pictures (see Figure 2). In other words, a lead is a picture that is captured by a group of electrodes (Emtresource 2017; Ekg-Flash-Cards 2017; John et al. 2006; Joshi et al. 2014).

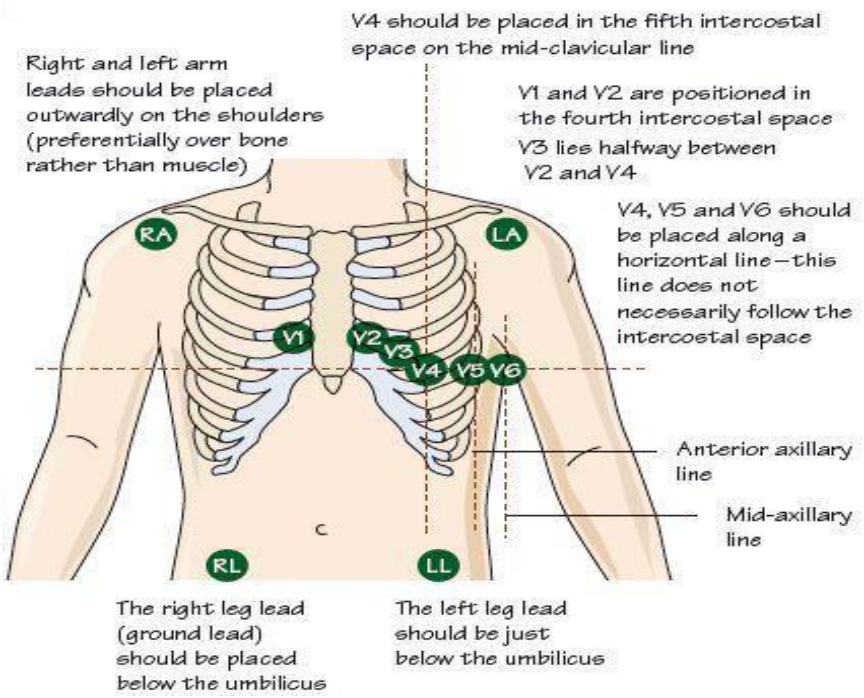


Figure 2. 12-Lead Position of ECG (Image adapted from (Khunti, 2014).

Mainly we obtain answers to two questions through an ECG recording:

a) how long the electrical fluctuation takes to cross by the heart?
b) how much electrical fluctuation cross by the heart?

By the first questioning, a doctor can determine whether the electrical fluctuation is normal or slow, fast or irregular. The second question helps to determine whether the parts of the heart are too large or are overworked.

ECG recording is a commonly performed test to detect cardiological diseases. Such popularity of ECG examination is due to its simplicity, effectiveness, and noninvasiveness. A patient will have no pain or risk associated with having an electrocardiogram. There may be some minor discomfort during the ECG stickers are being removed. The machine only records the ECG. It does not send electricity into the body. An ECG transfers a significant amount of knowledge about the formation of the heart and the use of its electrical conduction system. An ECG is done to

- check the heart's electrical activity,
- check the rate and rhythm of heartbeats,
- check the presence of any harm to the heart's muscle cells or conduction system,
- determine the cause of unexplained chest pain or pressure,
- find the cause of heart disease symptoms,
- check the size and position of the heart and heart chambers,
- determine the thickness of the heart chambers' walls,
- test the effects of cardiac medications,
- check any mechanical materials that are fixed in the heart, e.g., pacemakers,
- check the heart health when other diseases or conditions, e.g., high blood pressure, high cholesterol, a family history of heart disease, diabetes, etc.

An ECG recording can also show:

- Abnormal heartbeat (too fast, too slow, or irregular)
- Lack of blood flow to the heart muscle
- A heart that does not pump forcefully enough
- Too thick heart muscle or too big heart parts
- Birth defects in the heart
- Problems with the heart valves
- Inflammation of the heart surrounds.

(Joshi et al. 2014; Electrocardiogram 2 2017; akwmedical-what-ekg 2017).

However, ECG does not provide data on cardiac compression or pumping role. A normal heartbeat on ECG will show the timing of the top and lower chambers. During each heartbeat, the progression of depolarization can be monitored orderly for a healthy heart.

An ECG signal appears as an almost periodic signal. It can be segmented into heartbeats. Each heartbeat consists of five standard waves labeled with the letters P, Q, R, S and T. In some cases, U wave may also be observed. These waves indicate the depolarization and the repolarization phases of heart muscles. Additionally, five more

interweave timings named PR, PR segment, QRS, QT, ST segment are used. These intervals are indicated in Figure 3 below.

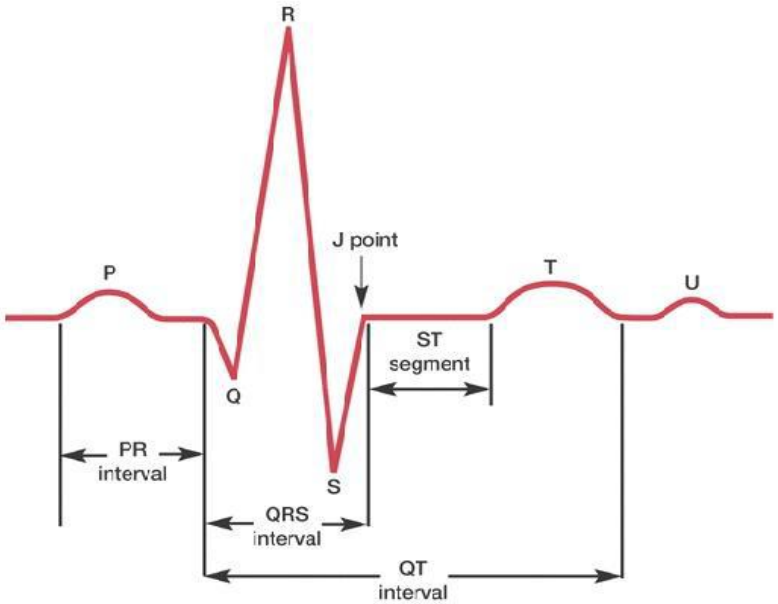


Figure 3. An ECG signal showing P-QRS-T and U waves (Image adapted from (Electrocardiogram 2, 2017).

Now we give a brief information about the role of these clinical features (Electrocardiogram 2 2017; Merone et al. 2017):

A sketch of a standard cardiac period with the associated waves of the ECG signal is shown Figure 4.

- **P wave:** The P wave shows the status of activation of the atria; it shows the time received by the beat to propagate to both atria. Atrial depolarization reaches from the SA node towards the AV node, and from the right atrium to the left atrium. P waves are usually low-amplitude features (amplitude equivalent to or smaller than 0.4 mV), and its span ranges from 60 ms to 120 ms. They are hard to detect but useful to precisely identify heart disorders, e.g., atrial flutter and to distinguish various cardiac arrhythmias, e.g., atrial tachycardia. The P wave is typically upright in most of the leads. An unusual P wave axis can indicate a change in the pacemaker site. If the duration of a P wave is unusually long, atrial enlargement may be present.
- **PQ stretch:** It is flat and independent of waves; it holds the time that elapses from the time when the atria start to be effective until the time in which the ventricles are actuated. The length of a regular PQ period ranges between 12 ms and 20 ms.

- **PR segment:** PR segment: The PR segment appears between the end of the P wave and the beginning of the QRS complex. It is a flat and usually isoelectric segment.
- **PR interval:** The PR interval begins from the opening of the P wave and ends at the opening of the QRS complex. This interval displays the time the electrical pulse gets to travel from the sinus node through the AV node. A normal PR interval is 120 seconds to 200 seconds in length. When the electrical impulse by bass the AV node, the PR interval becomes shorter than 120 ms and it is associated with junctional arrhythmias. When there is an AV nodal disfunction, the PR interval becomes longer as in atrioventricular blocks.
- **QRS complex:** The QRS complex describes the rapid depolarization of the right and left ventricles. It is composed of the Q, R and S waves. The QRS complex, the most significant wave of the ECG, regularly has a much larger amplitude than the P wave since ventricles have a large muscle bulk opposed to the atria. It can be easily detected, and it is often used to determine the heart rate. The duration of the QRS complex ranges from 60 ms to 90 ms. Overly wide (longer than 120 ms) QRS complex may imply the disruption of the heart's conduction system or ventricular rhythms. Metabolic issues may also cause wide QRS complex. An unusually high-amplitude QRS complex may describe left ventricular hypertrophy, and a low QRS complex may describe a pericardial diffusion or infiltrative myocardial disorder.
- **J-point:** The point at which the QRS complex ends and the ST segment starts is named the J-point. The presence of a separate J wave or Osborn flow at the J point is pathognomonic of hypothermia or hypercalcemia.
- **ST segment:** The ST segment, connecting the QRS complex and the T wave, describes the period from the end of ventricular depolarization to the beginning of ventricular repolarization. ST level shifts are significant markers of cardiac abnormalities.
- **ST stretch:** It is a long stretch, corresponding to the period where the ventricles contract and turn to rest, beginning later the S wave and finishing with the T wave. This stretch is about in the baseline of the ECG signal. The usual span of ST stretch extends from 230 ms to 460 ms. Its investigation can unveil ischemic difficulties.
- **T wave:** The T wave describes the repolarization of the ventricles (the time when the ventricles have ended their activation step, and they are available for a new contraction) and its duration extends between 100 ms and 250 ms. It is commonly elevated in all leads except a VR and V1. It is a recovery phase of the cardiac muscle. The shape of this wave carries a lot of information about cardiac abnormalities, which is why it is important to analyze its geometrical properties such as symmetry, asymmetry, and slope. The T wave provides knowledge of

cardiac hypertrophy, heart failure, and ischemic heart disorder. Inverted T waves can be a warning for myocardial ischemia, LVH, high intracranial pressure, or metabolic abnormalities. Sharp T waves can be an indication of hyperkalemia or myocardial injury.

- **Corrected QT interval (QTc):** The QT interval starts from the beginning of the QRS complex and ends at the end of the T wave. Its range varies with heart rate; therefore, most of the time breaking by the square root of the RR interval is done to get corrected QT interval (QTc). The QT interval is inversely proportional to heart rate: shortens at faster heart rates and expands at slower heart rates. Either a prolonged or an unusually short QTc interval may indicate a risk factor.
- **QT interval:** It denotes the electrical systole, i.e., the time corresponding to the series of depolarization and repolarization of ventricles. The width of this interval changes with the heart rate and, it ordinarily extends from 350 ms to 440 ms.
- **U wave:** The U wave is hypothesized to be generated by the repolarization of the interventricular septum. U wave span extends from 185 ms to 228 ms. It usually has a little amplitude, and even more regularly is absent. If the U wave is pronounced, hypokalemia, hypercalcemia, digoxin toxicity or hyperthyroidism may be suspected.

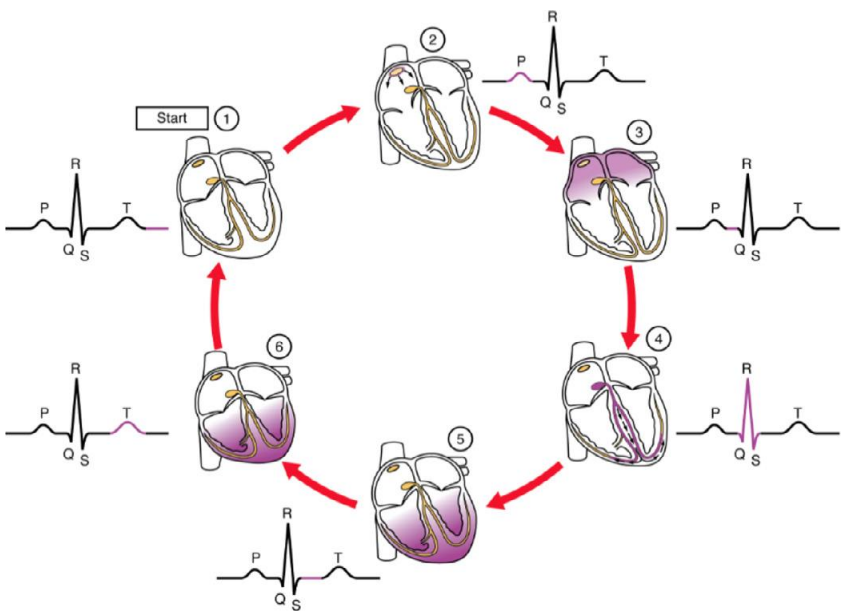


Figure 4. A sketch of a standard cardiac period with the associated waves of the ECG signal (one-lead). (Image adapted from (Merone et al. 2017)).

Frequency components of a QRS complex typically change from about 10 Hz to 25 Hz. So, to attenuate other signal segments and artifacts, e.g., P wave, T wave, baseline drift, and coupling noise, almost all QRS detection algorithms use a filtering stage before

the actual detection. While the attenuation of the P wave and T wave, as well as baseline drift, needs high-pass filtering, a low-pass filter regularly performs the elimination of coupling noise. In many algorithms, high-pass and low-pass filtering are used independently. Some of the earlier dated algorithms, such as (Fraden and Neumann 1980; Holsinger et al. 1971; Okada 1979), only use the high-pass filtering. The cleaned signals are then utilized for the formation of a feature signal in which the existence of a QRS complex is identified by comparing the feature against fixed or adaptive thresholds. Almost all algorithms use extra decision laws for the decline of false-positive detections (Kohler et al. 2002).

An algorithmic structure of automated detection for ECG was developed long time ago, and many algorithms have shared it. It is split into a preprocessing, feature extraction and a decision step. Frequently an extra processing section is used for the accurate decision of the temporary position of the expected candidate points. In this study, various algorithms are specified concerning their preprocessing steps, because most of the decision stages are rather heuristic and dependent on the preprocessing results.

AUTOMATED DETECTION FOR ECG

Signal Preprocessing

Due to the effective changes in the roles of the heart along with other organs, the ECG signals show time-varying as well as non-stationary replies. The heart's electrical signal is tiny, and at most of the time, it is combined with some other similar frequency signals to create an artifact. Later signal gain phase, signals are to be pre-processed. Signal pre-processing is also called Signal Enhancement. In general, noise and artifacts corrupt the collected signals. ECG signals are corrupted by various extra-cardiac noises like muscular, movement artifact and instrumentation noises. Typically, ECG signals are corrupted with three main sources of variability: artifacts and noise, intra-subject variability and inter-subject variability. These are explained in detail in (Merone et al. 2017; Lakshmi et al. 2014).

a) Artefacts and noise
   - Electrode material
   - Sensor locations
   - Power-line interference
   - Baseline drift
   - Movement artefacts

- Instrumentation

b) Intra-subject variability
   - Health status of the subject
   - Heart rate (HR) variability
   - Physical exercise
   - Affective status
   - Drugs
   - Long-term stability

c) Inter-subject variability
   - Thorax geometry and tissue characteristics.

Feature Extraction

Later getting the noise-free signals from the signal enrichment phase, primary features from the signals are obtained. For feature extraction from ECG signals, methods such as Genetic Algorithms, Adaptive Auto Regressive parameters, bilinear AAR, multivariate AAR, Fast Fourier Transformations, Independent Component Analysis, Principal Component Analysis, Wavelet Packet Decomposition, Wavelet Transformations can be listed (Lakshmi et al. 2014).

Significant points of an ECG define the intervals and amplitudes in an ECG, carrying most of the clinically beneficial knowledge. Thus, the progress of reliable and robust techniques for automated ECG description is a subject of extreme concern. The issue of automatic delineation of ECG has been broadly examined (Addison 2005). For automated analysis of the ECG, detection of P and T waves is also essential as the P wave represents atrial action and the T wave is connected to repolarization of ventricles. Due to the low slope and magnitude of P and T waves the analysis of P and T waves are more complicated (Lin 2012).

Correct detection of QRS is a major concern in many clinical situations. QRS detection gives an essential basis for instant heart rate calculation because most of the time the accuracy of instantaneous heart period calculation depends on the performance of QRS detection (Kohler et al. 2002; Chan et al. 2005; Chen et al. 2006). Thus, QRS complex is the most characteristic waveform of an ECG signal among all of its components. Due to its high amplitude, QRS detection becomes easier than of the other waves. Thus, most of the time it is used as a citation in the cardiac cycle (Addison 2005). However, to identify patterns of cardiac disorders, the detection of QRS complexes may not be sufficient, and the whole knowledge about its morphology may be required. This is also valid for the other waves: P, T and U waves (Madeiro et al. 2007).

Most of the time QRS detection is preferable to determine the heart movement, and as a reference for beat arrangement; ECG wave delineation presents basic characteristics (amplitudes and intervals) to be utilized in the subsequent automated analysis. It can be split into two principal groups of algorithms: QRS detection algorithms and wave delineation algorithms (Addison 2005).

The QRS complex is the most dominant characteristic of the ECG signal, and a cardiologist having the trained eye can quickly identify it. Unluckily, the issue of mechanization of this process is not easy because morphologies of QRS complexes vary broadly. The occupation of noise from many springs makes this problem more complicated. Moreover, other segments of ECG (P and T waves) can limit the detection of QRS complexes (Zidelmal et al. 2012). The QRS detection problem has been an investigation topic for almost four decades. A vast difference of algorithms has been introduced in the literature, for an extended list of studies see the references.

Algorithms on Automated Detection of ECG

Analysis of ECG signals is an essential assignment in the determination and discovery of many cardiac arrhythmias. Because ECG has a time-varying morphology subordinate to physiological situations and the occupation of noise, robust and reliable algorithm development for the detection of the P wave, QRS complex and T wave in ECG signal is a severe problem (Ghaffari et al. 2008).

In many QRS detectors, the slope of the R wave is a conventional signal feature used to locate the QRS complex. Such algorithms mostly use the signal derivative. Nevertheless, a derivative magnifies the unwanted higher frequency noise segments. Moreover, many abnormal QRS complexes with high amplitudes and prolonged durations such as ectopic ones can be avoided in a solely derivative method because of their almost low R wave slopes. To obtain strong performance other parameters from the signal such as morphology, width, and frequency component must also be used (Zidelmal et al. 2012).

Study of the electrical activity of the heart uses simple as well as complex algorithms of digital signal processing performed in electrocardiographs. We can usually classify these algorithms into three groups: time domain, frequency domain, and time-frequency domain. First two classes belong to traditional methods, which are active in many clinically tested treatments and analysis of ECG, detection of QRS complexes, detection of onset and offset of QRS complexes, investigation of variation of ST segments, analysis of the variability of heart rate, and others. New algorithms regularly work in the time-frequency domain and connect advantages of classical methods. They essentially enable frequency analysis with time knowledge about investigated features. For example, the

variations arising during depolarization of the myocardium during myocardial ischemia (Kicmerova 2009).

Using easy time domain analysis does not give adequately high-quality results because of its low sensitivity. It is generated by the comparatively small amplitude of the changes. Frequency analysis regularly raises sensation, but it cannot specify in which phase of the heart circle these changes arose (Kicmerova 2009).

There are many methods used in ECG wave delineation. More exhaustive details are given by (Karpagachelvi et al. 2010; Moses and Deisy 2015; Dewangan and Shukla 2010; Kar and Das 2011; Köhler et al. 2002; Lakshmi et al. 2014; Merone et al. 2017; Pahlm and Sörnmo 1984; Varshney et al. 2014; Reaz et al. 2006; Bravi et al. 2011). Some main and specific methods for ECG signals described here are: Digital Signal Analysis, Fuzzy Logic Methods, Artificial Neural Network, Hidden Markov Model, Genetic Algorithm, Support Vector Machines, Self-Organizing Map, Bayesian and other methods.

Thresholding

Thresholding is an easy but effective method. In its most fundamental form, each coefficient or data value is thresholded by comparing against threshold, if the coefficient is smaller than threshold, set to zero; oppositely it is blocked or modified. Replacing the little noisy coefficients by zero and in many cases by practicing some auxiliary transform on the result may lead to reconstruction with the essential signal characteristics and with less noise (Jansen 2001; Patil et al. 2012). Since the work of Donoho & Johnstone, there has been much research on determining thresholds. There are essentially three types of thresholding: Hard-Thresholding, Soft-Thresholding, and Semi-Soft Thresholding. Thresholding is a commonly used tool, it can be applied alone or accompanied by some other techniques as well. For further reading on this topic see the references (Christov 2004; Burte and Ghongade 2012; Rabbani et al. 2011; Akazawa et al. 1991; Reddy et al. 2009; Donoho 1995; Jansen 2001; Patil et al. 2012).

Derivative-Based Algorithms

Especially in the older algorithms, the high-pass filter is usually applied as a differentiator. This point out the usage of the characteristic steep slope of the QRS complex for its detection. Points in a QRS-complex can be characterized by large amplitude fluctuations. This characteristic can be revealed by the slope value (Xia et al. 2015). The detection of a QRS complex is achieved by comparing the feature against a threshold. Usually, the threshold levels computed are signal dependent (Kohler et al.

2002). Some examples for derivative-based algorithms can be found in (Madeiro et al. 2012; Ahlstrom and Tompkins 1983; Balda 1977; Fraden and Neumann 1980; Gustafson 1977; Holsinger et al. 1971; Morizet et al. 1981; Okada 1979). Some algorithms compute the second derivative as well (Ahlstrom and Tompkins 1983; Balda 1977; Kemmelings et al. 1994).

ECG Slope Criteria

The slope of an R wave is a signal feature commonly utilized to locate the QRS complex in several QRS detectors (Ahlstrom and Tompkins 1983). A real-time derivative or an analog circuit algorithm providing slope information is useful for implementation. On the other hand, a derivative by its nature increases the unwanted higher frequency noise elements. Further, many abnormal QRS complexes with high amplitudes and prolonged durations can be avoided in a solely derivative method due to their comparatively low R wave slopes. Therefore, the R wave slope single can be inadequate for decent QRS detection. To obtain proper performance extraction of other parameters from the signal such as amplitude, width, and QRS energy is recommended (Nygards and Sörnmo 1983; Ligtenberg and Kunt 1983; Pan and Tompkins 1985).

In many real-time ECG monitoring devices, Pan and Tompkins algorithm has been extensively implemented and allows well-recognized performances to accomplish the QRS detection (Lin 2012). Some of the papers on this method can be listed as (Algra and Zeenberg 1987; Chazal and Celler 1996; Daskalov et al. 1998; Daskalov and Christov 1999; Daskalov and Chiristov 1999a).

Low-Pass Differentiation-Based Methods

Low-Pass differentiation (LPD) method, a classical method, can also be used to detect the ECG wave boundaries (Laguna et al. 1990). The fiducial ECG points (onset, peak, and end) can be determined using a low-pass filter from the differentiated ECG. Simple implementation and robustness are the advantages of the LPD based methods. The major shortcoming is due to the differentiation being sensitive to noise. A different drawback is that the delineation standards need a priori knowledge on the waveform and the delineation performance rely on thresholds chosen arbitrarily according to trial values (Lin 2012).

(Laguna et al. 1994; Speranza et al. 1993; Meij et al. 1994; Laguna et al. 1990) use Low-Pass Differentiation based algorithms.

Algorithms Based on Digital Filters

Adaptive filters can adjust signal properties according to the special qualities of the examined signals. Noise elimination utilizing filters eliminates noise along with important knowledge. If the signal and noise are overlying, then filters will eliminate the signal of interest. The adaptive filters can overwhelm this problem. Adaptive interference canceling is a very effective approach to resolve the problem of signals and interferences with overlying spectra (Lakshmi et al. 2014).

Algorithms based on more complex digital filters were declared in (Börjesson et al. 1982; Dokur et al. 1997; Engelse and Zeelenberg 1979; Fancott and Wong 1980; Hamilton and Tompkins 1986; Keselbrener et al. 1997; Leski and Tkacz 1992; Ligtenberg and Kunt 1983; Nygards and Hulting 1979; Okada 1979; Pan and Tompkins 1985; Sörnmo et al. 1982; Sun et al. 1992; Suppappola and Sun 1994; Yu et al. 1985; Afonso et al. 1996; Afonso et al. 1995; Afonso and Tompkins 1995a; Afonso et al. 1999; Merino et al. 2015; Prasad and Varadarajan 2014). Adaptive Filtering is used in (Soria et al. 1998; Burte 2012; Manikandan and Soman 2012; Merino et al. 2015; Zhidong and Juan 2010; Xue et al. 1992).

Wavelet-Based QRS Detection

Wavelet Transformation was formed by Grossman and Morlet in 1984 and is utilized for feature extraction. The wavelet transform presents a representation of the signal in the time-scale domain, allowing the design of the temporary features of a signal at various resolutions; hence, it is a proper tool to examine the ECG signal, which is described by a cyclic existence of patterns with different frequency content (QRS complexes, P and T waves). Further, the noise and artifacts concerning the ECG signal also seem at various frequency bands, so having different contribution at the different scales (Martinez et al. 2004).

The wavelet analysis method consists of using a wavelet function, also named an analyzing wavelet or mother wavelet. The transient analysis is done by contracting and dilating of the same analyzing wavelet function. In wavelet transform, the original signal is described concerning a wavelet extension concerning the analyzing mother wavelet. Hence, we can implement the data operations using the corresponding wavelet coefficients. Because wavelet transforms measures the similarity between the wavelet basis function and the analyzed signal, it is essential to determine the suitable wavelet basis functions that are proper for the signal of interest that gives the great similarity index. Then the powerful similarity index will be the compact wavelet coefficients of the signal.

The wavelet function should have a special form that we would like to limit in the original signal. Then, one rule for deciding the wavelet function is it seems alike to the patterns of the original QRS waveform. Hence, our selection is based on the form of the QRS waveforms to be identified in the signal. As a result, we can get only essential signal compositions across the wavelet subbands with suitably elected wavelet functions. In the presence of noise, the successful separation of QRS complexes from noise can then be obtained.

The following method is performed in wavelet analysis:

1. Do multiscale decomposition of the signal utilizing wavelet functions
2. Do wavelet thresholding to remove QRS from noise at each subband
3. Locate the arrival times of QRS complexes at across the subbands
4. Combine the QRS feature vectors at different subbands

In multiscale wavelet decompositions, the higher scales match to the more "stretched" wavelets that capture the low-frequency contents of the signal. The more stretched wavelet will let the longer part of the signal to be covered, and so the rarer the signal features are measured by the wavelet coefficients. It was recognized that QRS complexes are best expressed at higher scales of the wavelet transforms, which our determination was done respectively.

The wavelet transform has been shown to be a valuable tool for non-stationary signal analysis (Boutaa 2006). Thresholding is utilized in wavelet domain to smooth out or to eliminate some coefficients of wavelet transform subsignals of the measured signal. The noise content of the signal is decreased, dramatically, within the nonstationary situation. The denoising method that implements thresholding in the wavelet domain has been introduced by Donoho (Zhang and Bao 2002; Zhang et al. 2005). It has been shown that the Donoho's method for noise decrease serves well for a broad class of one dimensional and two-dimensional signals. Other methods for threshold value estimators can be seen in (Donoho 1992; Chang et al. 2000; Taouli and Reguig 2013).

There has been a quite number of papers on this topic. Some of them are listed here: (Afonso and Tompkins 1995; Senhadji et al. 1995; Li and Tai 1995; Sahambi et al. 1997; Bahoura et al. 1997; Martinez et al. 2004; Abibullaev and Seo 2011; Martis et al. 2013; Kabir Shahnaz 2012; Kabir and Shahnaz 2012a; Faezipour et al. 2009; May et al. 1997; Inoue and Miyazaki 1998; Cellar et al. 1997; Crowe et al. 1992; Kokturk 1998; Khadra et al. 1997; Senhadji et al. 1990; Tuteur 1988; Afonso et al. 1999; Di-Virgilio et al. 1997; Gyaw and Ray 1994; Kadambe et al. 1999; Mallat and Hwang 1992; Rao 1997; Sayadi 2007 and Shamsollahi; Meyer et al. 2006; Sahoo et al. 2015; Schuck and Wisbeck 2003; Rabbani et al. 2011; Madeiro et al. 2012; Cuiwei et al. 1995; Banerjee et al. 2012; Bouaziz et al. 2014; Zidelmal et al. 2012; Saxena et al. 2002; Ieong et al. 2012; Oliveira and Cortez 2004; Abibullaev and Seo 2011; Benali et al. 2012; Chikamori et al. 1995;

Ruchita and Sharma 2010; Kadambe et al. 1999; Patil et al. 2012; Rute et al. 2014; Patil and Chavan 2012a; Acharya et al. 2005; Istepanian et al. 2001; Agante and Marques 1999; Kania et al. 2007; Özbay et al. 2011; Poornachandra 2008; Mohmoodabadi et al. 2005; Dib et al. 2011; Yu and Chou 2007; Addison 2005; Provaznik et al. 2000; Martinez et al. 2004; Yimman et al. 2004; Brouse et al. 2006; Nagendra et al. 2011; Ho et al. 2003; Alfaouri and Daqrouq 2008; Garg et al. 2010; Reddy et al. 2009; Sahambi et al. 1997a; Choouakri et al. 2011; Chan et al. 2008; Lannoy et al. 2008; Wai et al. 2009; Wiklund et al. 1997; Senhadji et al. 1990; Kadambe et al. 1992; Bradie 1996; Ye and Ouyang 1996; Tuzman et al. 1996; Chen et al. 2006; Hahoura et al. 1997; Ivanov et al. 1996; Gamero et al. 2002; Yochum et al. 2016; Romero et al. 2005; Ghaffari et al. 2008; Burte and Ghongade 2012; Xu et al. 2007; Ercelebi 2004; Minhas and Arif 2008; Sevindir et al. 2015; Sevindir et al. 2016; Sevindir et al. 2017a; Sevindir et al. 2017b; Sevindir et al. 2017c; Sevindir et al. 2018a; Sevindir et al. 2018b).

Filter-Bank Methods

Filter banks (FB) are nearly linked to wavelets. In a filter bank method of QRS detection, sub-bands at varying scales are connected to verify the positions of the local maxima (Afonso et al. 1999). The filter bank method is based on the fact that the QRS complex has a simultaneous appearance in the sub-bands, whereas other ECG waves and noise may not show this characteristic form. This is similar to the wavelet-based method for QRS detection. (Lin 2012) Their employment to QRS detection is stated in (Afonso et al. 1999). Studies on the idea and use of FB's is given in the literature (Vaidyanathan 1993; Strang and Nguyen 1996; Soman et al. 1993). An FB-based algorithm includes a collection of analysis filters; they decompose a signal into subband signals with uniform frequency bands. Downsampling continues for the subbands as well because the subband bandwidth is much smaller than that of the input signal. Processing can be done on the subbands according to a particular application. Furthermore, the subbands may be reproduced by a set of synthesis filters which will perfectly reconstruct the input signal. Filter Bank allows time and frequency-dependent processing to be performed at a computationally effective rate to analyze the ECG signal.

Some of the Filter banks applications are (Afonso et al. 1999; Borjesson et al. 1982; Engelse and Zeelenberg 1979; Dokur et al. 1997; Hamilton and Tompkins 1986; Ligtenberg and Kunt 1983; Pan and Tompkins 1985; Christov 2004).

Neural Network Methods

Behaviour of neural networks (NN) may be compared to the functionality of the human brain and structure of organic neurons (Macy and Pandya 1996). In the training

process in that the examples are presented; a neural network learns new information. A network works exactly only after it has processed a set of appropriate examples; thus, examples should be chosen very carefully. (Neural Networks 2015) Artificial neural networks (ANN) have been extensively applied in classification, nonlinear signal processing, and optimization. Their performance was demonstrated to be high to classical linear approaches in many applications (Kohler et al. 2002). ANN can be used to get information from complex or inaccurate data. Furthermore, they may be involved in the extraction of patterns and detection of processes which are extremely complicated. With the help of data acquired in a training process or any previous experience, neural networks can organize themselves and consequently represent information obtained in the preceding training period. Since ANN facilitate parallel processing, they can operate in real time. This strength may be further enhanced by using special hardware devices (Neural Networks 2015).

We can list four reasons to use an ANN as a classifier: (1) By iterative training, weights representing the solution can be found; (2) It has a simple structure for physical implementation; (3) It can easily map complex class distributions; and (4) Proper results for the input vectors can be obtained due to generalization property of the ANN (Dokur and Ölmez 2001; Iscan et al. 2009).

A neural network consists of neurons, which process information. Neurons are interconnected and work parallel. Basic distinction we can make between the types of networks is in patterns of connection between the units and the spread of the data. (Kicmerova 2009). Feed-forward networks spread data from input to output units, but there are no feedback connections. Classical examples of feed-forward networks are the perceptron and adaline (Krose and Smagt 1996). Recurrent networks as opposed to recurrent networks as counter to feed-forward networks have feedback connections and the dynamical properties of the networks are significant. Typical examples are the Elman, Jordan and Hopfield networks (Krose and Smagt 1996). We can also categorize the neural networks on their corresponding training algorithms: supervised networks and unsupervised networks. Characteristic example of neural network that is trained by unsupervised attitude is a self-organizing map (SOM), which generates a low-dimensional discrete representation of the input space of the training samples (Yin et al. 2007).

Some MLP applications can be listed as (Bishop 1995; Haykin 1999; Nadal and Bossan 1993; Prasad and Sahambi 2003; Yong et al. 2009; Yu and Chou 2008; Barro et al. 1998; Dokur et al. 1997; Fernandez et al. 1998; Ham and Han 1996; Hu et al. 1993; Laguna et al. 1996; Maglaveras et al. 1998; Maglaveras et al. 1998a; Mahalingam and Kumar 199;, Strintzis et al. 1992; Suzuki 1995; Trahanias and Skordalakis 1990; Vijaya et al. 1998; Xue et al. 1992; Tsipouras and Fotiadis 2004; Benali et al. 2012; Jiang and Kong 2007; Shah et al. 1992; Yu and Chou 2007; Bin et al. 2004; Martis and Chakraborty 2011; Xie et al., 1992).

In ECG signal processing, besides the multilayer perceptron (MLP), radial basis function (RBF) and learning vector quantization (LVQ) networks are mostly used methods (Kohler et al. 2002).

Artificial Neural Networks applications can be listed as (Lagerholm et al. 2000; Silipo and Marchesi 1998; Dokur et al. 1997; Bystricky and Safer 2002; Shende and Patel 2013; Barro et al. 1998; Abibullaev and Seo 2011; Dreiseitl and Ohno 2002; Basheer and Hajmeer 2000; Chaudhuri and Bhattacharya 2000; Fausett 1994; Haykin 1994; Hagan and Menhaj 1994; Dokur and Ölmez 2001; Song et al. 2005; Song and Lee 2006; Hu et al. 1993; Vijaya et al. 1998; Vijaya et al. 1997; Arbateni and Bennia 2014; Gothwal et al. 2011; Xue et al. 1992; Mehta and Lingayat 2007).

A probabilistic neural network (PNN) is a feed-forward neural network, derived from the Bayesian network and a statistical algorithm named Kernel Fisher Discriminant analysis (Banupriya and Karpagavalli 2014). PNN is a special type of radial basis-function networks. PNN consists of an input layer, a hidden layer, and an output layer. The input layer is solely a distribution layer. No computation is performed in this layer. The hidden layer is also called the pattern layer. Neurons in the hidden layer utilize multi-dimensional kernels to estimate the probability density function for classification. One of the most popular kernels is Gaussian function since it guarantees the convergence of the neural network (Rutkowski 2004; Yu and Chou 2007).

Some of the PNN applications in the field of ECG waveform classification can be listed as (Haykin 1996; Duda et al. 2001; Specht 1990; Mao et al. 2000; Rutkowski 2004; Banupriya and Karpagavalli 2014; Yu and Chou 2007; Vijaya et al. 1997).

Multi-Layer Perceptron

Multi-Layer Perceptron (MLP) is frequently used in biomedical signal processing (Lippmann et al. 1987; Miller et al. 1992). In the Multi-Layer Perceptron Neural Network (MLPNN), the network is arranged into three layers: input layer, hidden layer, and output layer. The advantage of MLPNN is that its fast operation, ease of implementation and requiring small training sets. The number of inputs denotes the number of selected features and, number of outputs denotes the number of formed classes. The number of neurons in the hidden layer of it estimates the complexity of an ANN. A large number of neurons in hidden layer causes to more complexity, while a smaller number of neurons in hidden layer causes classification errors. No specific criterion was defined for making this decision in the hidden layer. By using trial and error method, the number of neurons has to be decided (Garrett et al. 2003; Hekim 2012).

Some papers on MLP algorithm can be listed as (Lippmann 1987; Miller et al. 1992; Reilly et al. 1982; Golpayegani and Jafari 2009; Özbay et al. 2006; Kumari and Kumar

2013; Hassanpour and Boashash 2011; Hagan et al. 1996; Dokur and Ölmez 2001; Mai et al. 2011; Lippmann et al. 1987; Miller et al. 1992; Dokur and Olmez 2001).

Radial Basis Function Network

The Radial Basis Function Network is a universal approximator, with a concrete basis in the conventional approximation theory. RBFN is a well-liked substitute to the MLP, because of its more straightforward configuration and a more rapid training process (Du 2006). Learning is an approximation problem that is violently dependent to standard approximation methods, like generalized splines and regularization methods. RBFN has its beginning in the accurate interpolation of a set of data points in multidimensional space (Powell 1987). RBFN architecture is comparable to the traditional regularization networks (Poggio and Girosi 1990), in which the basic functions are the Green's functions of the Gram operator connected with the stabilizer. If the stabilizer shows radial symmetry, the basic functions are radially symmetric, and RBFN is reached (Du 2006; Alickovic and Subasi 2015).

Radial-based networks require more neurons than standard feedforward backpagation networks but can often be designed in the timeframe required to train standard feedforward networks. (Varshney et al. 2014).

The advantage of RBF networks over MLP networks is, as in the fuzzy-logic methods, the possibility to interpret the parameters. This makes the results more predictable and thus trustworthy. The LVQ network consists of three layers: an input layer, a competitive layer, and a linear layer. Unlike the MLP and RBF networks, which are trained by supervised learning algorithms, the LVQ network is adjusted in an unsupervised manner (Kohler et al. 2002).

(Du and Swamy 2006; Poggio and Girosi 1990; Powell 1987; Qin et al. 2005; Hassanpour and Boashash 2011; Karpagachelvi et al. 2010; Husain and Fatt 2007; Zayandehroodi et al. 2010; Alickovic and Subasi 2015; Varshney et al. 2014; Sadr et al. 2011) (Song et al. 2005; Alickovic and Subasi 2015; Özbay et al. 2006; Özbay et al. 2011; Tsipouras and Fotiadis 2004; Zong and Jiang 1998) are some of the research papers on the topic.

Bayesian Filtering

Bayesian filtering can be used for ECG classification as well. Recently, Bayesian filters were proposed for ECG denoising [26 in Sayadi] and filtering cardiac contaminants (Sayadi et al. 2010). Some of the research papers on this topic can be listed as (Mackay 1992; Sayadi et al. 2010; Oliveira et al. 2011; Lin et al. 2010; Sayadi and

Shamsollahi 2009; Popescuet al. 1998; Sameni et al. 2006; Alickovic and Subasi 2015; Barr et al. 2010; Sameni et al. 2007; Sayadi and Shamsollahi 2007; Popescu et al. 1998; McSharry et al. 2003; Friesen et al. 1990; Kohler et al. 2007).

Hidden Markov Models

A Hidden Markov Modeling (HMM) is a stochastic finite state automata. This model has been applied to signal analysis problems since the mid-1970. The importance of these models is firstly because of an automatic parameter estimation algorithm discovered by Baum and Eagon (Baum and Eagon 1967) in the late 1960's and afterwards it was applied to speech processing (e.g., Baker 1975; Jelinek 1976). Refinements in the theory and implementation have considerably enhanced this model. (e.g., (Bahl et al. 1983; Rabiner and Juang 1986; Lee 1988)) This model associates structural and statistical knowledge of the ECG signal parametric model (Seymore et al., 1999).

(Coast and Stern 1990; Clavier et al. 2002; Thomas et al. 2007; Coast et al. 1987; Coast 1988; Hughes et al. 2003; Rodrigo et al. 2006; Andreão et al. 2006; Nov´ak 2003) and (Cheng and Chan 1998) are some studies on the topic.

Mathematical Morphology

During the last decades, mathematical morphology has become a powerful technique for signal processing. It was originally applied to analyze geological images and biological specimens. Because of its rich theoretical framework, algorithmic efficiency, easy implementation on special hardware, and suitability for many shape-oriented problems, its popularity among the dominant image analysis methodologies has increased (Laguna et al. 1994; Donoho and Johnstone 1994; Stein 1981; Singh and Tiwari 2006). It has been successfully applied in numerous disciplines, such as mineralogy, histology, and medical diagnosis. (Taouli, Reguig 2013) Consequently, nowadays it offers many theoretical and algorithmic tools to and inspires new directions in many research areas from the fields of signal processing, image processing and computer vision, and pattern recognition (Tadejko and Rakowski 2007).

A morphological filter is a nonlinear signal transformation that locally modifies geometric features of a signal. It stems from the basic operations of a set-theoretical method for signal analysis, named mathematical morphology, which was introduced by Serra (Serra 1982). In morphological filtering, a signal is considered as a set, and its geometrical features are modified via morphologically convolving the signal with a structuring element. Altering the structuring element lets extraction of different types of information from the signal. The values of the structuring element define the shape of the

output waveform (Tadejko and Rakowski 2007; Chu-Song et al. 1999; Maragos 2000; Maragos and Schafer 1987; Maragos and Schafer 1990).

The basic idea of morphological signal processing is to modify the shape of a signal by transforming it via its interaction with another object, named structuring element. In practice, the structuring element should be more compact and should have a simpler shape than the original one. The fundamental operators of morphology transform include dilation, erosion, opening, and closing (Taouli and Reguig 2004; Taouli and Reguig 2006; Taouli and Reguig 2013). Opening and closing are derived operators defined in terms of erosion and dilation operations. In most of the applications, an opening is used to suppress peaks, while closing is used to suppress pits (Sun et al. 2005).

(Karpagachelvi et al. 2010; Tadejko and Rakowski 2007; Zong et al. 2003; Trahanias 1993; Zhang et al. 2009; Sun et al. 2005; Fei and Yong 2009) are some examples of research papers on the topic.

Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) decomposes a complex data set into its components called intrinsic mode functions (lMFs). This method is often useful for nonlinear and nonstationary processes such as ECG. The key point of EMD is to consider oscillations in signals at a very local level. The computation of EMD, not requiring any previously known value of the signal, is intuitive and adaptive with basic functions derived entirely from the data (Pal and Mitra 2010; Kabir and Shahnaz 2012; Kabir and Shahnaz 2012a; Sahoo et al. 2016).

Identifying the intrinsic oscillatory modes by their characteristic time scales in the signal empirically, and accordingly, decomposing the signal into intrinsic mode functions is the main task in EMD (Zhu et al. 2010).

A function can be used as an IMF if it satisfies the following two conditions:

a) In the entire data set, the number of local extrema and that of zero crossings must be equal to each other or different by at most one
b) At any point in the data set, both the mean value of the envelope defined by the local maxima and that defined by the local minima should be zero.

The systematic way to decompose the data into IMFs, known as the “sifting” process (Kabir and Shahnaz 2012). The goal of the sifting is to subtract the large-scale features of the signal repeatedly until only the fine-scale features remain (Narsimha et al. 2011).

(Pal and Mitra 2010; Zhu et al. 2010; Pal and Mitra 2012; Kim and Oh 2008; Huang et al. 1998; Kabir and Shahnaz 2012; Sahoo et al. 2016; Chacko and Ari 2012; Narsimha

et al. 2011; Slimane et al. 2010; Zhidong and Juan 2010; Pan et al. 2007; Weng et al. 2006) are examples.

k-Means Clustering

k-Means algorithm aims to divide the objects based on attributes or properties into some groups. The classification is made by minimizing the sum of the squares of the data and the distances between the corresponding cluster centroid. (Balouchestan et al. 2014).

Some of the related studies can be listed as (Selvakumar et al. 2012; Lin 1991; Donoso et al. 2013; Chew et al. 1967; Mehta et al. 2010; Kiranyaz et al. 2011; Duda et al. 2001; Xia et al. 2015; İşler and Narin 2012; Merino et al. 2015; Martis and Chakraborty 2011).

k-Nearest Neighbor Method

The k-Nearest Neighbor (k-NN) method is a non-parametric learning algorithm and a statistics-based tool for classification. These algorithm goals to find the most appropriately classified features that are most similar to the test characteristics and to verify the categories of test properties according to the categories of properties. The similarity between the two properties is measured by the Euclidean distance between them, where the decreasing distance indicates an increased similarity. In fact, it intends to find class labels for new unlabeled data points. The k-NN algorithm divides the data set into a test set and training set. For each line of the test set, the nearest k training kit objects are found and most of the data points determine the classification. If there are other data points for the nearest vector k, all candidates are in the k set. This algorithm consists of the following steps:

1) Training phase: In the training phase, the training examples are vectors in a multidimensional feature space (each with a class label); the feature vectors and class tags of the training vectors are stored. For a specific set of data points, remove the nearest neighbor using the distance function (Balouchestan et al. 2014).
2) Classification phase: Class of given number of data sets is made of the k nearest neighbor's labels (Balouchestan et al. 2014). In the classification phase, k is a user-defined constant, a query or test point (unlabelled vector) is classified by assigning a label, which is the most recurrent among the k training samples nearest to that query point. In other words, the k-NN method compares the query point or an input feature vector with a library of reference vectors, and the query

point is labeled with the nearest class of library feature vector. This way of categorizing query points based on their distance to points in a training data set is a simple, yet an effective way of classifying new points (Saini et al. 2013).

The k-NN is one of the statistical classification algorithms used for classifying objects based on closest training examples in the feature space. It is a lazy learning algorithm where the k-NN function is approximated locally, and all computations are delayed until classification. Although a training dataset is required, no actual model or learning is performed during the training phase; it is used only to populate a sample of the search space with instances whose class is known, for this reason, this algorithm is also known as the lazy learning algorithm. It means that the training data points are not used to make any generalization, and all the training data is needed during the testing phase (Saini et al. 2013).

When a sample whose class is unknown is presented for evaluation, the algorithm computes its k closest neighbors, and the class is assigned by voting among those neighbors. In k-NN algorithm, training phase is fast, but testing phase is costly regarding both time and memory (Thirumuruganathan 2010; Saini et al. 2013).

k-NN is a classification method based on closest training samples. An instance-based learning algorithm compare new problem instances with instances seen in training, which has been stored in memory. It is named instance-based because it constructs hypotheses directly from the training instances themselves. In this method, particular parameters need to be chosen base on which classification will be done. All samples present in the training set have their own value (Chitupe and Joshi 2013).

Some papers related to the topic can be listed as (Darko et al. 2007; Tang and Tseng 2009; Maleki et al. 2013; Zhan et al. 2006; Ming et al. 2004; Arif et al. 2009; Arif et al. 2010; Christov et al. 2005; Karimifard et al. 2006; Kutlu and Damla 2011; Balouchestani et al. 2014; Saini et al. 2013).

Least Square-Support Vector Machine

In some studies, the original SVM variant, known as the Least Square SVM, is a least square formulation of the problem. Suykens et al. (Suykens and Vandewalle 1999) suggested this. Conversely, standard SVM, the LS-SVM solution is usually not sparse. However, with the pruning and reduction techniques, a sparse solution is easily accessible. Depending on the number of training data set, either an iterative solver such as conjugate gradients methods (for large data set) can be used or direct solvers, in both cases with numerically reliable methods. For LS-SVM, there are many kernel function choices such as radial basis function, linear polynomial, sigmoid, b-spline, spline, etc.

However, the most used kernel functions are simple Gaussian, polynomial, and RBF functions (Tripathy et al. 2012).

(Suykens and Vandewalle 1999; Martis et al. 2013a; Tripathy et al. 2012) are only some examples on this topic.

QRS detection has been a research topic for more than 30 years. A wide diversity of algorithms has been proposed in the literature, for an extensive review, see (Kohler et al. 2002).

Automatic ECG heartbeat signal classification systems that may result in high performances are still attractive research field. Until now, several data mining techniques have been applied for ECG heartbeat signal classification. (Raghav and Mishra 2008) tried to use local fractal dimension for ECG signal-based heart disorders, and results achieved in this work were promising (Asl et al. 2008).

Principal Component Analysis

Principal Component Analysis is a linear dimensionality reducing method that looks for projection of the data into the directions of highest variability (Duda et al. 2001). It calculates the principal components, which are the basis vectors of directions in decreasing order of variability. Basis vector for the direction of highest variability is given by the first principal component. The basis vector for the next direction orthogonal to the first one is produced by the second principal component and so on. A percentage of total variability of the data is set as the threshold to select the number of principal components (Martis et al. 2013).

PCA was invented in 1901 by Karl Pearson and afterwards developed independently by Harold Hotelling in 1930 (Principal Component Analysis 2017). The PCA transforms the correlated vectors into linearly uncorrelated vectors which are called as “Principal Components” (Principal Component Analysis 2017; Gu et al. 2009). This is a classical method of Second Order Statistics. Covariance matrix is efficient for it. PCA helps in reduction of feature dimensions. Ranking will be done by using PCA based on the variability of the signal properties. This ranking helps in classification of the data. The application of PCA in a BCI system yields best classification results (Zeman 2000). The PCA is good but it is not as well as ICA (Draper et al. 2003).

PCA is s a feature extraction method as well as a pre-processing technique. It is a powerful tool for analyzing and for dimension reduction of data without loss of information (Mbeledogu et al. 2012). Using PCA the information present at all the time series multi-channel is extracted as principal components. By eliminating the artefacts and by forming the principal components PCA reduces the dimensions of signals (Prochazka et al. 2010).

Some PCA applications can be found in (Duda et al. 2001; Anthony and Terry 1997; Kalpana et al. 2013; Martis et al. 2012; Chawla et al. 2006a; Hao and Qing 2005).

Linear Discriminant Analysis

PCA involves a mathematical procedure that is based on principle component determination that determines the amount of variability present in the data set. Linear Discriminant Analysis (LDA), also known as Fisher's Discriminant Analysis, researches for directions in which higher class separation is achieved. LDA is a dimensionality reducing method. It deals with discrimination between classes, whereas PCA deals with the data in its entirety for the principal components analysis without paying attention to the underlying class structure. LDA searches for those vectors in the underlying space that best discriminate among classes (rather than those that best describe data as in PCA.) PCA seeks directions that are efficient for representation; LDA seeks directions that are efficient for discrimination (Toygar and Acan 2003). It is generally believed that algorithms based on LDA are superior to those based on PCA (Principal Components Analysis) (Aleix and Avinash 2001).

(Duda et al. 2001; Afonso et al. 1999; Toygar and Acan 2003; Martis et al. 2013; Balakrishnama et al. 1999; Song et al. 2005) are some of the LDA applications in the literature.

Independent Component Analysis

In signal processing, Independent Component Analysis (ICA), a special case of blind source separation, is a computational method for separating a multivariate data into additive subcomponents. It is a nonlinear dimensionality reduction method. It separates the artefacts from the signals into independent components based on the characteristics of the data without relying on the reference channels. We can use it also as a feature extraction method. Its forms the components that are independent to each other. From the components, essential features were extracted using ICA. An important application of ICA is Blind Source Separation (Torse et al. 2012).

Multiscale Principal Component Analysis (MSPCA) has been introduced as a defect detection method for the time series data (Bakshi 1998). This method associates the ability of PCA to extract the relationship among variables, then, to decorrelate the cross-correlation with that of wavelet analysis to decompose the time-series data into several frequency scales. Multiscale PCA reconstructs simplified multivariate signal, starting from a multivariate signal using a simple representation at each resolution level.

(Castells et al. 2007a; Comon 1994; Yuand Chou 2008; Taigang et al. 2005; Chawla et al. 2006a; Chawla et al. 2006b; Chawla et al. 2006c; Yu and Chou 2007; Martis et al. 2013; Hyvärinen 1999; Naik and Kumar 2011) are some short list on the topic.

Support Vector Machine

We can define SVM like that: it is a highly nonlinear and single layer network, which has higher generalization ability in the sense that it can categorize unseen data exactly (Gunn 1998; Christianini and Taylor 2000). It maximizes the distance between the patterns and the class separating hyper-plane. SVM minimizes the structural risk than the empirical risk. An objective function is formulated based on the distances to the class separating hyper-plane and the optimization process is realized. Generally, the patterns are not linearly separable, so first they were mapped into a high dimensional space using a proper kernel, and then the optimization was realized. For mapping the data into high dimensional space, various kernel transformations namely: quadratic, polynomial and radial basis function (RBF) were used.

Support Vector Machine (SVM) (Gunn 1998; Christianini and Taylor 2000) is a single layer highly non-linear network. It has higher generalization ability to categorize unknown data accurately. It minimizes the structural risk and maximizes the distance between the patterns to the class separating hyperplane at the same time. In general, if the patterns are not linearly separable, they are mapped to a higher dimensional space, where they are supposed to be linearly separable. Commonly quadratic, polynomial of order 3 and Radial Basis Function (RBF) kernels are used for kernel transformation.

We can list the algorithms based on SVM for ECG signals as (Gunn 1998; Christianini and Taylor 2000; Suykens and Vandewalle 1999; Melgani and Bruzzone 2004; Hsu and Lin 2002; Vapnik 1995; Vapnik 1998; Melgani and Bazi 2008; Ghoggali et al. 2009; Pasolli et al. 2009; Acır 2006; Asl et al. 2008; Besrour et al. 2008; Song et al. 2005; Zidelmal et al. 2013; Karpagachelvi et al. 2012; Varshney et al. 2014; Daamouche et al. 2012; Mehta and Lingayat 2008; Mehta and Lingayat 2007; Shen et al. 2012; Hao and Qing 2005; Osowski et al. 2004; Martis Chakraborty 2011; Kampouraki et al. 2009; Sevindir et al. 2018a; Sevindir et al. 2018b; Cetinkaya et al. 2018a, Cetinkaya et al. 2018b).

Fourier Transform Method

Fourier transform (FT), identified by Joseph Fourier in the 19th century (Adam), extracts the signal features by transforming a signal from time domain to frequency domain. Despite working quite well for stationary signals, FT cannot measure both time

and frequency. Some FT techniques may exhibit better performance in other cases (Merzagora et al. 2006). FT converts a time-varying signal into a histogram, which shows the various frequencies that compose the signal. After the transform is applied, an amplitude versus frequency graph indicating the relative presence of various frequencies in the original voltage reading is available (Khanacademy 2017). There are a number of papers on this topic. (Gothwal et al. 2011; Martinmaki et al. 2006) can be listed as examples.

The Discrete Cosine Transform (DCT) is a Fourier-related transform of roughly twice the length, operating on real data with even symmetry, where in some variants the input and/or output data are shifted by half a sample (Ahmed 2006). Some research papers on the topic are (Elliot and Rao 1982; Jalaleddme et al. 1990; Britanak and Rao 2002; Nikolajevic and Fettveis 2003; Ahmed 2006; Vernon et al. 2005; Martis et al. 2013a).

Autoregressive Model/Moving-Average Model

In signal processing and statistics, an autoregressive (AR) model is a representation of a sort of random process. It is utilized to describe certain time-varying processes such as natural, economical, medical. For an AR model, the output variable is dependent linearly on its previous values; therefore, a stochastic difference equation is obtained. Unitedly with the moving-average (MA) model, it forms ARMA and ARIMA models of time series.

These models can be used for computer-assisted recognition of cardiac diseases. AR model is the most frequently used method for nonstationary signals where parameters are supplied to the model. The difficulty lies in establishing the parameter model property. (Lakshmi et al. 2014). Many research papers are present on the subject: some of them can be listed as (AlMejrad 2010; Ofner et al. 2011; Geo et al. 2002; Dewangan and Shukla 2015, Lakshmi et al. 2014; Padmavathi and Ramakrishna 2015; Chen and Chen 2003; Chaturvedi and Yadav 2014).

Other Methods

Besides mentioned methods there are techniques such as: Hilbert Transform Methods (Rabbani et al. 2011; Madeiro et al. 2012; Manikandan and Soman 2012; Oliveira and Cortez 2004; Jing-Tian et al. 2007; Benitez et al. 2000; Benitez et al. 2001; Ghaffari and Homaeinezhad 2006; Nygards and Sörnmo 1983; Manriquez and Zhang 2008; Arzeno et al. 2008), local fractal dimension (Raghav and Mishra 2008; Yu and Chou 2007; Prasad and Sahambi 2003) Hermite functions (Lagerholm et al. 2000; Laguna et al. 1996)

nonlinear time-scale decomposition (Strumillo 2002; Suppappola and Sun 1994), multiplication of backward difference (Sun et al. 1992; Suppappola and Sun 1994), mathematical models (Sornmo et al. 1981; Sörnmo 1987; Murthy and Prasad 1992; Vila et al. 2000), S-Transform (Zidelmal et al. 2014; Stockwell 2006), genetic algorithms (Chris D Nugent et al. 2002; Poli et al. 1995; Cetinkaya et al. 2018c), etc.

ECG Databases

It is extremely crucial to evaluate a QRS detector algorithm using a standard arrhythmia database. There are several databases of ECG recordings available for analyzers to evaluate ECG. Most of these databases provide several essential demands: They include representative standard ECG signals with broad variations in ECG characteristics along with seldom-observed yet clinically significant signals. Two or more cardiologists working independently annotate the signals manually. Digitized computer-readable signals are held for researchers to perform a fully automated, strictly reproducible test in the digital domain allowing one to establish with certainty the effects of algorithm modifications on performance (Sanamdikar et al. 2015).

At present, the following ECG databases are available (Moody and Mark 2001; Goldberger et al. 2000; Mark et al. 1982; Moody and Mark 1990, physiotools 2017; physionet 2017):

- **The American Heart Association Database (AHA DB):** The database contains 80 Holter ECG recordings, 35 minutes each for evaluation of Ventricular Arrhythmia detectors.
- **The Massachusetts Institute of Technology-Beth Israel Hospital Arrhythmia Database (MIT/BIH DB):** The database was the first commonly available set of standard test material for evaluation of arrhythmia detectors. It contains 48 records, each including two-channel ambulatory ECG signals for 30 minutes duration selected from 24-hour Holter ECG recordings of 47 individuals.
- **The European Society of Cardiology ST-T Database (ESC DB):** The database contains 90 Holter ECG recordings from two hours each.
- **The Noise Stress Test Database (NST DB):** The database contains 12 Holter ECG recordings, 30 minutes each; supplied with the MIT DB.
- **The Creighton University Sustained Ventricular Arrhythmia Database (CU DB):** The database contains 35 records, 8 minutes each; supplied on the second edition of the MIT DB CD-ROM.
- **Common Standard for Quantitative Electrocardiography (CSE):** The European community in the year 1980, started a project under the leadership of late J. I.Willems with an aim of establishing “common standards for Quantitative

Electrocardiography (CSE)". The working party developed for the three CSE reference databases.

- **AAMI Standard:** MIT-BIH heartbeat types are combined according to Association for the Advancement of Medical Instrumentation (AAMI) recommendation. AAMI standard emphasize the problem of classifying ventricular ectopic beats (VEBs) from the non- ventricular ectopic beats.

All these databases represent a very much effort by many workers. Nonetheless, even these databases do not adequately represent the diversity of real-world ECG data.

CONCLUSION

Automated ECG analysis has been an attractive topic for researchers since 1980s. Although various number of algorithms have been proposed in the literature for extracting feature from an ECG signal, the search continues. Almost all the methods to develop these software algorithms comes with some shortcomings as well. Testing for real data is always a challenge as well. To increase accuracy usage of combination of two or more methods are very common. In the earlier years, the computational complexity and thus the performance of the algorithms was the main objective. Nowadays the detection performance is the principal purpose; the computational complexity becomes less significant than the detection rate. On the other hand, development of algorithms working battery-driven or wireless devices for automated ECG analysis is what is needed. Algorithms working on real life data, battery-driven or wireless devices for automated ECG analysis are only some part of the list.

REFERENCES

Abibullaev, B., Seo, H. D. (2011). "A new QRS detection method using wavelets and artificial neural networks." *Journal of Medical Systems* 35(4): 683–691.

Acharya, R. U., Bhat, P. S., Kannathal, N., Rao, A., Lim, C. M. (2005). "Analysis of cardiac health using fractal dimension and wavelet transformation." *ITBM-RBM* 26: 133–139.

Acır, N., (2006). "A support vector machine classifier algorithm based on a perturbation method and its application to ECG beat recognition systems." *Expert Syst. Appl*. 31: 150–158.

Adam I. *Complex Wavelet Transform: application to denoising*, Phd. Thesis, Polytechnic University of Timisoara Universite de Rennes, available at http://www.tc.etc.upt.ro/docs/cercetare/teze_doctorat/tezaFiroiu.pdf.

Addison, P. S., (2005). "Wavelet transforms and the ECG: a review." *Physiol. Meas*. 26: 155–199.

Afonso, V. X., Tompkins, W. J., Nguyen, T. Q., Trautmann, S., Luo, S., (1995). "Filter bank-based processing of the stress ECG." in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*., available: CD-ROM].

Afonso, V. X., Tompkins, W. J., (1995). "Detecting ventricular fibrillation: Selecting the appropriate time-frequency analysis tool for the application." *IEEE Engineering in Medicine and Biology Magazine* 14: 152-159.

Afonso, V. X., Tompkins, W. J., Nguyen, T. Q., Michler, K., Luo, S., (1996). "Comparing stress ECG enhancement algorithms: With an introduction to a filter bank based approach." *IEEE Eng. Med. and Biol., Mag*. 15(3): 37–44.

Afonso, V. X., Tompkins, W. J., Nguyen, T. Q., Luo, S., (1999). "ECG beat detection using filter banks." *IEEE Trans. Biomed. Eng*. 46: 192-202.

Agante, P. M., Marques de Sa, J. P., (1999). "ECG noise filtering using wavelets with soft-thresholding methods." *Coput. Cardiol*, 26: 535–538.

Ahlstrom, M. L., Tompkins, W. J., (1983). "Automated high-speed analysis of holter tapes with microcomputers." *IEEE Trans. Biomed. Eng.* 30: 651-657.

Akazawa, K., Motoda, K., Sasamori, A., Ishizawa, T., Harasawa, E., (1991). "Adaptive threshold QRS detection algorithm for ambulatory ECG." *Proceedings of 18th Annual Conference on Computer in Cardiology*: 445-448.

Akwmedical-What-Ecg, (2017). https://www.akwmedical.com/blog/what-ekg-machine-and-how-does-it-work Accessdate: 10.03.2017.

Aleix, M., Avinash C., (2001). "PCA versus LDA." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(2):228-233.

Alickovic, E., Subasi, A., (2015). "Effect of Multiscale PCA De-noising in ECG Beat Classification for Diagnosis of Cardiovascular Diseases." *Circuits Syst Signal Process* 34: 513–533.

Alfaouri, M., Daqrouq, K., (2008). "ECG signal denoising by wavelet transform thresholding." *Am J Appl Sci*. 5(3):276–81.

Algra, A., Zeelenberg, H. L. B. C., (1987). "An algorithm for computer measurement of QT intervals in the 24-hour ECG," in *Computers in Cardiology*. Los Alamitos, CA: IEEE Computer Society Press, 117–119.

AlMejrad, A. S., (2010). "Human Emotions Detection using Brain Wave Signals: A Challenging." *European Journal of Scientific Research* 44(4).

Andreão, R. V., Dorizzi, B., Boudy, J., (2006). "ECG Signal Analysis through Hidden Markov Models." *IEEE Transactions on Biomedical Engineering* 53(8).1541-1549.

Anthony, B., Terry, S., (1997). "The independent components of natural scenes are edge filters." *Vision Research* 37(23): 3327-3338.

Arbateni, K., Bennia, A., (2014). "Sigmoidal radial basis function ANN for QRS complex detection" *Neurocomputing* 145: 438–450.

Arif, M., Akram, M. U., Afsar, F. A., (2009). "Arrhythmia beat classification using pruned fuzzy k-nearest neighbor classifier soft computing and pattern recognition." in *SOCPAR '09*, Malacca, Malaysia, (December 4-7).

Arif, M., Akram, M. U., Minhas, F. A. A., (2010). "Pruned fuzzy K-nearest neighbor classifier for beat classification." *J. Biomed. Sci. Eng.* 3(4): 380–389.

Arzeno, N., De-Deng, Z., Poon, C., (2008). "Analysis of first-derivative based QRS detection algorithms." *IEEE Trans. Biomed. Eng*. 55(2): 478–484.

Ashley, E. A., Niebauer, J., (2004). *Conquering the ECG*, London: Remedica.

Asl, B. M., Setarehdan, S. K., Mohebbi, M. (2008). "Support vectormachine-based arrhythmia classification using reduced features of heart rate variability signal." *Artif. Intell. Med*. 44: 51–64.

Bahl, L. R., Jelinek, F., Mercer, U. L., (1983). "A maximum likelihood approach to continuous speech recognition." *IEEE Trans. Partern Anal. Mach. Intell*. 5(2): 179-190.

Bahoura, M., Hassani, M., Hubin, M., (1997). "DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis." *Comput. Meth. Programs Biomed*. 52: 35–44.

Balakrishnama, S., Ganapathiraju, A., Picone, J., (1999). "Linear Discriminant Analysis for Signal Processing Problems." *Institute for Signal and Information Processing*: 78-81.

Balda, R. A., (1977). *Trends in Computer-Processed Electrocardiograms*. Amsterdam: North Holland: 197-205.

Balouchestani M., Sugavaneswaran, L., Krishnan, S., (2014). "Advanced K-Means Clustering Algorithm for Large ECG Data Sets Based on K-SVD Approach" (2014) *9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP)*, Manchester, United Kingdom, (July 23-25).

Baker, I. K., (1975). *Stochastic Modeling for Automatic Speech Understanding*. New York: Academic: 521-542.

Bakshi, B. R., (1998). "Multiscale PCA with application to multivariate statistical process monitoring." *AIChE Journal* 44(7): 1596-1610.

Banerjee, S., Gupta, R., Mitra, M., (2012). "Delineation of ECG characteristic features using multiresolution wavelet analysis method." *Measurement* 45: 474–87.

Banupriya, C. V., Karpagavalli, S., (2014). "Electrocardiogram Beat Classification using Probabilistic Neural Network." *International Journal of Computer Applications, Machine Learning -Challenges and Opportunities Ahead, MLCONF*. San Francisco, USA, (14 November).

Barr, R. C., Nolte, L. W., Pollard, A. E., (2010). "Bayesian quantitative electrophysiology and its multiple applications in bioengineering." *IEEE Reviews in Biomed* 3: 155 –168.

Barro, S., Delgado, M. F., Sobrino, J. A. V., Regueiro, C. V., Sanchez, E., (1998). "Classifiying multichannel ECG patterns with an adaptive neural network." *IEEE Eng. Med. Biol. Mag*.17: 45-55.

Basheer, I. A., Hajmeer, M., (2000). "Artificial neural networks: fundamentals, computing, design, and application." *J. Microbiol. Methods* 43: 3-31.

Baum, L. E., Eagon, J. A., (1967). "An inequality with applications to statistical prediction for functions of Markov processes and to a model for ecology." *Bull. Amer. Math. Soc*. 73: 360-363.

Benali, R., Reguig F. B., Slimane Z. H., (2012). "Automatic classification of heartbeats using wavelet neural network." *Journal of medical systems* 36(2): 883–892.

Benitez, D. S., et al. (2000). "A new QRS detection algorithm based on the Hilbert transform" *Comput. Cardiol*. 27: 379–382.

Benitez, D. S., Gaydecki, P. A., Zaidi, A., Fitzpatrick, A. P., (2001). "The Use of the Hilbert Transform in ECG Signal Analysis." *Computers in Biology and Medicine* 31(5): 399–406.

Besrour, R., Lachiri, Z., Ellouze, N., (2008). "ECG beat classifier using support vector machine" in *Proceedings of the Third International Conference on Information and Communication Technologies: From Theory to Applications, ICTTA*, Damascus, Syria, (April 7-11).

Bin, M., Reaz, I, Wei, L. S., (2004). "Detection of the R wave peak QRS complex using neural network." *Info Comm Tech Theory Appl*. 3: 21–51.

Biosignal, 2017. https://en.wikipedia.org/wiki/Biosignal Access date: 10.07.2017.

Bishop, C. M., (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.

Bouaziz, F., Boutana, D., Benidir, M., (2014). "Multiresolution wavelet-based QRS complex detection algorithm suited to several abnormal morphologies." *IET Signal Process*. 8:774–82.

Borjesson, P. O., Pahlm, O., Sornmo, L., Nygards, M. E., (1982). "Adaptive QRS detection based on maximum a posteriori estimation." *IEEE Trans. Biomed. Eng*. 29(5):341–351.

Boutaa, M., (2006). *Analyze and quantification of the correlation of the rate of heartbeat with the various components of ECG signal*. Thesis of Magister in Electronics, University Aboubekr Belkaid, Biomedical Electronics Department.

Börjesson, P. O., Pahlm, O., Sörnmo, L., Nygards, M. E., (1982). "Adaptive QRS detection based on maximum a posteriori estimation." *IEEE Trans. Biomed. Eng*. 29: 341-351.

Bradie, B., (1996). “Wavelet packet-based compression of single lead ECG.” *IEEE Trans. Biomed. Eng*. 43:493–501.

Britanak, V. Rao, K. R., (2002). “A new fast algorithm for the unified forward and inverse MDCT/MDST computation.” *Signal Processing* 82: 433-459.

Brouse, C., Dumont, G. A., Herrmann, F. J., Ansermino, J. M., (2006). “A Wavelet approach to detecting electrocautery noise in the ECG.” *IEEE Eng. Med. Biol. Mag*.:76–82.

Burte, Ghongade, R., (2012). “Advances in QRS detection: Modified Wavelet energy gradient method.” *International Journal of Emerging Trends in Signal Processing* 1: 23-29.

Bystricky, W., Safer, A., (2002). “Modelling T-end in Holter ECG’s by 2-layer perceptrons,” in *Proc. Computers in Cardiology*. Los Alamitos, CA: IEEE Computer Society Press, 29:105–108.

Castells, F., Laguna, P., Sornmo, L., Bollmann, A., Roig, J. M., (2007). “Principal Component Analysis in ECG Signal Processing.” *Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing*, Article ID 74580. doi.org/10. 1155/2007/74580.

Cellar, B. Grace, Y. C. C., Phillips, C., (1997). “ECG analysis and processing using wavelets and other methods.” *Biomed. Eng. Appl. Basis Commun*. 9(2): 81-90.

Cetinkaya, S., Sevindir, K. H., Yazici, C., (2018a). “Comparison of kernel functions in arrhytmia classification.” *International Conference on Mathematics: Minisymposium on Approximation Theory & Minisymposium on Math Education*, Istanbul, Turkey, (July 3-6).

Cetinkaya, S., Demir, A., Sevindir, K. H., (2018b.) “Arrhythmia classification with support vector machines.” *International Marmara Science and Social Congress*, Kocaeli, Turkey, (November 23-25).

Cetinkaya, S., Sevindir, K. H., Yazici, C., (2018c.) “Decision Tree Based Arrhytmia Classification.” *Internatıonal Conference on Applıed Analysıs and Mathematıcal Modellıng (ICAAMM2018)*, Istanbul, Turkey, (June 20-24).

Chacko, A., Ari, S., (2012). “Denoising of ECG signals using Empirical Mode Decomposition based technique.” *IEEE-International Conference on Advances in Engineering, Science And Management*. Nagapattinam, Tamil Nadu, India. (March 30-31).

Chan, H. L., Chou, W. S., Chen, S. W., Fang, S. C., Liou, C. S., Hwang, Y. S., (2005). “Continuous and online analysis of heart rate variability.” *J. Med. Eng. Technol.*: 227–234.

Chan, A. D. C., Hamdy, M. M., Badre, A., Badee, V., (2008). “Wavelet distance measure for person identification using electrocardiograms.” *IEEE Trans. Instrum. Meas*. 57: 248–253.

Chang, S., Yu, B., Vetterli, M., (2000). "Adaptive wavelet thresholding for image denoising and compression." *IEEE Trans. Image Processing* 9: 1532-1546.

Chaturvedi, R., Yadav, Y., (2014). "Application of moving average filter in ECG denoising." *World Academics Journal of Engineering Sciences* 1: 1-4.

Chaudhuri, B. B., Bhattacharya, U., (2000). "Efficient training and improved performance of multilayer perceptron in pattern classification." *Neurocomputing* 34: 11-27.

Chawla, M., Verma, H. K., Kumar, V., (2006). "ECG modeling and QRS detection using principal component analysis." *Proceedings, IET Int conference, MEDSIP-06*, Glasgow, Scotland, UK. (July).

Chawla, M., Verma, H. K., Kumar, V., (2006). "Independent component analysis: a novel technique for removal of artifacts and base-line wander in ECG." *Proc., National Conference, CISCON-2006 MIT*, Manipal, 14–18, India.

Chawla, M., Verma, H. K., Kumar, V., (2006). "Modeling and feature extraction of ECG using independent component analysis." *Proceedings, IET Int. Conference APSCOM-2006*, Hongkong, (31 Oct–2 Nov).

Chazal, P., Celler, B., (1996). "Automatic measurement of the QRS onset and offset in individual ECG leads." presented at the *18th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc*., Amsterdam, The Netherlands. (October 31-November 3).

Chen, H. C. Chen, S. W., (2003). "A Moving Average based Filtering System with its Application to Real-time QRS Detection." *Computers in Cardiology* 30: 585−588.

Chen, S. W., Chen, H. C., Chan, H. L., (2006). "A real-time QRS detection method based on moving-averaging incorporating with wavelet denoising." *Computer methods and programs in biomedicine* 82(3): 187–195.

Cheng, W. T., Chan, K. L., (1998). "Classification of electrocardiogram using hidden Markov models." Engineering in Medicine and Biology Society, *Proceedings of the 20th Annual International Conference of the IEEE* 1: 143-146.

Chew, B. S., Chau, L., Yap, K. H., (2011). "Image based approach with k-mean clustering for the compression of human motion sequences." in *Circuits and Systems (ISCAS), IEEE International Symposium*. (May 15-18).

Chikamori, T., Takata, J., Furuno, T., Yabe, T., Matsumura, Y., Kitaoka, H., Doi, Y., (1995). "Usefulness of U-wave analysis in detecting significant narrowing limited to a single coronary artery." *Am J Cardiol.* 75: 508–511.

Christianini, N., Taylor, J. S., (2000). *An Introduction to Support Vector Machines and Other Kernel Based Learning Methods.* Cambridge University Press, Cambridge, MA.

Christov, I. I., (2004). "Real time electrocardiogram QRS detection using combined adaptive threshold." *Biomedical Engineering*: 3-28.

Chitupe, A. R., Joshi, S. A., (2013). "Data Classification Algorithm Using k-Nearest Neighbour Method Applied to ECG Data." *IOSR Journal of Computer Engineering* 14(4): 13-21.

Christov, I., Jekova, I., Bortolan, G., (2005). "Premature ventricular contraction classification by the Kth nearest neighbors rule." *Physiol. Meas*. 26: 123–130.

Coast, D. A., (1998). *Cardiac arrhythmia analysis using hidden Markov models*. Ph.D. thesis, Dep. Elect. Comput. Eng., Carnegie Mellon Univ., Pittsburgh, PA.

Choouakri, S. A., Reguig, F. B., Ahmed, A. T., (2011). "QRS complex detection based on multi wavelet packet decomposition." *Appl Math Comput.* 217: 9508–9525.

Chu-Song, C., Ja-Ling, W., Yi-Ping, H., (1999). "Theoretical aspects of vertically invariant gray-level morphological operators and their application on adaptive signal and image filtering." *IEEE Trans. on Signal Processing* 47(4):1049-1060.

Clavier, L., Boucher, J. M., Lepage, R., Blanc, J. J., Cornily, J. C., (2002). "Automatic P-wave analysis of patients prone to atrial fibrillation." *Med. Biol. Eng. Comp*. 40(1): 63–71.

Coast, D. A., Stern, R. M., Cano, G. G., Briller, S. A., (1987). "Cardiac arrhythmia analysis using hidden Markov models." in Proc. Ninth Annu. Cons., *IEEE Eng. Med. Biol. Soc*., 901-903.

Coast, A. D., Stern, R., (1990). M. "An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Model." *IEEE Transactions on Biomedical Engineering* 37: 826-836.

Comon, P., 81994). "Independent component analysis, A new concept?" *Signal Processing* 36: 287-314.

Crowe, J. A., Gibson, N. M., Woolfson, M. S., Somekh, M. G., (1992). "Wavelet transform as a potential tool for ECG analysis and compression." *J. Biomed. Eng*. 14(3): 268-272.

Cuiwei, L., Chongxun, Z., Changfeng, T., (1995). "Detection of ECG characteristic points using wavelet transforms." *IEEE Trans. Biomed. Eng.* 42: 21–28.

Daamouche, A., Hamami, L., Alajlan N., Melgani, F., (2012). "A wavelet optimization approach for ECG signal classification." *Biomedical Signal Processing and Control* 7: 342– 349.

Darko, F., Denis, S., Mario, Z., (2007). "Human movement detection based on acceleration measurements and k-NN classification." in *EUROCON, International Conference on "Computer as a Tool"*. Warsaw, Poland, (September 9-12).

Daskalov, I. K., Dotsinsky, I. A. and Christov, I. I., (1988). "Developments in ECG acquisition, preprocessing, parameter measurement and recording." *IEEE Eng. Med. Biol. Mag*. 17: 50–58.

Daskalov, I., Christov, I., (1999). "Electrocardiogram signal preprocessing for automatic detection of QRS boundaries." *Med. Eng. Phys*. 21: 37–44.

Daskalov I. K., Christov I. I., (1999). "Automatic detection of the electrocardiogram T-wave end." *Med. Biol. Eng. Comput*. 37(3): 348–353.

Delgado, M., F., Ameneiro, S. B., (1998). "MART: A multichannel ART-based neural network." *IEEE Trans. Neural Networks* 9: 139-150.

Dewangan, N. K., Shukla, S. P., (2015). "A Survey on ECG Signal Feature Extraction and Analysis Techniques." International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering 3(6).

Di-Virgilio, V., Francaiancia, C., Lino, S., Cerutti, S., (1997). "ECG fiducial points detection through wavelet transform." in *1995 IEEE Eng. Med. Biol. 17th Ann. Conf. 21st Canadian Med. Biol. Eng. Conf*., Montreal, Quebec, Canada, (September 20-23).

Dib, N., Benali, R., Slimane, Z. H., Bereksi-Reguig, F., (2011). "Delineation of the complex QRS and the T-end using wavelet transform and surface indicator." *The 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*, Come D'Or, Tipaza, Algeria, (May 09-11).

Dokur, Z., Olmez, T., Yazgan, E., Ersoy, O. K., (1997). "Detection of ECG waveforms by neural networks." *Med. Eng. Phys*. 19(8): 738-741.

Dokur, Z., Ölmez, T., (2001). "ECG beat classification by a novel hybrid neural network." *Computer Methods and Programs in Biomedicine* 66: 167–181.

Donoho, D. L., (1992). "De-noising by soft-thresholding." *IEE Trans. Inform. Theory* 41(3): 612-627.

Donoho, D. L., Johnstone, I. M., (1994). "Ideal spatial adaptation via wavelet shrinkage*." Biometrika* 81: 425-455.

Donoho, D. L., (1995). "De-noising by soft-tresholding." *IEEE Trans. Information Theory* 41(3): 613-627.

Donoso, F. I., Figueroa, R. L., Lecannelier, E. A., Pino, E. J., Rojas, A. J., (2013). "Clustering of atrial fibrillations based on surface ECG measurements." in *Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE/EMBC*, Osaka, Japan, (July 3-7).

Draper, B. A., Baek, K., Bartlett, M. S., Beveridge, J. R., (2003). "Recognizing Faces with PCA and ICA." *Computer Vision and Image Understanding* 91:115-137.

Du, K. L., Swamy, M. N. S., (2006). *Neural Networks in a Softcomputing Framework.* Springer, London.

Duda, R. O., Hart, P. E., Stork, D. G., (2011). *Pattern Classification*. 2nd ed., Wiley, New York.

Dutta, S., Chatterjee, A., Munshi, S., (2010). "Correlation technique and least square support vector machine combine for frequency domain-based ECG beat classification." *Med. Eng. Phys*. 32: 1161–1169.

Dreiseitl, S., Ohno-Machado, L., (2002). "Logistic regression and artificial neural network classification models: a methodology review." *J. Biomed. Inform.* 35: 352-359.

Ekg-Flash-Cards, (2017). https://quizlet.com/80139892/ekg-flash-cards/ Access date: 10.07.2017

Electrocardiogram, (2017) http://www.webmd.com/heart disease/electrocardiogram Access date: 10.03.2017

Electrocardiogram 2, (2017) http://house.wikia.com/wiki/Electrocardiogram Access date: 10.07.2017

Elliot, D. F., Rao, K. R., (1982). *Fast Transforms, Algorithms, Analysis and Applications*. New York, Academic.

Emtresource, (2017). http://www.emtresource.com/resources/ecg/12-lead-ecg-placement/ Access date: 10.03.2017.

Engelse, W. A. H., Zeelenberg, C., (1979). "A single scan algorithm for QRS detection and feature extraction." in *Proc. of Comput. in Cardiol*.:37–42.

Ercelebi, E., (2004). "Electrocardiogram signals de-noising using lifting-based discrete wavelet transform." *Comput Biol Med*. 34(6): 479–93.

Faezipour, M., Tiwari, T. M., Saeed, A., Nourani, M., Tamil, L. S., (2009). "Wavelet-based denoising and beat detection of ECG signal." in *IEEE/NIH Life Science Systems and Applications Workshop* (LiSSA 2009), NIH Campus, Bethesda, Maryland, USA, (April 9-10).

Fancott, T., Wong, D. H., (1980). "A minicomputer system for direct high-speed analysis of cardiac arrhythmia." in 24 h ambulatory ECG tape recordings, *IEEE Trans. Biomed. Eng*. 27:685-693.

Fausett, L., (1994). *Fundamentals of Neural Networks Architectures, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ.

Fei, Z., Yong, L., (2009). "QRS detection based on multiscale mathematical morphology for wearable ECG devices in body area networks." *IEEE Trans. Biomed. Circuits Syst*. 3: 220–228.

Fraden, J., Neumann, M. R., (1980). "QRS wave detection" *Med. Biol. Eng. Comput.* 18: 125-132.

Friesen, G. M., Jannett, T. J., Jadallah, M. A., Yates, S. L., Quint, S. R., Nagle, H. T., (1990). "A comparision of the noise sensitivity of nine QRS detection algorithms." *IEEE Trans. Biomed. Eng*. 37(1): 85–98.

Garg, G., Singh, V., Gupta, J. R. P., Mittal, A. P., (2010). "Optimal algorithm for ECG denoising using discrete wavelet transforms." *IEEE Int Conf Comput Intell Comput Res*.:1–4.

Garrett, D., Peterson, D. A., Anderson, C. W., Thaut, M. H., (2003). "Comparison of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification." *IEEE Trans. on Neural Systems And Rehabilitation Engg.* 11(2):141-144.

Gamero, L. G., Vila, J., Palacios, F., (2002). "Wavelet transform analysis of heart rate variability during myocardial ischaemia." *Med Biol Eng Computer* 40: 72-78.

Geo, D., Srinivasan, N., Krishnan, S. M., (2002). "Cardiac arrhythmia classification using autoregressive modeling." *BioMedical Engineering* On Line 1:5.

Ghaffari, A., Golbyani, H., Ghaseni, M., (2008). "A new mathematical based QRS detector using continuous wavelet transform." *Comput Electric Eng*. 34: 81–91.

Ghoggali, N., Melgani, F., Bazi, Y., (2009). "A multiobjective genetic SVM approach for classification problems with limited training samples." *IEEE Transactions on Geoscience and Remote Sensing* 47: 1707–1718.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C. H., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., Stanley, H. E., (2000). "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals." *Circulation* 101(23): 215-220.

Golpayegani, G. N., Jafari, A. H., (2009). "A novel approach in ECG beat recognition using adaptive neural fuzzy filter." *J. Biomedical Science and Engineering* 2: 80-85.

Gothwal, H., Kedawat, S., Kumar, R., (2011). "Cardiac arrhythmias detection in an ECG beat signal using fast fourier transform and artificial neural network." *Biomed. Sci. Eng*. 4: 289–296.

Gu, F., Greensmith, J., Oates, R., Aickelin, U., (2009). "PCA 4 DCA: The Application of Principal Component Analysis To The Dendritic Cell Algorithm." In *9th Annual Workshop on Computational Intell. (UKCI),* London, United Kingdom, (September 7-9).

Gunn, S., (1998). *Support Vector Machines for Classification and Regression*. Technical Report, University of Southampton, May.

Gustafson, D., (1977). *Automated VCG interpretation studies using signal analysis techniques*. R-1044 Charles Stark Draper Lab.

Gyaw, T. A., Ray, S. R., (1994). "The wavelet transforms as a tool for recognition of biosignals." *Biomed. Sci. Instrum*. 30: 63-68.

Hagan, M. T., Menhaj, M. B., (1994). "Training feedforward networks with the Marquardt algorithm." *IEEE Trans. Neural Netw*. 5(6): 989-993.

Hagan, M. T., Demuth, H. B., Beale, M. H., (1996). *Neural Network Design.* PWS Publishing, Boston, first edition.

Hahoura, M., Hassani, M., Hubin, M., (1997). "DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis." *Comput. Methods Prog. Biomed* 52: 35–44.

Ham, F. M., Han, S., (1996). "Classification of cardiac arrhythmias using fuzzy ARTMAP." *IEEE Trans. Biomed. Eng*. 43: 425-430.

Hamilton, P. S., Tompkins, W. J., (1986). "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmiac database." *IEEE Trans. Biomed. Eng.* 33: 1157-1165.

Hao, Z., Qing, Z. L., (2005). "ECG analysis based on PCA and support vector machines." *IEEE Trans*.: 743–747.

Hassanpour, H., Boashash, B., (2011). "A Time-Frequency Approach For EEG Spike Detection." in *Iranica Journal of Energy & Environment* 2 (4): 2079-2115.

Haykin, S., (1994). *Neural Networks: A Comprehensive Foundation*. Macmillan, New York.

Haykin, S., (1999). *Neural Networks: A Comprehensive Foundation* second ed., Prentice Hall.

Hekim, M., (2012). "ANN-based classification of EEG signals using the average power based on rectangle approximation window." *Przegląd Elektrotechniczny* [*Electrical Review*], 88(8):210-215.

Ho, C. Y. F., Ling, B. W. K., Wong, T. P. L., Chan, A. Y. P., Tam, P. K. S., (2003). "Fuzzy multiwavelet denoising on ECG signal." *Electro Lett.* 39(16): 1163–4.

Holsinger, W. P., Kempner, K. M., Miller, M. H., (1970). "A QRS preprocessor based on digital differentiation." *IEEE Trans. Biomed. Eng*. 18:121-217.

Hsu, C. W., Lin, C. J., (2002). "A comparison of methods formulticlass support vector machines." *IEEE Trans. Neural Netw*. 13(2): 415-425.

Hu, Y. H., Tompkins, W. J., Urrusti, J. L., Afonso, V. X., (1993). "Applications of artificial neural networks for ECG signal detection and classification." *J. Electrocardiology* 26: 66-73.

Huang, N. E., Shen, Z., Long, S. R., Wu, M., Shih, H. H., Zheng, Q., Yen, N. Tung, E. E., Liu, H. H., (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis." *Proc. Roy. Soc. Lond. A, Math. Phys. SCi.* 454: 903-995.

Hughes, N. P., Tarassenko, L., Roberts, S. J., (2003). "Morkov Models for Automated ECG Interval Analysis." *IEEE Transactions on Biomedical Eng.* 50(3): 289-294.

Husain, H., Fatt, L. L., (2007). "Effecient ECG Signal Classification Using Sparsely Connected Radial Basis Function Neural Network." *6th WSEAS International Conference on Circuıts, Systems, Electronics, Control & Signal Processing*, Cairo, Egypt, (Dec 29-31).

Hyvärinen, A., (1999)." Fast and robust fixed-point algorithms for independent component analysis." *IEEE Transactions on Neural Networks* 10(3): 626–634.

Ieong, C. I., Mak, P. I., Lam, C. P., Dong, C., Vai, M. I., Mak, P. U., Pun, S. H., Wan, F., Martins, R. P., (2012). "A QRS detection processor using quadratic spline wavelet transform for wireless ECG acquisition in 0.35-CMOS." *IEEE Trans. Biomed. Circ. Syst.* 6 (6): 586–595.

Illanes-Manriquez, A., Zhang, Q., (2008). "An algorithm for robust detection of QRS onset and offset in ECG signals." *Computers in Cardiology* 35: 857–60.

Inoue, H., Miyazaki, A., (1998). "A noise reduction method for ECG signals using the dyadic wavelet transform." In *(1997) Int. Tech. Conf. Circuits/ Systems, Computers and Communications (ITC-CSCC'97)*, Okinawa, Japan, (July 14-16).

Iscan, Z., Yuksel, A., Dokur, Z., Korurek, M., Olmez, T., (2009). "Medical image segmentation with transform and moment-based features and incremental supervised neural network." *Digital Signal Processing* 19: 890–901.

Istepanian, R. S. H., Hadjileontiadis, L. J., Panas, S. M., (2001). "ECG data compression using wavelets and higher order statistics methods." *IEEE Trans. Inf. Technol. Biomed* 5: 108–115.

Ivanov, P. C., Rosenblum, M. G., Peng, C. K., Mietus, J., Havlin, S., Stanley, H. E., Goldberger, A. L., (1996). "Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis." *Nature* 38: 3173-3184.

İşler Y., Narin, A., (2012). "WEKA Yazılımında k-Ortalama Algoritması Kullanılarak Konjestif Kalp Yetmezliği Hastalarının Teşhisi." *SDU Journal of T echnical Sciences* 4(2): 21-29.

Jalaleddme, S. M. S., Hutchens, C. G., Strattan, R. D., Coberly, W. A., (1990). "ECG Data Compression Techniques - A Unified Approach." *IEEE Trans. Biomed. Engin* 37 (4):329-343.

Jansen, M., (2001). *Noise reduction by Wavelet Thresholding*. 161, Springer Verlag.

Jelinek, F., (1976). "Continuous speech recognition by statistical methods." *Proc. IEEE* 64: 532-556.

Jiang, W., Kong, G. S., (2007). "Block-based neural networks for personalized ecg signal classification." *IEEE Trans. Neural Networks* 18(6): 1750–1761.

Jing-tian, T., Qing, Z., Yan, T., Bin, L., Xiao-kai, Z., (2007). "Hilbert-Huang Transform for ECG De-Noising." in: *Proceedings of 1st International Conference on Bioinformatics and Biomedical Engineering*: 664–667.

John, M. M., Mithilesh, K. D., Anil, V. Y. D. B., Girish, N. and Cesar, A., (2006). "Value of the 12-lead ECG in wide QRS tachycardia." *Cardiology Clinics* 24: 439-451.

Joshi, A. K., Tomar, A., Tomar, M., (2014). "A Review Paper on Analysis of Electrocardiograph (ECG) Signal for the Detection of Arrhythmia Abnormalities." International Journal of Advanced Research in Electrical, *Electronics and Instrumentation Engineering* 3(10): 12466-12475.

Kabir, M. A., Shahnaz, C., (2012). "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains." *Biomedical Signal Processing and Control* 7: 481– 489.

Kabir, M. A., Shahnaz, C., (2012). "Comparison of Ecg Signal Denoising Algorithms in EMD and Wavelet Domains." *IJRRAS* 11(3):499-516.

Kadambe, S., Murray, R., Boudreaux-Bartels, G. F., (1992). "The dyadic wavelet transform-based QRS detector." *Proc. 26th Asil. Conf. Signal Systems Comput*., Pacific Grove, CA.: 130–134,

Kadambe, S., Murray, R., Boudreaux-Bartels, G. F., (1999). "Wavelet transform-based QRS complex detector." *IEEE Trans. Biomed. Eng*. 46: 838-848.

Kalpana, V., Hamde, S. T., Waghmare, L. M., (2013). "ECG feature extraction using principal component analysis for studying the effect of diabetes." *Journal of Medical Engineering & Technology* 37(2): 116-126.

Kampouraki, A., Manis, G., Nikou, C., (2009). "Heartbeat time series classification with support vector machines." *IEEE Trans Inf Technol Biomed*. 13: 512-518.

Kania, M., Fereniec, M., Maniewski, R., (2007). "Wavelet denoising for multi-lead high resolution ECG signals." *Meas. Sci. Rev*. 7(2): 30–33.

Kar, A., Das, L., (2011). "A Technical Review on Statistical Feature Extraction of ECG signal." *IJCA Special Issue on "2nd National Conference- Computing, Communication and Sensor Network" CCSN.*

Karimifard, S., Ahmadian, A., Khoshnevisan, M., (2006). "Morphological heart arrhythmia detection using hermitian basis functions and kNN classifier." in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology*: 4489–4492.

Karpagachelvi, S., Arthanari, M., Sivakumar, M., (2010). "ECG Feature Extraction Techniques - A Survey Approach." (IJCSIS) *International Journal of Computer Science and Information Security* 8(1):76-80.

Karpagachelvi, S., Arthanari, M., Sivakumar, M., (2012). "Classification of electrocardiogram signals with support vector machines and extreme learning machine." *Neural Comput & Applic*. 21: 1331–1339.

Kemmelings, J. G. C., Linnenbank, A. C., Muilwijk, S. L. C., Sippens- Groenewegen, A., Peper, A., Grimbergen, C. A., (1994). "Automatic QRS onset and offset detection for body surface QRS integral mapping of ventricular tachycardia." *IEEE Trans. Biomed. Eng.* 41: 830–836.

Khunti, K., (2014). "Accurate interpretation of the 12-lead ECG electrode placement: A systematic review." *Health Education Journal* 73(5): 610-623.

Kicmerova, D., (2009). *Methods for Detection and Classification in Ecg Analysis*. Phd Thesis, Brno University of Technology, Brno, 130 pages.

Kim, D., Oh, H. S., (2008). "EMD: Empirical Mode Decomposition and Hilbert Spectral Analysis." *The R Journal* 1(1): 40-46.

Keselbrener, L., Keselbrener, M., Akselrod, S., (1997). "Nonlinear high pass filter for R-wave detection in ECG signal." *Med. Eng. Phys.* 19(5): 481-484.

Khadra, L., Al-Fahoum, A. S., Al-Nashash, H., (1997). "Detection of life-threatening cardiac arrhythmias using the wavelet transformation." *Med. Biol. Eng. Comput.* 35(6): 626-632.

Khanacademy (2017). https://www.khanacademy.org/test-prep/mcat/physical-sciences-practice/Access date: 10.07.2017.

Kiranyaz, S., Ince, T., Pulkkinen, J., Gabbouj, M., (2011). "Personalized long-term ECG classification: a systematic approach." *Expert Syst. Appl.* 38(4): 3220–3226.

Kohler, B., Hennig, C., Orglmeister, R., (2002). "The principles of software QRS detection." *IEEE Eng. Med. Biol*. 21(1): 42–57.

Kohler, B. U., Hennig, C., Orglmeister, R., (2007). "The principles of software QRS detection." *IEEE Eng. in Med. and Biology Magazine:* 42–57.

Kokturk, G., (1998). "A real-time simulated QRS detection system constructed using wavelet filtering technique." in *Proc. IEEE-SP Int. Symp. Time-Frequency and Time-Scale Analysis*, PA, Pittsburgh: 281-284.

Kumari, V., Kumar, R., (2013). "Cardiac arrhythmia prediction using improved multilayer perceptron neural network." *International Journal of Electronics, Communication & Instrumentation Engineering Research and Development* (IJECIERD) 3(4): 73- 80.

Kutlu, Y., Damla, K., (2011). "Amulti-stage automatic arrhythmia recognition and classification system." *Comput. Biol. Med*. 41: 37–45.

Krose, B., Smagt, P., (1996). *An Introduction to Neural Networks*. First published in 1997 by UCL Press.

Lakshmi, M. R., Prasad, T. V., Prakash, C. V., (2014). "Survey on EEG Signal Processing Methods." *International Journal of Advanced Research in Computer Science and Software Engineering* 4(1): 84-91.

Lannoy, G., Frenay, B., Verleysen, M., Delbeke, J., (2008). "Supervised ECG delineation using the wavelet transform and Hidden Markov models." In: *The Proceedings of IFMBE* 22: 22–25.

Lagerholm, M., Peterson, C., Braccini, G., Edenbrandt, L., Sornmo, L., (2000). "Clustering ECG complexes using hermite functions and self-organizing maps." *IEEE Transactions on Biomedical Engineering* 47: 838-848.

Laguna, P., Thakor, N. V., Caminal, P. R., Yoon, H. R., Luna, A. B., Marti, V., Guindo, J., (1990). "New algorithm for QT interval analysis in 24-hour Holter ECG: Performance and applications." *Medical and Biological Engineering and Computing* 28(1): 67–73.

Laguna, P., Jane, R., Olmos, S., Thakor, N. V., Rix, H., and Caminal, P., (1996). "Adaptive estimation of QRS complex wave features of ECG signal by the Hermite Model." *Medical and Biological Engineering and Computing*: 58–68.

Laguna, P., Jané, R., Caminal, P., (1994). "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database." *Comput. Biomed. Res*. 27(1): 45–60.

Lee, K. F., (1988). *Large-vocabulary speaker-independent continuous speech recognition: the SPHINX system*. Ph.D. thesis. Dep. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA.

Leski, J., Tkacz, E. (1992). "A new parallel concept for QRS complex detector." in *Proc. 14th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society* 2: 555-556.

Li, C., Zheng, C., Tai, C., (1995). "Detection of ECG characteristic points using wavelet transforms." *IEEE Trans. Biomed. Eng.* 42: 21–28.

Ligtenberg, A., Kunt, M., (1983). "A robust-digital QRS-detection algorithm for arrhythmia monitoring." *Comput. Biomed. Res*. 16: 273-286.

Lin, S., (1991). "Comparison of kohonen feature map against Kmean clustering algorithm with application to reversible image compression." in *Circuits and Systems Conference Proceedings, International Conference*, China, 2: 808-811.

Lin, C., Mailhes, C., Tourneret, J. Y., (2010). "P- and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler." *IEEE Transactions on Biomedical Engineering* 57(12): 9–2840.

Lin, M. K., (2012). *Evaluating the Acceptance of Mobile Technology in Healthcare: Development of a Prototype Mobile ECG Decision Support System for Monitoring Cardiac Patients Remotely*. Ph.D. thesis, Dep. Accounting, Economics and Finance, Univ. of Southern Queensland, Toowoomba, Australia.

Lippmann, R. P., (1987). "An introduction to computing with neural nets." *IEEE ASSP Magazine*: 4–22.

Mackay, D., (1992). "A practical Bayesian framework for backpropagation networks." *Neural Computation* 4: 448-472.

Macy, R. B., Pandya, A. S., (1996). *Pattern Recognition with Neural Networks in C++*. Florida: CRC Press LLC.

Madeiro, J. P., Cortez, P. C., Oliveira, F. I., Siqueira, R. S., (2007). "A new approach to QRS segmentation based on wavelet bases and adaptive threshold technique." *Medical Engineering & Physics* 29(1): 26– 37.

Madeiro, J. P., Cortez, P. C., Marques, J. A. L., Seisdedos, C. R., Sobrinho. C. R., (2012). "An innovative approach of QRS segmentation based on first-derivative, Hilbert and wavelet transforms." *Med Eng Phys*. 4(9): 1236–1246.

Maglaveras, N., Stamkopoulos, T., Diamantaras, K., Pappas, C., Strintzis, M., (1998). "ECG pattern recognition and classification using non-linear transformations and neural networks: A review." *Int. J. Med. Informatics* 1(3): 191-208.

Maglaveras, N., Stamkopoulos, T., Pappas, C., Strintzis, M., (1998). "ECG processing techniques based on neural networks and bidirectional associative memories." *J. Med. Eng. Technol*. 22(3): 106-111.

Mahalingam, N., Kumar, D., (1997). "Neural networks for signal processing applications: ECG classification." *Australas. Phys. Eng. Sci. Med.* 20(3): 147-151.

Mai, V., Khalil I., Meli, C., (2011). "ECG Biometric Using Multilayer Perceptron and Radial Basis Function Neural Networks." *33rd Annual International Conference of the IEEE EMBS Boston*, Massachusetts USA, (August 30 - September 3).

Mallat, S., Hwang, W. L., (1992). "Singularity detection and processing with wavelets." *IEEE Trans. Inform. Theory* 38:617-643.

Maleki, M., Eroglu, K., Aydemir, O., Manshoori, N., Kayikcioglu, T., (2013). "A new method for selection optimum k value in k-NN classification algorithm." in Signal Processing and Communications Applications Conference (SIU), Haspolat, Turkey, (April 24-26).

Manikandan, M. S., Soman, K. P., (2012). "A novel method for detecting R-peaks in electrocardiogram (ECG) signal." *Biomed. Signal Process*. Control 7: 118–128.

Mao, K. Z., Tan, K. C., Ser, W., (2000). "Probabilistic Neural-Network Structure Determination for Pattern Classification." *IEEE Transactions on Neural Networks* 11(4): 1009-1016.

Maragos, P., Schafer, R., (1987). "Morphological filters-part i: Their set-theoretic analysis and relations to linear shift invariant filters." *IEEE Trans. on Acoustic, Speech and Signal Processing, ASSP* 35(8):1153-1169.

Maragos, P., Schafer, R., (1990). "Morphological systems for multidimensional signal processing." *Proceedings of the IEEE* 78(4):690-710.

Mark, R. G., Schluter, P. S., Moody, G. B., Devlin, P. H., Chernoff, D., (1982). "An annotated ECG database for evaluating arrhythmia detectors." *IEEE Transactions on Biomedical Engineering* 29(8).

Maragos, P., (2000). *Morphological signal and image processing*. CRC Press.

Martinez, J. P., Almeida, R., Olmos, S., (2004). "A Wavelet-Based ECG Delineator: Evaluation on Standard Databases." *IEEE Transactions on Biomedical Engineering*, 51(4):570-581.

Martinmaki, K., Rusko, H., Saalasti, S., Kettunen, J., (2006). "Ability of short-time Fourier transform method to detect transient changes in vagal effects on hearts: a pharmacological blocking study." *Am J Physiol Heart Circ Physiol*. 290: 2582-2589.

Martis, R. J., Chakraborty, R. J., (2011). "Arrhythmia Disease Diagnosis using Neural Network, SVM, and Genetic Algorithm-Optimized k-Means Clustering." *Journal of Mechanics in Medicine and Biology* 11(4): 897–915.

Martis, R. J., Acharya, U. R., Mandana, K. M., Ray, A. K., Chakraborty, C., (2012). "Application of principal component analysis to ECG signals for automated diagnosis of cardiac health." *Expert Syst. Appl.* 39: 11792–11800.

Martis, R. J., Acharya, U. R., Min, L. C., (2013). "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform." *Biomedical Signal Processing and Control* 8: 437-448.

Martis, R. J., Acharya, U. R., Lim, C. M., Suri, J. S., (2013). "Characterization of ECG beats from cardiac arrhythmia using discrete cosine transform in PCA framework." *Knowledge-Based Systems* 45: 76–82.

May, C., Hubing, N., Hahn, A. W., (1997). "Wavelet transforms for electrocardiogram processing." *Biomed. Sci. Instrum*. 33: 1-6.

Mbeledogu, N. N., Odoh, M., Umeh, M. N., (2012). "Stock Feature Extraction Using Principal Component Analysis." (2012) *Int. Conf. on Computer Tech. and Sc., IPCSIT* 47.

McSharry, P. E., Clifford, G. D., Tarassenko, L., Smith, L., (2003). "A dynamical model for generating synthetic electrocardiogram signals." *IEEE Trans. Biomed. Eng*. 50(3) 289 –294.

Meij, S., Klootwijk, P., Arends, J., Roelandt, J., (1994). "An algorithm for automatic beat-to-beat measurement of the QT-interval." in *Computers in Cardiology: IEEE Computer Society Press*, Bethesda, MD, USA, (September 25-28).

Mehta, S., Lingayat, N., (2007). "Comparative study of QRS detection in single lead and 12-lead ECG based on entropy and combined entropy criteria using Support Vector Machine." *J. Theoretical Appl. Inform. Technol.* 3(2): 8–18.

Mehta, S., Lingayat, N., (2008). "Svm-based algorithm for recognition of QRS complexes in electrocardiogram." *IRBM* 29(5): 310–317.

Mehta, S. S., Shete, D. A., (2010). "K-means algorithm for the detection and delineation of QRS-complexes in Electrocardiogram." *IRBM* 31: 48-54.

Melgani, F., Bruzzone, L., (2004). "Classification of hyperspectral remote sensing images with support vector machine." *IEEE Trans. Geosci. Remote Sens*. 42(8): 1778-1790.

Melgani, F., Bazi, Y., (2008). "Classification of electrocardiogram signals with support vector machines and particle swarm optimization." *IEEE Transactions on Information Technology in Biomedicine* 12(5): 667–677.

Merino, M., Gómez, I. M., Molina, A. J., (2015). "Envelopment filter and k-means for the detection of QRS waveforms in electrocardiogram." *Med. Eng. Phys.* 37(6): 605–609.

Merone, M., Soda, P., Sansone, M., Sansone C., (2017). "ECG databases for biometric systems: A systematic review." *Expert Systems with Applications* 67: 189-202.

Merzagora, A. C., Bunce, S., Izzetoglu, M., Onaral, B., (2006). "Wavelet analysis for EEG feature extraction in deception detection." *Proceedings of the 28th IEEE EMBS Annual Int. Conf.* New York City, USA, (Aug 30-Sep 3).

Meyer, C., Gavela, J. F., Harris, M., (2006). "Combining algorithms in automatic detection QRS complexes in ECG signals." *IEEE Transactions on Information Technology in Biomedicine* 10(3):468-475.

Miller, A. S., Blott, B. H., Hames, T. K., (1992). "Review of neural network applications in medical imaging and signal processing." *Med. Biol. Eng. Comput.* 30: 449-464.

Ming, L., Zong, Y. B., Fang, T. X., (2004). "A new approach to determine the parameters of dissimilarity function for the evidence-theoretic k-NN classification rule." in *Signal Processing, Proceedings of ICSP '04, 7th International Conference*, Beijing, China, (Aug 31-Sep 4).

Minhas, F. A. A., Arif, M. 2008. "Robust electrocardiogram (ECG) beat classification using discrete wavelet transform." *Physiological Measurement* 29: 555–570.

Moody, G. B., Mark, R. G., (1990). "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it." *Computers in Cardiology* 17:185-188.

Moody, G. B., Mark, R. G., (2001). "The impact of the MIT-BIH Arrhythmia Database." *IEEE Eng in Med and Biol*. 20(3):45-50

Mohmoodabadi, S. Z., Ahmadian, A., Abolhasani, M. D., (2005). "ECG feature extraction using daubechies wavelets." *Proc. of the fifth IASTED International Conference*, Benidorm, Spain, (September 7-9).

Morizet-Mahoudeaux, P., Moreau, C., Moreau, D., Quarante, J. J., (1981). "Simple microprocessor-based system for on-line ECG arrhythmia analysis." *Med. Biol. Eng. Comput*. 19(4): 497-501.

Moses, D., Deisy C., (2015). "A survey of data mining algorithms used in cardiovascular disease diagnosis from multi-lead ECG data." *Kuwait J. Sci*. 42(2): 206-235.

Murthy, I., Prasad, G. D., (1992). "Analysis of ECG from pole-zero models." *IEEE Trans. Biomed. Eng.* 39: 741–751.

Nadal, J., Bossan, M., (1993). "Classification of cardiac arrhythmia based on principal components analysis and feed forward neural networks." in *IEEE Proceedings on Computers in Cardiology*: 341–344.

Nagendra, H., Mukherjee, S., Kumar, V., (2011). Application of wavelet techniques in ECG signal processing: an overview, *Int J Eng Sci Technol*., 3(10):7432–7443.

Naik, G. R., Kumar, D. K., (2011). "An overview of independent component analysis and its applications." *Informatica* 35(1): 63–81.

Narsimha, B., Suresh, E., Punnamchandar, K., Reddy, M. S., (2011). "Denoising and QRS detection of ECG signals using Empirical Mode Decomposition." in: *Proceedings of International Conference on Communications and Signal Processing*, 439–442.

Neural Networks, 2017. http://www.forexadvisor.net/neuralnetwork.htm Access Date: 10.07.2017

Nikolajevic, V., Fettweis, G., (2003). "A new Recursive Algorithm for the Unified Forward and Inverse MIDCT/MIDST." *Journal of VLSI Signal Processing* 9(34): 203-208.

Nov´ak, D., (2003). *Electrocardiogram Signal Processing using Hidden Markov Models*. Ph.D. Thesis, Czech Technical University in Prague Faculty of Electrical Engineering Department of Cybernetics Technick´a

Nugent, C. D., Lopez, J. A., Smith, A. E., Black, N. D., (2002). "Prediction models in the design of neural network-based ECG classifiers: A neural network and genetic programming approach." *BMC Medical Informatics and Decision Making* 2:1-6.

Nygards, M. E., Hulting, J., (1979). "An automated system for ECG monitoring." *Comput. Biomed. Res*. 12: 181-202.

Nygards, M., Sornmo, L., (1981). "A QRS delineation algorithm with low sensitivity to noise and morphology changes." *Comput. Cardiol*. : 347-350.

Nygards, M., Sörnmo, L., (1983). "Delineation of the QRS complex using the envelope of the ECG." *Med. Biol. Eng. Comput*. 21: 538–547.

Ofner, P., Muller-Putz, G. R., Neuper, C., Brunner, C., (2011). "Comparison of Feature Extraction Methods for Brain-Computer Interfaces." *Proc. of the 5th Int. BCI Conf. (2011)*, Graz, Austria, (September 22-24).

Okada, M., (1979). "A digital filter for the QRS complex detection." *IEEE Trans. Biomed. Eng*. 26: 700-703.

Oliveira, F. I., Cortez, P. C., (2004). "A QRS detection based on hilbert transform and wavelet bases, in: Machine Learning for Signal Processing." Proceedings of the (2004) 14th IEEE Signal Processing Society Workshop, *IEEE*: 481–489.

Oliveira, L., de, Andreao, R., Sarcinelli, M., (2011). "Premature Ventricular beat classification using a dynamic Bayesian network." in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc*. :4984–4987.

Osowski, S., Hoai, L. T., Markiewicz, T., (2004). "Support vector machine-based expert system for reliable heartbeat recognition." *IEEE Transactions on Biomedical Engineering* 51(4): 582-589.

Özbay, Y., Ceylan, R., Karlik, B., (2006). "A fuzzy clustering neural network architecture for classification of ECG arrhythmias." *Comput. Biol. Med.* 36: 376–388.

Özbay, Y., Ceylan, R., Karlik, B., (2011). "Integration of type-2 fuzzy clustering and wavelet transform in a neural network-based ECG classifier." *Expert Syst*. Appl. 38: 1004-1010.

Padmavathi, K., Sri Ramakrishna, K., (2015). "Classification of ECG signal during Atrial Fibrillation using Autoregressive modeling." *Procedia Computer Science* 46: 53 – 59.

Pal, S., Mitra, M., (2010). "QRS complex detection using empirical mode decomposition-based windowing technique." in: *International Conference on Signal Processing and Communications*: 1–5.

Pal, S., Mitra, M., (2012). "Empirical mode decomposition-based ECG enhancement and QRS detection." *Computers in Biology and Medicine* 42: 83–92.

Pan, J., Tompkins, W. J. 1985. "A real-time QRS detection algorithm." *IEEE Trans. Biomed. Eng*. 32: 230-236.

Pan, N., I., Un, M. P., Hang, P. S., (2007). "Accurate removal of baseline wander in ECG using empirical mode decomposition." in: *Proceedings of NFSI—Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference of Functional Biomedical Imaging* :177–180, (October 12-14).

Patil, M., et al. (2012). "Evaluation of QRS complex based on DWT coefficients analysis using daubechies wavelets for detection of myocardial ischemia." *Journal of Mechanics in Medicine & Biology* 210: 273-290.

Patil, P. B., Chavan, M. S., (2012). "A wavelet-based method for denoising of biomedical signal, in: Proceedings of International Conference on Pattern Recognition." *Informatics and Medical Engineering*: 278–283.

Pasolli, E., Melgani, F., Donelli, M., (2009). "Automatic analysis of GPR images: a pattern recognition approach." *IEEE Transactions on Geoscience and Remote Sensing* 47: 2206–2217. physionet (2017). https://www.physionet.org/physiobank/ database/mitdb/ Access Date: 10.03.2017 physiotools (2017).

Poggio, T., Girosi, F., (1990). "Networks for approximation and learning." *Proc. IEEE* 78(9): 1481–1497.

Poli, R., Cagnoni, S., Valli, G., (1995). "Genetic design of optimum linear and nonlinear QRS detectors." *IEEE Trans. Biomed. Eng*. 42(11): 1137–1141.

Poornachandra, S., (2008). "Wavelet-based denoising using sub band dependent threshold for ECG signals." *Digital Signal Processing*: 49-55.

Popescu, M., Criestea, P., Bezerianas (1998). "High Resolution ECG Filtering using Adaptive Bayesian Wavelet Shrinkage." In *Proceeding Conference of Computers in Cardiology*: 401-404.

Powell, M. J. D., (1987). *Radial basis functions for multivariable interpolation, in Algorithms for Approximation*. ed. by J. C. Mason, M. G. Cox Clarendon Press, Oxford.

Prasad, G. K., Sahambi, J. S., (2003). "Classification of ECG arrhythmias using multiresolution analysis and neural networks." in *Conference on convergent technologies for Asia-Pacific region (TENCON 2003)* 1: 227–231 pp.

Prasad, S. T., Varadarajan, S., (2014). "Analysis of ECG Using Filter Bank Approach." *Int. Journal of Engineering Research and Applications* 4(1): 186-190.

Principal Component Analysis, (2017). http://en.wikipedia.org/wiki/Principal_component _analysis Access Date: 10.07.2017.

Prochazka, A., Mudrova, M., Vysata, O., Hava, R., Araujo, C. P. S., (2010). "Multi-Channel EEG Signal Segmentation AND Feature Extraction." *Intelligent Engineering Systems (INES)*. Las Palmas, Spain, (May 5-7).

Provaznik, I., Kozumplik, J., Bardonová, J., Bardonová, Z. 2000. "Wavelet transform in ECG signal processing." in: *EuroConference, BIOSIGNAL (2000)*, Brono, Czech Republic.

Qin, Z. Chen, J., Liu, Y., Lu, J., (2005). "Evolving RBF Neural Networks for Pattern Classification." *Lecture Notes in Computer Science* 3801: 957–964.

Rabbani, H., Mahjoob, M. P., Farahabadi, E., Farahabadi, A., (2011). "R Peak Detection in Electrocardiogram Signal Based on an Optimal Combination of Wavelet

Transform, Hilbert Transform, and Adaptive Thresholding." *Journal of Medical Signals & Sensors* 1(2): 91-98.

Rabiner L. R., Juang, B. H., (1986). "An introduction to hidden Markov models." *IEEE Acoust., Speech, Signal Process. Mag*.:4-16.

Raghav, S., Mishra, A. K., (2008). "Fractal feature based ECG arrhythmia classification." in *Proceedings of IEEE TENCON*. Hyderabad, India, (November 19-21).

Rao, K. D., (1997). "Dwt based detection of R-peaks and data compression of ECG signals." *IETE J. Res*. 43(5): 345-349.

Reaz, M. B. I., Hussain, M. S., Mohd-Yasin, F., (2006). "Techniques of EMG signal analysis: detection, processing, classification and Applications." *Biol. Proced. Online* 8(1): 11-35.

Reddy, G. U., Muralidhar, M., Varadarajan, S., (2009). "ECG de-noising using improved thresholding based on Wavelet transform." *Int. J Comput Sci Network Security* 9(9):221–225.

Reilly, D. L., Cooper, L. N., Elboum, C., (1982). "Neural model for category learning." *Biol. Cybernet* 45: 35-41.

Rodrigo, V., Dorizzi, B., Boudy, J., (2006). "ECG signal Analysis through Hidden Markov Models." *IEEE Transactions on Biomedical Eng.* 53(8):1541-1549.

Romero Legarreta, I., Addison, P. S., Reed, M. J., Grubb, N. R., Clegg, G. R., Robertson, C. E., Watson, J. N., (2005). "Continuous wavelet transform modulus maxima analysis of the electrocardiogram: beat-to beat characterization and beat-to-beat measurement." *Int. J. Wavelets Multiresolution Inf. Process* 3: 19–42.

Ruchita, G., Sharma, A. K., (2010). "Detection of QRS complexes of ECG recording based on wavelet transform using Matlab." *Int. J. Eng. Sci.* 2(7): 3038–3044.

Rute, A., et al. (2014). "Fetal QRS detection and heart rate estimation: A wavelet-based approach." *Physiological Measurement* 35: 1723-1735.

Rutkowski, L., (2004). "Adaptive Probabilistic Neural Networks for Pattern Classification in Time-Varying Environment." *IEEE Transactions on Neural Networks*, 15(4):811-827.

Sabah, M. A., (2006). "Ecg Signal Compression Using Combined Modified Discrete-Cosine and Discrete-Wavelet Transforms." Journal of Engineering Sciences, Assiut University 34(1): 215-226.

Sadr, A., Mohsenifar N., Okhovat, R. S., (2011). "Comparison of MLP and RBF neural networks for Prediction of ECG Signals." *IJCSNS International Journal of Computer Science and Network Security* 11(11):124-128.

Sahambi, J. S., Tandon, S., Bhatt, R. K. P., (1997). "Using wavelet transform for ECG characterization." *IEEE Eng. Med. Biol*. 16(1): 77–83.

Sahambi, J. S., Tandon, S. M., Bhatt, R. K. P., (1997). "Using wavelet transforms for ECG characterization, an on-line digital signal processing system." *IEEE Eng. Med. Biol*. 16: 77–83.

Sahoo, S., Das, T., Sabut, S., (2016). "Adaptive Thresholding based EMD for Delineation of QRS Complex in ECG Signal Analysis." *IEEE*: 500-504.

Sahoo, S. K., Subudhi, A., Kanungo, B., Sabut, S. K., (2015). "Feature extraction of ECG signal based on wavelet transform for arrhythmia detection." IEEE Int Conference on Electrical, Electronics, Signals, *Communication and Optimization* (EESCO):1-5.

Saini, I., Singh, D., Khosla, A., (2013). "QRS detection using k-nearest neighbor algorithm (KNN) and evaluation on standard ECG databases." *J. Adv. Res*. 4 (4): 331–344.

Sameni, R., Shamsollahi, M. B., Jutten, C., (2006). "Multi-Channel Electrocardiogram Denoising Using a Bayesian Filtering Framework." *Proc Comput Cardiol.* : 17–20.

Sameni, R., Shamsollahi, M. B., Jutten, C., Clifford, G. D., (2007). "A nonlinear Bayesian filtering framework for ECG denoising." *IEEE Trans. Biomed. Eng.* 54(12): 2172–2185.

Sameni, R., Shamsollahi, M. B., Jutten, C., (2008). "Model-based Bayesian filtering of cardiac contaminants from biomedical recordings." *Physiological Measurement* 29: 595-613.

Sanamdikar, S. T., Hamde, S. Asutkar, V. G., (2015). "A literature review on arrhythmia analysis of ECG Signal." *International Research Journal of Engineering and Technology* 2 (3), 307-312.

Saxena, S., Kumar, V., Hamde, S., (2002). "QRS detection using new wavelets." *J. Med. Eng. Technol*. 26(1): 7–15.

Sayadi, O., Shamsollahi, M. B., (2007). "Multiadaptive BionicWavelet Transform: Application to ECG Denoising and BaselineWandering Reduction." *EURASIP Journal on Advances in Signal Processing*, Article ID 41274, DOI:10.1155/ 2007/41274.

Sayadi, O., Shamsollahi, M. B., (2009). "A model-based Bayesian framework for ECG beat segmentation." *Physiological Measurement* 30: 335–52.

Sayadi, O., Shamsollahi, M. B., Clifford, G. D., (2010). "Robust detection of premature ventricular contractions using a wave-based Bayesian framework." *IEEE Trans. Biomed. Eng.* 57(2): 353–362.

Schuck Jr, A., Wisbeck, J. O., (2003). "QRS detector pre-processing using the complex wavelet transform." *Proc 25 Ann Inter Conf. IEEE EMBS* Cancun Mexico, (September).

Selvakumar, J., Lakshmi, A., Arivoli, T., (2012). "Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and fuzzy C-mean algorithm." in *Advances in Engineering, Science and Management (ICAESM), International Conference*: 186-190.

Senhadji, L. Bellanger, J. J., Carrault, G., Goatrieux, J. L., (1990). "Wavelet analysis of ECG signals." *Annu. Conf. IEEE Engineering in Medicine and Biology Society* 12: 811-812.

Senhadji, L., Carrault, G., Bellanger, J. J., Passariello, G., (1995). "Comparing wavelet transforms for recognizing cardiac patterns." *IEEE Engineering in Medicine and Biology Magazine* 14: 167-173.

Serra, I., (1982). "Image analysis and mathematical morphology." New York Academic.

Sevindir, K. H., Cetinkaya, S., Yazici, C., (2015). "Wavelet transform based noise removal from ECG signal for accurate heart rate detection using ECG." (2015) *Medical Technologies National Conference (TIPTEKNO)*, Bodrum, Turkey, (October 15-18).

Sevindir, K. H., Cetinkaya, S., Yazici, C., (2016). "QRS Detection Based on Wavelet Transform" *The 8th International conference on Image Processing, Wavelet and Applications* (IWW2016), Istanbul, Turkey, (September 22-24).

Sevindir, K. H., Cetinkaya, S., Yazici, C., (2017) a. "On Choosing a Suitable Wavelet for ECG Artefact Reduction." *International Conference on Applied Analysis and Mathematical Modeling ICAAMM17*, Istanbul, Turkey, (July 3-7).

Sevindir, K. H., Cetinkaya, S., Yazici, C., (2017) b. "On Problems of ECG Feature Extraction." *International Conference on Applied Analysis and Mathematical Modeling ICAAMM17*, Istanbul, Turkey, (July 3-7).

Sevindir, K. H., Cetinkaya, S., Yazici, C., (2017) c. "Comparison among Wavelets for Preprocessing of ECG Signal." *9th International Conference on Image Processing, Wavelet and Applications*, Kars, Turkey, (November 5-8).

Sevindir, K. H., Cetinkaya, S., Yazici, C., (2018) a. "QRS Detection via machine learning algorithms." *International Conference on Mathematics: Mini-symposium on Approximation Theory & Minisymposium on Math Education*, Istanbul, Turkey, (July 3-6).

Sevindir, K. H., Cetinkaya, S., Yazici, C., (2018) b. "ECG feature extraction via wavelet transform and machine learning algorithms." *J. BAUN Inst. Sci. Technol*. 20(1):94-109. doi: 10.25092/baunfbed.413705.

Seymore, K., Mc Callum, A., Rosenfield, R., (1999). "Learning Hidden Markov Model Structure for Information Extraction." In *workshop on Machine Learning for Information Extraction.*

Shah, S., Palmieri, F., Datum, M., (1992). "Optimal filtering algorithms for fast learning in feedforward neural Networks." *Neural Netw. Five*: 779–787.

Shen, C. P., et al. (2012). "Detection of cardiac arrhythmia in electrocardiograms using adaptive feature extraction and modified support vector machines." *Expert Syst. Appl*. 39: 7845–7852.

Shende, S., Patel, R., (2013). "Review of Face Detection using PCA and ANN Techniques." *International Journal on Advanced Computer Theory and Engineering* 2(5): 2319–2526.

Silipo, R., Marchesi, C., (1998). "Artificial Neural Networks for Automatic ECG Analysis." *IEEE Transactions on Signal Processing* 46(5): 1417-1425.

Singh, B. N., Tiwari, A. K., (2006). "Optimal selection of Wavelet basis function applied to ECG signal denoising." *Digital Signal Processing* 16: 275-287.

Slimane, Z.- E. H., Nait-Ali, A., (2010). "QRS complex detection using empirical mode decomposition." *Digit. Signal Process*. 20 (4): 1221–1228.

Soman, A. K., Vaidyanathan, P. P., Nguyen, T. Q., (1993). "Linear phase para-unitary filter banks: Theory, factorizations and designs." *IEEE Trans. Signal Processing* 41: 3480–3495.

Song, M. H., Lee, K. J., Cho, K. J., Yoo, S. K., (2005). "Support vector machine based arrhythmia classification using reduced features." *Int. J. Control Automat. Syst*. 3(4): 571–579.

Song, M. H., Lee, K. J., (2006). "Real-time classification of heartbeats using least square acceleration filter for ambulatory monitoring." *Computers in Cardiology* 33: 565–567.

Soria-Olivas, E., Martínez-Sober, M., Calpe-Maravilla, J., Guerrero- Martínez, J. F., Chorro-Gascó, J., Espí-López, J., (1998). "Application of adaptive signal processing for determining the limits of P and T waves in an ECG." *IEEE Trans. Biomed. Eng.* 45:1077–1080.

Sornmo, L., Borjesson, P. O., Nygards, M., Pahlm, O., (1981). "A method of evaluating QRS shapes features using a mathematical model for the ECG." *IEEE Trans. Biomed. Eng*. 28(10): 713–717.

Sörnmo, L., Pahlm, O., Nygards, M. E., (1982). "Adaptive QRS detection in ambulatory ECG monitoring: A study of performance." in *Computers in Cardiology*. Long Beach, CA: IEEE Computer Society: 201–204.

Sörnmo, L., (1987). "A model-based approach to QRS delineation." *Comput. Biomed. Res*. 20: 526–542.

Specht, D. F., (1990). "Probabilistic neural networks." *Neural Networks* 3(1): 109-118.

Speranza, G., Nollo, G., Ravelli, F., Antolini, R., (1993). "Beat-to beat measurement and analysis of the R-T interval in 24 h ECG Holter recordings." *Med. Biol. Eng. Comput*. 31(5): 487–494.

Stein, C. M., (1981). "Estimation of the mean of amultivariate normal distribution." *Ann. Statist* 9: 1135-1151.

Strang G., Nguyen, T. Q., (1996). *Wavelets and Filter Banks*. Wellesley, MA: Wellesley-Cambridge Univ. Press.

Strintzis, M. G., Stalidis, G., Magnisalis, X., Maglaveras, N., (1992). "Use of neural networks for electrocardiogram (ECG) feature extraction, recognition and classification." *Neural Netw. World* 3(4): 313-327.

Stockwell, R. G., (2007). "A basis for efficient representation of the S-transform." *Digital Signal Processing* 17:371-393.

Strumillo, P., (2002). "Nested median filtering for detecting T-wave offset in ECGs." *Electron. Lett*. 38 (14): 682–683.

Sun, Y., Suppappola, S., Wrublewski, T. A., (1992). "Microcontroller-based real-time QRS detection." *Biomed. Instrum. Technol*. 26(6): 477-484.

Sun, Y. Chan, K. L., Krishnan, S. M., (2005). "Characteristic wave detection in ECG signal using morphological transform." *BMC Cardiovascular Disorders* 5 (September (28)).

Suppappola, S., Sun, Y., (1994). "Nonlinear transforms of ECG signals for digital QRS detection: A quantitative analysis." *IEEE Trans. Biomed. Eng.* 41: 397-400.

Suykens, J. A. K., Vandewalle, J., (1999). "Least square support vector machine classifiers." *Neural Processing Letters* 9: 293–300.

Suzuki, Y., (1995). "Self-organizing QRS-wave recognition in ECG using neural networks." *IEEE Trans. Neural Networks* 6: 1469-1477.

Tadejko, P., Rakowski, W., (2007). "Mathematical Morphology Based ECG Feature Extraction for the Purpose of Heartbeat Classification." *6th International Conference on Computer Information Systems and Industrial Management Applications, CISIM* '07:322-327.

Taigang, H., Clifford, G., Lionel, T., (2005). "Application of independent component analysis in removing artifacts from the electrocardiogram." *Neural computing and applications*: 1–19.

Tang, P. H., Tseng, M. H., (2009). "Medical data mining using BGA and RGA for weighting of features in fuzzy k-NN classification." in *Machine Learning and Cybernetics, International Conference*: 3070-3075.

Taouli, S. A., Bereksireguig, F., (2004). "Application of the morphological filter in the processing of the electrocardiogram signal ECG." *CNIEE 04*, Oran, Algeria: 22-23.

Taouli, S. A., Bereksireguig, F., (2006). "Application of the operator's morphology for the detection of the waves the ECG signal." *International conference on Electrical engineering ICEE*, Oran, Algeria.

Taouli, S. A., Reguig, F. B., (2013). "ECG Signal Denoising by Morphological Top-Hat Transform." *Global Journal of Computer Science and Technology Software & Data Engineering* 13(5): 1-11.

Thirumuruganathan, S., (2010). *A detailed introduction to K-nearest neighbor (KNN) algorithm*. https://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/, Access (October 2017).

Thomas, C., Rose, Charpillet, F., (2007). "A support system for ECG segmentation based on Hidden Markov Models." Annual International Conference of the IEEE Engineering in Medicine and Biology Society, *IEEE Engineering in Medicine and Biology Society*, 3228– 2007, (January 31).

Torse, D. A., Maggavi, R. R., Pujari, S. A., (2012). "Nonlinear Blind Source Separation for EEG Signal Pre-processing in Brain-Computer Interface System for Epilepsy." *Int. J. of Comp. Applications* 50(14).

Toygar, O., Acan, A., (2012). Face Recognition Using PCA, LDA and ICA Approaches on Colored Images, *Journal Of Electrical & Electronics Engineering* 3(1): 735-743.

Trahanias, P., Skordalakis, E., (1990). "Syntactic pattern recognition of the ECG." *IEEE Trans. Pattern Anal. Machine Intell*. 12: 648-657.

Trahanias, P. E., (1993). "An approach to QRS complex detection using mathematical morphology." *IEEE Transactions on BME* 40: 201–205.

Tripathy, R. K., Acharya, A., Choudhary, S. K., (2012). "Gender Classification from ECG Signal Analysis using Least Square Support Vector Machine." *American Journal of Signal Processing* 2(5): 145-149.

Tsipouras, M. G., Fotiadis, D. I., (2004). "Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability." *Computer Methods and Programs in Biomedicine* 74: 95–108.

Tuteur, F. B., (1988). "Wavelet transformations in signal detection." in *Proc. ICASSP 88: 1988 Int. Conf. Acoustics, Speech, and Signal Processing*, New York: 1435-1438.

Tuzman, A., Acosta, M., Bartesaghi, R., Hobbins, T., (1996). "Wavelet-based compression of Holter ECG signals." in P*roc. Annu. Int. Conf. IEEE EMBS.* Amsterdam, Netherlands, (October 31-November 3).

Vaidyanathan, P. P., (1993). *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall.

Vapnik V. N., (1995). *The nature of statistical learning theory*. Springer, New York: 187

Vapnik V. N., (1998). *Statistical learning theory*. Wiley, New York.

Varshney, M., Chandrakar, C., Sharma, M., (2014). "A Survey on Feature Extraction and Classification of ECG Signal." *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* 3(1):6572-6576.

Vernon, A. Allen, Belina, J., (1996). "ECG Data Compression using the Discrete Cosine Transform." *Proceedings Computers in Cardiology*, Durham, NC, USA: 687-690, (October 11-14).

Vila, J., Gang, Y., Presedo, J., Delgado, M. F., Malik, M., (2000). "A new approach for TU complex characterization." *IEEE Trans. Biomed. Eng*. 47: 764–772.

Vijaya, G., Kumar, V., Verma, H. K., (1997). "Artificial Neural Network Based Wave Complex Detection in Electrocardiograms." *Int. J. Of System Science* 28: 125-132.

Vijaya, G., Kumar, V., Verma, H. K., (1998). "ANN-based QRS-complex analysis of ECG." *J. Med. Eng. Technol.* 22(4): 160-167.

Wai, P. M., Yuanjin, Z., Bin, Z., Liu, X., Sheng, W. Y., (2009). "A real-time ECG QRS detection ASIC based on wavelet multiscale analysis." in *Proc. IEEE Asian Solid-State Circuits Conf*.:293–296, Taipei, Taiwan, (November 16-18).

Weng, B., Blanco-Velasco, M., Barner, K. E., (2006). "Baseline wander correction in ECG by the empirical mode decomposition." in: *Proceedings of IEEE 32nd Annual Northeast Bioengineering Conference*: 135–136.

Wiklund, U., Akay, M., Niklasson, U., (1997). Short-Term Analysis of Heart-Rate Variability by Adapted Wavelet Transforms. *IEEE Engineering and Medicine in Biology*: 113–118, Easton, PA, USA, (April 1-2).

Xia, Y., Han, J., Wang, K., (2015). "Quick detection of QRS complexes and R-waves using a wavelet transform and K-means clustering." *Bio-Medical Materials and Engineering* 26:1059–S106.

Xie, Q. Z., Hu, Y. H., Tompkins, W. J., (1992). "Neural-network based adaptive matched filtering of QRS detection." *IEEE Trans. Biomed. Eng*. 39: 317-329.

Xu, W., Li, L., Zou, S., (2007). "Detection and Classification of Micro calcifications Based on DWT and ANFIS." In: *The 1st International Conference on Bioinformatics and Biomedical Engineering*, ICBBE (2007), USA, (July 6–8).

Ye, D., Ouyang, X., (1996). "Application of wavelet analysis in detection of fetal ECG." in *Proc. Annu. Int. Conf. IEEE EMBS.* Amsterdam, Netherlands, (October 31-November 3).

Yimman, S., Hinjit, W., Desyoo, P., Bholsithi, W., Dejhan, K., (2004). "Wavelet packet transform applications on QRS detection." in: *Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology*, pp. 445–448, Rome, Italy, (December 18-21).

Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X., (2007). "Intelligent Data Engineering and Automated Learning - IDEAL." *8th International Conference*, Birmingham, UK, (December 16-19) Springer, p. 1174.

Yochum, M., Renaud, C., Jacquir, S., (2016). "Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT." *Biomed. Signal Process Control* 25: 46–52.

Yong, Z., Wenxue, H., Yonghong, X., (2009). "ECG beats feature extraction based on geometric algebra computational intelligence and software engineering." in *International Conference on Computational Intelligence and Software Engineering*, (CiSE 2009), 1–3 pp.

Yu, B. C., Liu, S., Lee, M., Chen, C. Y., Chiang, B. N., (1985). "A nonlinear digital filter for cardiac QRS complex detection." *J. Clin. Eng.* 10: 193-201.

Yu, S. N., Chou, K. T., (2007). "Integration of independent component analysis and neural networks for ECG beat classification." *Expert Syst. Appl.* 34(4):2841-2846.

Zayandehroodi, H., Mohamed, A., Shareef, H. Mohammadjafari, M., (2010). "Performance Comparison of MLP and RBF Neural Networks for Fault Location in Distribution Networks with DGs." *IEEE International Conference on Power and Energy* (PECon2010), Nov29 - Dec1, pp. 341-345.

Zeman, T., (2000). *BSS - Preprocessing Steps for Separation Improvement*, May Available at http://noel.feld.cvut.cz/vyu/prj-czs-asi/zeman/pca.pdf

Zhan, Y., Chen, H., Zhang G., (2006). "An optimization algorithm of K-NN classification, in Machine Learning and Cybernetics." *International Conference*: 2246- 2251, Dalian, China, (August 13-16).

Zhang, L., Bao, A., (2002). "Edge detection by scale multiplication in wavelet domain." *Pattern Recognition Letters* 23(14): 1771-1784.

Zhang, L., Bao, A., Wu, X., (2005). "Canny edge detection enhacement by scale multiplication." *IEEE Trans. On Pattern Analysis and Machine Intelligence* 27(9): 1485-1490.

Zhang, F., et al. (2009). "QRS Detection Based on Multi-Scale Mathematical Morphology for Wearable ECG Device in Body Area Networks." *IEEE Trans. on Biomedical Circuits and Systems* 3(4): 220-228.

Zhidong, Z., Juan, L., (2010). "Baseline wander removal of ECG signals using empirical mode decomposition and adaptive filter." in: *Proceedings of International Conference on Bioinformatics and Biomedical Engineering*: 1–3, Chengdu, China, (June 18-20).

Zidelmal, Z., Amirou, A., Adnane, M., Belouchrani, A., (2012). "QRS detection based on wavelet coefficients." *Comput. Meth. Prog. Biomed.* 107(3):490–496.

Zidelmal, Z., Amirou, A., Ould-Abdeslamb, D., Merckleb, J., (2013). "ECG beat classification using a cost sensitive classifier." *Comp. Methods Programs Biomed*. 11(3): 570–577.

Zidelmal, Z., Amirou, A., Moukadem, A., Dieterlen, A., (2014). "QRS detection using S-transform and Shannon energy." *Comput. Methods Programs Biomed*. 116: 1–9.

Zong, W., Jiang, D., (1998). "Automated ECG rhythm analysis using fuzzy reasoning." *IEEE Conference, Computers in Cardiology*: 69 – 72, Cleveland, OH, USA, (September 13-16).

Zong, W., Moody, G. B., Jiang, D., (2003). "A robust open-source algorithm to detect onset and duration of QRS complexes." in *Proc. of Comput. In Cardiol.* 737–740, Thessaloniki Chalkidiki, Greece, (September 21-24).

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 9

COORDINATES OF NEUROSURGERY IN 3-DIMENSIONAL CUBIC SYSTEM: ASSESSMENT OF BRAIN THROUGH THE PRINCIPLES OF GEOMETRY

İ. Doğan[1,*] and T. Dündar[2]
[1]Ankara University, Faculty of Medicine, Department of Neurosurgery, Ankara, Turkey
[2]Bezmi Alem University, Faculty of Medicine, Department of Neurosurgery, İstanbul, Turkey

ABSTRACT

In this report, the use of geometrical terms and analysis in health sciences as well as in neurosurgery are examined and documented. Our focus of interest and study was based on the issues related to the applied fields of 3-dimensional Cartesian system to neurosurgery and its clinical and anatomic usage in neurosurgical procedures. For this purpose, we have done an elaborate literature review and significant examples were presented and discussed in the lights of data obtained from 3-dimensional radiological images created by using Osirix software program (Pixmeo SARL, Switzerland). According to these data, well-known but disregarded importance and implementation of geometry in neurosurgery have been underlined and clarified.

Keywords: neurosurgery, geometry, mathematic, Cartesian system, coordinate system

* Correaponding Author's E-mail: ihsandogan@ankara.edu.tr.

INTRODUCTION[1]

If we hypothesized that the human brain was placed on an imaginary intersection area that could also be named as "zero point" on a fictional line, there would be two separates but originally similar natures in each open-ended side of this point (Figure 1). As it gets further from brain throughout the "plus" direction, the depths of expanding space will describe the term the positive or outer infinity. In the opposite route, to the "minus" direction, beneath the DNA double helix of a cell will generate the negative or inner infinity. According to this hypothetical point of view, brain can be considered that it is centered in an enlarging circle (major cosmos) within its continuously decreasing cellular compositions (minor cosmos) inside (Figure 1).

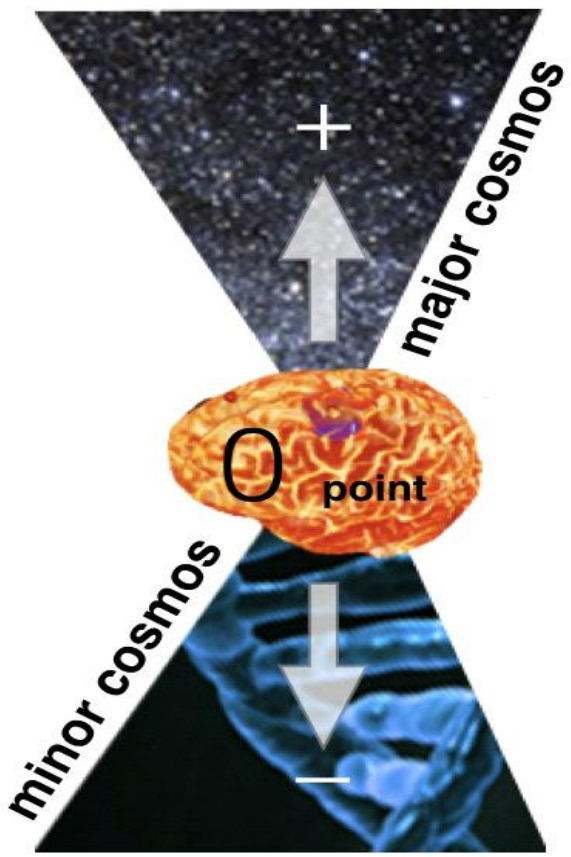


Figure 1. An illustrative summary of hypothetical environment in which brain is placed between major and minor cosmosis.

From a perspective of a neurosurgeon; brain is not only consisted of anatomical structures such as ridges (gyri), grooves (sulcuses), channels (ventricles) and corridors

[1] All the figures in this article are original and not published before.

(cisterns) within fluid inside (cerebrospinal fluid) but also enclosing a complex internal gray and white matter networks and roads which are responsible for signal producing and transmission to surrounding organs via spinal cord and peripheric nervous system (1-3, 5, 6, 12). All of these macro and micro neuroanatomical and histologic/genetic structures of brain need to be explored, understood and explained in terms of their settlements, functionality, connectivity and disorders until all unknowns about central nervous system are clarified.

The history of brain investigations is as old as the history of humanity. Existence of universe and everything associated with nature are explained by the language of mathematics. Examining and studying the amazing brain, which is still compromised of unknowns, have always been one of the human's favorite interest and sensation subjects. Brain is not only a guide for human being in understanding the universe and feeding the desire to travel to unreachable parts of space, but also a medium in solving the internal biological system. During the brain's maturation and evolution period, human beings started to question their existence and reached the facts via their brain.

In this paper, we aimed to discuss the neuroanatomy, neurological pathologies and neurosurgical procedures from the geometric perspective. We focused on three-dimensional Cartesian system and mentioned about where this basic and main geometric model is applicable and convenient primarily in neurosurgeons' daily practice.

MATERIAL AND METHODS

After an extensive literature review, we determined the significant and main topics in neurosurgery such as in which geometry is basically related to the issue, explains the configuration of an anatomical structure or directs the surgical planning. After determination of these issues, we created three-dimensional radiological images of these pathologies, normal anatomic structures or surgical devices by using Osirix software program (Pixmeo SARL, Switzerland). The 3-dimensional images were superpositioned with Cartesian coordinate system and the mathematical relationship and similarity were emphasized according to these images.

RESULTS

Nomenclature in anatomy, inner ear structures, neurosurgical procedures such as pedicular screw placement, endoscopic surgical approaches, and ventricular drainage techniques created the basis of our findings. We figured our discussion according to these issues and geometric-medical harmonies.

DISCUSSION

All through the history of human's searching and reaching aim to the unknowns and reality, mathematics has become the essential tool and common language for all scientists. While mathematic having been used in physical sciences, engineering and life sciences; it is also indispensable in health sciences and medicine in contrary to what was thought in general meaning. In addition, geometry, as a sub-branch of mathematics has an important role in describing the anatomical configuration and construction of central nervous system.

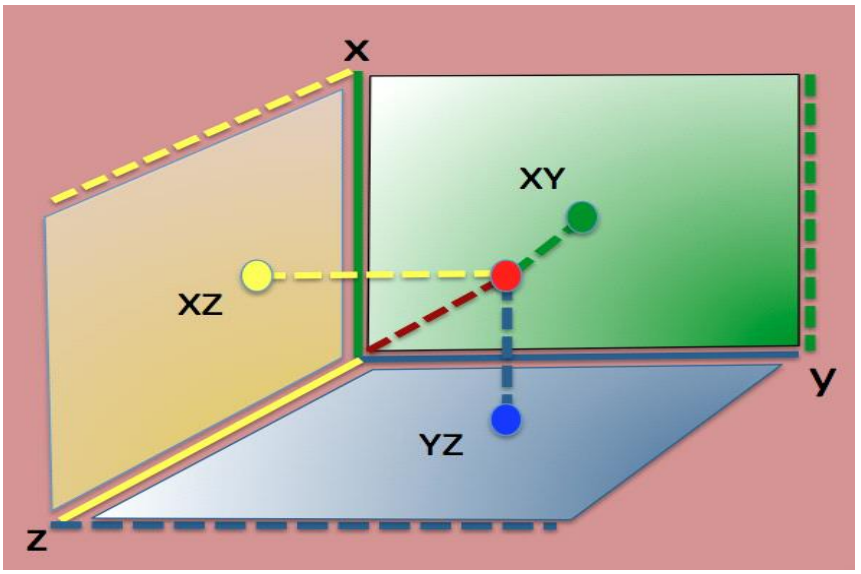


Figure 2. Three-dimensional Cartesian (cubic) system used to define a location of a point (red dot) in space. (Green, yellow, blue surfaces are XY, YZ, XZ planes respectively and green, blue, yellow dots are shown as projection of red dot in each plane).

Three-dimensional cubic (coordinate) system lays the foundation for analytic geometry (Figure 2) (7). Both the location of a point itself alone and position of this point relative to the other points in space are described by placing this point in this 3-dimensional coordinate system. Three geometric axis and planes, which are perpendicular to one another form this configuration (Arıncı et al. 1987, 1989, 1997, Doğan, 2011, Drake et al. 2010). Each of these three geometric planes is also termed as cartesian system and in contradistinction to coordinate system, this system defines a location of a point in a 2-dimensional plane (Kaplan, 2010). In summary, these three 2-dimensional cartesian systems each of which is indicating different planes constitute a one 3-dimensional coordinate system.

At first sight, it seems like there is no relevance between health sciences with this basic geometrical concept. Yet, it should be known that anatomical nomenclature has

been established on this system (Arıncı et al. 1987, 1989, 1997). Nomenclature of the surfaces and tips of organs, the position and proximity of the organs, neural and vascular structures, etc., relative to each other or inside human body, the bones' directions of movement are the main examples of the use of coordinate system in medicine where the human being standing on an anatomical position is situated on a coordinate system (Drake, 2010). Thus, all anatomic nomenclature is established based on this rationale (Figure 3A, B) (Arıncı, 1997). Additionally, the same principles are also acceptable for head and neck, brain anatomy more specifically, from geometric point of view and this coordinate system can be modified and specialized for studies about brain itself and issues related to the brain are also currently adopted in this scope to neurosurgical procedures and pathologies (Figure 3C-D)(1,2,5,6). Whereas 3-dimensional coordinate system consists of 3 axes (x, y and z axes) and 3 planes (x', y' and z' planes); 3-dimensional anatomical coordinate system consists of 3 axes (transverse, vertical and sagittal) and 3 planes (axial, coronal and sagittal) respectively. In addition to these, if a body is accepted as being a compound of two symmetric parts, the plane sectioning the body into left and right two symmetric parts is named as median plane (Drake, 2010).

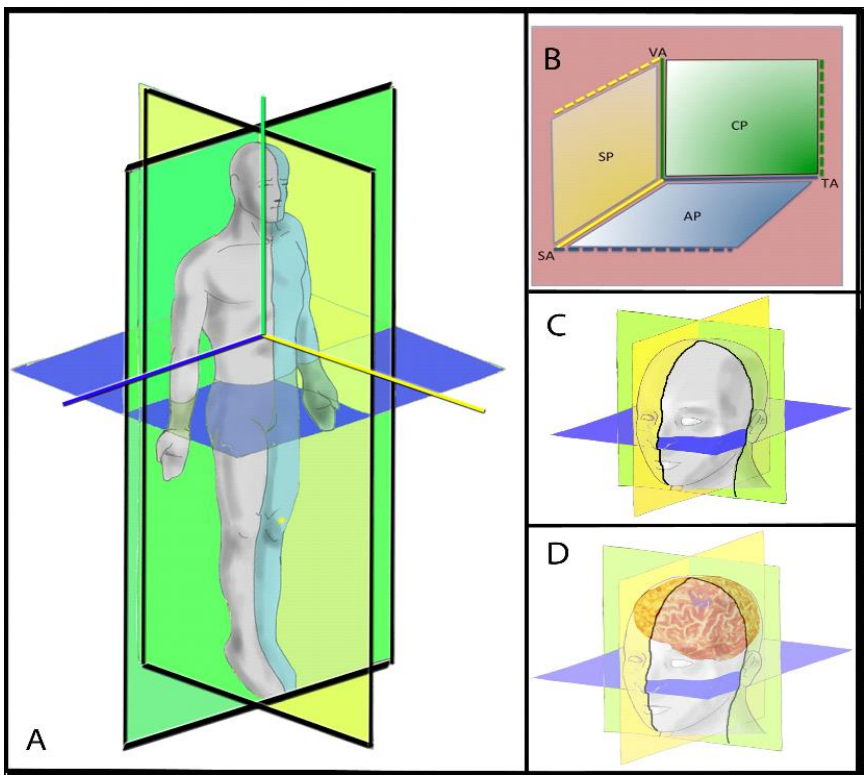


Figure 3. A. Anatomical position in medicine. B. Three-dimensional anatomical planes and axis. C-D. Adaptation and representation of 3-d anatomical planes and axis on head and neck (C) and brain (D). (Green surface, yellow surface, blue surface are shown as coronal plane, sagittal plane, axial plane; green line, yellow line, blue line are shown as vertical axe, sagittal axe, and transverse axe respectively).

The anatomical location/direction-related terms such as medial-lateral, superior-inferior, distal-proximal, ventral-dorsal, anterior-posterior are all defined and used according to the posture of human in anatomical position (Arıncı, 1997). In this position, anatomical coordinate system is adapted from geometrical coordinate system and is set up for human (Arıncı, 1997, Putz, 1993). From this point of view, it should be clearly understood that mathematics is creating the basis of anatomical terminology from a geometrical perspective. As in human body, this basic 3-dimensional coordinate system can be adopted and specialize to brain itself (Figure 3D). The positions of neural and vascular structures according to each other and the locations of these structures according to the brain can be defined and termed to this basic principle.

As an example, three semicircular canals, located on temporal bone, which are responsible for the sensation of head movements and functioning for giving information to the brain about the position of head, are the positioned state of coordinate system in our body. The configuration of these canals together is similar in shape with the three planes of coordinate system (Figure 4) (Cömert et al., 2014).

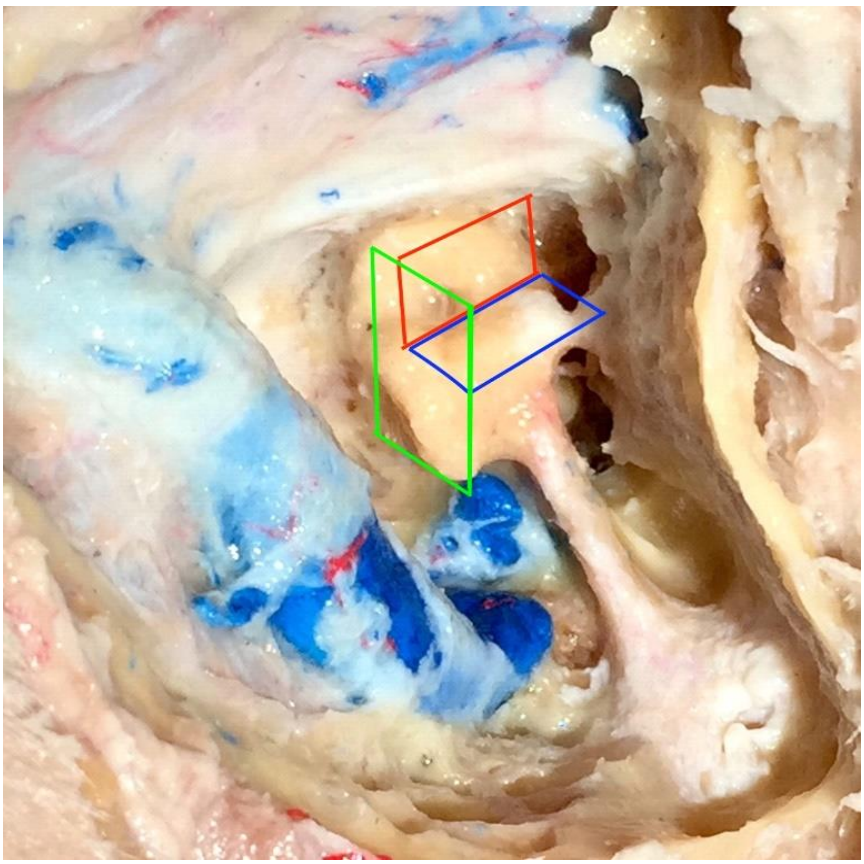


Figure 4. Demonstration of three semicircular canals inside inner ear mimicking the same positions of three geometric planes.

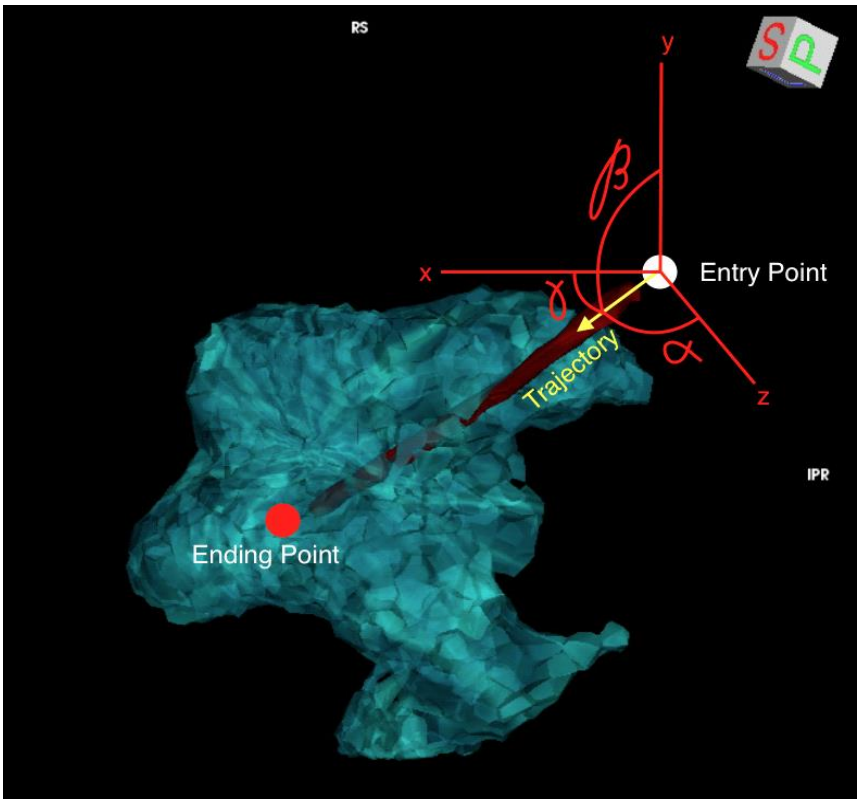


Figure 5. Three-dimensional radiological model of lateral ventricules and shunt catheter (red rod). Angulation of catheter is shown with different angles (α, β, ϒ) according to the 3 anatomical axes.

As in defining the normal brain anatomy, geometry also helps us to understand and clarify the central nervous system problems (pathologies) and show us to solve (treat) them.

In patients with hydrocephalus, placement of ventricular catheter or shunt is a curable treatment option. In this procedure, determination of trajectory of shunt catheter from its entry point on skull to endpoint inside the ventricul for a proper and effective placement is a crucial step. The preoperative determination and calculation of this procedure is done as placing the brain in cubic system with its beginning and ending points (Figure 5) (Stein et al., 2007).

In addition, in spinal procedures, pedicul screw placement is performed by taking into account of the angulation the srew has with each of the 3 anatomical planes. As shown in figure 6, for correct placement of the screw, all three angulations should be considered and measured for correct position and effective treatment (Vaccaro et al., 1995).

Furthermore, there are some studies focusing on comparing different surgical skull base approaches in terms of surgical exposure, range of motion of surgeon's hand, line of vision of a surgeon under microscope. For this purpose, 3-d coordinate system is used as an objective tool for quantitative results (Li et al., 2016, Uğur et al., 2013).

Endoscopic procedures are another area of use of geometrical principles and 3 planes and axis. For the purpose of reaching the pathology inside the deeper structures of brain, angulation of endoscopic device with each 3 planes should also be measured on radiological 3-dimensional images superposed with anatomical coordinate system (Figure 7) (Kulkarni et al., 2000,2010, Lam et al. 2014).

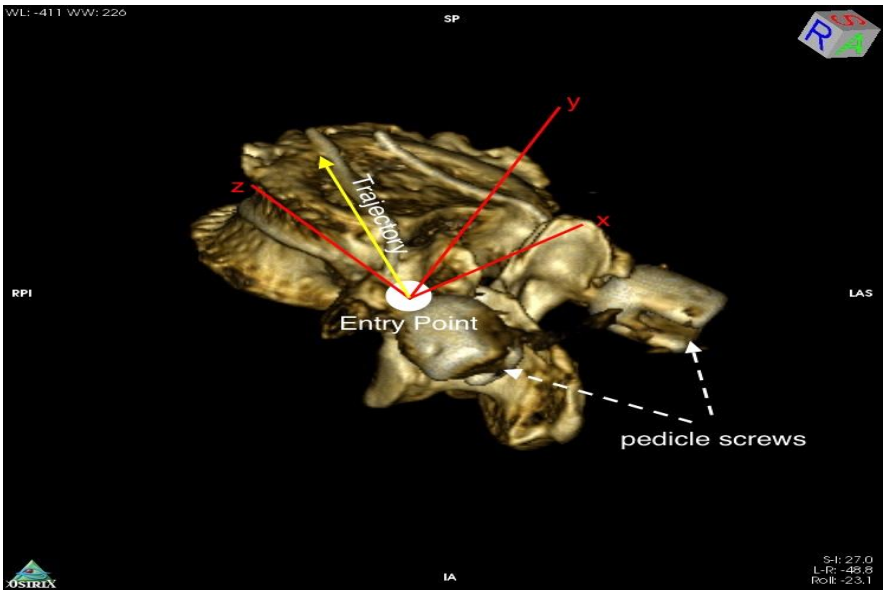


Figure 6. Three-dimensional radiological view of a vertebra and screws. Angular relationship of screw trajectory inside the vertebral body with anatomical axis is shown.

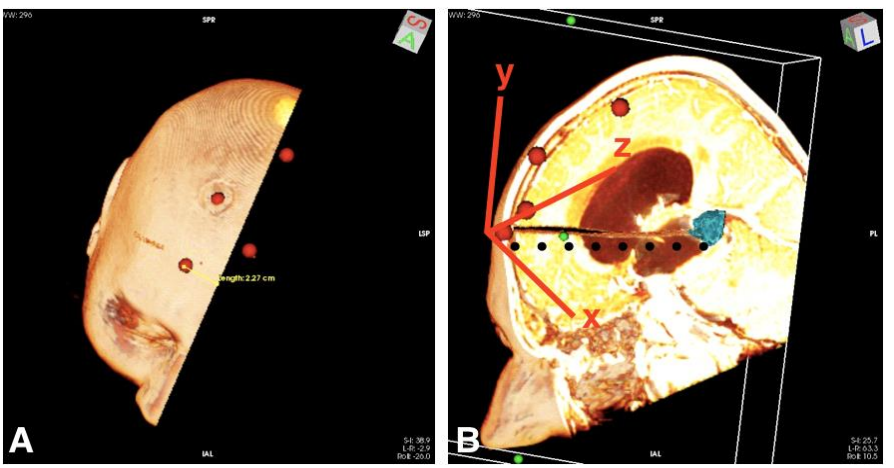


Figure 7. A. Demonstration of the entry points (red dots) over head surface corresponding to intersection of each three axis. B. Trajectory of endoscope is illustrated with black dots and its relationship with 3 axes. (Blue irregular shape is shown as pathology at the end of the trajectory line.).

As it is tried to be shown by giving conspicuous examples in this brief article, mathematical/geometrical modeling can also be encountered in health sciences and it is

the basis of anatomical nomenclature and principles. Neurosurgery, as a subdivision of medicine, is also advancing under the guidance of the mathematic.

RESULTS

Beside all these concrete examples from mathematics regarding its adaptation to neurosurgery and medicine in our article, it should always be kept in mind that health is an equation that needs to be solved in a philosophical manner. Whereas one side of the equation is formed by diseases (problems), the other side of it should include the treatments (solutions). Just like how mathematics proves that, there will always be a solution when a problem exists; it will teach us life is an equation continuing in a balanced manner and tell should there be a disease, the cure will be provided.

REFERENCES

Arıncı K., Elhan A., (1987). *Structure of the Human Body (Morphologic information)* Turkish Historical Society Publications, Ankara, 67p.

Arıncı K., Elhan A., (1989). *Anatomy Terms Guide* 7. Edition. Ankara: Güneş Publications, 3-12pp.

Arıncı K, Elhan A., Anatomy. In: Arıncı K., Elhan A., (1997). *Vol1*. 2. Edition. Güneş Publications, Ankara, 89-160pp.

Cömert E., Cömert A., Çay N., Tunçel Ü., Tekdemir İ., (2014). Surgical anatomy of infralabyrinthine approach. *Otolaryngol Head Neck Surg*. (2):301–307.

Doğan, İ., (2011). *The nomenclature and axis of intracranial arteries along their axis.* Medical Thesis, Ankara University School of Medicine, Ankara, 116s.

Drake R.L., Vogl W., Mitchell A.W.M., (2010). The body. *Gray's Anatomy for Students*, 2nd ed. Churchill Livingstone Elsevier, London, 4-53pp.

Kaplan E., (2010). *Geometry 9 Textbook*, 9th school. Paşa Publications, Ankara 2-54pp.

Kulkarni A.V., Drake J.M., Armstrong D.C., Dirks P.B., Imaging correlates of successful endoscopic third ventriculostomy, *J. Neurosurg*. 92 (2000) 915–919.

Kulkarni A.V., Drake J.M., Kestle J.R., Mallucci C.L., Sgouros S., Constantini S., Predicting who will benefit from endoscopic third ventriculostomy compared with shunt insertion in childhood hydrocephalus using the ETV Success Score, *J. Neurosurg. Pediatr*. 6 (2010) 310–315.

Lam S., Harris D., Rocque B.G., Ham S.A., Pediatric endoscopic third ventriculostomy: a population-based study, *J. Neurosurg. Pediatr*. 14 (2014) 455–464.

Li R.C., Liu J.F., Li K., Qi L., Yan S.Y., Wang M.D., Xie W.F., (2016). Localization of Anterosuperior Point of Transverse-sigmoid Sinus Junction Using a Reference

Coordinate System on Lateral Skull *Surface Chin. Med. J.* (Engl). (15):1845-9. doi:10.4103/0366-6999.186633

Putz R,. Pabst R., (1993). *Sobotta atlas of human anatomy*. Vol. 1. 20th ed. München: Urban &Schwarzenberg. [Translation editor: Elhan A., (1994). *Sobotta insan anatomisi atlası*. Cilt 1. Beta Basım A.Ş. İstanbul], 385- 389.

Stein S.C., Guo W., (2007). A mathematical model of survival in a newly inserted ventricular shunt. *J. Neurosurg*. 107 (6 Suppl): 448–454.

Ugur H.C., Dogan I., Kahilogullari G., AlBeyati E.S., Ozdemir M., Kayaci S., (2013). New practical landmarks to determine sigmoid sinus free zones for suboccipital approaches: An anatomical study. *J. Craniofac Surg.* (24):18158. doi:10.1097/SCS. 0b013e3182997ff8.

Vaccaro AR., Rizzolo SJ., Allardyce TJ., et al. Placement of pedicle screws in the thoracic spine. Part (I): Morphometric analysis of the thoracic vertebrae. *J. Bone Joint Surg. Am.* (1995). 77: 1193–9.

Vaccaro AR., Rizzolo SJ., Balderston RA., et al. Placement of pedicle screws in the thoracic spine. Part (II): An anatomical and radiographic assessment. *J. Bone Joint Surg. Am.* (1995). 77: 1200–6.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 10

UNDERSTANDING THE BRAIN: COMPUTATIONAL NEUROSCIENCE

A. S. Mutluer[1], I. Doğan[2] and T. T. Dündar[3,*]
[1]Bezmialem Vakif University, Faculty of Medicine, Istanbul, Turkey
[2]Ankara University, Faculty of Medicine, Department of Neurosurgery, Ankara, Turkey
[3]Bezmialem Vakif University, Faculty of Medicine, Department of Neurosurgery, Istanbul, Turkey

ABSTRACT

The thirst for knowledge led to the formation and development of sciences. The mysteries of the universe led the humankind to ask questions related to its environment and life, unsurprisingly, is at the center of those questions. The whole efforts were for understanding and it will be always for it. Mathematics is the most popular and valid principle for humankind. By definition, mathematics is the study of seeking out patterns in the micro and macro universe and use them to formulate new conjectures. Theoretical modeling is important for understanding what nervous system do, defining how they function and discovering particular ways in process operation. We just describe a small part of it in this chapter. Understanding the human brain will take many years in our century. We hope basic sciences will be in progress until finally figure out the whole mechanism of the human brain in this century.

Keywords: brain, computational neuroscience, mathematics

The noblest pleasure is the joy of understanding.
Leonardo Da Vinci

* Corresponding Author's E-mail: drtolgadundar@hotmail.com.

INTRODUCTION

Life is a beautiful, magical, and the most properly tiny part of existence for a human being. Compared to the age of the universe, the history of mankind is as brief as the life of autumn grass. According to our current understanding, the universe is existing by some solid and complex laws. From the beginning of the human era, understanding the universal laws of existence has always been a remarkable and meaningful part of creation. Laws of the universe have always been there, and they were waiting to be discovered and understood by a perceptible mind. Mankind created a new concept for the universe with their curiosity and passion to understanding. They classified and categorized the laws of existence as mathematics, physics, biology, and geometry.

Mathematics has become the key to understand the mysteries of the universe. Mathematics is proved to enable people to understand the order and the structure of the universe. Today, mathematics constitutes the basis of contemporary science: physics uses mathematical models to explain the relativity, chemistry relies on the power of mathematics to explain the law of energy conversion. Regardless of how chaotic it seems, there is an order in the universe and mathematics is the key to discover this order. It is, thus, only natural to observe the ubiquity of mathematics. A knowledge of mathematics enables one to understand the harmony and order in a star cluster, in a living microorganism, in music and enjoy it to its fullest.

In our era, the brain is the biggest mysterious for humanity. Being the center of perception in the human body with a lot of extra features makes brain precious and worth to understand. By definition, the human brain is the center of the nervous system and it process and regulates data by constantly incoming information from some 100 billion neurons, each with thousands of synapses connection regulated by a huge range of molecular interactions. Afferent and efferent nerves regulate the body and transmit the data created by the interaction with the environment. The continuous flow and process of information is the essential mission for the central nervous system in a living human. Understanding the brain, henceforth, is the key to have more control human life and nature. We are meaningful in the entire universe with our perception and intelligence by our brains. There are billions of neurons in a chaotic communicated with each other, and the result of this complex communication and incessant data flow bring out a human perception with unrivaled intelligence. Yet, as all we know, technology and science are used to evolve something in our world incredibly.

Some function and abilities can be created again by technology with some huge extra performance and energy like a transform to supercars from horses. It is a good metaphor for instance. A horse can be described as carriage and it can make use of speed in terms of their functions. Then we evolved that function with super cars that they have hundreds of horsepowers by taking inspiration from biological evolution (horses). Anatomically and structurally different, but functionally they are better and incomparable. If we can

create again a functional brain in technology world without limits and bindings to our bodies, it will have processing capacity more than hundreds of brains. That is our expectation from artificial intelligence and it is only possible with understanding the brain and its mechanism well. Understanding the brain, functionally and physiologically give us an opportunity to fight with neurodegenerative diseases and more complex problem with the nervous system also.

In the last centuries, understanding the brain become a hot topic of science for humanity. Many studies and research programs in pursuit of knowledge about the brain are based on a multidisciplinary approach (Brea and Gerstner, 2016; Kesner and Rolls, 2015).

A number of scientific fields, like mathematics, physics, medicine, biology, physiology, computer science, are working together to collect and produce data about the brain. In the early 19th century, mathematics was used in attempts to explain the brain. Mathematics was the first step in the pursuit of understanding the brain. Following this, in the last few decades, mathematical models of brain acquired popularity as computer science and technology become more involved. Latest analytics shows that in the last 50 years the annual research articles including the word 'brain' increased from less than 3000 per year in 1960 to more than 60 000 in 2010 (Bressan and Blackwell, 2014).

Nevertheless, humanity still does not have enough knowledge and the answer to the key questions to understand the brain. There is a need to describe the brain from the molecular level through to cells, synaptic connections, neuronal microcircuits and functional system and areas of the nervous system by taking help from interdisciplinary science fields. Although the existing literature on the field can cover only a limited part of the mysteries, there are some emerging initiatives considered to be promising.

The Human Brain Project (HBP) is one of those large initiatives with its great budget amounts to one billion euro (ten-year project) and aims at a totally new research method for understanding the brain and for developing new technologies based on this understanding. HBP supported and financed by European Nations and have radical ideas about brain studies. According to neuroscientist HBP is not defining the brain and producing new information from nothing. HBP is building the brain from big data. Big data has collected all information about the brain until today. The HBP is based on developing six Information & Communication Technology (ICT) Platforms, dedicated respectively to Neuro-informatics, Brain Simulation, High-Performance Computing (HPC), Medical Informatics, Neuromorphic Computing and Neuro-robotics (Bressan B., Blackwell W., 2014).

Computing revolution in science helps to various ambitious science field in neuroscience. HBP is one of them that requires major transformation like rapid growth in the capabilities of high-performance computing technologies (Rewriting Life a Detailed 3-D Atlas of a Human Brain, 2013). In addition to that, Big Data and the internet are one

of the major transformations in Information Technologies (IT) that will accelerate the development of the HBP (Calimera A, Macii E, Poncino M., 2014).

In the last few years, scientists have imaged the anatomy of a brain at peerless resolution in the name of BigBrain project. A detailed 3D-atlas of a human brain will allow scientists to explore the anatomy of a single brain in three dimensions, Figure 1. The BigBrain has been created as a part of the Human Brain Project and it will guide the scientists to further discoveries as greater detailed brain map for human brain activities as never before.

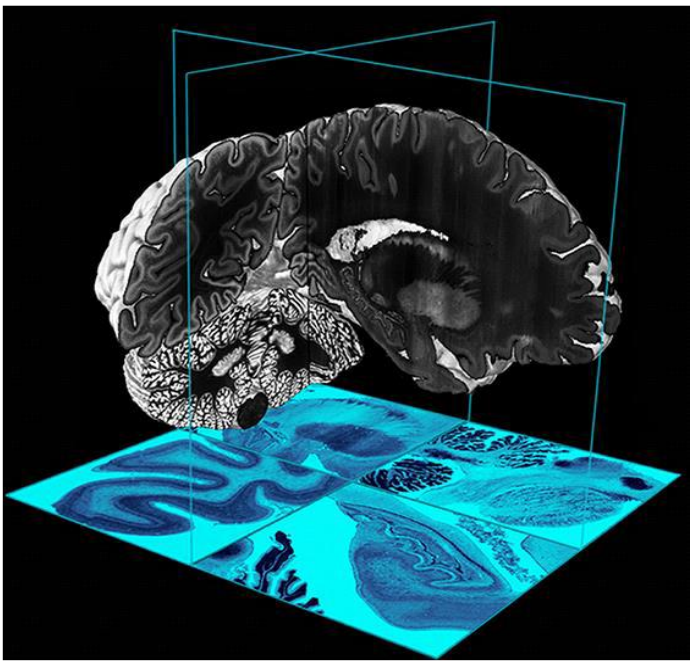


Figure 1. A new detailed 3D Brain Map reveals the human brain in the greatest detail ever. Here, a detailed look at the brainstem and cerebellum from the BigBrain Project (Katrin Amunts et al., 2013).

The creators of the BigBrain, led by Katrin Amunts at the Research Centre Jülich and the Heinrich Heine University Düsseldorf in Germany, imaged the healthy brain of 65-year-old woman cadaver by using an MRI machine. Then, the brain was cut it into 7,404 slices, each just 20 micrometers thick, with microtome after embedding into paraffin wax process, Figure 2. Each micro-slice was digitally scanned and inbuilt in slide frames. As they said, if you want to describe a city, you will need a city map. In parallel to this, brain mapping plays a key role in understanding the brain. Big Brain project is a great achievement about brain mapping.

The last developments about the brain and neuroscience are mentioned. From here onwards, the mathematics and computational revolution based on computational neuroscience will be presented.

Figure 2. Researchers used a tool called a microtome to cut a brain into slices 20 micrometers thick (Image courtesy of Amunts, Zilles, Evans, et al.), (Nature News Whole human brain mapped in 3D, 2013).

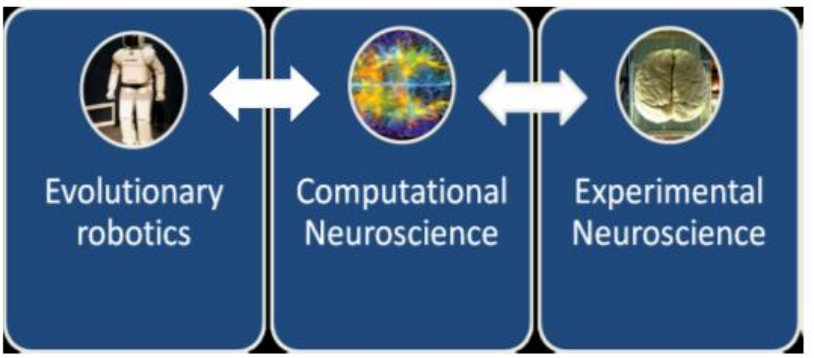


Figure 3. Computational neuroscience (CN) is the natural source of an experimental research area. Formulating and testing hypotheses can be possible with the help of computational neuroscience. In addition, they can be utilized to make expectations that can consequently be tried tentatively. The outline of such models is an imaginative procedure compelled by test information.

COMPUTATIONAL NEUROSCIENCE (THEORETICAL NEUROSCIENCE)

Theoretical modeling is important to understand what nervous system do, defining how they function and discovering particular ways in process operation. Neuroscience comprehends molecular and cellular studies to human psychology and neurophysiology. Computational neuroscience emboldens the interdisciplinary information exchange for a better understanding of the nervous system and brain function. By definition,

computational neuroscience (theoretical neuroscience) is a multidisciplinary approach that connects various fields of neuroscience, mathematics, computer sciences, physics, electric electronics, cognitive sciences, and psychology (Catani and Sandrone, 2015). Computational neuroscience has a key role in producing brain models and employing biological and technological approaches to understand the brain, Figure 3.

The early historical roots of the computational neuroscience field can be commemorated with a few popular names such as Louis Lapicque, Hodgkin & Huxley, Hubel & Wiesel, and David Marr. Lapicque described integrate and fire model of the neuron in his article *Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation* (1907). That model is still one of the most popular and the simplest models in computational neuroscience for both cellular and neural networks studies. Later, Hodgkin & Huxley (1952) developed the voltage clamp method and created the first mathematical model of the action potential of a neuron. Hubel and Wiesel (1962) found that neurons in the essential visual cortex, the main cortical zone to process data originating from the retina, have situated open fields and are sorted out in sections (Human Brain Project). David Marr's (1977) work centered on the collaborations between neurons, recommending computational ways to deal with the investigation of how practical gatherings of neurons inside the hippocampus and neocortex interface, store, prepare, and transmit data. Wilfrid Rall (1989) has made an initial with the first multi-compartmental model using cable theory on computational modeling of a more realistic biophysical neuron.

COMPUTATIONAL NEUROSCIENCE HAS MAJOR TOPICS

Single neuron modeling sensory processing, memory and synaptic plasticity, behaviors of networks-cognition, discrimination, and learning, consciousness, computational clinical neuroscience, development, axonal patterning, and guidance. The basis of computational neuroscience with defining a few topics of neuron model will be mentioned in this chapter. When neuron models are described, an attempt to express mathematical equations will be made as simple as possible.

UNDERSTANDING TO NEURON CELL

We just started with the mathematics for understanding the conductance of neuron and neurotransmitter information flows between neurons and functional interactions. We need to have a better understanding of the neuron that fundamental structures of the nervous system. A single neuron is working like a diode principally. There are two basic

missions of a neuron: generation of electricity and secretion of chemicals to the synaptic cleft. Single neuron dendrite that receive inputs from other neurons and carries the neuronal output to other cells with their axons. The structure of dendrites allows receiving inputs from many other neurons through synaptic connections. We can see diagrams of neurons in Figure 4. Physical features of neurons can be described on mouse neurons. Cortical neurons of the mouse brain typically send out a total of about 40 mm of an axon and have approximately 4 mm of total dendritic cable in their branched dendritic structures. The axon makes approximately 180 synaptic connections with other neurons. The soma of typical cortical neurons has a diameter range about 10 to 50 μm.

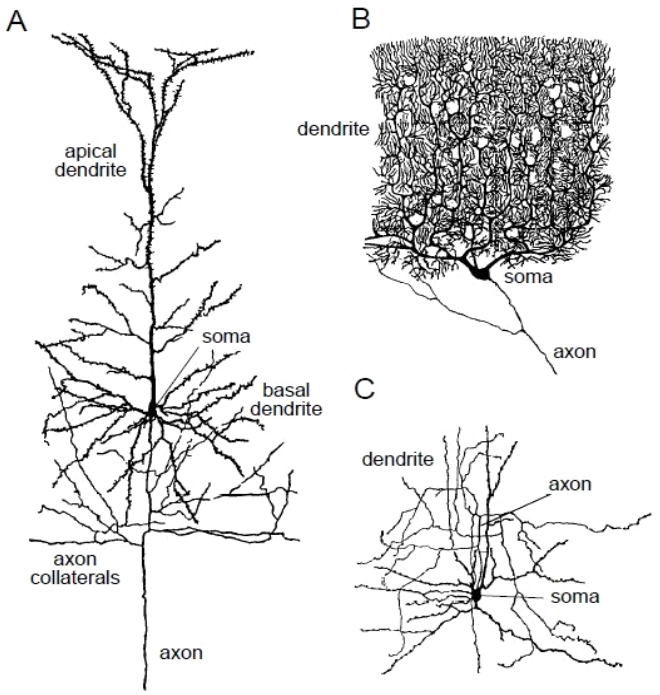


Figure 4. Diagrams of three neurons. A) A cortical pyramidal cell. These are the primary excitatory neurons of the cerebral cortex. Pyramidal cell axons branch locally, sending axon collaterals to synapse with nearby neurons, and also project more distally to conduct signals to other parts of the brain and nervous system. B) A Purkinje cell of the cerebellum. Purkinje cell axons transmit the output of the cerebellar cortex. C) A stellate cell of the cerebral cortex. Stellate cells are one of a large class of cells that provide inhibitory input to the neurons of the cerebral cortex. To give an idea of scale, these figures are magnified about 150 fold (Dayan and Abbott, 2000).

Besides all these, neurons have physiological specializations like featured membrane-spanning ion channels (predominantly sodium (Na+), potassium (K+), calcium (Ca2+), and chloride (Cl−)) opening to the synaptic cleft that allows to ions to move into and out of the cell structures (Figure 5). Ion channels control that ion flow between inside and outside of neuron membrane by opening and closing reflexes in response to voltage changes and both internal and external stimulants. Neuron cells can be functional with these ion channels that secrete to synaptic cleft from synaptic vesicles.

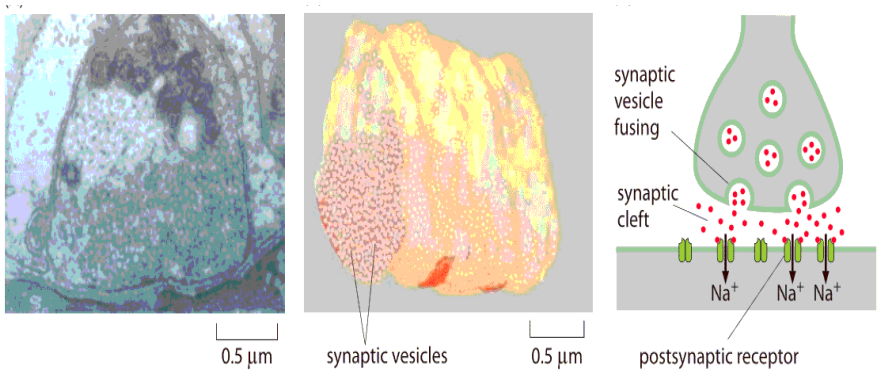


Figure 5. Structure of a neuromuscular junction. (A) Electron microscopy image of a nerve terminal and its synapse with a neighboring cell in a neuromuscular junction. (B) Cryo-electron microscopy reconstruction image of a fraction of the presynaptic neuron showing the synaptic vesicles it harbors for future release. (C) Schematic of a synapse. Note that the synaptic cleft, vesicles, etc. are not drawn to scale (Rizzoli and Betz, 2005).

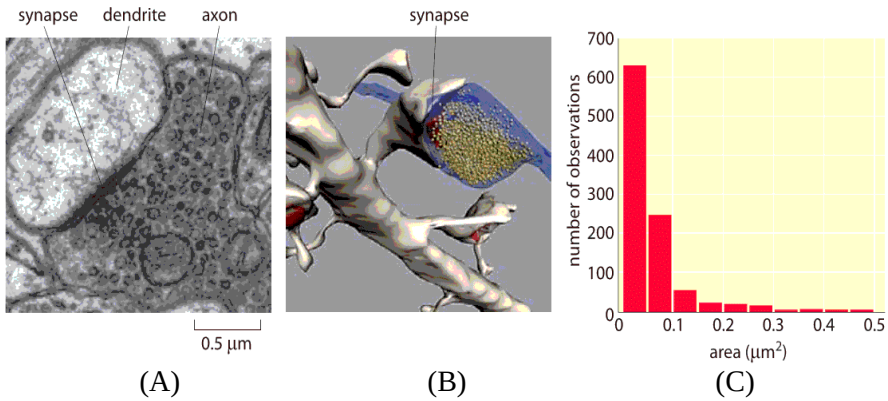
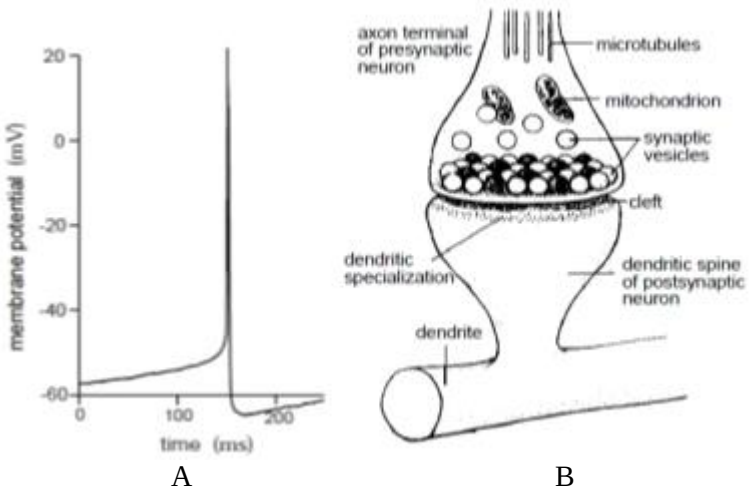


Figure 6. Size of synapses in the brain. (A) Elctron microscopy image of a synapse between an axon and a dendrite. (B) Reconstruction of a synapse-like that shown in (A) illustrating the synaptic vesicles. (C) Distribution of synapse sizes as measured using electron microscopy. (Figures courtesy of Linnaea Ostroff) (Cell Biology by the Numbers, How big is synapse?, 2017).

Synapses between neuron cells make possible to propagation of information from one neuron to the next one. As seen in Figures 5 and 6, the synapse consists of a presynaptic and a postsynaptic terminal with a synaptic cleft between them. As we said, data transmission between the neuron cells is partly electrical and partly chemical. An action potential when crossing between the neurons, it causes a short-term change in the transmembrane potential - from its highly negative resting value it moves to a nearly equal positive potential. When the action potential reaches the synapse, it initiates the vesicle fusion and subsequent release of neurotransmitters induces channel gating in the neighboring cell (Rizzoli and Betz, 2005).Under the resting state, the potential inside the polarized neuron cell membrane is -70 mV approximately. Potential of neuron membrane whenever the reaches to -50 mV fire action get occur, Figure 7.



A B

Figure 7. A) An action potential recorded intracellularly from a cultured rat neocortical pyramidal cell. B) Diagram of a synapse. The axon terminal or bouton is at the end of the axonal branch seen entering from the top of the figure. It is filled with synaptic vesicles containing the neurotransmitter that is released when an action potential arrives from the presynaptic neuron. Transmitter crosses the synaptic cleft and binds to receptors on the dendritic spine, a roughly 1 μm long process extending from the dendrite of the postsynaptic neuron. Excitatory synapses onto cortical pyramidal cells form on dendritic spines as shown here. Other synapses form directly on the dendrites, axon, or soma of the postsynaptic neuron (Dayan and Abbott, 2000).

INTEGRATE AND FIRE NEURON MODEL

Here, the neuron cell is described with its biophysical features. A typical neuron cell fire an action potential when the membrane potential reaches the threshold level, like -50 to -55 mV. After this, fire mechanism the membrane potential follows the circle of depolarization and hyper-polarization loops in a routine way relative to the threshold value. According to well-known voltage-dependent K+ and Na+ channels mechanism, an action potential can occur whenever the membrane potential of the model neuron reaches to threshold value Vth. After the firing ceremony, the potential returns to a value 202V reset.

Integrate and fire neuron model is the most known computational neuroscience model for both cellular and neural network studies. This mathematical model is presented in the year by Louis Lapicque (1907). The membrane potential of a neuron is described in terms of the synaptic input and the injected current data by integrate and fire neuron model. Regardless of its old style and simplicity, the model of integrate and fire neuron model is the most useful and simple description of the neuronal activity. For understanding the whole mechanism, it is necessary to know Nerst equations for channel gates and some simple biophysical conductance and potential laws for small tubes and

molecules. In this chapter, the basic model equations for integrate and fire neuron model will be mentioned.

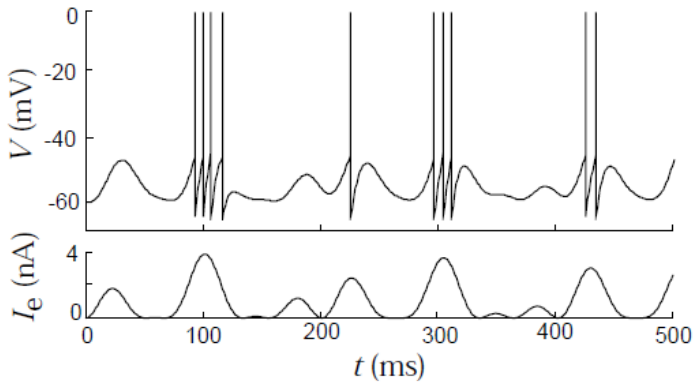


Figure 8. A passive integrate-and-fire model driven by a time-varying electrode current. The upper trace is the membrane potential and the bottom trace the driving current. The action potentials in this figure are simply pasted onto the membrane potential trajectory whenever it reaches the threshold value. The parameters of the model EL = Vreset = −65 mV, Vth = −50 mV, τm = 10 ms, and Rm = 10 MΩ (Catani and Sandrone, 2015).

The all modeling mechanism is about the threshold and subthreshold membrane potential values that made a model of integrate and fire model to the simplest model between all of the models. In the simplest model of integrate and fire neuron models, all active membrane conductance are ignored, including, for the moment, synaptic inputs, and the entire membrane conductance is modeled as a single passive leakage term. This model version is called the passive or leaky integrate and fire neuron model. Resting membrane potential values can have little fluctuations in this model. Neuronal conductance is approximately steady, and the passive integrate and fire neuron model assumes that this constancy holds over the all subthreshold range. This approximation can be reasonable for some neurons and others it is not. The model neuron behaves like an electric circuit established by resistors and capacitors, and the membrane potential is determined by an equation with (Hubel and Wiesel, 1962).

$$\text{im} = \bar{g}(\text{VL-EL}) \tag{1}$$

$$\text{cm}\frac{dV}{dt} = -\bar{g}(\text{V} - \text{EL}) + IeA \tag{2}$$

It is convenient to multiply equation 2 by the specific membrane resistance rm, which in this case is given by rm = 1/ḡL10. This cancels the factor of ḡL on the right side of the equation and leaves a factor cmrm = τm on the left side, where τm is the membrane time constant of the neuron. The electrode current ends up being multiplied by rm/A which is the total membrane resistance Rm. (Catani and Sandrone, 2015) Thus, the basic equation of the passive integrate and fire passive integrate and fire models is

$$\tau m = EL - V + RmIe \quad (3)$$

Whenever the V value s reaches the threshold levels V^{th}, an action potential is fired and the potential is reset to Vreset, by the rules of equation 3. Equation 3 demonstrates that when Ie = 0, the membrane potential relaxes exponentially with time constant τm to V = EL. Thus, EL is the resting potential of the model cell (Dayan et al., 2000).

The firing rate of an integrate-and-fire model in response to a constant injected current can be computed analytically14. When Ie is independent of time, the subthreshold potential V(t) can easily be computed by solving equation 2 and is

$$V(t) = EL + RmIe + (V(0) - EL - RmIe)\exp(-t/\tau m) \quad (4)$$

where V(0) is the value of V at time t = 0.

CONCLUSION

There are more detailed models of neuron with simple equations. In this chapter, we only could cover a part of models that attempt to explain the human brain in the division of neuron. Understanding the human brain will take many years in our century. In this effort, mathematics has been the key instrument to understand the universe and understanding the brain is no different. It is used as the basis of many models and projects varying from Lapicque's integrate and fire neuron model to Human Brain Project. We hope other fields of science will support mathematics to figure out the whole mechanism of the human brain in this century. In this sense, such contemporary developments in the field, like the BigBrain Project, is quite promising as they are capable of assisting us in our pursuit of understanding the brain and eventually the universe.

REFERENCES

Brea, J., and Gerstner, W., (2016). Does computational neuroscience need new synaptic learning paradigms? *Current Opinion in Behavioral Sciences* pp: 61-66.

Bressan, B., and Blackwell, W., (2014). *From Physics to Daily Life: Applications in Biology, Medicine, and Healthcare*, First Edition, CERN, Geneva, Switzerland. 179-191pp.

Calimera, A., Macii, E., and Poncino, M., (2014). The Human Brain Project and neuromorphic computing. *Funct. Neurol*. (2013) Jul-Sep; 28(3):191-6. doi: 10.11138/FNeur/2013.28.3.191.

Catani, M., Sandrone, S., and Brain, Renaissance. (2015) *From Vesalius to Modern Neuroscience*. Oxford University Press, 152-157 pp.

Cell Biology by the Numbers. (2017). *How big is synapse?* Available from http://book. bionumbers.org/how-big-is-a-synapse/Accessdate:03.12. 2017.

David, M., and Tomaso, P., (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, 15:470-488.

Dayan, P., and Abbott, L. F., (2000). *Theoretical Neuroscience, Computational and Mathematical Modeling of Neural Systems*, The MIT Press Cambridge, London pp: 1-42, 156:199 (Dayan et al. 2000).

Hubel, D. H., and Wiesel, T. N., (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." *J. Physiol*. London, (160) (1): 106-54pp.

Katrin, Amunts., Claude, L., Louis, B., and Hartmut, M., Timo, D., Marc-Étienne, R., Sebastian, B., Pierre-Louis, Bazin., Lindsay, B., Lewis, A. M. O., Nadim, J. S., Thomas, L., Karl, Z., and Alan, C. E., (2013). BigBrain: An Ultrahigh-Resolution 3D Human Brain Model. *Science:* Vol. 340, Issue 6139, pp. 1472-1475.

Kesner, R. P., and Rolls, E. T., (2015). A computational theory of hippocampal function, and tests of the theory: *New developments Neuroscience and Bio-behavioral Reviews* pp: 92-147.

Lapicque, L., (1907). "Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation." ["Quantitative research on electrical excitation of nerves treated as a polarization"] *J. Physiol. Pathol. Gen*. 9: 620-635 (Lapicque 1907).

Nature News. (2017). *Whole human brain mapped in 3D*, Available from https://www. nature.com/news/whole-human-brain-mapped-in-3d-1.13245. Access date: 03.23. 2017.

Report, Human Brain Project. (2012). Available from https://www.humanbrainproject.eu/ documents/ 10180/ 17648/ TheHBPReport_ LR.pdf/ 18e5747e-10af-4bec-9806-d03aead57655. Access date: 03.25.2017.

Rewriting, Life. (2017). *Detailed 3-D Atlas of a Human Brain (2013)*, Available from https://www.technologyreview.com/s/516196/a-detailed-3-d-atlas-of-a-human-brain. Access date: 03.25.2017.

Rizzoli, S. O., and Betz, W. J. (2005). *Nat. Rev. Neurosci.*, pp: 6:57.

Wilfrad, R., (1989). Cable theory. In: Koch C., Segev I. (eds.) *Methods in neuronal modelling: from synapses to networks.* 1st edn. MIT press, Cambridge, MA, pp 9-62.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 11

BRAIN TUMORS AND FUTURE THERAPEUTIC OPTIONS

T. T. Dündar[1,*], İ. Doğan[2] and A. S. Mutluer[3]
[1]Bezmialem Vakif University, Medicine Faculty, Department of Neurosurgery, Istanbul, Turkey
[2]Ankara University, Medicine Faculty, Department of Neurosurgery, Ankara, Turkey
[3]Bezmialem Vakif University, Medicine Faculty, Istanbul, Turkey

ABSTRACT

Brain tumors are still a *devastating* and highly fatal disease in both children and adults. Brain tumors consist of primary and metastatic brain tumors. Glioblastoma (GBM; WHO grade IV astrocytoma) is the most aggressive and the most common primary brain tumor that is resistant to treatment. Furthermore, glioblastoma is heterogeneous and may have significant different vascularization. Standard treatment options include surgical resection, radiotherapy and chemotherapy. Despite all effective treatment combinations, the average survival time is under 15 months. Thus, more effective treatment targets are needed. The terms glioblastoma multiforme stem cell, cancer stem cell and brain tumor stem cell are used interchangeably. These cells are responsible for resistance to both radiotherapy and chemotherapy, tumor recurrence and the heterogenous tumor structure. In the near future, genetical, epigenetical and molecular factors concerning tumor stem cell, its microenvironmental atmosphere or the immune system response to these will be constituting the larger part of brain tumor treatment aims. In this chapter, we aimed at presenting current information about brain tumor stem cells as well as new treatment options.

Keywords: neural stem cells, glioblastoma, glial tumor stem cells, brain tumor stem cell

* Corresponding Author's E-mail: drtolgadundar@hotmail.com.

1. INTRODUCTION

The concept of brain tumor includes primary and metastatic tumors of intracranial area. Brain tumors are still a *devastating* and highly fatal disease in both children and adults. Metastases commonly arise within the brain parenchyma, where tumor cells travel from another part of the body. Metastases are the most frequent neoplasm in brain tumors, occurring approximately ten times more often than primary brain tumors (Singh et al., 2017). Primary brain tumors, known as gliomas are brain tumors that derive from normal stromal (glial) cells of the brain. For example, astrocytomas derive from astrocytes, oligodendrogliomas from oligodendrocytes and ependymomas from ependymal cells (Singh et al., 2017). There is also a glioma grading system that distinguishes, astrocytomas, by four World Health Organization (WHO) grades (I, II, III, and IV); and oligodendrogliomas and oligoastrocytomas, by two grades (II and III) (Louis et al., 2016).

Among the these tumors, Glioblastoma (GBM; World Health Organization grade IV astrocytoma is the most common primary brain tumour, and the most aggressive in nature, with an annual incidence of 5.26 per 100,000 people. Gold-standard treatment of GBM includes tumour resection, radiotherapy and chemotherapy. Despite the currently available treatment options, most patients succumb to the disease some 12 months to 18 months after diagnosis, while for recurrent GBM, life expectancy is ~6 months. Only about 30% of patients achieve 2- year survival and fewer than 10% survive more than 3 years. (1, 2)Even with advancements in intra-operative brain imaging, which have assisted with gross resections, new chemotherapies, and more focused radiation treatments, the prognosis for GBM remains extremely poor with survival rates of 33% for one year and a five year survival rate of 5% (Gardeck et al., 2017).

Despite extensive molecular and genetic analyses of gliablastoma there is still a standing need for more effective therapies. New options of treatment of brain tumors are explained by the molecular biology of tumor stem cells and their signaling pathways. The first tumor stem cell theory was proposed by Virchow 150 years ago. According to Virchow, tumors might orginate from immature cells. In the following years, Cohnheim and Durante supported the stem cell concept that adult tissues still contain dormant immature cells which could activate the specific conditions to give rise to a tumor. 70 years after these developments, Makino introduced the nomenclature of “tumor stem cells” defining them as “a small subpopulation of cells”. At last, GSC were identified directly from patient-derived tumors in brain by 2000s. Today, it is well-known that the poor prognosis and recurrence in brain tumors are mainly due to intratumoral tumor stem cells. This has to do with the resistance of stem cells both to chemotherapy and to radiotherapy.

Temozolomide is rapidly converted into its reactive format, 5-3-(methyl)-1-(triazen-1-yl) imidazole-4-carboxamide, at physiological pH, causing DNA damage through

methylation of the O6-position of guanines, blocking DNA replication and inducing the death of tumor cells (Glaser et al., 2017). Due to the fact that currently available therapies result in such a short survival span urges extensive exploration of alternative treatments (Jovčevska et al., 2017).

CSCs have been isolated from cancer to be analyzed and later used to screen for stem cell-specific biomarkers in tumor cells, particularly surface biomarkers. Cell-surface markers are generally cell membrane-surface antigens to which antitumor drugs can easily bind, consequently increasing the therapeutic efficiency of the drug. Membrane surface markers are more meaningful than nuclear or cytoplasmic antigens in targeted tumor therapy (Glaser et al., 2017). Therefore, an understanding of cancer development and progression requires the elucidation of collective properties of cells within a tissue and their interaction with the microenvironment.

2. GLIABLASTOME STEM CELL MARKERS AND PATHWAYS

Gliablastome Stem Cells have been characterized by several markers such as Nestin, CD133, Musashi-1, and CD15 etc. Among these, *CD133* (PROM1, AC133, and Prominin-1): A pentaspan membrane glycoprotein, CD133 has been used as a biomarker for glioblastoma stem cells. CD133+ cells are capable of self-renewal and recapitulation.

Nestin: Nestin is an intermediate filament (IF) protein. Nestin's expression correlates highly with ―stemness‖. Nestin plays roles in the organization of the cytoskeleton, cell signaling, organogenesis, cell metabolism, proliferation and migration.

Integrin-a6: Integrin-a6 is a member of the heterodimer integrin family and is a laminin member of the extracellular matrix protein family. Integrin-a6 can be used as a marker of neural stem cells and the expression of integrin-a6 can be used to detect the tumorigenic potential of normal neural stem cells (Glaser et al., 2017).

Musashi-1: Musashi-1 is an RNA-binding protein which inhibits the translation of mRNAs believed to be involved in the differentiation of NSCs and increasing evidence points towards its involvement in tumorigenesis.

CXCR4: Neurons express some receptors that regulate neuronal signaling and survival, including CXCR4. The CXCR4 signaling pathway is the key regulator of many essential biological processes, such as stem cell motility, differentiation switch, angiogenesis, apoptosis, and lymphocyte homing.

CD44: CD44 is one of the most prominent stem cell markers and a well-known cancer stem cell marker. It represents the polymorphic family of cell adhesion molecules that seem to play the key role in the mechanism of tumor invasion and metastasis.

CD15, CD44, CD90, Sox 2 which are also cell surface markers that are used to isolate from GSCs. Numerous aberrantly expressed genes and signaling pathways have been identified as important in GBM biology, such as *EGFR*, *INK4a/ARF*, *NF1*,

PDGFRA/IDH1, *P53*, *IDH1*, *RB1*, and *ERBB2*, PTEN/AKT/PI3K, Sonic Hedgehog/GLI, Notch, P53/MDM2/ MDM4/p14ARF, and p16INK4a/CDK4/RB1 pathways (Dimov et al., 2011).

Sonic Hedgehog Pathway, Notch Pathway, Wnt/beta-Catenin pathway and their receptor activities play an important role in the regulation of GSC's and have also been found to be altered or overexpressed in GSCs. Normal stem cells exhibit similar markers too. These three pathways appear to be crucial GSCs factors and thus attractive therapeutic targets.

2.1 Notch Signaling Pathway

The Notch pathway seems to play a critical role in linking angiogenesis, response to hypoxia, and cancer stem cell self-renewal, presenting as a promising therapeutic target. Notch signaling also regulates stem cells. This pathway enables homeostasis in the adult tissue, including the brain tissue. Notch signaling pathway is very active in brain tumors too.

2.2 Hedgehog Signaling Pathway

Hedgehog secreted signaling proteins are critical for embryonic tissue development and postnatal tissue homeostasis. Hedgehog pathway appears to be activated in a small population of cells that have been proposed to have stem or progenitor-like features. Moreover, the Hedgehog pathway is frequently activated with IDH *(+) type.*

Aberrant activation of the several pathways of stem cells has been implicated in the growth of malignancies in a role that is largely attributed to the action of the pathway on stem or progenitor cells.

3. IMMUNOTHERAPY

Immunotherapy for gliablastome includes signalling pathways and activities of niches as well as the activation of stem cell's autophagy. Immune and viral therapy based on stem cells that include activated dendritic cells, adoptive T cells, cytokines, and vaccines or viral vectors are used for this purpose. Immune system is excited toward stem cell or a small *structure* of the stem cell's *plasma membrane.*

Active and passive immunotherapies are used for this purpose. Both of them can either be non-specific or specific. Specific immunotherapies are the ones that induce a

particular immune response with a specific antigen, whereas non-specific immunotherapies are those that induce a general, non-specific immune response. Active immunotherapy means that the the host's immune system is activated with part of tumor stem cell such as through cancer vaccines, cellular immunotherapy**.** In passive immunotherapy, the host's immune response is initiated by external antibodies or other immune components such as monoclonal antibodies, Cytokines. These are usually generated in the laboratory.

Another modality is gene therapy. The aim of gene therapy is to provoke an immune response to destroy tumor cells. In the context of gene therapy, the antitumor response depends on the cell destruction by the virus and the induction of a systemic immune response by the local environment.

Non-viral Vectors and viral vectors can be used for this aim. Non-viral gene transfection and expression techniques applied to tumor stem cells may be generally categorized as physical (e.g., electroporation) and chemical (e.g., lipofection). Genes of many different sizes have been tried to be transferred to tumor stem cells.

These techniques often cannot achieve satisfactory transfection efficiencies. As for Retroviral and Lentiviral Vectors, the common types of vectors used for tumor stem cell engineering are single – stranded RNA retroviruses and lentiviruses. Vectors try to insert viral genes into the host DNA, theoretically enabling long-term expression of the inserted genes. Other types of viruses are adeno and herpes simplex viruses. Either the adenoviral vector containing thymidine kinase gene (AdV-tk), or herpes simplex thymidine kinase gene (HSV-tk) vectors may contribute to the therapy by activating apoptosis and caspase systems within the tumor cells. But, studies are limited in this area (Gardeck et al., 2017, Mitsui et al., 2017).

4. MATHEMATICAL MODELS OF BRAIN TUMOR STEM CELLS

Mathematical models can prove useful for deriving a detailed understanding about brain tumor. Quantitative descriptions of brain tumors with stem cell's *biological* mechanisms can supervise new theurapatic options and may present integrated empirical evidences. Tumoral progression must make quantitative description of treatment's dynamics and resistance. Moreover, mathematical modelling can be testable theories on several factors. These models may make investigations quantitative and predictive, and hypotheses falsifiable. There are limited studies about mathematical identifications of brain tumor biology, especially about brain tumor stem cells (Yan et al., 2017, Altrock et al., 2015, Bogdańska et al., 2017). Among these, Yan et al. bring 3D mathematical modeling of glioblastoma to forefront (Yan et al., 2017). Yan et al. developed a 3D mathematical model which contributes to the growth of gliablastoma cells and its response to cancer therapies. This model starts with gliablastoma stem cells (GSC, $\partial_{GSC)}$,

committed progenitor GBM cells (GCP, ∂_{GCP}), terminally differentiated GBM cells (GTD, $\partial_{GTD)}$, vascular endothelial cells generated by transdifferentiation of GSC (GEC, ∂_{GEC}), dead GBM cells (DC, ∂_{D}), fraction of total tumor cells (∂_{T}). The formula is:

$$\partial_{T} = \partial_{GSC} + \partial_{GCP} + \partial_{GTD} + \partial_{GEC} + \partial_{D.}$$

This *representation* is quite simple yet so important. At last, their study shows us that a well- prepared mathematical model can be used for clinical analyses, just like in clinical trials or statistical analyses. Finally, mathematical models of brain tumors will propose new experiments, suggest different treatment modalities and alter risk prognoses with new *developments* of molecular and genetic studies.

5. NANOTECHNOLOGY

The development of new technologies based on nanometersized particles (nanotechnology) for cancer treatment has been extensively investigated in the last decade and this approach shows potential benefit for glioma diagnosis and treatment. Few clinical trials using nanotherapies to target gliablastoma have been conducted. This section aims at focusing on nanometersized particles at gliablastoma and giving brief information about nanotechnology.

An alternative approach for discovering proteins with specificity for GBM is offered by using nanobodies, which can both include stem cell fragments and be used as a means for the transportation of these to the target tissue. Lately, nanoparticles have been much widely studied, because of their high drug-loading capacity and protection against chemical and enzymatic degradation. Nanoparticles include enormous medical potential compared with conventional drug delivery methods (Glaser et al., 2017).

This novel approach novel to the therapy of brain tumors require the involvement of a specific ligand (e.g., an antibody or antibody fragment, synthetic peptide, natural ligand) of stem cell or the stem sell's microenvironment, which has affinity for an endocytic receptor expressed on the endothelial cell surface, to the chemotherapeutic drug or to a drug-loaded nanocarrier. Binding to the targeted receptor induces intracellular signaling cascades, mediating invagination and formation of membranebound vesicles in the cell interior (Glaser et al., 2017).

Another advance of nanotechnology that can be used is the engineering of the cell genome. For this purpose, positively charged and degradable polymers, including chitosan, poly (beta-amino esters), poly (amidoamines), and many other cationic polymers have been used, because of their cationic nature, which allows complexation with negatively charged molecules like DNA or RNA. This activity may drive the stem cell's life circle.

Additionally, nature of nanoparticles offers benefits for diagnosis of brain tumors. Inorganic nanoparticles are better to be applied for imaging and drug delivery purposes, because their synthesis is easily tunable and reproducible. For example, injectable superparamagnetic iron oxide nanoparticles can be used as contrast agents for magnetic resonance imaging, and gold nanoparticles can be used to carry a conjugated drug.

CONCLUSION

Vitality is of great importance in order to understand the fundamental mechanisms driving stem cells. If we can manage to figure out the life cycle mechanisms of stem cells with or without biological nanotechnology, then we can control the growth of cancer and of many other diseases.

REFERENCES

Altrock, PM., Liu, LL., and Michor, F. (2015).The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer* 15:730-45.

Bogdańska, MU., Bodnar, M., Piotrowska, MJ., Murek, M., Schucht, P., Beck, J., Martínez-González, A., and Pérez-García, VM. (2017). A mathematical model describes the malignant transformation of low grade gliomas: Prognostic implications. *PLoS One*. Aug 1; 12(8):e0179999. Doi: 10.1371/journal.pone. 0179999. eCollection 2017.

Dimov, I., Tasić-Dimov, D., Conić, I., and Stefanovic, V. (2011). Glioblastoma multiforme stem cells. *ScientificWorld Journal*. Apr 19; 11: 930-958.

Gardeck, AM., Sheehan, J., and Low, WC. (2017). Immune and viral therapies for malignant primary brain tumors. *Expert Opin. Biol. Ther.* Apr; 17(4):457-474.

Glaser, T., Han, I., Wu, L., and Zeng, X. (2017). Targeted Nanotechnology in Glioblastoma Multiforme. *Front Pharmacol.* Mar 31; 8:166.

Jovčevska, I., Zupanec, N., Urlep, Ž., Vranič, A., Matos, B., Stokin, CL., Muyldermans, S., Myers, MP., Buzdin, AA., Petrov, I., and Komel, R. (2017). Differentially expressed proteins in glioblastoma multiforme identified with a nanobody-based anti-proteome approach and confirmed by OncoFinder as possible tumor-class predictive biomarker candidates. *Oncotarget.* Jul 4; 8(27):44141-44158.

Louis, DN., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, WK., Ohgaki, H., Wiestler, OD., Kleihues, P., and Ellison, DW. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol*. 131(6):803-820.

Mitsui, K., Ide, K., Takahashi, T., and Kosai, KI. (2017). Viral Vector-Based Innovative Approaches to Directly Abolishing Tumorigenic Pluripotent Stem Cells for Safer Regenerative Medicine. *Mol. Ther. Methods Clin. Dev*. Mar 18; 5:51-58.

Razpotnik, R., Novak, N., Čurin Šerbec, V., and Rajcevic, U. (2017).Targeting Malignant Brain Tumors with Antibodies. *Front Immunol*. Sep 250; 8:1181.

Singh, M., Bakhshinyan, D., Venugopal, C., and Singh, SK. (2017). Preclinical Modeling and Therapeutic Avenues for Cancer Metastasis to the Central Nervous System. *Front Oncol*. 19; 7:220. Doi: 10.3389/fonc.2017.00220. eCollection 2017.

Wang, J., Ma, Y., and Cooper, MK. (2013). Cancer stem cells in glioma: challenges and opportunities. *Transl. Cancer Res.* 1; 2(5):429-441.

Yan, H., Romero-López, M., Benitez, LI., Di, K., Frieboes, HB., Hughes, CCW., Bota, DA., and Lowengrub, JS. (2017). 3D Mathematical Modeling of Glioblastoma Suggests That Transdifferentiated Vascular Endothelial Cells Mediate Resistance to Current Standard-of-Care Therapy. *Cancer Res.* 1; 77(15):4171-4184.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 12

A NOVEL APPROACH TO NEURO-FUZZY SYSTEM ALGORITHMS: MODELLING FOR THE CLASSIFICATION OF MS SUBGROUPS IN MRI

Y. Karaca*
University of Massachusetts Medical School, Worcester, MA, US

ABSTRACT

Neuro-fuzzy systems can find the parameters related to a fuzzy system by employing approximation techniques from neural networks. They are preferred to be used owing to their human-like reasoning since these systems enable the modelling of complex processes which include human experts and non-expert humans. Given that, these systems come to the fore as available tools to create sound and robust models related to the modelled process. Nero-fuzzy systems have been extensively used in many fields to solve various problems, such as regression, pattern recognition and other types of estimation. Accordingly, the aim of this study is to develop a clinical decision support system for the classification of Multiple Sclerosis (MS) subgroups based on the data obtained from Magnetic Resonance Imaging (MRI) and Expanded Disability Status Scale (EDSS). The MS dataset in this study includes a total of 120 patients as well as 19 healthy individuals (who make up the control group, not afflicted by MS). Unlike the earlier works in the literature, in this study, artificial Neuro Fuzzy Inference System (ANFIS) algorithms were used, and accordingly fuzzy rules related to MS subgroups were obtained for the classification of patients and healthy individuals. The following steps were applied for this purpose: (i) K-means clustering algorithm was applied to the MS dataset. (ii) the matrix of the centroid cluster was obtained for the MS subgroups

* Corresponding Author's E-mail: yeliz.karaca@ieee.org, yeliz.karaca@umassmemorial.org.

(RRMS, SPMS, PPMS, and Healthy individuals). (iii) Fuzzy Inference System Algorithm (FIS) was applied on the matrix of the centroid cluster and as a result of this application, fuzzy rules were obtained. Fuzzy rules were used for the applications of ANFIS Back Propagation algorithm and ANFIS Hybrid-based algorithm. (iv) Classification was done for the MS subgroups and healthy individuals. Consequently, this study has aimed at providing new direction to those involved in medical field, applied sciences and engineered fields to attain higher levels of accuracy.

Keywords: neuro-fuzzy systems, ANFIS, K-means, learning, neural network, MS subgroups, clinical decision support system, fuzzy rules

Key phrases: Neuro-fuzzy Systems, human-like reasoning, Fuzzy rules, ANFIS Back Propagation algorithm, health decision-making, Hybrid-based algorithms, classification accuracy performance by ANFIS Back Propagation algorithm and Hybrid-based algorithm for MS subgroups

1. INTRODUCTION

Fuzzy neural networks, also known as neuro-fuzzy systems, are learning machines which find the parameters of a fuzzy system. To do this, a neuro-fuzzy system exploits the techniques of approximation from the neural networks. Regarding the relation of fuzzy systems with neural networks, it can be noted that both neural networks and fuzzy systems share similarities in that they can be used to solve a problem. When the certain problem is expressed by plenty number of observed cases, neural networks can come into play and have a significant role in the resolution of the problem. A neuro-fuzzy system is reliant on an underlying fuzzy system, and this underlying system is trained by means of data driven learning method (Ahtiwash et al., 2002). The learning procedure is controlled to assure the the underlying fuzzy system's semantic properties. The use of neuro-fuzzy systems is extensive in many fields, including but are not limited to medicine, biology, applied sciences, biomedical engineering, computational intelligence and design for the identification, prediction and control of systems. One advantage related to fuzzy systems is that they have human-like reasoning (Balbinot and Favieiro, 2013).

Multiple Sclerosis (MS) is a disease that harms the myelin which is the protective tissue that surrounds the nerve fibres conveying messages in the spinal cord and brain (Haider et al., 2016). MS occurs due to the physical impairment of the myelin layer (Harmouche et al., 2015), (Sand, 2015), (Hagens et al., 2016).

MS patients who have a benign type can remain intact relatively for long years following the first attack. Benign MS is often associated with less severe symptoms (Giacomini, 2016). MS has a number of subgroups. The subgroups used in the MS dataset of this study are as follows: Relapsing Remitting MS (RRMS) demonstrates

relapses that lead to the acute deterioration of neurological functions following the partial or full recovery regardless of the fact that how far the disorder has advanced (Giacomini, 2016). Another subgroup is Secondary progressive Multiple Sclerosis (SPMS). The attacks seen in this subgroup may manifest some improvements. (Hagens et al., 2016). In the early phase, it may be possible to see a period with a full recovery after the attacks. Primary progressive Multiple Sclerosis (PPMS) is another group in MS whose attacks advance in a constant and slow way following the onset of the disease. There are no clear recoveries or relapses. This is the severe phase of MS which may progress slowly or rapidly and the possibility of recovery is low. Nevertheless, patients may experience some slight improvements (Karaca, et al., 2017), (Selvi et al., 2015).

Machine learning approaches are used to assist medical and clinical tools for the diagnosis of MS subgroups. Many studies exist in the literature aimed at classifying MS subgroups and healthy individuals as well as providing computer-based decision mechanism in the relevant field. (Stamile et al., 2015) performed a classification regarding three different clinical types with 65 MS patients. They configured the connection features in complex structure by means of the Graph theory approach. (Karaca et al., 2017) worked on the MRI and EDSS scores that belonged to a total of 139 individuals, among whom 120 were afflicted with a subgroup of MS (76 RRMS, 38 SPMS and 6 PPMS patients) and 19 were healthy individuals. CIKA algorithm as proposed by the authors yielded higher accuracy than the other mostly used non-linear Machine Learning classifiers such as artificial neural network, single kernel as well as other machine learning methods. In another relevant study, (Zhang et al., 2015) worked on the MRI that belonged to three different datasets. The data of their study were obtained from the patients as well as healthy individuals. They carried out a study on SWT coefficient values, applying the Principle Component Analysis feature selection method to the dataset obtained. In addition, the study by (Patel et al., 2009) addresses the importance of the representation of all kinds of knowledge and high-level systems modelling regarding basic Artificial Intelligence in medicine research. The authors also state that if exploitation of knowledge is handled effectively in building decision making tools and extracting information, this will prove to be very important in the future.

There exist studies in the literature which made use of Fuzzy methods. The study by (Garro et al., 2016) aimed at proposing a methodology for the classification of DNA microarray. Their proposed method performed a feature selection process which was based on a swarm intelligence algorithm. Their purpose was to find a subset of genes that best describe a disease. They also tested the relevance of genes so that the correct classifications of disease samples. A study by (Mathur et al., 2016) proposed to implement ANFIS for the prediction of the in-socket residual limb temperature. They used Gaussian processes for machine learning. They also compared the actual and predicted data, and the results of that study show that both modelling techniques have

comparable performance metrics and therefore they can be utilized in an efficient way for a non-invasive temperature monitoring.

Some of the relevant studies that used Fuzzy methods are as follows: (Kumar et al., 2016) carried out a segmentation of the MRI, to which the K-means algorithm was applied. They applied a logical AND gate for their work. They compared the accuracy rates for K-means algorithm and that of the Fuzzy C Means algorithm. A recent study by (Kadhim, 2018) is concerned with back pain. A decision support system for intelligent diagnosis of back pain was proposed by the author. In the study, he used a fuzzy-neuro technique to assist physicians in their diagnoses. In a relevant study by (Karaca et al., 2018) K- means and fuzzy C means (FCM) algorithms were applied on the stroke dataset. 2D multifractal denoising techniques (Bayesian (mBd), Nonlinear (mNold) and Pumping (mPumpD)) were used for the classification purposes.

In this study, MRI data were used for three different years of each MS patient (there were at least three years between the first and second MRI scans and there were maximum 8 or 10 years between the second and third and MRI scans. Apart from the MRI data, the patients' EDSS scores were also used (regarding the disability of the patients belonging to the aforementioned duration in years). Artificial rules were generated in the machine environment through the processing of the data pixels obtained from different parts of the brain, which are the brain stem, corpus colosseum, periventricular and upper cervical regions.

As for the methodological aspects, this study made use of artificial Neuro Fuzzy Inference System (ANFIS) algorithms. Different solution methods can be applied in a fuzzy rule based system. The most important of these methods are Mamdani and Sugeno models. In this study, the ANFIS structure uses the Sugeno Fuzzy model. Accordingly, fuzzy rules related to MS subgroups were obtained for the classification of patients and healthy individuals. In this regard, this study is different from the earlier works in the literature (Mathur et al., 2016), (Kumar et al., 2016), (Kadhim, 2018), (Karaca et al., 2018). For the purpose of the study, the following steps specified were employed: Firstly, K-means clustering algorithm was applied to the MS dataset (with MS subgroups and healthy individuals whose data have been based on the MRI and EDSS scores). Secondly, a centroid cluster centres matrix was obtained for the related MS subgroups which are RRMS, SPMS, PPMS as well as Healthy individuals. Next, Fuzzy Inference System Algorithm (FIS) was applied on the centroid cluster centres matrix. Finally, fuzzy rules were obtained and these fuzzy rules were used for the applications of ANFIS Back Propagation algorithm and Hybrid-based algorithm. Consequently, the classification for the MS subgroups and healthy individuals was carried out.

To recap, this study revealed a classification performance which is higher when compared to other studies in the literature concerning the classification of MS subgroups. This is one contribution of this study. Another contribution is that a highly comprehensive MS dataset was used. Realistic fuzzy rules were inferred after the

calculation of centroid values for disease subgroups and healthy individuals were carried out from the MS dataset. The aim is to infer the fuzzy rules through the ANFIS back propagation and ANFIS hybrid based learning algorithms for the classification of MS subgroups and healthy individuals.

Centroid matrix was obtained from the MS dataset. This centroid matrix was utilized as the initial value so as to extract rules in the fuzzy system. The ANFIS back propagation algorithm and the hybrid based learning algorithm were applied on the dataset with the rules obtained. Accordingly, the classification performance of the ANFIS back-propagation algorithm and that of the hybrid-based learning algorithms were compared. ANFIS back propagation and hybrid based learning were chosen because they yielded efficient results since fuzzy rules were obtained therefrom.

The organization of the study is as follows: Section 2 addresses Materials and Methods, Section 3 covers Results and Discussion concerning the study and finally Section 4 is the Conclusion.

2. MATERIALS AND METHODS

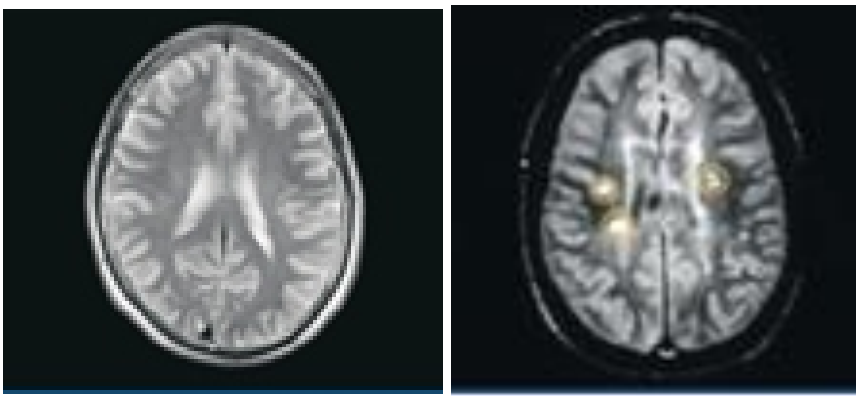
2.1. Patient Details

This study has been conducted based on data obtained from subjects aged 20 – 65 with definite clinical MS diagnosis based on McDonald criteria (Polman et al., 2011), (Karaca et al., 2015). The total number of the MS patients was 120 (RRMS (76), SPMS (38) and PPMS (6)). 19 of the subjects in the dataset were healthy individuals who were not afflicted by MS (as the control group). The dataset was obtained from the patients who had been registered at Hacettepe University Faculty of Medicine, Departments of Neurology and Radiology and Primary Magnetic Resonance Imaging centre (Ankara, Turkey). For the determination of the level of disability in MS patients the Extended Disability Status Scale (EDSS) was employed. The lesions in the three regions of the brain were studied, which were involved in the brainstem, corpus callosum-periventricular, including the upper cervical region. While observing the course of the disease, there were at least three years between the first and second MRI scans, and there was a maximum period of 8 or 10 years between the second and third and MRI scans (Karaca et al., 2015).

2.1.1. Magnetic Resonance Imaging (MRI)

MRI is an important tool for brain and spinal cord images (Karaca et al., 2017), (Karaca et al., 2015). Through strong magnetic fields, it assists the process of revealing the inflamed or injured tissue parts in the central nervous system. (Figure 1(a)), (Figure 1 (b)) presents an MRI for a patient with MS and a healthy individual, respectively. MRIs

were taken on 1,5-tesla MR machines (Magnetom - Siemens Medical Systems, Integra Achieve-Philips or/and GE Healthcare) (Karaca et al., 2017), (Karaca, et al., 2015). The lesions on T2-weighted turbo spin-echo (TSE) sequences were handled in metric units.



(a) Healthy individual MRI (b) MS individual MRI

Figure 1. MRI taken from (a) healthy individual and (b) MS patient (Karaca and Cattani, 2018).

2.1.2. Expanded Disability Status Scale (EDSS)

EDSS is based on the measurements of 8 functional systems in the central nervous system which include voluntary movements, brain stem, vision, brain, cerebellum, sensation, intestines, bladder and exhaustion (Kurtzke, 1983), (Meyer-Moocket et al., 2014).

These systems are scored based on the level of each disability. These scores might vary from 0 to 5 or 6, 0 referring to normal situation and 5 – 6 referring to a very high level of severity. This functional system (FS) adds movement and daily life limitations to system grades and completes 20 steps in EDSS (Kurtzke, 1983), (Meyer-Moocket al., 2014) (for further details see (Karaca et al., 2017), (Karaca and Cattani, 2018)).

2.2. Methods

The main aim of the study is to infer the fuzzy rules through the ANFIS back propagation and hybrid based learning algorithms for the classification of MS subgroups and healthy individuals. Therefore, classification of the MS subgroups and healthy individuals was done by generating fuzzy rules from the comprehensive MS dataset (based on MRI and the EDSS scores of the patients). The method utilized for this purpose has the following steps which are specified below:

(1) K-means clustering algorithm was applied to the MS dataset.
(2) The matrix of the centroid cluster was obtained for the MS subgroups (namely RRMS, SPMS, PPMS) and Healthy individuals (control group).
(3) Fuzzy Inference System Algorithm (FIS) was applied on the centroid cluster centres matrix. Consequently, fuzzy rules were obtained. These fuzzy rules were utilized for the ANFIS Back Propagation algorithm and Hybrid-based algorithm applications.
(4) Classification for the MS subgroups and healthy individuals was performed in terms of accuracy rates.

The experimental results and figures of the study have been obtained through Matlab (The MathWorks, 2018).

2.2.1. K-Means Clustering Algorithm

K-means algorithm method is the simplest unsupervised learning method that solves the clustering problem (Han and Kamber, 2006), (Karaca and Cattani, 2018). The K-Means algorithm is a cyclic partitioning algorithm that continuously updates clusters and continues until the optimal solution is reached. From the partitioning methods, pieces of data are grouped according to the number of pre-given sets ($k < n$) (Xu and Wunsch, 2005). The K-means algorithm starts with the determination of cluster centres. Classification of samples outside the centre is done according to their distances. New cluster centres are determined according to the classification made. Calculations are repeated for new cluster centres as long as the data become in a stable state (Jain, 2010), (Karaca and Cattani, 2018).

$C = \{c_1, c_2, \ldots, c_k\}$ is random points in the k number chosen from among the items or the identification can be made through the average of all the items. The proximity of each data in test cluster to the selected central points is calculated with cosine similarity $X = \{x_1, x_2, \ldots, x_k\}$. Each data is included into the cluster where there is the nearest central point in (Equation 1) (Jain, 2010).

$$\cos(x_i, center(c_k)) = \frac{x_k}{\|x_k\|} \times \frac{center(c_k)}{\|center(c_k)\|} \tag{1}$$

The dimension of the dataset matrix where data for RRMS, SPMS, PPMS, and Healthy individuals kept in this study is 139 x 228. It is represented by (i = {1, 2, 3, ..., n}, j = {1, 2, 3, ..., k}) (n = 139, k = 228).

The central points of the clusters formed are changed with the average values of all the items in that cluster (Equation 2) (Jain, 2010).

$$center(c_k) = \frac{\sum_{i=1}^{n_k}(x_i)}{n_k} \quad (2)$$

2.2.2. Neuro Fuzzy Systems

Neural-fuzzy systems are is a combination of neural networks and fuzzy systems. The combination is in such a way that neural networks or their algorithms are utilized in order to determine the parameters of a fuzzy system. By this, it is meant that the primary goal of a neural-fuzzy approach is to create or enhance a fuzzy system in an automatic through neural network methods (Kruse and Nauck, 1998). Approaches related to neuro-fuzzy systems manifest learning capabilities, which means they utilize algorithms that allow them to determine the parameters from the training data in iterative processes. There are several technical advantages to using neuro fuzzy systems over using classic systems. First of all, there is the ease of development since fuzzy systems yield an intuitive way for the assessment of nonlinear functions. Another benefit is seen as the ease of adaptation. Methodologically, there benefits to use neuro-fuzzy systems in modelling as well. Firstly, fuzzy systems enable the modelling of imprecise yet rational behaviour of human experts. Owing to their properties, neuro-fuzzy systems enable the modelling of complex processes that include both non-expert humans and human experts (Teodorescu and Yamakawa, T, 1996). Hence, nero-fuzzy systems play a guiding role for problems that could not be resolved by human intuition.

In this chapter, in the current system used, blurred numerical information is gathered together with fuzzy verbal information, resulting in a situation reduced to the entrance of the general fuzzy system. Based on this, the present study identifies the behaviour of the system by transforming input variables that constitute inputs to nonlinear MS disease, output variables, namely RRMS, PPMS, SPMS subgroups, and Healthy individuals. It is thus possible to control the system so that the information base is exposed to non-linear transformations to arrive at the estimates of the subset of the desired disease classification.

2.3. Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS is a member of the family of fused neuro fuzzy system, belonging to a family of a hybrid system and including the properties of both fuzzy logic and neural networks. Such fuzzy system is included in a framework that is adaptive in nature. Fuzzy-neural network logic is based on the principle of blending the “learning and finding the best” skills of Artificial Neural Network (ANN) and the decision-making skill of a human as well as providing specialist knowledge of fuzzy logic (Zadeh, 1965), (Takagi and Sugeno,

1985). Fuzzy cluster is identified with the membership value. One element is either a member of a set or not in classical concept of sets. In fuzzy logic, cluster membership value μ ranges between 0 and 1. 0 refers to not belonging to the set and 1 definitely denotes being a member of that set. The membership functions of the set can be defined with standard functions like triangle, trapezoid and gauss curve. They can also be formed in different functions. Modelling with fuzzy logic has certain stages which are called fuzzification, rule base and defuzzification (Zadeh, 1965), (Takagi and Sugeno, 1985). In the fuzzification stage, the membership degrees of inputs for the available fuzzy clusters are identified by forming the membership functions. Using the membership functions and "if-then" rules, Fuzzy Inference Stage includes associating one fuzzy cluster with another one. Defuzzification is the transformation process that translates the output fuzzy value into an exact output value. Fuzzy logic compiles all the relations formed among the input and output fuzzy clusters on the rule basis. By doing so, it is a sum of stages that ensure the system act with single output manner (Tanaka and Sugeno, 1992). Different solution methods can be applied in a fuzzy rule based system. In this study, the ANFIS structure uses the Sugeno Fuzzy model. A mixed learning algorithm has been used in the ANFIS structure where the Curve Gradient and the Least Squares Method (LSM) are used together.

The typical fuzzy rule in Sugeno fuzzy model has the following form:

$$IF \quad x = A \text{ and } y = B \text{ THEN } z = f(x, y) \tag{3}$$

If the A and B input fuzzy sets are $z = f(x, y)$, it is the function that gives the output dependent on *x* and *y*. The end value is found by getting the weighted average over the membership values obtained from the rules of the outputs obtained from fuzzy rules. ANFIS modelling has been used in this study for the updating of parameters that belong to the Neuro-Fuzzy mixed network. According to (Equation 3), *x* and *y* in the study are identified by lesion diameter sizes from the patients' MRI and EDSS scores. The general configuration of the neuro-fuzzy mixed structure that sums up the applications formed in this study is as the one given in (Figure 3). In the fuzzy section of this structure, membership functions are formed with generalized bell-shaped membership function (gbellmf). α and *k* coefficients obtained from membership functions and inputs have been applied as input to the mixed learning algorithm.

As can be seen from (Figure 2), ANFIS structure is defined with 6 layers functionally applicable to the Sugeno fuzzy inference in this study.

Layer 1 denotes the input nodes where each node's input value is transmitted to the other layers. On Layer 2, the fuzzy clusters of each node are denoted. The output of the nodes on this layer denotes the membership degree depending on the input samples and

the membership function used. Membership degrees are obtained from these nodes as (Equation 4).

$$o_i^2 = \mu_A(\text{1st Centroid Cluster})$$
$$o_{i+2}^2 = \mu_{Di} \text{ (4th Centroid Cluster)}, \quad i = 1, 2, 3, 4 \tag{4}$$

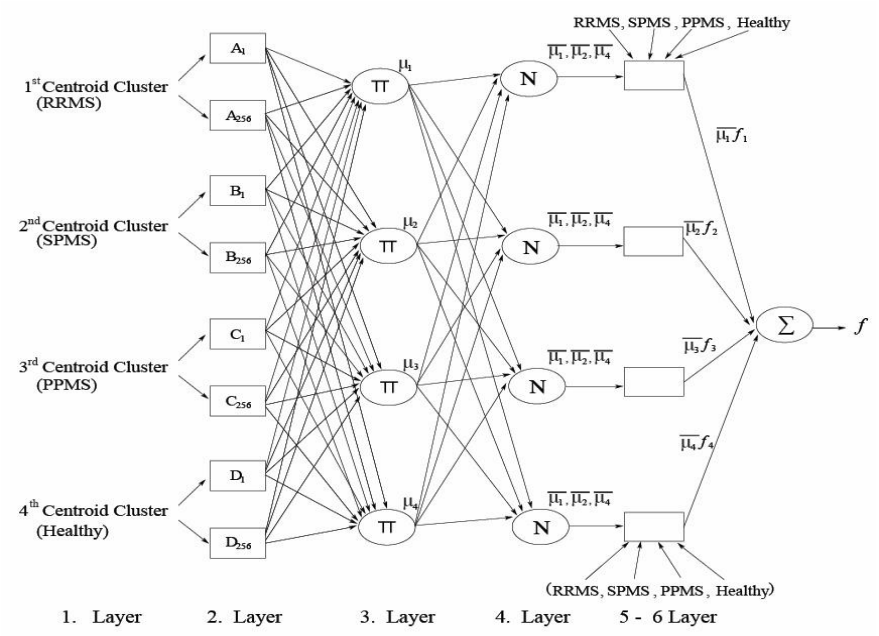


Figure 2. Sugeno-type fuzzy inference equivalent ANFIS architecture (Elmas, 2007).

Four different node outputs are written in (Equation 5, Equation 6) since the RRMS, SPMS, PPMS subgroups and Healthy individual clusters have four different centroid cluster inputs. There are 228 nodes in total for 4 inputs. In each node, maximum 1, minimum 0 bell curve membership functions are used as the membership function, and the result function is presented in (Equation 5, Equation 6).

$$\mu_{Ai}(x) = \frac{1}{1 + \left| \frac{x - \mu_i}{\sigma_i} \right|^2} \tag{5}$$

$$\mu_{Ai}(x) = \exp\left[-\left(\frac{x - \mu_i}{\sigma_i} \right)^2 \right] \tag{6}$$

μ_i and σ_i denote the mean and standard deviation of the bell-shaped membership function, respectively.

The multiplication of all the data with input through Π in Layer 3 is seen in (Equation 7). The output of each node signifies the firing level of a rule. Node function has been formed in this study for the nodes in this layer.

$$o_i^3 = \overline{\mu_I}, \quad i = 1,2,3,4 \tag{7}$$

Each node is labelled with *N* on Layer 4. The normalized firing level of a rule is equal to the sum of firing levels of all the rules for *i*. node, *i*. rule as (Equation 8).

$$o_i^4 = \overline{\mu_I} \tag{8}$$

On Layer 5 for each i node, the result weights value is calculated. $\overline{\mu_i}$ is the normalized firing level as (Equation 9).

$$o_i^5 = \overline{\mu_i} f_i \tag{9}$$

Layer 6 is the sum of the values taken from Layer 5 through $\sum$. The result obtained is the real f output of the system as (*Equation 10*).

$$o_i^6 = f = \sum_i \overline{\mu_i} f_i \tag{10}$$

In the present study, training and test accuracy results have been obtained by applying the rules formed through ANFIS back propagation and hybrid learning algorithms.

2.3.1. ANFIS Hybrid Learning Algorithm

Neuro-fuzzy is a mathematical model whose most important characteristic is adaptive learning (Şen, 2004), (Elmas, 2007), (Jang, 1997). It is a new model based on the learning power of neural networks in fuzzy logic system. Input-output fuzzy membership functions and heuristic fuzzy logic rules are formed with rule-forming from the data. Hybrid learning algorithm is one algorithm that accelerates the learning process. The training set of each cluster is denoted with *X* matrix on each line. (Equation 11) is used for the performance evaluation of the training set comprised of lesion diameter size (from MRI) and EDSS score.

$$X = (x_1(i), x_2(i-2), x_3(i-3), x_4(i-4)) \tag{11}$$

Y vector denotes the RRMS, SPMS, PPMS, and Healthy target values through the training of training set in the study.

(Equation 12) and (Equation 13) rank all the rules in non-linear passive dynamic ranking as a vector.

$$A = \left(x(i), \beta(i), \delta(i)\right) \tag{12}$$

$$XA = Y \tag{13}$$

(Equation 14) helps minimize A through least squares (LSE) prediction method and *A** squared error is an effective method used in ANFIS training.

$$A^* = \left(X^T X\right)^{-1} V^T Y \tag{14}$$

Membership functions of Adaptive Inference System are updated through least squares method (Table 1).

Table 1. Two phases in the hybrid learning

Parameter	**Forward Phase**	**Backward Phase**
Premise Parameter	Specific	Gradient descending
Consequent Parameter	Least– Square	Specific

In the next step, input data and functions are calculated for each node. In the Forward phase, the error is calculated according to (Figure 3) and in the backward phase, again according to (Figure 3), error rates are reflected back from the output, and progresses up until the input layer. In the final stage, parameters are updated. (Figure 4) presents a summary of hybrid based learning (Şen, 2004), (Elmas, 2007), (Jang et al., 1997), (Loganathan and Girija, 2014).

Data given in the input as can be seen in (Figure 4) is calculated based on the iteration value. The training of data is conducted in two phases. In the first one, all the clusters in the training set are propagated back and consequent parameters are determined in line with the iterative Least Square Estimation (LSE). Antecedent parameters remain constant. In the second phase, hybrid learning algorithm performs in the forward phase.

2.3.2. ANFIS Back Propagation Learning Algorithm

The error value obtained from the back propagation learning algorithm is reflected backwards up until the input layer. The required weight variables are identified. The aim is to ensure the *E* error criterion to reach zero for all the input samples at the end of the learning procedure. The error in the network output is the difference between the desired value *d* and actual value *f* in (Equation 14) (Alrashed et al., 2018), (Şen, 2004).

$$e = d - f \tag{15}$$

The error criterion is calculated based on (Equation 16), the applying least squares method to this error value.

$$E = \frac{1}{2} e^2 \tag{16}$$

Using error criterion, backward propagation is performed layer by layer through the stages stated below:

According to (Figure 3), on Layer 6, the weight value is not identified, the reflection of the error in output to the later output is calculated based on (Equation 17).

$$\delta^6 = -\frac{\delta E}{\delta f} = e \tag{17}$$

On Layer 5, it is necessary to calculate the result variables by identifying them as (Equation 18).

$$\begin{aligned} &\Delta p_i = \delta \overline{\mu_i} \left(\text{1st Centroid Cluster}\right) \\ &... \\ &\Delta q_i = \delta \overline{\mu_i} \left(\text{4th Centroid Cluster}\right) \end{aligned} \tag{18}$$

On Layer 4, the error value reflected from the output layer is calculated in (Equation 19).

$$\delta_i^4 = \delta^6 f_i \tag{19}$$

On Layer 3, the error value reflected from the output layer is calculated in (Equation 20).

$$\delta_i^3 = \delta_i^4 = \left[\frac{a - \mu_i}{a^2} \right] \tag{20}$$

On Layer 2, both the error value and values belonging to input membership functions are calculated in (Equation 21).

$$\begin{aligned} &\delta_{i+2}^2 = \delta^3 \mu_{A_i} \\ &\mu_{A_i} = \exp[-(\frac{x - \mu_i^2}{\sigma_i})] \end{aligned} \tag{21}$$

Change in the medium points of bell-shaped membership functions are calculated in (Equation 22).

$$\Delta\sigma_i = 2\left[\frac{(x - \mu_i)}{\sigma_i^3} \right] \delta_i^3 \mu_i \tag{22}$$

The membership values of other clusters have been calculated in the same manner.

Change in the standard deviations of bell-shaped membership functions is calculated based on (Equation 23).

$$\Delta\sigma_i = 2\left[\frac{x - \mu_i}{\sigma_i^2} \right] \delta_i^3 \mu_i \tag{23}$$

3. EXPERIMENTAL RESULTS AND DISCUSSION

The two algorithms which are ANFIS Back Propagation Learning algorithm and hybrid learning algorithms were applied on the MS dataset. K-Means algorithm was used for obtaining C centroid matrix values so that ANFIS algorithm could be applied as the input data.

As can be seen from (Figure 3(a)) and (Figure 3(b)) the steps for the fuzzy rule derivation from K-means centroid matrix are summarized:

- **Step 1:** Initially, classification is made for RRMS, SPMS, PPMS, and Healthy clusters within the MS dataset (139 x 228). Idx = mean (MultipleSclerosis, 4) code. 139 x1 dimension idx vector is processed, including the cluster indices for each record. Afterwards, initial values are assigned to the squared centroid based on cosine similarity. Finally, with [idx,C] = kmeans (Multiple Sclerosis, 4) code,

centroid values are obtained for RRMS, SPMS, PPMS subgroups, and Healthy clusters (4 x 228).

- **Step 2:** Rules were formed through Sugeno Fuzzy Inference System from the MS centroid matrix. 256 rules were obtained by FIS.
- **Step 3:** Input membership functions were identified by means of generalized bell-shaped membership function (gbellmf).
- **Step 4:** For each node input values on Layer 1 of ANFIS architecture, the centroid values matrix as training dataset was transmitted to other layers according to Figure 2, and applied to back propagation and hybrid learning algorithms. 50 iterations were made for the training procedure for the back propagation and hybrid based algorithms. Later, the test procedure was carried out and a comparison was carried out between the two algorithms.

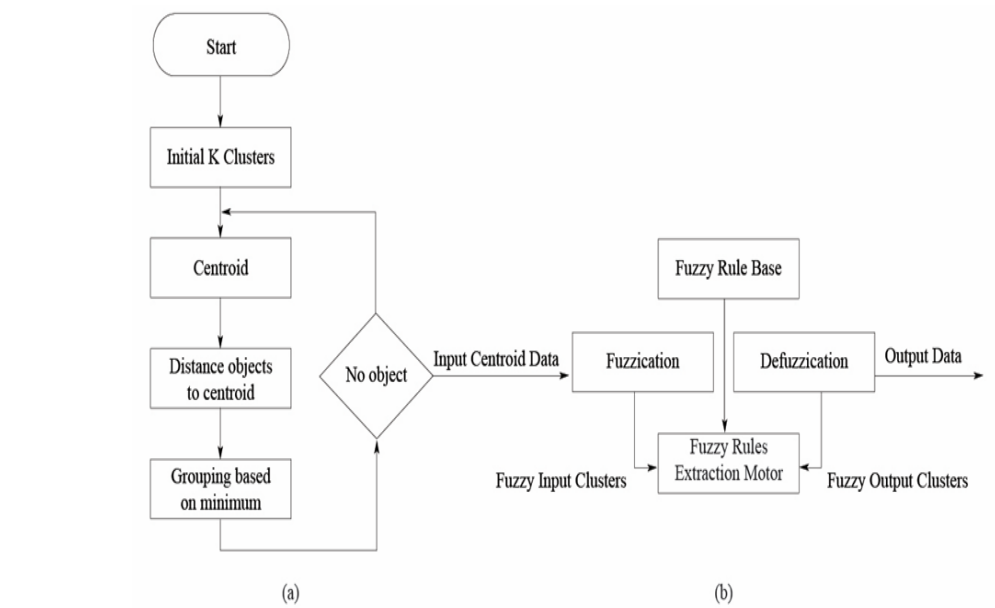


Figure 3. Flow process for obtaining the MS centroid matrix from K-Means algorithm and ANFIS configuration as obtained from Fuzzy rules.

3.1. Application of Centroid Matrix to ANFIS Based Algorithms: Hybrid Based Learning Algorithm and Backpropagation Learning Algorithm

By identifying the initial fuzzy inference system, each node multiplication on Layer 3 has been performed as the training procedure for ANFIS. The output of each node on this layer brings out the firing level of a rule and here generalized fuzzy function has been used. Normalized firing level is calculated and for Layer 5, result variables and actual output value are calculated.

In this study, the MS dataset includes 76 RRMS, 38 SPMS, 6 PPMS subgroups and 19 Healthy individual clusters (based on lesion diameter size as obtained from the MRI and EDSS scores.) 70% and 30% have been allotted for the training and test procedures of the dataset, respectively. A generalized bell-shaped membership function (gbellmf) has been used for the inference of FIS rules based on two algorithms.

As can be seen in (Figure 4), a graph is presented for Sugeno type FIS system based on AND method, defining the rules.

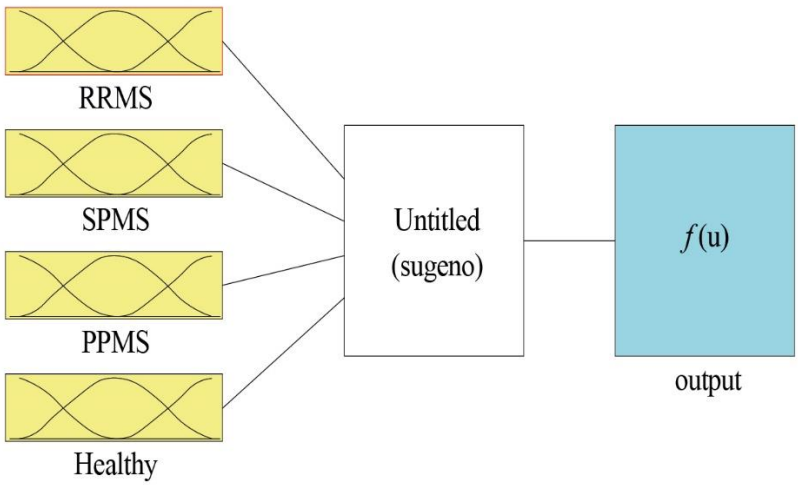


Figure 4. Fuzzy Inference System (FIS) Sugeno Type Block Diagram.

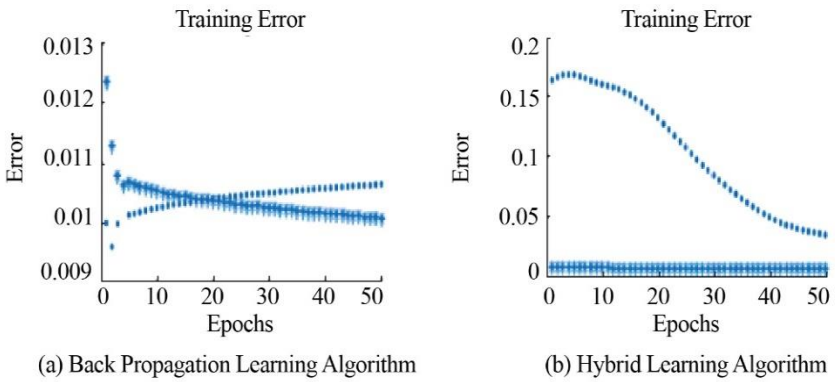


Figure 5. Error Results by Epoch Graph for (a) Back Propagation Algorithm (b) Hybrid Learning Algorithm

The parameters applied for training and test procedures in the experimental studies are as indicated in (Table 2) (Figure 5(a)) presents training (denoted with ‘o’) and test (denoted with ‘*’) dataset error graph in Back Propagation Algorithm. Error rate is 0.11 at Epoch 50 according to mean squared error MSE. Designated epoch number is attained and ANFIS training is completed at Epoch 2. In (Figure 5(b)) error rate is 0.007 at Epoch

50 based on LSE. The designated epoch number is reached and ANFIS training is completed at epoch 2.

Table 2. ANFIS algorithms training and test parameters

ANFIS Algorithms	Training & Testing Parameters
Hybrid based learning algorithm Backpropagation learning algorithm	Epoch number is 50 Number of nodes is 551 Number of linear parameters is 256 Number of nonlinear parameters is 48 Total number of parameters is 304 Number of training data pairs is 160 Number of test data pairs is 68

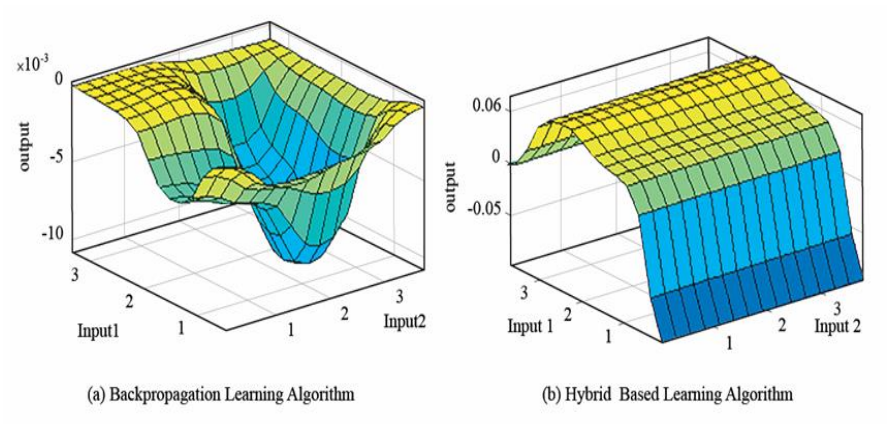


Figure 6. Input 1 (RRMS centroid values) – Input 2 (SPMS centroid values)
rules surface result for (a) Backpropagation Learning algorithm (b) Hybrid Based Learning algorithm.

In (Figure 6), rule structure between Input 1 (RRMS centroid values) and Input 2(SPMS centroid values) and output data based on (a) Backpropagation Learning algorithm (b) Hybrid Based Learning algorithm is defined according to surface depiction.

Table 3. ANFIS algorithms' classification accuracy rate results

ANFIS Algorithms	Accuracy Rate
Hybrid based learning algorithm	99.3%
Back propagation learning algorithm	89%

Classification accuracy results obtained using the ANFIS Hybrid based learning algorithm and the Backpropagation learning algorithm are provided in (Table 3) and (Figure 6).

CONCLUSION

Neuro-fuzzy systems preferable owing to their human-like reasoning because these systems allow the modelling of complex processes. They are capable of finding the parameters related to a fuzzy system by employing approximation techniques from neural networks. Among many fields, these approaches are also utilized to assist medical and clinical tools for the diagnosis of medical problems, including that of MS subgroups which make up the scope of this study. These systems come to the fore as available tools to create sound and robust models related to the modelled process. In this study which is based on medical data, MS dataset is a comprehensive one that is made up of the individuals' MRI (belonging to three different years) and EDSS scores (belonging to those years). 120 individuals in the study are MS patients and 19 of them are healthy individuals. This study is different from the earlier works in the literature (Stamile et al., 2015), (Karaca et al., 2017), (Zhang et al., 2015), (Garro et al., 2016), (Kadhim, 2018), (Karaca et al., 2018) since artificial Neuro Fuzzy Inference System (ANFIS) algorithms were used, and based on this, fuzzy rules with respect to MS subgroups were obtained for the classification of the subgroups of MS and healthy individuals. Briefly, the contribution of the study can be summarized in this way: A classification performance for MS subgroups and healthy individuals was revealed and this performance has proven to be higher when compared to other studies in the literature. The other contribution is that a very comprehensive MS dataset was used. In addition, realistic fuzzy rules were inferred after the calculation of centroid values for disease subgroups and healthy individuals were carried out from the MS dataset. Consequently, the aim has been to infer the fuzzy rules through the ANFIS Backpropagation and ANFIS Hybrid based learning algorithms for the classification of MS subgroups and healthy individuals. Besides this, it has also been intended to be of assistance to those engaged in medical field for accurate classification of the diseases. As for future direction, the dataset can be expanded and applications can be carried out on more comprehensive datasets so that higher classification accuracy rates can be attained and the method can shed light on diseases other than MS.

This centroid matrix has been utilized as the initial value so as to extract rules in the fuzzy system. The rules which have been obtained have been applied on the ANFIS Backpropagation algorithm and the Hybrid based learning algorithm. As the next step, the MS Centroid matrix is assigned as the initial value to the FIS Sugeno system so that the fuzzy rules could be formed. Those rules that have been formed have been applied to

the Neuro fuzzy inference system Backpropagation algorithm and Hybrid based learning algorithm. Accordingly, the classification has been compared in terms of classification test accuracy rate. The hybrid learning algorithm has been able to yield a higher accuracy rate compared to the back propagation algorithm in terms of learning fuzzy rules related to the disease as per the results derived from the training procedure. Consistent with this, the results obtained have demonstrated that hybrid based learning algorithm has yielded 10.3% more accuracy rate when compared to the back propagation algorithm for the testing process. The hybrid-based method has proven be an alternative to the back propagation algorithm for the studies to be conducted in the future. All in all, the fuzzy rule base system formed in this study can be utilized by the specialists as a tool for assistance in their diagnostic processes as regards the MS disease.

ACKNOWLEDGMENTS

The author is sincerely thankful to Prof. Dr. Rana Karabudak, MD and her team at Hacettepe University Medical Faculty, Neurology and Radiology Department for their support. The author is also appreciative of the Turkish Neurological Association for all of their contribution.

REFERENCES

Ahtiwash, O. M., Abdulmuin, M. Z. and Siraj, S. F. (2002). A neural-fuzzy logic approach for modeling and control of nonlinear systems. In *Proceedings of the IEEE Internatinal Symposium on Intelligent Control* IEEE, 270-275.

Alrashed, A. A., Gharibdousti, M. S., Goodarzi, M., de Oliveira, L. R., Safaei, M. R. and Bandarra Filho, E. P. (2018). Effects on thermophysical properties of carbon based nanofluids: Experimental data, modelling using regression, ANFIS and ANN. *International Journal of Heat and Mass Transfer*, 125: 920-932.

Balbinot, A., Favieiro, G. (2013). A neuro-fuzzy system for characterization of arm movements. *Sensors*, *13*(2): 2613-2630.

Elmas, Ç. (2007). Yapay Zeka Uygulamaları [Artificial Intelligence Applications]. *Seçkin Yayıncılık*, 321-349.

Garro, B. A., Rodríguez, K. and Vázquez, R. A. (2016). Classification of DNA microarrays using artificial neural networks and ABC algorithm. *Applied Soft Computing*, 38: 548 - 560.

Giacomini, P. S. (2016). Case Studies in Multiple Sclerosis. *Adis, Switz*. 3 - 26.

Hagens, M., van Berckel, B. and Barkhof, F. (2016). Novel MRI and PET Markers of Neuroinflammation in Multiple Sclerosis. *Current opinion in neurology*, 29 (3): 229-236.

Haider, L., Zrzavy, T., Hametner, S., Höftberger, R., Bagnato, F., Grabner, G., Trattnig, S., Pfeifenbring, S., Brück W., Lanssmann, H. (2016). The topograpy of demyelination and neurodegeneration in the multiple sclerosis brain. *Brain*, 139 (3): 807 - 815.

Han J., Kamber, M. (2006). *Data Mining Concepts and Techniques*. Morgan Kauffmann Publishers, Elsevier, 451- 454.

Harmouche, R., Subbana, N. K., Collins, D. L., Arnold, D. L. and Arbel, T. (2015). Probabilistic Multiple Sclerosis Lesion Classification Based on Modeling Regional Intensity Variability and Local Neighborhood Information. *IEEE Trans. on Biomedical Engineering*, 62 (5): 1281 - 1292.

Jain, A. K. (2010). Data Clustering: 50 Years beyond K-means, *Pattern Recognition Letters*, 31 (8): 651-666.

Jang, J. S. R., Sun, C. T. and Mizutani, E. (1997). Neuro-Fuzzy and Soft Computing; A Computational Approach to Learning and Machine Intelligence. *IEEE Transactions on Automatic Control*, 42 (10): 1482 - 1484.

Kadhim, M. A., (2018). FNDSB: A fuzzy-neuro decision support system for back pain diagnosis. *Cognitive Systems Research*, 52, 691-700.

Karaca, Y., Cattani, C. (2018). *Computational Methods for Data Analysis*. Walter de Gruyter GmbH & Co KG.

Karaca, Y., Cattani, C., Moonis, M. and Bayrak, Ş. (2018). Stroke Subtype Clustering by Multifractal Bayesian Denoising with Fuzzy C Means and K-Means Algorithms. *Complexity*, 1-15.

Karaca Y., Cattani, C. (2017). Clustering Multiple Sclerosis Subgroups with Multifractal Methods and Self-Organizing Map Algorithm. *Fractals*, 25 (4): 1-10. 1740001.

Karaca, Y., Osman, O. and Karabudak, R., (2015). Linear Modeling of Multiple Sclerosis and its Subgroubs. *Turkish Journal of Neurology*, 21 (1): 7-12.

Karaca, Y., Zhang, Y., Cattani, C. and Ayan, U. (2017). The Differential Diagnosis of Multiple Sclerosis Using Convex Combination of Infinite Kernels. *CNS & neurological disorders drug targets*, 16 (1): 36 - 43.

Kumar, PS. S., Dharun, V. S. (2016). Hybrid Brain MRI Segmentation Algorithm Based on K-means Clustering and Texture Pattern Matrix", *International Journal of Applied Engineering Research*, 11 (6): 4343-4348.

Kurtzke, J. F., (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, 33 (11): 1444-1452.

Kruse R., Nauck D. (1998). Neuro-Fuzzy Systems. *In Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications*. NATO ASI Series

(Series F: Computer and Systems Sciences), Springer, Berlin, Heidelberg, 162, 230-259.

Loganathan, C., Girija, V. (2014). Investigations on Hybrid Learning in ANFIS. *International Journal of Engineering Research and Applications*, 4 (10): 31-37.

Mathur, N., Glesk, I., Buis, A, (2016). Comparison of adaptive neuro-fuzzy inference system (ANFIS) and Gaussian processes for machine learning (GPML) algorithms for the prediction of skin temperature in lower limb prostheses. *Medical engineering & physics*, 38 (10): 1083 - 1089.

Mathur, N., Glesk, I. and Buis, A. (2016). Comparison of adaptive neuro-fuzzy inference system (ANFIS) and Gaussian processes for machine learning (GPML) algorithms for the prediction of skin temperature in lower limb prostheses. *Medical engineering & physics*, 38(10), 1083-1089.

Meyer-Moock, S., Feng, Y. S., Maeurer, M., Dippel, F. W. and Kohlmann, T. (2014). Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC neurology*, 14 (1): 58.

Patel, V. L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Berthold, M. R., Bellazzi, R. and Abu-Hanna, A. (2009). The coming of age of artificial intelligence in medicine. *Artificial intelligence in medicine*, 46 (1): 5-17.

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thomson, A. J., Waubant, E., Weinshenker B. and Wolinsky, J. S. (2011). Diagnostic Criteria for Multiple Sclerosis: (2010) Revisions to the Mcdonald Criteria. *Annals of neurology*, 69 (2): 292-302.

Sand, I. K. (2015). Classification, Diagnosis, and Differential Diagnosis of Multiple Sclerosis. *Multiple Sclerosis*, 28 (3): 193 - 205.

Selvi, M., Phil, K. M. and Maheswari, M. C. A. (2015). A Review on Brain Tumour Segmentation, *International Journal of Advanced Research in Computer and Communication Engineering*, 4 (10): 233 - 239.

Stamile, C., Kocevar, G., Hannoun, S., Durand-Dubief, F. and appey-Marinier, D. (2015). A graph based classification method for multiple sclerosis clinical forms using support vector machine. In *Medical learning meets medical imaging*, Springer, Cham, 57-64,

Şen, Z., (2004). Mühendislikte Bulanık Mantık ile Modelleme Prensipleri [Fuzzy Logic Modeling Principles in Engineering], *Su Vakfi,* 261 - 350.

Takagi, T., Sugeno, M. (1985). Fuzzy Identification of Systems and Its Applications to Modeling and Control. *IEEE Trans. on Systems, Man, and Cybernetics*, (1): 116-132.

The MathWorks. (2018). *MATLAB (R2018b).* The MathWorks, Inc., Natick, MA.

Teodorescu, H. N., Yamakawa, T. (1996). Neuro-fuzzy systems: hybrid configurations. In *Fuzzy logic* Vieweg + Teubner Verlag. 267-298.

Tanaka, K., Sugeno, M. (1992). Stability Analysis and Design of Fuzzy Control Systems, *Fuzzy Sets and Systems*, 45 (2): 135-156.

Xu, R., Wunsch, D. (2005). Survey of Clustering Algorithms, *IEEE Transactions on Neural Networks*, 16 (3): 645 – 678.

Zadeh, L. A., (1965). Fuzzy sets, *Information and Control*, 8(3): 338 - 353.

Zhang, Y., Dong, Z., Liu, A., Wang, S., Ji, G., Zhang, Z. and Yang, J. (2015). Magnetic Resonance Brain Image Classification via Stationary Wavelet Transform and Generalized Eigenvalue Proximal Support Vector Machine, *Journal of Medical Imaging and Health Informatics*, 5 (7): 1395-1403.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 13

WAVELET TRANSFORM WITH ANALYZED SIGNALS: DETECTION OF BREAST CANCER WITH SVM KERNELS ALGORITHMS

Y. Karaca[1,*] and Z. Noore[2]
[1]University of Massachusetts Medical School, Worcester, MA, US
[2]Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

ABSTRACT

Wavelet-based applications provide significant improvements in dealing with complex information, patterns and signals. Among the comlex information, it is important to determine the most significant and efficient attributes by eliminating the redundant information. Related to this purpose, wavelet-based applications are used extensively as instantaneous frequency estimators in various fields. Among many various diseases whose diagnoses are hard to perform, breast cancer is regarded as one of these types of diseases. In addition, breast cancer (BC) is one of the most frequent malignancies in women. This paper deals with Breast Cancer (WDBC) dataset comprised of data of 699 patients (with 10 attributes taken from the dataset). The subjects are made up of benign and malignant groups. Our study is concerned with the identification of significant and efficient attributes among 10 of the attributes in the WDBC dataset for the classification of benign and malignant types of breast cancer using mammograms. For this aim, the following steps have been applied: (i) 1D Continuous Wavelet Transform (CWT) (db4, level 5) analysis has been carried out and applied on the dataset. Based on the significant signals analysed, attributes were obtained. As a result of this application, a new dataset has been formed, named 1D CWT WDBC dataset (699 x 5). (ii) Support Vector Machine (SVM) kernels algorithms (Gaussian radial basis function (RBF), linear, polynomial, sigmoid kernels) have been applied on the 1D WDBC dataset (699 x 10) and

* Corresponding Author's E-mail: yeliz.karaca@ieee.org, yeliz.karaca@umassmemorial.org.

1D CWT WDBC dataset (699 x 5). Finally, the classification accuracy rates concerning breast cancer have been compared. For the first time in the literature, the comparison has been made through the application of SVM kernels algorithms on the significant attributes with respect breast cancer that were selected by 1D Continuous Wavelet Transform (CWT) (db4, level 5) methods. The results of the study have yielded that when SVM is applied on these two datasets, 1D CWT WDBC dataset (699 x 5) has higher accuracy rate in terms of classification. The study also reveals that the identification of the significant attributes in the 1D CWT WDBC dataset (699 x 5) prove to be of crucial value with regard to classification accuracy rate.

Keywords: continuous wavelet transform, SVM kernels, mammograms, signals, breast cancer, linear kernel

Key phrases: Determining the most significant signal by 1D Wavelet Continuous Transform, Classification accuracy performance by Support Vector Machine (SVM) kernels algorithms, Estimate, Gaussian radial basis function (RBF), linear kernel, polynomial kernel, sigmoid kernel, signals, detection of breast cancer

1. INTRODUCTION

The continuous wavelet transform (CWT), in mathematics, is a formal tool which can provide a complete representation of a signal by letting the translation and scale parameter of the wavelets vary in a continuous way. Wavelet-based applications provide significant improvements for dealing with complex information patterns and signals. Therefore, they are used extensively as instantaneous frequency estimators. Owing to these properties, wavelet transform have been recently employed particularly in partial differential equation solving, detection, texture analysis, filter design, business information analysis, electrocardiogram (ECG) analysis and Electroencephalography (EEG) data analysis, to name just a few.

Support vector machines are also a highly preferred tools since they can produce a significant level of accuracy with less computation power. Support Vector Machines (SVM) are used for both regression and classification tasks. Yet, their wide use covers classification purposes for a number of different problems. Kernel methods or kernel machines stemmed from the process employ the mathematical techniques in order to attain maximal flexibility, generality, and performance. The use of these algorithms is highly diverse, including but not limited to neural networks, statistics, classic machine learning, functional analysis, optimization, time-series prediction and information retrieval (Cristianini and Scholkopf, 2002).

Breast cancer is the most prevalent cancer type seen among women (Million Women Study Collaborators, 2003). In the medical world, taking this threat into consideration, there has been a significant need to develop methods for the classification of suspicious

areas in mammograms, which will assist radiologists so that they can enhance the efficacy of screening programs. (Eltoukhy et al., 2011). It is important to note that the early detection as well as the accurate diagnosis of this disease can provide a long survival for the patients (Chen et al., 2011). One of such methods is 1D continuous wavelet analysis method that is capable of revealing the interrelation between the continuous attributes in the dataset (Lamarque et al., 1996). Support Vector Machine is another fundamental effective method for data groups that are linearly separable (Vapnik, 1995). As for decision making mechanisms, computer aided mechanisms have also become a viable option since they yield more accurate results in the detection of malignancies in the mammograms (Forrest et al., 1996). SVM kernels have also been among the ways used as machine learning methods for computer aided decision making processes.

The increasing number of breast cancer cases has made early diagnosis more significant. Accordingly, there exist various methods for age-based early diagnosis. Yet, mammography is the most frequently used one (Suckling et al., 1994), (Ell et al., 2007), (Zhen et al., 2001), (Miller et al., 2014).

To cite relevant studies in the literature, it can be stated that the studies are concerned with the application of the most significant and efficient attributes as obtained from different means that employ machine learning methods. Such an approach used in the studies has been able to suggest alternative ways in medicine for the diagnosis of breast cancer for the tumours as benign or malignant. The first we can mention was done by (Salama et al., 2012) who utilized Principle Component Analysis method on data with three different mammogram datasets for feature minimization. The applications were done by decision tree (J48), Multi-Layer Perception, Sequential Minimal Optimization, Naive Bayes, and Instance Based for K-Nearest neighbour. Another relevant study was done by (Peng et al., 2010). The authors applied a filter and wrapper method on the dataset obtained from mammogram data and were able to do breast cancer diagnosis with an accuracy rate of 99.50%.

Concerning the studies that address particularly Support Vector Machines, which are among machine learning systems, the following works can be be mentioned: (Zhang et al., 2016) worked on the mammogram images of 200 individuals (100 healthy ones and 100 breast cancer patients). The authors developed an application based on such features as architectural distortion, calcification, speculated masses, ill-defined masses, circumscribed masses, and asymmetry obtained from the image data. (Ganesan et al., 2014) put forth an application with discrete wavelet analysis and Spherical Wavelet Transform methods for the mammogram images of individuals who had benign and malignant tumour types. Classification was performed by Linear Discriminant Classifier, Quadratic discriminant Classifier, Nearest Mean Classifier, Support Vector Machines as well as Parzen Classification. Another study related to SVM was done by (Chen et al., 2011). The authors made the proposition of a rough set (RS) based support vector

machine classifier (RS_SVM) for the diagnosis of breast cancer diagnosis in their study, and their experimental results indicate that the proposed RS_SVM achieves a very high classification accuracy and also detects five informative features' combination. These can be of important value for physicians in their diagnoses of breast cancer. Related to the SVMs, one further study by (Zheng et al., 2014) intended to diagnose breast cancer based on extracted tumour features. They developed a hybrid of K-means and support vector machine (K-SVM) algorithms. The K-means algorithm was used for recognizing the hidden patterns of benign and malignant tumours separately. SVM was utilized to obtain the new classifier to differentiate the incoming tumours. The methodology they proposed improved the relevant accuracy to 97.38%. Considering the necessity of developing machine learning algorithms to guide prediction process, (Turki et al., 2018) used boosting, and developed three computational methods for the purpose of increasing support vector machines' performance. The authors used several gene expression data with respect to the following cancer types: breast cancer, pheochromocytomas and paragangliomas, bladder cancer, gastric cancer and oral cancer.

It is also important to cite studies which made use of Wavelet methods. (Beura et al., 2015) proposed a mammogram classification scheme for the classification of breast tissues (normal, benign or malignant). Feature matrix was formed using Gray-Level Co-Occurrence Matrix (GLCM). Apart from the suggested scheme, the competent schemes were also simulated for a comparative analysis. Another study was done by (Mohanalin and Beena mol., 2014) who intended to use entropy based concepts such as Tsallis and Shannon for the computation of threshold from wavelet coefficients. The modified wavelet coefficients would improve the mammogram and simplify it as a simple two-class problem. They tested their proposed algorithms and confirmed them by two experiments. (Eltoukhy et al., 2011) presented a method for the diagnosis of breast cancer in digital mammogram images. They used wavelet or curvelet and multi-resolution representations to convert the mammogram images into a long vector of coefficients. For this, they constructed a matrix by putting wavelet or curvelet coefficients of each image in row vector. As a method, they developed feature extraction method reliant on the statistical t-test method. A dynamic threshold was administered for their aim which was the optimization of the number of features that could attain the highest classification accuracy rate possible. In addition to this, SVM was also used for the classification of normal and abnormal tissues as well as distinguishing between the malignant and benign tumours. The classification accuracy rates they obtained show that the method proposed could contribute to the accurate detection of breast cancer which is of crucial importance for treatment and survival rates.

The primary aim in this study is to determine the most significant and efficient attributes by eliminating the redundant information while obtaining the attributes based on the significant signals analysed with 1D CWT. Based on the scale in the 1D CWT analysis, the most important attributes have been identified by selecting the most relevant

attributes within the dataset. For this aim, 1D CWT method has been used for the determination of the most significant attributes within the WDBC dataset that would detect breast cancer. The WDBC dataset, comprising a total of 699 patients with benign or malignant tumour type, includes 10 attributes (Sample code number, Uniformity of Cell Size, Uniformity of Cell Shape, Mitoses, Clump thickness, Bare Nuclei, Normal Nucleoli, Single Epithelial cell size, Marginal adhesion and Bland Chromatin). 1D Continuous Wavelet Transform (CWT) method was applied on this WDBC dataset. Attributes were obtained based on the significant signals that have been analysed. This application yielded the dataset named as 1D CWT WDBC (699 x 5). Afterwards, Support Vector Machine (SVM) kernels algorithms (Gaussian radial basis function (RBF), linear, polynomial, sigmoid kernels) were applied on these datasets WDBC (699 x 10) and 1D CWT WDBC (699 x 5). As a final step, comparison was performed with regard to classification accuracy rates for breast cancer. When other related recent studies are reviewed (Salama et al., 2012), (Zhang et al., 2016); (Ganesan et al., 2014), (Zheng et al., 2014), it has been seen that for the first time in the literature, the comparison has been made through the application of SVM kernels algorithms on the significant attributes related to breast cancer which were selected by 1D Wavelet Continuous methods. The results of the study have yielded that when SVM is applied on these two datasets, 1D CWT WDBC dataset (699 x 5) has a higher accuracy classification rate. The study also reveals that the identification of the significant attributes in the 1D CWT WDBC dataset (699 x 5) prove to be of crucial value with regard to classification accuracy. It has also been shown that 1D Wavelet is an important method in determining the significant attributes in the dataset. The results have been obtained following the application of SVM Kernels on the two datasets, namely WDBC and 1D CWT WDBC, and it has been seen that the 1D CWT WDBC dataset that is made up of the significant attributes yields higher results with regard to classification accuracy.

The organisation of the paper is as follows: Section 2 presents Materials and Methods, Section 3 provides Results and Discussion. Conclusion of the study is provided in Section 4.

2. MATERIALS AND METHODS

2.1. Patient Details

Mammography is a particular type of breast imaging which uses low-dose x-rays for the purpose of detecting cancer early – prior to the time when women experience the symptoms – when it is the most treatable (Wolberg et al., 1990). 699 individuals were taken under observation in the Wolberg of Wisconsin University Hospital. A total of 699

breast cancer patients with benign (458) and malignant (241) types of tumour were included.

The individuals included in this dataset were diagnosed with benign (see Figure 1(a)) and malignant tumour type (see Figure 1(b)).

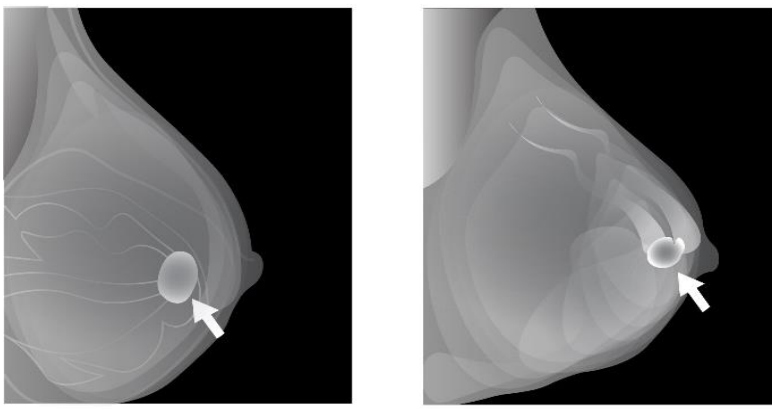


Figure 1. Benign (a) – Malignant (b) mammogram.

Table 1 presents the details of the dataset along with the attributes.

Table 1. WDBC dataset description

Dataset	**Attributes**
WDBC dataset (699 x 10)	Sample code number Uniformity of Cell Size Uniformity of Cell Shape Mitoses Clump thickness Bare Nuclei Normal Nucleoli Single Epithelial cell size Marginal adhesion Bland Chromatin

2.2. Methods

In this paper, the SVM kernels algorithms were applied on the WDBC dataset (699 x 10) which consists of 10 attributes (see Table 1). The aim for doing so is the determination of the most significant attributes in the dataset. The following steps were carried out for the relevant purpose:

1. 1D CWT method (db4, level 5) was utilized to determine the most significant attributes within the WDBC dataset (see Table 1 for details related to the dataset). The dataset generated from this application was named 1D CWT WDBC (699 x 5).
2. Support Vector Machine (SVM) kernels algorithms (Gaussian radial basis function (RBF), linear, polynomial, sigmoid kernels) were applied on the datasets which are WDBC (699 x 10) and 1D CWT WDBC dataset (699 x 5).
3. A comparative analysis was carried out related to the classification accuracy rates for the breast cancer based on the SVM kernels alogirthms' results.

For the computations and generating of the figures in the study, Matlab was utilized (The MathWorks. 2018).

2.2.1. 1D Continuous Wavelet Transform

Continuous Wavelet Transform is a method that enables the gradual splitting of signals into high and low frequency components (Karaca et al., 2017), (Siddiqui 2003), (Kumar and Balakrishnan, 2016), (Nakayama, 2017). Fourier transformation is another method used in frequency analysis, and it is applied based on the assumption that the signals' frequency components undergo no change in time. During the transformation process, signal is defined by the wavelets which are obtained by the shifting and scaling of a wavelet $\psi(t)$ (Siddiqui, 2003). The wavelet equation scaled is provided in Equation (1) with coefficients *a* shifting and *b* scaling coefficient.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{b}\right) \tag{1}$$

It is important for the wavelets to be chosen to fulfil the criterion presented in Equation 2. Here $\psi(w)$ is the Fourier transformation of $\psi_{a,b}(t)$ (Mousa, 2005), (Nakayama, 2017).

$$\int_{\infty}^{-\infty} \frac{|\psi(w)|^2}{w} d_w < \infty \tag{2}$$

The decomposition is done gradually and in each stage, details that carry the high frequency components and approach signals that carry the low frequency components are obtained. After the initial phase, the decomposition is applied to the approach signals. This study has transformed each parameter used for the diagnosis of the disease distinctively into 1D space. Each parameter has been decomposed into pieces that carry

the Mexican hat function (mexh) frequency bands as can be seen in Equation 3 (Karaca et al., 2017).

$$\psi(a,b) = \frac{1}{\pi\sigma^2}\left(1-\frac{a^2+b^2}{2\sigma^2}\right)e^{-\frac{a^2+b^2}{2\sigma^2}} \tag{3}$$

In this study, records of the patients are represented by the parameter *a*. The attributes obtained from the patients' mammograms (Sample Code Number, Uniformity of Cell Size, Uniformity of Cell Shape, Mitoses, Clump Thickness, Bare Nuclei, Normal Nucleoli, Single Epithelial Cell Size, Marginal Adhesion, Bland Chromatin) are represented by the parameter *b*.

2.2.2. Support Vector Machine (SVM) Kernels Algorithms

Support vector machines are machine learning algorithms which can solve problems related to classification, using a flexible representation of the class boundaries and applying automatic complexity control. The aim is to reduce overfitting. They are popular since they are convenient to use and have a high level of generalization performance. Besides these, the same algorithm can solve a different range of problems with little tuning required.

Support Vector Machines are types of learning algorithms that do not have the need of joint distribution function for any data (Liu, 2017), (Karaca et al., 2017), (Karaca and Cattani, 2018). SVM is capable of doing transformations to the high dimensional feature space where it will carry out linear classification from the original input space. Hence, rather than finding the multiplication values of all the values repeatedly, by utilizing the kernel functions, we directly place the value in the kernel function and the value in the feature space is found in this way.

Support Vector Machines ensure a linear splitting of high dimensional data via the help of a kernel function represented mathematically in the following way (Karaca and Cattani, 2019):

$K(x_i, x_j) = \phi(x_i)\phi(x_j)$ (Dheeba et al., 2017), (Dheeba et al. 2011), (Vapnik, 1995), (Huang et al., 2017) (Karaca et al., 2018).

Gaussian RBF Kernel: $K(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}}$ (4)

Linear Kernel: $K(x_i, x_j) = (x_j + 1)$ (5)

Polynomial Kernel: $K(x_i, x_j) = (x_i.x_j + 1)^h$ (6)

Sigmoid Kernel: $K(x_i, x_j) = \tanh(\kappa x_i . x_j - \delta)$ (7)

As this study has been carried out on a dataset with two classes, data can be split from one another on a linear plane. $\phi(x_i)$, x_i represents the WDBC dataset (699 x 10) and 1D CWT WDBC dataset (699 x 5). The parameter c represents our classes, which are benign and malignant. The SVM algorithm was addressed in two ways in this study. In the first way, the WDBC dataset (699 x 10) was applied to the SVM (Gaussian RBF, linear, polynomial, sigmoid) kernels; and a classification accuracy rate was obtained on 10-fold cross validation method. In the second way, 1D Wavelet method was applied on the WDBC dataset, and 1D CWT WDBC dataset (699 x 5) was obtained by determining the efficient and significant attribute. SVM kernels were applied on this dataset obtained. Finally, the SVM kernels were compared in terms of classification accuracy rates for these two datasets.

3. EXPERIMENTAL RESULTS AND DISCUSSION

For this study, the relevant dataset is comprised of a total of 699 breast cancer patients (with benign and malignant tumour types). 10 attributes were determined as the most significant ones. Accordingly, the following steps have been applied on the dataset:

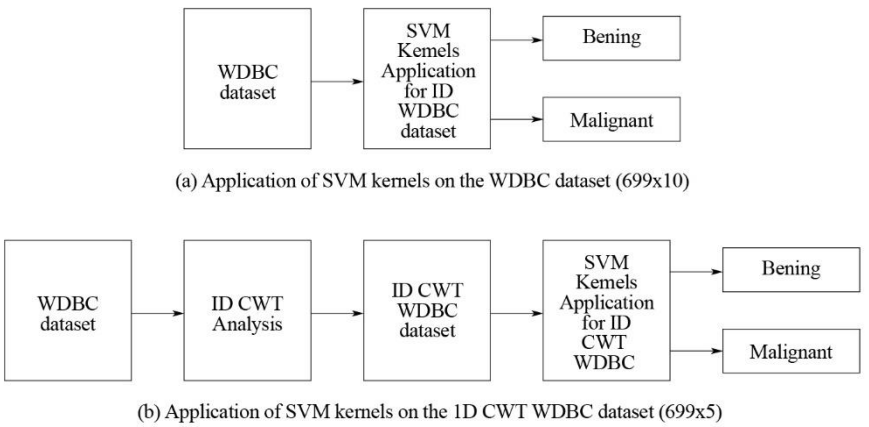


Figure 2. Classification of breast cancer with the application of SVM Kernels algorithms (a) for WDBC dataset (b) for 1D CWT WDBC dataset.

Step 1: 1D CWT method (db4, level 5) was used for the determination of the most significant attributes within the WDBC dataset (see Table 1). The dataset generated from this application was named 1D CWT WDBC (699 x 5) (see Figure 2 (a)).

Step 2: SVM kernels algorithms (Gaussian radial basis function (RBF), linear, polynomial, sigmoid kernels) were applied on these datasets: WDBC (699 x 10) and 1D CWT WDBC dataset (699 x 5) (see Figure 2 (b)).

The applications mentioned above have been depicted in the relevant Figures below.

Step 3: A comparison was performed with respect to the classification accuracy rates concerning breast cancer.

3.1. Application of 1D Continuous Wavelet Transform on 1D CWT WDBC Dataset

The range numerical representation by region as per the results obtained from the application of 1D continuous wavelet transform on the WDBC dataset is provided below (see Table 2)

Table 2. Ranges numerically represented by the regions on the *x*, *y* axes in Figure 3

Ranges of the numbers of patients (*x* axis)	**Tumour related to Breast Cancer (*y* axis)**
[1- 241]	Malignant
[242 - 699]	Benign

Figure 3 shows the determination of the most significant and efficient attributes by eliminating the redundant information with 1D CWT analysis. The most important attributes have been identified by selecting the most relevant attributes within the relevant scales based on the WDBC dataset. For this aim, 1D CWT method has been used for the determination of the most significant attributes within the WDBC dataset that would detect breast cancer. Based on the density scales of the signals, all the attributes involved as per the breast cancer tumour (benign and malignant) and ranges of the numbers of patients (*x* and *y* axis) are depicted as follows:

(a) Sample code attribute at *y* axis displays that the density occurs in the regions coloured in blue and pink. This shows that the benign part is dense, which is at *x* axis. It is seen that although the density occurs in the benign part more, it is an important attribute for the malignant type as well (see Figure 3 (a)).

(b) Uniformity of Cell Size attribute is at *y* axis and the Malignant is at *x* axis. It is seen that there exists a density in regions coloured in blue and pink. Uniformity of cell size attribute is seen to be dense as shown in yellow colour. Accordingly, it is a more important attribute for the malignant type compared to the benign type for the tumour. (See Figure 3 (b)).

(c) Uniformity of Cell Shape attribute is at *y* axis and the malignant is at *x* axis. The density obtained is seen to be in blue and pink colour. Uniformity of cell shape attribute is seen to be dense as shown in yellow colour. Accordingly, it is a more important attribute for the malignant type compared to the benign type. (See Figure 3 (c)).

(d) Mitoses attribute is at *y* axis and malignant is at *x* axis. The density is shown to be more in blue, pink and turquoise colours. Mitoses attribute is seen to be dense as shown in yellow colour. Accordingly, this attribute is a more important attribute for the malignant type compared to the benign type (see Figure 3 (d)).

(e) Clump Thickness attribute is at *y* axis and malignant is at *x* axis. The density is shown to be more in regions which have been coloured in blue and pink. Clump thickness attribute is seen to be dense as shown in green coloured regions. Accordingly, clump thickness attribute is an important attribute both for the malignant and benign types (see Figure 3 (e)).

(f) Bare Nuclei attribute is at *y* axis and malignant is at *x* axis. The dense distribution is shown in regions which are coloured in blue, pink and green. Based on this, it has been shown that bare nuclei is an important attribute with respect to the malignant type (see Figure 3 (f)).

(g) Normal Nucleoli attribute (at *y* axis) and malignant (at *x* axis) have the density in regions coloured in blue, pink and green. Accordingly, normal nucleoli attribute proves to be a more significant attribute for the malignant type than for the benign type (see Figure 3 (g)).

(h) Single Epithelial Cell Size attribute (at *y* axis) and malignant (at *x* axis) have the density distributed in blue, pink and green coloured regions. It is seen dense as shown in the green region for the benign type. Accordingly, single epithelial cell size is a significant attribute for both the malignant and benign types (see Figure 3 (h)).

(i) Marginal Adhesion attribute (at *y* axis) and malignant (at *x* axis) have a dense distribution in regions coloured in blue and green. Accordingly, marginal adhesion is a more important attribute for the malignant type compared to the benign type (see Figure 3 (i)).

(j) Bland Chromatin attribute (at *y* axis) and malignant (at *x* axis) have dense distribution in regions shown in blue and green colours. Density is observed in green and yellow coloured regions for the benign type. Bland chromatin seems to be an important attribute for both the malignant and benign types of breast cancer tumour (see Figure 3 (j)).

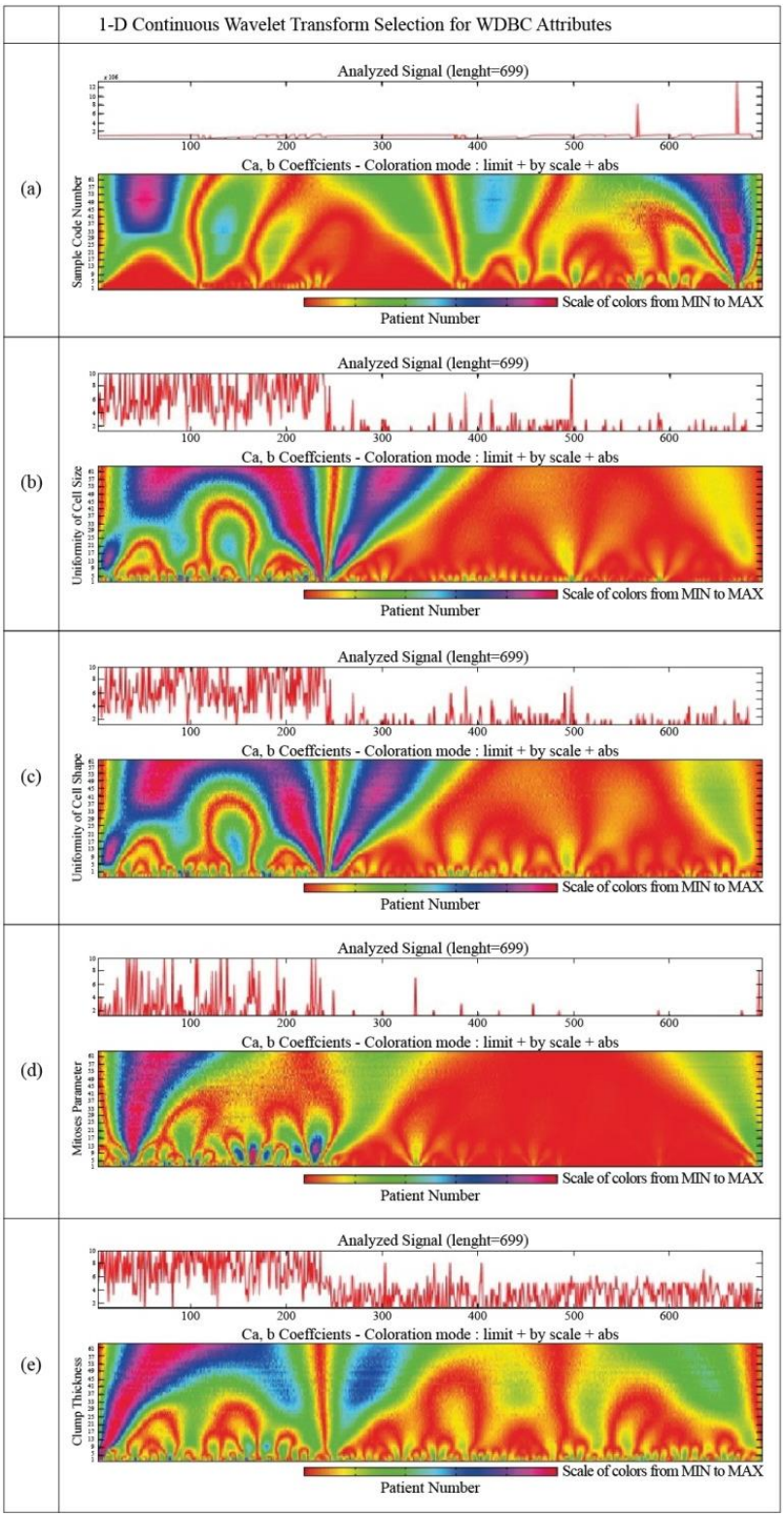


Figure 3. (Continued).

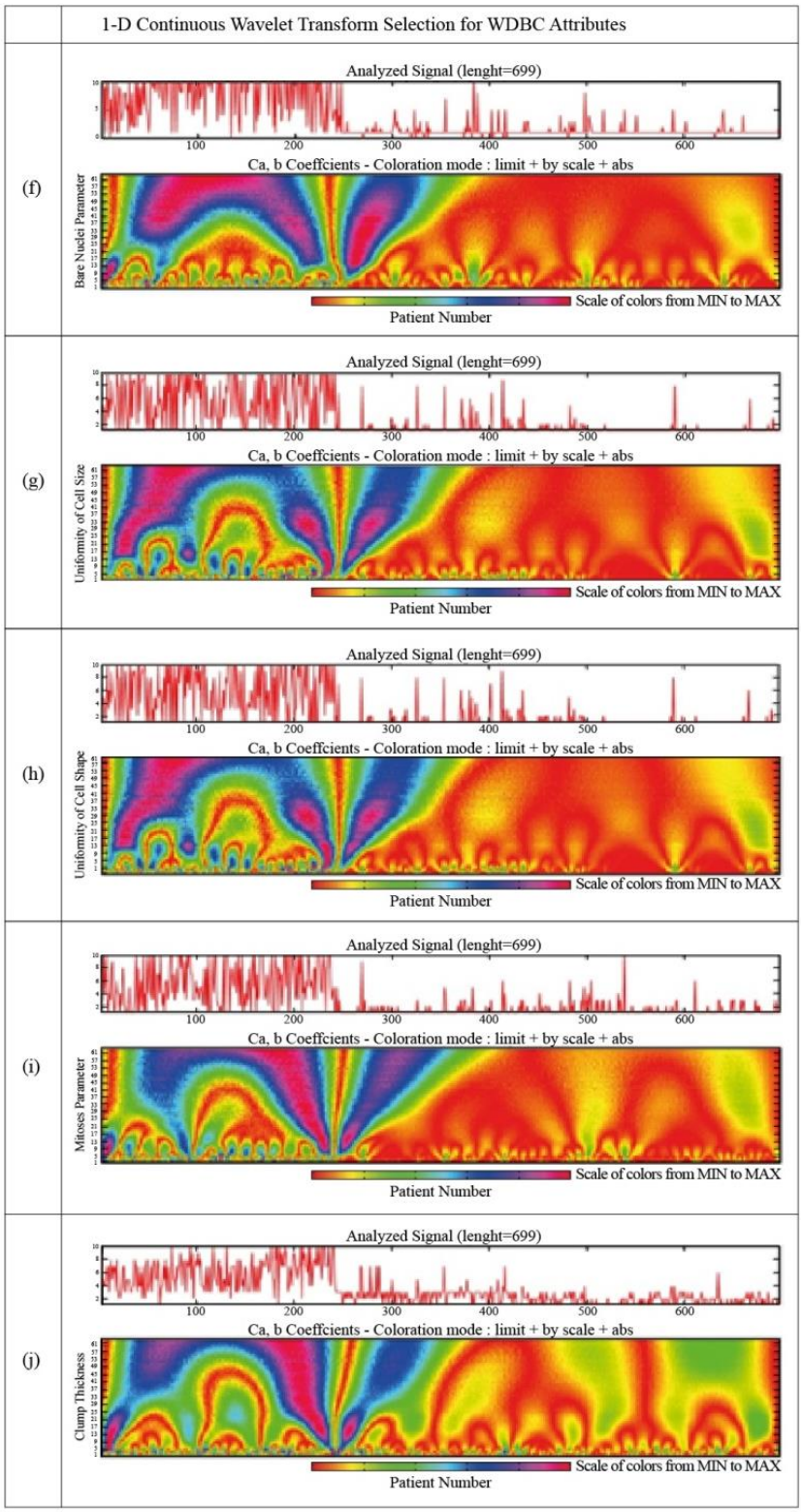


Figure 3. Application of 1D CWT Wavelet Analysis to the WDBC dataset.

Following the 1D CWT analysis carried out on the WDBC Dataset, the significant attributes for the classification of benign and malignant types were found as follows: Sample code number, mitoses, clump thickness, single epithelial cell size and bland chromatin. The dataset obtained accordingly has been named as 1D WDBC with significant attributes (699 x 5).

3.2. Application of SVM Kernels Algorithms on WDBC Dataset and 1D CWT WDBC Dataset

The outcomes derived from the experimental analyses were dealt with in the following ways: SVM kernel algorithms (namely Gaussian RBF, linear, polynomial, sigmoid kernels) were first of all applied on the WDBC dataset. The same SVM kernel algorithms mentioned above were applied on the 1D CWT WDBC dataset (699 x 5). The five significant attributes of the 1D CWT WDBC dataset (699 x 5) were selected from the WDBC dataset (699 x 10) (see Table 1).

The results obtained from the application of SVM algorithm kernels (Gaussian radial basis function (RBF), linear, polynomial, sigmoid kernels) in this study were found as per the parameters of each SVM kernel algorithms type. Table 3 presents the details of the parameters in detail.

Regarding the Sigmoid Kernel, the Gaussian RBF kernel type was utilized in the study. σ plays an important role in the kernel performance. σ Value should be aligned with attention depending upon the existing problem. If an extreme value is assigned for the σ, the exponential value acts almost linearly and loses value for the non-linear data that is big sized (Han et al., 2011). Therefore, there will be deficiency of function regulation, and the determination threshold is going to be sensitive for noise. Gaussian RBF kernel was chosen for the WDBC dataset which was not classified linearly. Hence, it was regarded to be the appropriate kernel in the WDBC dataset. σ value was taken as 3.

Table 3. SVM algorithm Kernel Parameters and Types

Parameters	SVM Kernel Types			
	Linear	**Poly**	**RBF**	**Sigmoid**
Degree (poly)	1.00	2.00	1.00	3.00
Gamma in kernel function (poly/rbf/sigmoid)	0			
Coefficient in kernel function	0			
Tolerance of termination criteria (eps)	0.001			
C(Complexity Cost)	1			
Compute probability estimates	1			
Use shrinking heuristics	1			
Data normalization	1			

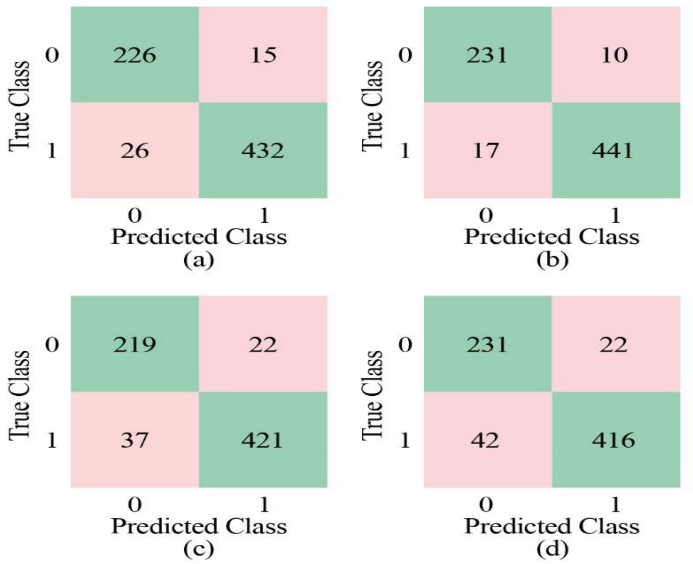


Figure 4. Application of 10-fold Cross Validation SVM Kernels on the WDBC dataset (a) Gaussian RBF (b) Linear (c) Polynomial (d) Sigmoid Confusion Matrix.

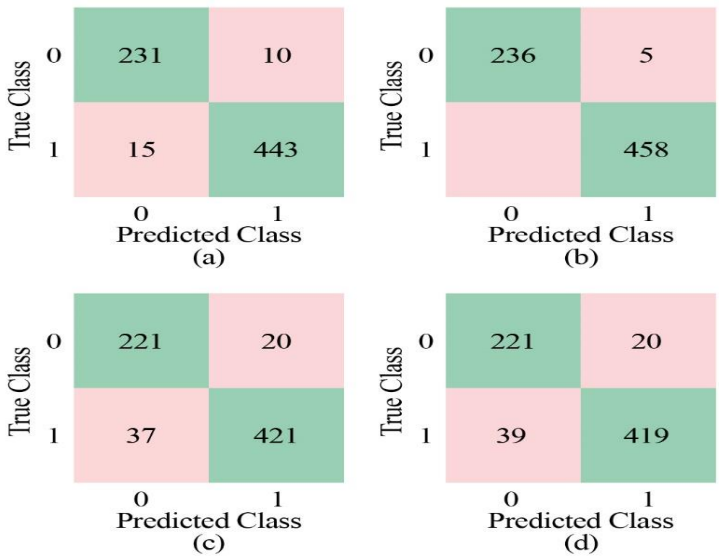


Figure 5. Application of 10-fold Cross Validation SVM Kernels on the 1D CWT WDBC dataset (a) Gaussian RBF (b) Linear (c) Polynomial (d) Sigmoid Confusion Matrix.

Polynomial kernel is an appropriate kernel for the non-stationary dataset. Sigmoid Kernel is considered as the Multi-layered Perceptron. It is used as the activation function in artificial neurons. This kernel was utilized in this study because the WDBC dataset proved to be appropriate for multi-layer learning.

This study has the main purpose of classifying types of tumours (benign and malignant) related to breast cancer by the SVM algorithm. It also aims to compare the

classification accuracy rates (between WDBC dataset (699 x 10) and 1D CWT WDBC dataset (699 x 5)) concerning the applications of the SVM algorithm.

Classification was performed through application of SVM (Gaussian RBF, linear, polynomial, sigmoid kernel types) on the WDBC dataset as can be seen in Figure 4.

As can be seen from Figures 4 and 5, the class that has label 0 refers to the malignant tumour type, and the class with label 1 refers to the benign tumour type in the Confusion Matrices. In line with the 10-fold cross validation method and the results obtained, linear Kernel yields better classification accuracy results than the other kernel types. Comparative analyses have shown this result. As it can be seen in Table 4, the SVM classification accuracy rates for the 1D CWT WDBC dataset are better than those for the WDBC dataset. In both datasets, classification accuracy results of the linear Kernel are higher than the accuracy rates of the other kernels.

Table 4. Classification Accuracy Results for the WDBC and 1D CWT WDBC datasets with the application of SVM Kernels

Datasets	Type of Tumour related to Breast Cancer	SVM Kernels			
		Gaussian RBF	Linear	Polynomial	Sigmoid
WDBC dataset (699 x 10)	Benign	93.7%	95.8%	90.8%	90.8%
	Malignant	94.3%	96.2%	91.9%	90.8%
1D CWT WDBC dataset (699 x 5)	Benign	95.8%	**100%**	91.7%	91.7%
	Malignant	96.7%	**97.9%**	91.9%	91.4%

The results were derived after the application of SVM Kernels on the two datasets, namely WDBC (699 x 10) and 1D CWT WDBC (699 x 5), and it has been seen that the 1D CWT WDBC dataset which is comprised of the significant attributes has yielded higher results in terms of classification accuracy (see Table 4).

CONCLUSION

The primary aim in this study has been to facilitate the determination process of significant attributes in medical dataset, which is a complicated matter encountered frequently in medicine as well as other fields. Accordingly, our aim in the study has been to determine the most efficient and significant by eliminating the redundant information while obtaining the attributes based on the significant signals analysed with 1D CWT. Based on the scale in the 1D CWT analysis, the most important attributes have been identified by selecting the most relevant attributes within the dataset. For this aim, 1D CWT method has been employed to do the determination of the most significant

attributes within the WDBC dataset that would detect breast cancer. Attributes were obtained based on the significant signals that have been analysed.

The main contribution of this paper is that SVM kernels algorithms (Gaussian radial basis function (RBF), linear, polynomial, sigmoid kernels) were applied on 1D CWT WDBC dataset (699 x 5) and WDBC dataset (699 x 10) with respect to breast cancer (benign and malignant tumour types). As a result, comparison between the two datasets was performed to yield high classification accuracy rates based on multi-staged analyses. In this regard, it has been done for the first time in the literature when other related other works have been reviewed, (Salama et al., 2012), (Zhang et al., 2016), (Ganesan et al., 2014), (Zheng et al., 2014). Accordingly, its contribution has been in two ways. The results obtained from this study show that when SVM is applied on these two datasets, 1D CWT WDBC dataset (699 x 5) has a higher accuracy rate in terms of classification. The study also reveals that the identification of the significant attributes (bland chromatin, clump thickness, single epithelial cell size, mitoses and sample code number) in the 1D CWT WDBC dataset (699 x 5) prove to be highly important for the aim of achieving classification accuracy. It should be noted that not only for breast cancer but also for other diseases, accurate classification of the disease related elements carry vital importance to attain more reliable, robust and safer diagnosis processes as well as treatment procedures. With this remark, we intend to provide future direction for studies and research to be carried out for life-threatening diseases.

REFERENCES

Ahtiwash, O. M., Abdulmuin, M. Z. and Siraj, S. F. (2002). A neural-fuzzy logic approach for modeling and control of nonlinear systems. In *Proceedings of the IEEE Internatinal Symposium on Intelligent Control*, IEEE, 270-275.

Beura, S., Majhi, B. and Dash, R. (2015). Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing*, *154*, 1-14.

Cattani, C. (2007). *Wavelet and Wave Analysis as applied to Materials with Micro or Nanostructure*, *World Scientific*, 74: 137 – 143.

Cristianini, N., Scholkopf, B. (2002). Support vector machines and kernel methods: the new generation of learning machines. *Ai Magazine*, *23*(3): 31-31.

Chen, H. L., Yang, B., Liu, J. and Liu, D. Y. (2011). A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis, *Expert Systems with Applications*, *38*(7): 9014 - 9022.

Dheeba, J., Jaya, T. and Singh, N. A. (2017). Breast cancer risk assessment and diagnosis model using fuzzy support vector machine based expert system. *Journal of Experimental & Theoretical Artificial Intelligence*, *29*(5): 1011-1021.

Dheeba, J., Selvi, S. T. (2011.) Classification of malignant and benign microcalcification using SVM classifier. In *Emerging Trends in Electrical and Computer Technology (ICETECT), International Conference on IEEE*, 686-690.

Ell, K., Vourlekis, B., Lee, P. J. and Xie, B. (2007). Patient navigation and case management following an abnormal mammogram: a randomized clinical trial. *Preventive medicine*, 44(1): 26-33.

Eltoukhy, M. M., Faye, I. and Samir, B. B. (2011). A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation, *Computers in Biology and Medicine*, *42*(1): 123-128.

Forrest, A. Patrick., Helen, J. Stewart., Dawn, Everington., Robin, J. Prescott., Colin, S. McArdle., Adrian, N. Harnett., David, C. Smith., David George, W. and Scottish Cancer Trials Breast Group. (1996). Randomised controlled trial of conservation therapy for breast cancer: 6-year analysis of the Scottish trial. *The Lancet*, *348*(9029): 708-713.

Ganesan, K., Acharya, U. R., Chua, C. K., Min, L. C. and Abraham, T. K. (2014). Automated diagnosis of mammogram images of breast cancer using discrete wavelet transform and spherical wavelet transform features: a comparative study. *Technology in cancer research & treatment*, 13(6): 605-615.

Han, J., Pei, J. and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier, 408 – 416.

Huang, M. W., Chen, C. W., Lin, W. C., Ke, S. W. and Tsai, C. F. (2017). SVM and SVM Ensembles in Breast Cancer Prediction. *PloS one*, 12(1).

Karaca, Y., Cattani, C. (2019). Model Estimation with SVM Kernel Types for Cognitive Diagnostics. *Advanced Mathematical Models & Applications.* Jomard Publishing. 4(1), 5-14.

Karaca, Y., Cattani, C. (2018). *Computational Methods for Data Analysis*. Walter de Gruyter GmbH & Co KG.

Karaca, Y., Sertbaş, A. and Bayrak, Ş. (2018). Classification of Erythematous-Squamous Skin Diseases Through SVM Kernels and Identification of Features with 1-D Continuous Wavelet Coefficient. In *International Conference on Computational Science and Its Applications* . Springer, Cham, 107-120.

Karaca, Y., Aslan, Z., Cattani, C., Galletta, D. and Zhang, Y. (2017). Rank determination of mental functions by 1D wavelets and partial correlation. *Journal of medical systems*, 41(1): 2.

Karaca, Y., Zhang, Y. D., Cattani, C. and Ayan, U. (2017). The differential diagnosis of multiple sclerosis using convex combination of infinite kernels. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)*, 16(1): 36-43.

Karaca, Y., Aslan, Z. and Siddiqi, A. H. (2017). 1D Wavelet and Partial Correlation Application for MS Subgroup Diagnostic Classification. *In Industrial Mathematics and Complex Systems* Springer, Singapore, 171-186.

Kumar, S. M., Balakrishnan, G. (2016). Wavelet and Symmetric Stochastic Neighbor Embedding based Computer Aided Analysis for Breast Cancer. *Indian Journal of Science and Technology*, 9(47).

Liu, W., Fowler, J. E. and Zhao, C. (2017). Spatial Logistic Regression for Support-Vector Classification of Hyperspectral Imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(3): 439-443.

Lamarque, C. H., Robert, F. (1996). Image analysis using space-filling curves and 1D wavelet bases. *Pattern Recognition*, 29(8): 1309-1322.

Mohanalin, J., Beena mol, M. (2014). A new wavelet algorithm to enhance and detect microcalcifications. *Signal Processing*, 105, 438-448.

Miller, A. B., Wall, C., Baines, C. J., Sun, P., To, T. and Narod, S. A. (2014). Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial., *Bmj*, 348-366.

Million Women Study Collaborators. (2003). Breast cancer and hormone replacement therapy in the Million Women Study. *The Lancet*, 362(9382): 419-427.

Mousa, R., Munib, Q. and Moussa, A. (2005). Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. *Expert systems with Applications*, 28(4): 713-723.

Nakayama, R. (2017). Computer-Aided Detection and Differentiation of Breast Cancer on Mammograms. *In Image-Based Computer-Assisted Radiation Therapy*, 41-66.

Peng, L., Yang, B. and Jiang, J. (2009). A novel feature selection approach for biomedical data classification, *J. Biomed Inform*, 179(1): 809-819.

Peng, Y., Wu, Z. and Jiang, J. (2010). A novel feature selection approach for biomedical data classification. *Journal of Biomedical Informatics*, 43(1): 15-23.

Salama, G. I., Abdelhalim, M. and Zeid, M. A. E. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. *Breast Cancer (WDBC)*, 32(569): 2.

Siddiqui, A. H. (2003). Wavelet Method for Partial Differential Equations and Image Processing, Numerical Methods. *Wavelet Methods and Image Processing*, 1 - 60.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S. and Taylor, P. (1994). The mammographic image analysis society digital mammogram database. *InExerpta Medica*. *International Congress Series*, 1069: 375-378.

The MathWorks. (2018). *MATLAB (R2018b).* The MathWorks, Inc., Natick, MA.

Turki, T., Wei, Z. (2018). Boosting support vector machines for cancer discrimination tasks. *Computers in biology and medicine*, 101: 236-249.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.

Wolberg, W. H., Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *In Proceedings of the National Academy of Sciences*, 87: 9193-9196.

Zhang, Y. D., Wang, S. H., Liu, G. and Yang, J. (2016). Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform. *Advances in Mechanical Engineering*, *8*(2): 1687814016634243.

Zhen, L., Chan, A. K. (2001). An artificial intelligent algorithm for tumor detection in screening mammogram. *IEEE transactions on medical imaging*., *20*(7), 559-67.

Zheng, B., Yoon, S. W. and Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications*, 41(4): 1476-1482.

In: Mathematical Modeling of Real World Problems ISBN: 978-1-53616-267-7
Editors: Z. Aslan, F. Dökmen, E. Feoli et al.

Chapter 14

DIAGNOSING PARKINSON'S DISEASE BY MACHINE LEARNING METHODS USING FORCE SIGNALS

Fatih Aydın[1,*] and Zafer Aslan[2]
[1]Department of Software Engineering, Kirklareli University, Kirklareli, Turkey
[2]Department of Computer Engineering, Istanbul Aydin University, Istanbul, Turkey

ABSTRACT

Parkinson's disease (PD) is the second most commonly seen neurodegenerative disease in the world, after Alzheimer's disease. Therefore, the diagnosis of PD is of great significance. In this study, we proposed a method to diagnose Parkinson's disease, i.e., to distinguish subjects with PD from healthy subjects. First of all, we obtained the data from PhysioNET database, which are composed of vertical force signals applied to the ground by the subjects. Then, we applied the signal processing methods (such as Fourier transform, wavelet transform, and Hilbert-Huang transform) to these data. Later, we derived features from these transformed signals. We needed to carry out feature selection since tens of thousands of features are formed for each signal transforming process. The OneRAttributeEval algorithm in WEKA tool was used for feature selection. Besides, the feature selection method called the WrapperSubsetEval algorithm was applied to the selected features once more during classification. Thereby, a two-layer feature selection process was performed and only a dozen of features were selected from among tens of thousands of features. Based on the results of the experiments we have performed, parameters providing the highest classification accuracy are as follows: the signal processing method is wavelet transform, the wavelet type is daubechies3 and the classification algorithm is the aggregating one-dependence estimators (A1DE) algorithm. Consequently, the accuracy rates yielded by the proposed methods are 90.91%, 93.33%, 93.94% and 93.94% respectively, according to 2- fold Cross Validation (CV), 5- fold CV,

* Corresponding Author's E-mail: fatih.aydin@klu.edu.tr.

10-fold CV and leave-One-Out Cross Validation (LOOCV) methods. Accordingly, our results are more reliable than and highly accurate when compared to previous studies.

Keywords: Parkinson's disease, machine learning, feature selection, Hilbert-Huang transform, wavelet transform, Fourier transform

INTRODUCTION

Machine learning is a field of study which covers automatic calculation procedures based on logical or binary operations learning a task from a set of examples (Michie, Spiegelhalter, & Taylor, 2009). There are many studies on the application of machine learning to the field of medicine. These studies concentrate specifically on medical diagnosis (Deo, 2015). In the medicine field, the use of machine learning methods instead of traditional methods allows the creation of very robust systems (Aydın, 2017). This is because in traditional systems, the rules are provided to computers by us. However, with the use of machine learning algorithms, instead of rules, only data are provided to computers. Thus, the task of extracting rules from data is left for the algorithms to perform. As a result, the systems that are developed end up learning the rules rather than memorizing them (Obermeyer & Emanuel, 2016).

The field of neurodegenerative diseases is one of the fields in which machine learning algorithms are applied to medical diagnosis. The disease which has the second highest prevalence among neurodegenerative diseases is Parkinson's disease (PD). PD was first described in the article titled "An Essay on the Shaking Palsy" by James Parkinson in 1817 (Adams, Victor, & Ropper, 1997). The distinctive characteristic of PD is bradykinesia. Bradykinesia is defined as the deceleration of motions. In the early period of the disease, the most significant symptoms of bradykinesia are hardships encountered in gait, speech and sitting (Gibberd, 1986). In recent years, significant progress has been made in the diagnosis and treatment of motion-related problems. Nowadays, gait analysis technology allows the analysis of joint and muscle functions by converting them into digital data during normal and pathological gait. Hereby, the effects of treatments such as physiotherapy, medication and surgery applied to patients on gait can be objectively evaluated. By gait analysis, gait and motion troubles which human-experts cannot resolve can easily be analyzed and muscles and joints leading to the problem can be intervened with (Özaras & Yalçın, 2002).

This paper is on distinguishing subjects with PD from healthy subjects. The data used in this study were obtained from PhysioNET database and PhysioNet is supported by the National Institute of General Medical Sciences (NIGMS). The effort, through which the data were obtained, was supported in part by grants from the National Institutes of Health, National Parkinson Foundation, and the Parkinson's Disease Foundation and the

data were collected at the Laboratory for Gait & Neurodynamics, Movement Disorders Unit of the Tel-Aviv Sourasky Medical Center. The data includes the vertical ground reaction force records of subjects as they walked at their usual, self-selected pace for approximately two minutes on level ground. Thereby, one can investigate the force record as a function of time and location, derive measures that reflect the center-of-pressure as a function of time, and determine timing measures (e.g., stride time, swing time, etc.) for each foot as functions of time. Thus, one can study the stride-to-stride dynamics and the variability of these time series (Goldberger et al., 2000). In this study, we used raw data, i.e., the vertical ground reaction force records. We used machine learning algorithms which had various learning approaches for pattern discovery in data mining because patterns in these data/signals are unclear. Moreover, we used signal processing methods (e.g., Fourier transform, wavelet transform and Hilbert-Huang transform) for transforming the signals, and we derived tens of thousands of features. Later on, we reduced the number of these features down to hundreds or even dozens by using feature selection methods. As a result, we developed a method which yields highly accurate results by using a few features.

RELATED WORKS

The same data were used in all of the studies mentioned below as the data used in this study. Hence, even if related to the diagnosis of PD, we have not included here any studies which do not share the same dataset with this study. In addition, the comparison of our results with those of the following studies may be found in the results section.

In a study conducted by Lee and Lim (Lee & Lim, 2012), a method was proposed to distinguish individuals with idiopathic PD from healthy individuals, by relying on the gait characteristics of patients with idiopathic PD and deriving features based on wavelet transform. In the study, neural networks with weighted fuzzy membership function were used as the classification algorithm. In addition, 40 features were obtained by using statistical approaches. The highest classification accuracy rate among the methods used in the study was 77.33%.

In a study conducted by Daliri (Daliri, 2013), a new approach was suggested for distinguishing individuals with PD from healthy individuals. This method relies on digitizing the reaction force applied to the ground by the subjects during normal gait and measuring it. These measures were carried out using eight sensors under each foot. The absolute values of difference between the force measures were computed for each sensor and time. Then, short-time Fourier transform was applied to these signals. Thereby, various features were derived from the spectrum of signals. The histogram of these features was computed and the bin selection was performed using the feature discriminant ratio method. Finally, chi-square distance between the reduced histograms

was computed and a kernel for support vector machine was formed for classification. The classification accuracy was obtained as 91.20% in the study.

In a study conducted by Ertuğrul et al. (Ertuğrul, Kaya, Tekin, & Almalı, 2016), a new method was introduced to distinguish between healthy subjects and those with PD. The methods that were used in the study were the shifted one dimensional local binary patterns method and machine learning algorithms to pinpoint local changes in the signal. The highest classification accuracy rate of the study was 88.89%.

In a study conducted by Jane et al. (Jane, Nehemiah, & Arputharaj, 2016), a new approach was proposed to distinguish healthy individuals from those with PD. The mean accuracy of the classification process carried out by Q-back propagated time delay neural network was 91.53%.

In another study carried out by Zeng (Zeng et al., 2016), a new approach was presented for distinction between healthy subjects and those with PD. Radial Basis Function Network (RBF Network) algorithm was used as the classification algorithm in the study and the mean classification accuracy of the study was 93.78%.

In a study conducted by Shrivastava et al. (Shrivastava, Shukla, Vepakomma, Bhansali, & Verma, 2017), a new study was proposed for distinguishing healthy subjects from those with PD. A comparative analysis of various algorithms inspired from nature was performed for selecting the best features and increasing the classification accuracy. The results of the experiment demonstrated that the Binary Bat algorithm has better performance than traditional methods such as the Particle Swarm Optimization, Genetic Algorithms and the Modified Cuckoo Search Algorithm. In conclusion, the mean classification accuracy using an artificial neural network was obtained as 93.12%.

DATASETS USED IN EXPERIMENTS

91 machine learning algorithms in WEKA tool were applied to the Parkinson dataset in PhysioNET (Goldberger et al., 2000) database (Frenkel-Toledo et al., 2005a, 2005b; Hausdorff et al., 2007; Yogev et al., 2005). This dataset includes gait measurements obtained from 93 patients with idiopathic PD and 72 healthy control individuals. The dataset also involves the records of the vertical ground reaction force that the subjects applied to the ground for about two minutes at the ground level. Under the feet of each subject, there were eight sensors (Ultraflex Computer Dyno Graphy, Infotronic Inc.) which measure force (Newton unit) as a function of time. 100 samples per second were recorded by digitizing the output of each 16 detectors. Moreover, the records involve two signals which reflect the sum of the outputs of the eight detectors for each foot. The descriptive information regarding this dataset is shown in Table 1.

Table 1. Descriptive information for the Parkinson dataset (Goldberger et al., 2000)

Subjects	Number of data	Age	Weight	Height	Gender	
					M	F
PD	93	66.30 ± 9.50	72.40 ± 11.96	167.44 ± 8.59	58	35
Control	72	63.68 ± 8.70	72.69 ± 12.42	168.25 ± 8.58	40	32

Feature Creation and Selection

The process of feature creation by applying statistical analysis to unstructured or raw data is a significant step for the discipline of machine learning. At the same time, the process of feature creation is one of the challenging areas in the field of machine learning. Apart from the feature creation process, feature selection is also an important sub-topic for the field of machine learning. Feature selection refers to the selection of a subset of features used in the model-making process. There are three main purposes for the use of feature selection: to make the interpretation of models easier (James, Witten, Hastie, & Tibshirani, 2013), to reduce the training time of classifiers, and to avoid errors resulting from variance (Bermingham et al., 2015). In this study, we performed feature creation and selection over the Parkinson dataset. These processes are explained below in detail.

Signal Transforms and Feature Creation

In order to be used effectively in the field of machine learning, raw data needs to be transformed. For this purpose, signal transform methods are used. Signal transform methods are fairly effective, especially in time series. In this study, various signal transform methods were applied over the Parkinson dataset and statistical features were derived for it. These methods are briefly mentioned below.

Fourier Transform

In his work titled *The Analytic Theory of Heat* (1822), French scientist Joseph Fourier (1768-1830) stated that any convergent and integrable function can be represented by a trigonometric series (Brown & Churchill, 1993). Fourier transform is an integral transform which transforms a signal from time domain into frequency domain. In other words, Fourier transform gives us the frequency components of a time signal.

Wavelet Transform

Fourier transform is a useful tool to analyze the frequency components of a signal. However, if Fourier transform is taken across the entire time axis, the exact moment when a certain frequency emerges cannot be determined. To put it briefly, Wavelet

transform was developed in order to overcome this problem. Wavelet transform is formed by limited-time small wavelets.

The base functions of wavelet transform are called “wavelets”. Wavelets are defined by a scale and wavelet function. Wavelet transform is a method developed in order to make up for the shortcomings of Fourier transform when it comes to non-stationary signals. This analysis method allows the optimum time-frequency resolution in all frequency ranges because of the fact that it has varying window sizes which are wider for low-frequency signals and narrower for high-frequency signals (Morlet, 1981; Prochazka & Storek, 1995). Furthermore, this method is resistant to noise and applied to non-stationary signals easily.

Hilbert-Huang Transform

Traditional data analysis methods are based on the assumption that a signal is linear and stationary, whereas Wavelet analysis and Wagner-Wille distribution (Flandrin, 1999; Gröchenig, 2001) are based on the assumption that a signal is non-stationary but linear. In addition to these methods, there are many other non-stationary time series analysis methods (Kantz & Schreiber, 2004; Tong, 2002). However, the assumption of these methods is that the system is stationary and stable. But, signals are not both linear and stable in many real systems (natural or human-made). Analyzing non-linear and non-stationary signals according to an assumption engenders more problems than the solutions it provides. The necessary condition to represent non-linear and non-stationary data is to have an adaptive base. An a priori defined function cannot be based on a strong foundation. It is required to construct an adaptive, i.e., data-dependent, and a posteriori defined method in order to resolve that problem (Huang, 2005). This problem is resolved by the method called Hilbert-Huang transform (HHT) developed by Huang et al. (Huang et al., 1998; Huang, Long, & Shen, 1996; Huang, Shen, & Long, 1999). This method enables both non-linear and non-stationary signals to be analyzed. The comparative summary of Fourier transform, wavelet transform and HHT are shown in Table 2 (Huang & Wu, 2008).

HHT consists of two parts: Empirical mode decomposition (EMD) and Hilbert spectral analysis (HSA). HHT is potentially suitable for non-linear and non-stationary data analysis, particularly in time-frequency-energy demonstrations. Moreover, HHT has been tested and gained validity, but only empirically (Huang, 2005).

EMD is an essential method for dealing with data from non-stationary and non-linear processes. Contrary to almost all previous methods, this new method is data-based, a posteriori-defined, intuitional and adaptable. The process of decomposing modes is based on the assumption that all data consists of different intrinsic oscillation modes.

Table 2. The comparative summary of Fourier transform, Wavelet transform and HHT (Huang & Wu, 2008)

	Fourier	**Wavelet**	**HHT**
Basis	a priori	a priori	a posteriori adaptive
Frequency	convolution over global domain, uncertainty	convolution over global domain, uncertainty	differentiation over local domain, certainty
Presentation	energy in frequency space	energy in time-frequency space	energy in time-frequency space
Non-linearity	No	no	Yes
Non-stationarity	No	yes	Yes
Feature extraction	No	discrete, no; continuous, yes	Yes
Theoretical base	complete mathematical theory	complete mathematical theory	Empirical

An intrinsic mode function (IMF) represents a simple oscillation mode corresponding to a simple harmonic function, but to express it in more general terms: instead of a constant amplitude and frequency as a simple harmonic component, IMF has an amplitude and frequency which vary as a function of time. The original signal is the sum of these IMFs and the last monotonic tendency (Colominas, Schlotthauer, & Torres, 2014; Huang et al., 1998).

The local nature of EMD can generate oscillations in very different scales for a certain mode, or oscillations in similar scales in different modes. This is not a desired situation. It is desired that similar scales are preferred for each mode. This outcome of EMD leads to a problem known as "mode mixing". In order to eliminate this problem as well as many other shortcomings such as resistance to noise, the improved complete ensemble empirical mode decomposition with adaptive noise (the improved CEEMDAN) algorithm was developed by Colominas et al. (Colominas et al., 2014).

We downloaded MATLAB functions pertaining to the improved CEEMDAN and EMA methods from: http://perso.ens-lyon.fr/patrick.flandrin/emd.html and, http://www.bioingenieria.edu.ar/grupos/ldnlys/.

The objective of HHT is to reveal the time-frequency-energy relation of time series. The method also depicts non-stationary data locally. In order to calculate Instantaneous frequency (IF) and amplitudes, and to describe a signal locally, it uses Hilbert transform (HT) instead of a Fourier or Wavelet-based transform.

Feature Selection

One of the factors which significantly affect the model-making process of machine learning algorithms is the feature set. There are three general approaches for feature selection; namely, Filter approach, Wrapper approach, and embedded approach. Filter approach uses the general characteristics of training data independently of machine learning algorithms, i.e., statistical tests are used for the purpose of evaluating features (Bolón-Canedo, Sánchez-Maroño, & Alonso-Betanzos, 2013). Wrapper approach discovers the relation between the level of relevance and the best feature subset selection

and searches for an optimal feature subset adapted to the machine learning algorithm at hand (Kohavi & John, 1997). And the last approach to feature selection is based on the selected feature subsets which allow the classification algorithms used in the experiments to deliver a good learning performance. This approach resembles the wrapper approach. But, the learning process has an effect on the search process in the embedded approach. This situation decreases the computational cost and excessive tendency for data.

Many feature selection methods have been proposed in the literature. But, studying these methods in a comparative manner is a very challenging task. Finding the strengths of feature selection methods without any knowledge regarding the related features before the real dataset is very difficult. This is because datasets can include many unrelated and redundant features. Therefore, the performance of the feature selection method is dependent on the performance of the learning method. There are many performance criteria in the literature, such as accuracy, computer sources and the speed of feature selection. Many researchers acknowledge that there is no such thing as "the best method" in feature selection (Bolón-Canedo et al., 2013). Therefore, new feature selection methods are used in trying to overcome particular problems, using different strategies.

It is required to select the features belonging to the dataset before machine learning algorithms are applied to the Parkinson dataset. This is because performing data mining over tens of thousands of features is a fairly difficult process. In order to perform the data mining process quickly, it is required to select the features giving the greatest amount of information (or the information with the highest classification accuracy) from among all the features generated for the Parkinson dataset. The feature selection method used for this purpose is *OneRAttributeEval* algorithm. Thereby, features giving little information are eliminated while features giving more information are selected. In the course of feature selection, *OneRAttributeEval* algorithm was run using 10-fold cross validation (CV) strategy. Consequently, the percentages of classification accuracies for each of the features were obtained. Because of 10-fold cross validation, the classification accuracy of each feature is obtained by dividing these accuracy ratios by ten. Thus, results in the form of $(mean + deviation)$ are obtained for each feature. Then, when selecting the features, those corresponding to the condition $(mean - deviation \geq 70)$ are selected. Thereby, only hundreds of features among tens of thousands of features are selected.

The Process of Preparing the Datasets

The machine learning algorithms used in the experiments are the algorithms in WEKA tool. These algorithms are grouped into seven categories: Bayes (Bayesian learning), functions (function learning), lazy (instance or lazy learning), misc (miscellaneous learning approaches), rules (rule learning), trees (decision tree learning) and meta (meta or ensemble learning).

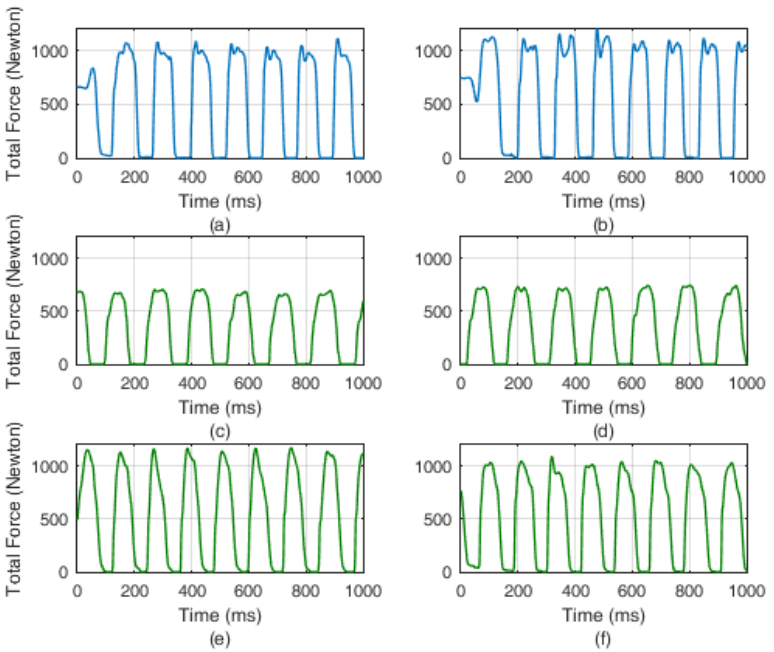


Figure 1. (a and b) control subjects, (c and d) and (e and f) the time-dependant change of forces applied to the ground for left and right feet belonging to two parkinsonian individuals.

A sample reaction force record for the Parkinson dataset obtained from PhysioNET database is shown in Figure 1.

It is necessary to create features belonging to the dataset before machine learning algorithms are applied to the Parkinson dataset. Besides, the raw version of the signal and the interaction of signals of each sensor were also used while creating features. Then, the raw signals that belong to the subjects and the interaction signals were transformed by signal transform methods. Fourier, wavelet and Hilbert-Huang transform were used in the transformation of the signals. Features obtained after each transform process were selected using the feature selection algorithm called *OneRAttributeEval*. Thus, features giving little information were eliminated while features giving more information were selected. In the course of selecting features, *OneRAttributeEval* algorithm was run using 10-fold cross validation method. As a result, the correct classification percentages of each feature were obtained. The correct classification ratio is the mean of ten folds because of the fact that 10-fold cross validation method was applied in the experiments. Thereby, results in the form of $(mean \pm deviation)$ were obtained for each feature. Features corresponding to the rule $(mean - deviation \geq 70)$ were selected. Consequently, only hundreds of features among tens of thousands of features were chosen. How features are created from the Parkinson dataset and the process of selecting relevant features among those are elucidated below in detail.

Let us designate 16 signals obtained from each foot of the subjects as: L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8, R_1, R_2, R_3, R_4, R_5, R_6, R_7 and R_8. Then, let us designate two signals

indicating the sum of signals that belong to each foot as: T_1 and T_2. Thus, we have 18 signals for each subject in total. Finally, we have obtained new signals by applying various basic functions to those 18 signals. Functions applying to these signals are:

$$f_1(x,y) = x + y,\ f_2(x,y) = x - y \text{ and } f_3(x,y) = |x - y|$$

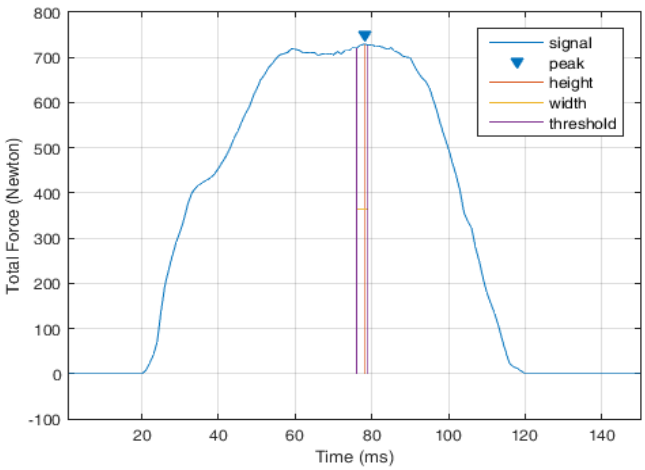


Figure 2. The peak analysis of force signal applied to the ground for the right foot that belongs to a Parkinsonian subject.

For instance, when these functions are applied for L_1 and L_2, three new signals would be formed, respectively as: L_1+L_2, L_1-L_2 and |L_1-L_2|. When these 18 basic signals interact according to these three functions, without any repetition, 459 new signals are obtained. Thus, with the 18 basic signals and 459 new signals, we ended up with 477 signals in total for each subject. After the signals were obtained, they were transformed using signal transform methods. Finally, the peak analyses of 477 raw signals and the transform signals obtained by transforming them were performed. The peak analysis that belongs to the right foot signal of a Parkinsonian subject is shown in Figure 2.

As a result of peak analysis, four vectors are obtained; namely, maximum extreme values, their locations, widths of the peaks and peak-height values of the signals. After performing the peak analyses, statistical features are obtained by applying the statistical functions to these four vectors that belong to each signal. By using the raw versions of some transform signals, the statistical features are obtained as well. The statistical functions applied to these signals are shown in Table 4. While some magnitudes in Table 4 are indicated by mathematical notation, others are explained as MATLAB functions.

The first transform method applied to the signals was Fourier transform. Through Fourier transform, a signal in time domain is transformed into frequency domain. With regard to this case, the frequency domains which belong to a control subject and a parkinsonian subject are shown in Figure 3. With 477 signals for each subject, the

number of features obtained by using Fourier transform was 85860 for each subject. Then, only 24 features from among these 85860 features were selected according to the condition($mean - deviation \geq 70$), using *OneRAttributeEval* algorithm.

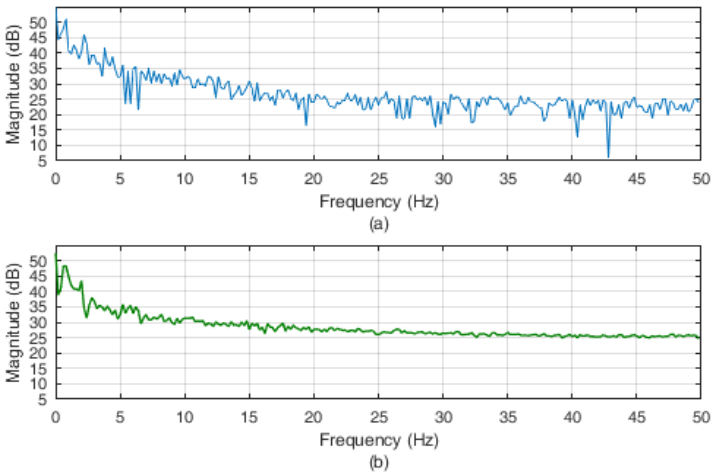


Figure 3. One-sided magnitude spectrum of the right foot signals which belong to (a) Control subject and (b) parkinsonian subject.

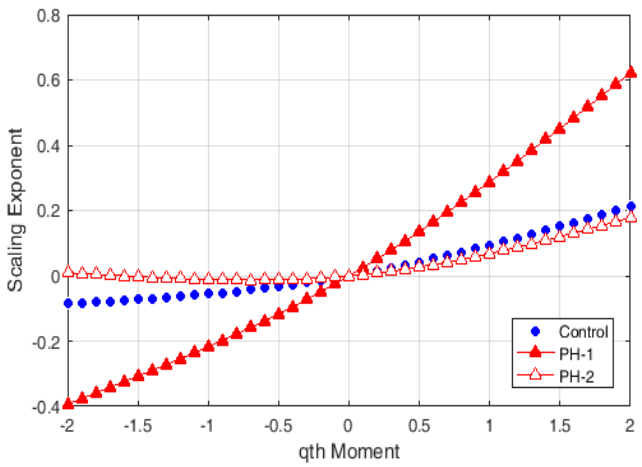


Figure 4. The scaling exponents of right foot signals of the Control subject (Control) and two parkinsonian subjects (PH-1 and PH-2) – q^{th} moment graph.

The second transform method applied to the signals was wavelet transform. Through it, time-frequency-scale resolution of a signal is obtained. In this study, seven wavelet families were used for the wavelet transform of the signals. These wavelet families are: Gaus, biorsplines, coiflets, daubechies, dmeyer, haar, and symlets. Gaus wavelet provides continuous wavelet transform whereas the others provide discrete wavelet transform.

With 477 signals for each subject, the number of features obtained by using signal transform by a second degree Gaussian wavelet was 92061. Then, only 23 features from

those 92061 features were selected according to the condition ($mean - deviation \geq 70$) using *OneRAttributeEval* algorithm. The scaling force of force signals that belonged to the subjects was also computed. Scaling force is a scale which indicates whether a signal is mono-fractal or multi-fractal. The scaling exponents - q^{th} moment graph of force signals belonging to the control subject and two parkinsonian subjects are shown in Figure 4. Accordingly, these signals are multi-fractal because the graph according to q^{th} moment of the scaling exponents of force signals that belong to the subjects in Figure 4 is a curvilinear function. Otherwise, it would be a mono-fractal signal. The wavelet coherence of the force signal that belongs to a parkinsonian subject according to the time signal is shown in Figure 5. On the other hand, the wavelet coherence of the force signal that belongs to a control subject according to the time signal is shown in Figure 6.

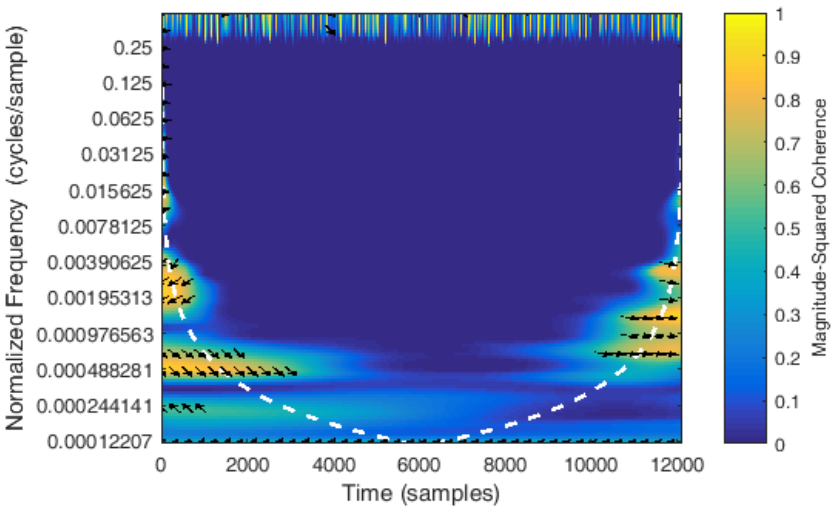


Figure 5. Time-dependent wavelet coherence of right foot signal of a parkinsonian subject.

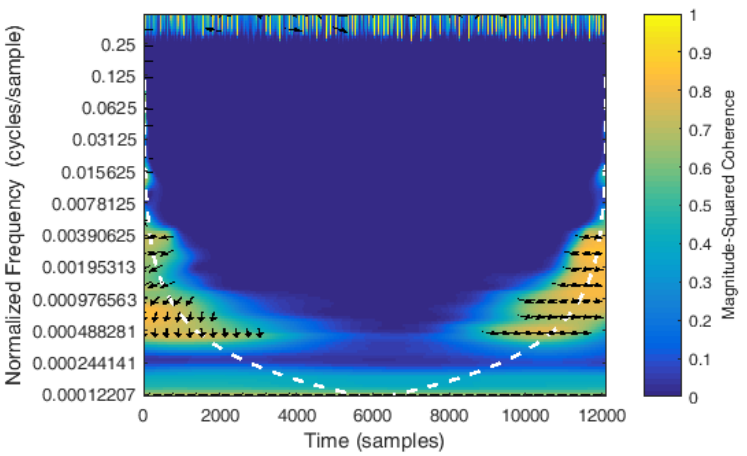


Figure 6. Time-dependent wavelet coherence of right foot signal of a control subject.

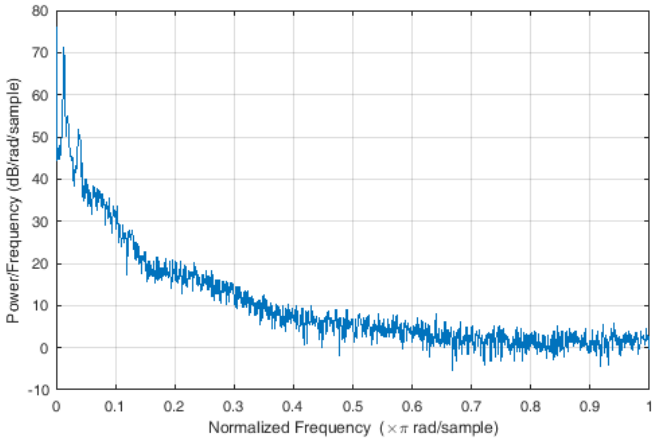


Figure 7. The Welch power spectral density estimate of the right foot signal of a parkinsonian subject.

The wavelet coherence is a measure which demonstrates the relation between two signals. While the force signals of each subject diverge from each other, the time function remains the same for all. Our purpose here is to measure the interrelations of other signals according to a constant signal.

The total number of features obtained for each subject for BiorSplines3.3, Coiflets2, Dmeyer, Haar, Symlets2, Daubechies2, Daubechies3, Daubechies4, Daubechies5, Daubechies6 and Daubechies7 wavelets was 278091. Then, features were selected for each wavelet type from among these 278091 features according to the condition $(mean - deviation \geq 70)$ using *OneRAttributeEval* algorithm. The numbers of features obtained by BiorSplines3.3, Coiflets2, Dmeyer, Haar, Symlets2, Daubechies2, Daubechies3, Daubechies4, Daubechies5, Daubechies6 and Daubechies7 wavelets were respectively 146, 147, 149, 114, 126, 125, 142, 172, 132, 130, and 168.

The final transform method applied to the signals was HHT. Before creating features by HHT, the Welch power spectral density estimates of force signals were computed, and then, features were derived from these density estimates. The Welch power spectral density estimate of the right foot signal of a parkinsonian subject is shown in Figure 7. A signal is decomposed into its IMF's by HHT. Then, IFs are obtained using these IMFs. With 477 signals for each subject, the number of features obtained by using the HHT versions of these signals was 694512. Then, only 135 features were selected from among these 694512 features according to the condition $(mean - deviation \geq 70)$ by using *OneRAttributeEval* algorithm.

Upon obtaining features by using the signal transform methods and completing the selection process of high-information yielding features from among these features, training sets are created in such a structure as can be used by the WEKA tool. All machine learning algorithms that were applied to the Parkinson dataset have been run by the 10-fold cross validation method unless otherwise indicated.

RESULTS AND DISCUSSION

Two baseline predictors (random predictor and majority predictor, i.e., ZeroR algorithm) were applied to the Parkinson dataset before the application of machine learning algorithms. The results of two baseline predictors are shown in Figure 8. The detailed results are shown in Table 5 as well. According to these results, ZeroR classifier has higher classification accuracy compared to a random predictor over all the Parkinson datasets to which signal transform methods were applied. This is a result that is in accordance with the nature of these two predictors. The classification accuracy of other machine learning algorithms over these datasets is supposed to be higher than the predictions of these two predictors. Otherwise, the predictions to be made would not be acceptable predictions.

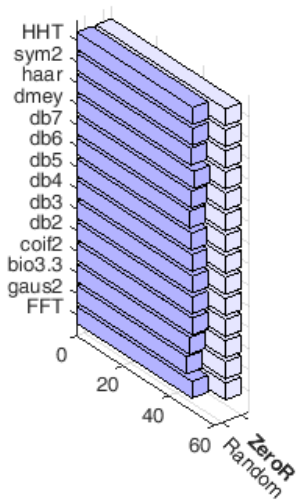


Figure 8. The classification accuracy of two baseline predictors over the Parkinson dataset to which various signal-transforms were applied.

91 machine learning algorithms in WEKA have been applied to the Parkinson dataset to which the 0-1 normalization method had been applied. Feature selection process was performed before the application of machine learning algorithms to these datasets. The feature selection method called the WrapperSubsetEval in WEKA was used for feature selection (in WEKA, all of the data were selected as training data, i.e., cross validation method was not performed. But, it was not deemed necessary to apply the cross validation method once again because the WrapperSubsetEval algorithm carries out 10-fold cross validation method). This method selects the feature set with which the selected machine learning algorithm gives the highest rate of accuracy as well as the lowest rate of error. It uses a variety of methods for feature selection. In this study, the BestFirst algorithm was preferred as the feature search method. The fact that the BestFirst

algorithm creates feature sets with higher accuracy was effective in the making of this choice. In addition, forward search was performed for the search process. Although the wrapper approach creates the feature set with the highest-accuracy, it is inadequate in terms of time cost. The wrapper method takes a long time particularly for algorithms of which the training process lasts long. Our general approach for this situation is to use the feature set with the highest-accuracy on the relevant Parkinson dataset for machine learning algorithms as well, the training and search time of which lasts long.

As with benchmark datasets, the machine learning algorithms were also compared with regard to learning approaches for the Parkinson datasets. Machine learning algorithms in WEKA are divided into seven main categories: bayes, functions, lazy, meta, misc, rules, and trees.

The first group of machine learning algorithms applied to the Parkinson datasets was the group of algorithms which have the Bayesian learning approach. Of these algorithms, the Bayesian Network (BN) algorithm has the highest classification accuracy over the Parkinson dataset to which HTT was applied, with an accuracy rate of 93.94%. Also, Aggregating One-Dependence Estimators (A1DE) algorithm has the highest classification accuracy over the Parkinson dataset to which Daubechies3 wavelet was applied, with an accuracy rate of 93.94%. Other algorithms with a Bayesian learning approach have yielded lower classification accuracy rates over the Parkinson datasets. The details regarding these results are shown in Table 6.

The second group of machine learning algorithms applied to the Parkinson datasets was the group of algorithms with the function learning approach. Of these algorithms, Linear Discriminant Analysis (LDA) algorithm has the highest classification accuracy over the Parkinson dataset to which Daubechies3 wavelet was applied, with an accuracy rate of 93.33%. The others have obtained lower classification accuracy rates over the Parkinson datasets. The details relevant to these results are shown in Table 7.

The third group of machine learning algorithms applied to the Parkinson datasets was the group of algorithms with the instance learning approach. Of these, Instance Based-k (IBk) algorithm has the highest classification accuracy over the Parkinson dataset to which HHT was applied, with an accuracy rate of 91.52%. IBk algorithm has the highest classification accuracy by 91.52% over the Parkinson dataset to which Daubechies4 wavelet was applied. The other algorithms have lower classification accuracies. The details regarding the results are shown in Table 8.

The fourth group of machine learning algorithms applied to the Parkinson datasets was the group of algorithms which have miscellaneous learning approaches. Of these algorithms, HyperPipes algorithm has the highest accuracy over the Parkinson dataset to which Daubechies4 wavelet was applied, with an accuracy rate of 89.70%. Other algorithms have lower classification accuracy rates. Detailed results are shown in Table 9.

Table 3. The comparison of the results of the proposed method and other studies over the same Parkinson dataset

Studies	Data type	Signal transform/Feature creation method	Classification algorithm	k-fold cross validation	Accuracy rate (%)	Mean accuracy rate (%)
The proposed method	Vertical force signals applied to the ground	Daubechies3 and statistical methods	A1DE	Use all data as training set	93.94 (exclusive of the mean)	93.03
				2-fold CV	90.91	
				5- fold CV	93.33	
				10- fold CV	93.94	
				LOOCV	93.94	
(Zeng et al., 2016)	Vertical force signals applied to the ground	Statistical methods	Radial based function neural network	Use all data as training set	99.40 (exclusive of the mean)	93.78
				2- fold CV	91.57	
				5- fold CV	96.39	
				LOOCV	93.37	
(Shrivastava et al., 2017)	Vertical force signals applied to the ground	Binary bat algorithm	Artificial neural network	CV	93.08 (Ga's data)	93.12
					96.35 (Ju's data)	
					89.93 (Si's data)	
(Jane et al., 2016)	Vertical force signals applied to the ground	-	Q-back propagated time delay neural network	10-fold CV	93.10 (Ga's data)	91.53
					91.67 (Si's data)	
					89.66 (Ju's data)	
(Daliri, 2013)	Vertical force signals applied to the ground	Short-time Fourier transform and statistical methods	SVM (Chi-squared distance kernel)	50% of the dataset is for training, the rest of it is for test (the process was repeated 20 times)	91.20	91.20
(Ertuğrul et al., 2016)	Vertical force signals applied to the ground	Statistical methods	Shifted one-dimensional binary patterns	10-fold CV	88.89	88.89
(Lee & Lim, 2012)	Vertical force signals applied to the ground	Wavelet transform and statistical methods	Neural network with weighted fuzzy membership functions	-	77.33	77.33

The fifth group of machine learning algorithms applied to the Parkinson datasets was the group of algorithms using a rule-based learning approach. Of these, the Decision Table algorithm has the highest accuracy over the dataset with HHT by 90.30%. Other

rule-based algorithms have lower classification accuracy rates compared to it. The details regarding the results are shown in Table 10.

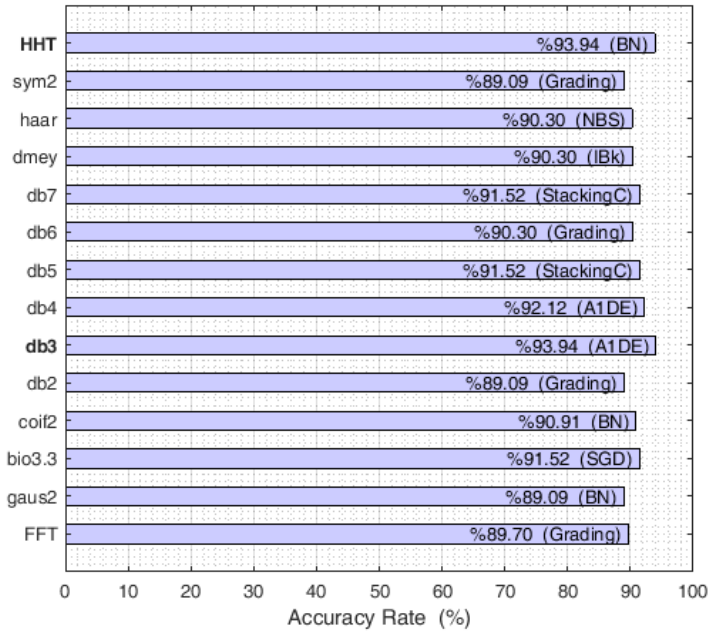


Figure 9. The highest classification accuracy rates of machine learning algorithms in WEKA over each Parkinson dataset to which various signal transforms are applied and the relevant machine learning algorithms.

The sixth group of machine learning algorithms applied to the Parkinson datasets was the group of algorithms using a decision tree learning approach. J48 (also known as C4.5) algorithm is the one with the highest classification accuracy over the Parkinson dataset with Daubechies3 by 90.30%. The others yielded lower classification accuracy rates. The details concerning these results are shown in Table 11.

The last group of machine learning algorithms applied to the Parkinson datasets was the group of algorithms using the ensemble/meta learning approach. Of these, END, Filtered Classifier, MultiClass Classifier, OrdinalClass Classifier, Weighted Instances Handler Wrapper (WIHW), MultiScheme and StackingC algorithms have the highest classification accuracy over the Parkinson dataset with Daubechies3, with an accuracy rate of 93.94%. END, Filtered Classifier, MultiClass Classifier, OrdinalClass Classifier, WIHW and Grading algorithms have the highest classification accuracy over the Parkinson dataset with HHT as well, with an accuracy rate of 93.94%. The accuracy rates of others are lower when compared to these results. The details are shown in Table 12 and Table 13.

The highest classification accuracy rates of the machine learning algorithms in WEKA over each Parkinson dataset as well as the machine learning algorithms which yield them are shown in Figure 9.

The table featuring the comparative results of our method and other studies over the Parkinson dataset (Frenkel-Toledo et al., 2005a, 2005b; Hausdorff et al., 2007; Yogev et al., 2005) obtained from PhysioNET (Goldberger et al., 2000) database is shown in Table 3. According to these results; the classification accuracy rate of our study and those of two other studies (Shrivastava et al., 2017; Zeng et al., 2016) which correspond to 93% approximately, are higher when compared to the rest. However, in one (Shrivastava et al., 2017) of these two studies the data was used separately, without combining three datasets that actually make up the original Parkinson dataset. The classification accuracy rates of the study conducted by Zeng et al. [11] are 91.57%, 96.39% and 93.37% respectively, according to 2- fold CV, 5- fold CV and leave-one-out cross validation (LOOCV) methods. Besides, when using all data as the training set, its accuracy rate is 99.40%. The accuracy rates pertaining to the proposed method are 90.91%, 93.33%, 93.94% and 93.94% respectively, according to 2- fold CV, 5- fold CV, 10-fold CV and leave-one-out cross validation (LOOCV) methods. When using all data as the training set, its accuracy rate is 93.94%. Accordingly, the standard deviation of the results regarding the study of Zeng et al. is 2.44, whereas ours is 1.6. Hereby, we can suggest that our results are a little more reliable in comparison as the standard deviation is lower. Furthermore, when all data is used as the training set, its standard deviation turns out to be 3.44, whereas ours is 1.44.

CONCLUSION

In this study, we proposed a method to distinguish individuals with PD from healthy ones by using the gait data obtained from PhysioNET database. First, we transformed the original signals using signal processing methods and generated tens of thousands of features from the transformed signals using statistical magnitudes. Later, we applied two feature selection methods to the features in a two-phased manner. As a result of the experiments we conducted, we obtained the highest classification accuracy over the dataset with daubechies3 using A1DE classifier. We can conveniently suggest that, when the results are considered, our method is more stable in comparison with the other studies although the difference of mean classification accuracy between our closest competitor and our method is only 0.75%. Moreover, we have been planning to significantly increase the classification accuracy using a new ensemble learning algorithm that we have been developing.

ACKNOWLEDGMENTS

We would like to express our thanks to Istanbul Aydın University for the support it has provided to this study as well as to Mahir Rasulov and Ahmad Babanlı for their kind contribution of feedbacks.

APPENDIX A: STATISTICAL MAGNITUDES

Table 4. The statistical magnitudes applied to the signals

Statistical magnitudes	**Formulas / Functions**
Maximum	max(x)
Minimum	min(x)
Mean	$\frac{1}{N}\sum_{i=1}^{N} x_i$
Standard deviation	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2}$
Skewness	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}\right)^3}$
Kurtosis	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^2}$
Median	median(x)
Range	max(x) - min(x)
Mean absolute deviation	mad(x, 0);
Median absolute deviation	mad(x, 1)
Sum	sum(x)
Shannon entropy	wentropy(x,'shannon')
Logarithmic entropy	wentropy(x,'log energy')
Generalized vector 1-norm	norm(x, 1)
Euclidean vector norm	norm(x)

APPENDIX B. THE DETAILED CLASSIFICATION RESULTS

Table 5. The classification results of the baseline algorithms over the Parkinson dataset

Transform Type	Baseline	
	Random Predictor[1]	**Majority/ZeroR**
	50.24 ± 8.76	56.36
ın2	48.00 ± 3.58	56.36
lines3.3	49.15 ± 5.38	56.36
s2	50.24 ± 3.03	56.36
hies2	50.24 ± 6.52	56.36
hies3	50.42 ± 4.78	56.36
hies4	50.18 ± 6.88	56.36
hies5	50.00 ± 4.14	56.36
hies6	49.33 ± 5.91	56.36
hies7	50.00 ± 4.43	56.36
r	50.97 ± 5.65	56.36
	49.70 ± 6.70	56.36
s2	50.42 ± 4.73	56.36
	50.55 ± 8.30	56.36
	49.96	56.36
Rank	0	14

periments were performed ten times.

Table 6. The classification results of the Bayesian learning algorithms over the normalized Parkinson dataset

Transform Type	Bayesian Learning						
	A1DE	**A2DE**	**BayesNet**	**Bayesian Logistic Regression**	**NaiveBayes**	**NaiveBayes Simple**	**NaiveBayes Updateable**
	89.09	88.48	**89.09**	84.24	86.67	85.45	86.67
ın2	86.67	88.48	**89.09**	86.06	86.67	86.67	86.67
ines3.3	87.88	86.06	**90.91**	86.67	87.88	89.09	87.88
2	87.27	90.30	**90.91**	87.88	87.88	88.48	87.88
hies2	**88.48**	86.67	**88.48**	87.88	87.27	86.06	87.27
hies3	**93.94**	90.91	92.12	87.27	89.70	89.70	89.70
hies4	**92.12**	90.30	91.52	86.06	86.67	86.67	86.67
hies5	88.48	**89.09**	87.88	87.27	87.88	87.88	87.88
hies6	**90.30**	89.09	89.70	83.03	88.48	85.45	88.48
hies7	89.70	90.30	**91.52**	86.67	89.09	89.09	89.09
r	85.45	87.27	**87.88**	87.27	**87.88**	**87.88**	**87.88**
	89.70	89.70	89.70	85.45	87.88	**90.30**	87.88
2	**88.48**	86.67	**88.48**	87.88	87.27	86.06	87.27
	89.70	89.70	**93.94**	86.06	87.88	89.09	87.88
	89.09	88.79	**90.09**	86.41	87.79	87.71	87.79
ank	6	1	**9**	0	1	2	1

Table 7. The classification results of the Function learning algorithms over the normalized Parkinson dataset

orm Type	Function Learning																
	FLDA	Kernel Logistic Regression[2]	LDA	libLINEAR	libSVM	Logistic	MLP Classifier[2]	Multilayer Perceptron[2]	Multilayer PerceptronCS[2]	QDA	RBF Classifier[2]	RBF Network[2]	SGD	Simple Logistic[2]	SMO	SPegasos[2]	Voted Perceptron
	81.82	86.67	84.85	86.06	86.06	86.67	82.42	81.82	81.82	83.64	83.64	78.18	85.45	84.85	85.45	83.64	83.64
n2	81.82	87.27	84.85	86.06	86.67	87.27	83.03	80.61	80.61	84.24	84.85	79.39	84.85	85.45	85.45	81.82	83.64
ines3.3	87.27	86.67	89.70	84.24	87.27	88.48	87.88	87.88	87.88	86.06	85.45	75.15	91.52	90.91	87.88	84.85	86.67
2	85.45	85.45	86.06	87.88	87.27	85.45	82.42	82.42	82.42	87.27	84.85	76.97	88.48	84.85	86.67	84.85	85.45
hies2	85.45	86.67	88.48	84.24	87.88	83.03	79.39	80.61	80.61	83.03	86.06	84.85	83.64	86.67	86.67	84.24	85.45
hies3	86.06	90.30	93.33	89.09	87.88	87.27	86.06	87.27	87.27	87.27	84.24	81.82	88.48	87.88	88.48	90.30	86.67
hies4	90.30	88.48	86.67	88.48	84.24	86.67	87.88	87.88	87.88	88.48	89.09	82.42	86.67	87.88	88.48	89.09	86.67
hies5	85.45	84.85	86.67	86.67	86.06	86.67	82.42	84.85	84.85	86.67	86.06	79.39	87.88	86.06	86.67	85.45	86.06
hies6	86.67	87.88	87.88	84.24	83.64	87.27	83.64	83.03	83.03	82.42	85.45	75.76	83.64	83.03	83.64	85.45	84.24
hies7	89.09	89.70	84.85	86.67	88.48	89.70	87.88	83.64	83.64	89.09	87.27	75.76	88.48	86.06	86.06	86.67	86.67
r	90.30	83.64	84.85	84.85	85.45	87.27	83.64	80.61	80.61	83.64	85.45	79.39	86.06	86.06	86.67	84.24	89.70
	88.48	83.64	88.48	89.09	90.30	86.06	86.67	80.00	80.00	85.45	84.85	78.18	86.06	83.03	87.88	84.24	86.67
2	85.45	86.67	88.48	84.24	87.88	85.45	79.39	80.61	80.61	83.03	86.06	84.85	83.03	86.67	87.88	84.24	85.4
	87.27	80.61	87.27	87.27	88.48	87.27	83.03	80.00	80.00	86.06	85.45	80.61	87.88	81.21	89.09	83.64	86.67
	86.49	86.32	87.32	86.36	86.97	86.75	83.98	82.95	82.95	85.45	85.63	79.48	86.58	85.76	86.93	85.19	85.98
ank	2	4	4	0	1	3	0	0	0	0	0	0	3	0	1	0	0

se learner used in feature selection is the base learner that yields the highest accuracy for the signal transform applied. For example, the base learner that was selected for er transform from among the Function learning algorithms is the Logistic classifier.

Table 8. The classification results of the Instance/Lazy learning algorithms over the normalized Parkinson dataset

Transform Type	Instance/Lazy Learning					
	IBk	**IbkLG[2]**	**Kstar[2]**	**LocalKnn[2]**	**LWL**	**RseslibKnn[2]**
	81.21	81.21	76.97	80.00	79.39	80.61
n2	83.64	83.64	80.00	82.42	79.39	84.24
ines3.3	83.64	83.64	83.64	80.61	79.39	83.03
2	87.27	87.27	81.82	85.45	82.42	86.06
hies2	84.24	84.24	77.58	80.00	80.00	79.39
hies3	87.27	87.27	80.61	83.03	82.42	83.03
hies4	91.52	91.52	84.85	83.64	82.42	86.06
hies5	90.91	90.91	79.39	86.06	80.61	81.82
hies6	88.48	88.48	86.06	83.03	82.42	80.61
hies7	87.88	87.88	80.00	81.21	79.39	84.24
r	90.30	90.30	85.45	83.64	80.61	81.21
	84.85	84.85	76.97	80.61	82.42	79.39
2	84.24	84.24	77.58	80.00	80.00	79.39
	91.52	91.52	80.00	80.61	83.03	84.85
	86.93	86.93	80.78	82.17	80.99	82.42
ank	13	13	1	0	0	1

Table 9. The classification results of the miscellaneous learning algorithms over the normalized Parkinson dataset

Transform Type	-			
	CHIRP	**FLR**	**HyperPipes**	**VFI**
	82.42	72.73	77.58	65.45
n2	**81.21**	72.73	78.18	65.45
ines3.3	82.42	80.61	82.42	**83.03**
2	83.64	83.03	**86.67**	86.06
hies2	83.64	78.79	83.64	**84.24**
hies3	86.06	77.58	86.67	**87.27**
hies4	81.82	80.61	**89.70**	87.27
hies5	83.03	82.42	82.42	**84.24**
hies6	83.64	80.00	**84.24**	79.39
hies7	84.24	80.61	**86.06**	85.45
r	83.03	83.03	83.64	**84.24**
	83.03	79.39	81.82	**86.06**
2	80.00	78.79	83.64	**84.24**
	84.85	81.21	89.09	**87.27**
	83.07	79.40	**83.98**	82.12
ank	2	0	4	**8**

Table 10. The classification results of the Rule-based learning algorithms over the normalized Parkinson dataset

Transform	Rule-based Learning												
	Conjunctive Rule	Decision Table	DTNB[2]	FURIA[2]	JRip	MODLEM[2]	NNge	OLM	OneR	PART	Ridor	RoughSet[2]	Vfdr
	80.61	86.67	84.24	79.39	80.61	78.18	83.03	72.73	78.79	81.82	79.39	74.55	86.06
ın2	80.61	84.85	81.21	81.82	83.64	78.79	83.03	75.15	78.79	83.03	83.03	82.42	86.67
lines3.3	81.21	81.82	82.42	78.18	83.64	78.79	83.03	78.18	78.79	86.67	83.64	80.00	86.67
;2	80.61	82.42	80.00	79.39	84.85	78.79	82.42	78.79	80.00	82.42	84.85	78.79	87.27
hies2	80.61	83.64	81.82	80.00	83.64	81.21	81.21	83.64	79.39	83.64	83.03	76.97	87.88
hies3	80.61	88.48	87.27	81.82	86.67	78.18	84.24	79.39	80.00	89.70	83.03	78.18	87.88
hies4	81.21	85.45	83.64	84.85	86.67	83.03	82.42	81.82	79.39	89.70	80.00	81.82	86.67
hies5	80.00	86.06	80.00	80.00	83.03	80.61	82.42	80.61	79.39	83.03	83.64	80.00	88.48
hies6	80.00	86.67	86.06	80.00	83.64	80.61	82.42	81.21	79.39	82.42	84.24	84.24	88.48
hies7	80.61	86.67	83.64	80.61	82.42	78.79	84.85	83.64	78.18	84.85	83.03	79.39	89.09
r	80.61	81.82	79.39	79.39	82.42	78.79	83.03	80.00	78.18	87.88	83.03	75.76	86.67
	81.82	84.85	79.39	80.61	88.48	77.58	83.03	81.21	78.79	87.27	84.24	84.24	89.09
;2	80.61	83.64	81.82	83.03	83.03	81.21	81.21	83.64	79.39	83.64	83.03	76.97	87.88
	78.18	90.30	86.67	81.82	81.82	80.00	81.82	78.79	80.00	86.67	84.85	86.06	87.88
	80.52	85.24	82.68	80.78	83.90	79.61	82.73	79.91	79.18	85.20	83.07	79.96	87.62
ank	0	2	0	0	0	0	0	0	0	4	0	0	9

Table 11. The classification results of the Decision tree learning algorithms over the normalized Parkinson dataset

orm Type	Decision Tree Learning																
	ADTree	BFTree	CDT	Decision Stump	Extra Tree	FT[2]	Hoeffding Tree	J48	J48 Consolidated	J48graft	LADTree[2]	LMT[2]	NBTree[2]	Random Forest[2]	RandomTree	REPTree	SimpleCart[2]
	83.03	79.39	81.21	77.58	76.97	84.24	86.06	81.21	83.64	80.61	79.39	83.03	78.18	81.82	77.58	80.00	78.18
ın2	83.03	79.39	81.82	77.58	81.21	82.42	86.67	83.64	82.42	83.03	81.82	83.64	81.82	84.85	79.39	81.21	80.00
lines3.3	83.64	84.24	84.24	79.39	78.79	84.24	86.67	83.64	86.67	84.24	76.97	84.24	76.97	80.00	84.24	81.82	72.73
;2	86.06	83.03	84.24	79.39	74.55	81.21	87.27	83.03	82.42	85.45	75.76	83.64	76.97	81.21	81.21	83.64	76.97
hies2	83.03	85.45	83.03	79.39	79.39	81.21	87.88	83.64	81.82	83.03	78.79	86.06	80.61	83.03	80.00	81.21	75.76
hies3	89.09	84.24	82.42	79.39	81.82	81.21	87.88	90.30	87.88	87.27	84.24	83.03	86.67	84.24	81.82	81.82	81.21
hies4	86.06	86.06	80.61	81.21	78.79	80.00	86.67	86.06	88.48	84.85	85.45	83.03	83.03	84.85	86.06	81.82	86.06
hies5	81.82	84.85	83.64	79.39	79.39	82.42	88.48	79.39	81.21	83.03	72.73	84.24	78.79	83.64	84.24	83.03	75.15
hies6	85.45	83.64	80.00	80.00	80.00	79.39	88.48	85.45	81.82	86.06	80.00	82.42	79.39	83.03	80.00	80.00	72.73
hies7	87.88	83.03	84.24	80.00	80.61	84.85	89.09	80.61	83.64	83.64	78.79	82.42	80.61	84.24	82.42	78.79	75.15
r	83.64	86.06	81.21	80.00	78.79	80.61	86.67	85.45	84.85	84.85	78.79	81.82	84.85	82.42	84.24	83.03	75.15
	87.88	85.45	83.03	78.79	81.82	77.58	89.09	89.70	86.06	86.06	82.42	80.61	83.03	82.42	84.85	80.61	85.45
;2	83.03	85.45	83.03	79.39	76.97	81.21	87.88	84.85	81.82	83.03	78.79	86.06	81.21	85.45	80.00	81.21	75.76
	89.70	82.42	82.42	79.39	76.97	80.61	87.88	85.45	89.70	88.48	89.09	81.21	84.85	86.67	84.85	84.24	80.61
	85.24	83.76	82.51	79.35	79.01	81.51	87.62	84.46	84.46	84.55	80.22	83.25	81.21	83.42	82.21	81.60	77.92
ank	1	0	0	0	0	0	10	2	3	0	0	0	0	0	0	0	0

Table 12. The classification results of the Ensemble/Meta learning algorithms over the normalized Parkinson dataset (the first group)

Transform	Ensemble/Meta Learning												
	AdaBoost-M1[3]	Bagging[3]	Classification via Cluctering[4]	Classification via Regression[4]	Dagging[3]	Decorate[3]	END[3]	Filtered Classifier[3]	Iterative Classifier Optimizer[4]	Logit Boost[4]	MultiBoostAB[3]	MultiClass Classifier[3]	MultiClass Classifier Update-able[4]
	85.45	86.67	77.58	77.58	82.42	88.48	89.09	89.09	79.39	78.79	84.24	89.09	83.64
ın2	86.67	87.88	81.82	81.82	85.45	85.45	88.48	88.48	78.79	78.79	86.06	88.48	81.21
lines3.3	84.85	86.06	81.82	82.42	80.61	91.52	91.52	84.85	77.58	84.85	87.27	91.52	91.52
;2	84.24	86.67	84.24	82.42	85.45	87.88	90.91	90.91	82.42	80.61	86.06	90.91	84.24
hies2	88.48	87.27	79.39	80.61	80.00	83.64	88.48	88.48	80.61	81.21	86.06	88.48	81.21
hies3	87.88	89.09	80.61	82.42	83.64	89.70	93.94	93.94	83.64	83.64	89.70	93.94	83.64
hies4	89.09	89.70	78.79	82.42	80.61	89.70	92.12	92.12	85.45	86.06	90.30	92.12	80.61
hies5	90.30	89.09	80.61	79.39	84.24	90.30	90.91	76.97	78.79	80.00	90.91	90.91	82.42
hies6	89.09	83.64	80.00	81.21	81.21	88.48	90.30	90.30	83.03	81.82	87.88	90.30	82.42
hies7	87.27	89.09	81.82	78.79	86.67	89.70	91.52	91.52	83.03	84.24	90.91	91.52	80.00
r	89.09	87.27	79.39	78.79	81.21	90.30	90.30	81.21	81.21	81.21	89.09	90.30	80.61
	84.85	89.09	86.06	81.82	87.27	89.70	90.30	87.88	83.64	81.82	87.88	90.30	83.03
;2	88.48	87.27	79.39	80.61	80.00	83.64	88.48	88.48	80.61	81.21	86.06	88.48	81.21
	84.85	89.70	82.42	77.58	84.24	89.09	93.94	93.94	82.42	83.03	89.09	93.94	81.82
	87.19	87.75	81.00	80.56	83.07	88.40	90.74	88.44	81.47	81.95	87.97	90.74	82.68
tank	0	0	0	0	0	2	9	6	0	0	0	9	1

ıse learner which classified the given signal transform with the highest accuracy rate was used. Wrapper method was used for feature selection. The default base learner ısed for Wrapper method as a base learner.

t parameters were used. Wrapper method was used for feature selection. The default base learner was used for Wrapper method as a base learner.

Table 13. The classification results of the Ensemble/Meta learning algorithms over the normalized Parkinson dataset (the second group)

Transform	Ensemble/Meta Learning													
	Ordinal Class Classifier[3]	Raced Incremental LogitBoost[4]	Random Committee[4]	Randomizable Filtered Classifier[3]	Random SubSpace[3]	RealAda-Boost[3]	Rotation Forest[3]	Threshold Selector[3]	Weighted Instances Handler Wrapper[3]	Grading[5]	Multi-Scheme[5]	Stacking[5]	StackingC[5]	Vote[5]
	89.09	56.36	76.36	77.58	87.27	88.48	84.85	87.88	89.09	89.70	89.09	88.48	87.88	89.09
ın2	88.48	56.36	81.21	78.18	87.88	88.48	83.03	86.06	88.48	88.48	89.09	88.48	87.88	89.09
ines3.3	91.52	56.36	84.24	85.45	83.03	83.64	90.30	91.52	91.52	89.09	90.30	91.52	91.52	91.52
2	90.91	56.36	81.21	78.18	86.06	90.30	83.03	84.24	90.91	88.48	90.91	87.27	90.91	86.67
hies2	88.48	56.36	78.79	81.82	83.64	87.88	80.00	81.82	88.48	89.09	87.88	88.48	88.48	88.48
hies3	93.94	56.36	83.64	82.42	89.70	84.85	86.06	89.70	93.94	93.33	93.94	92.12	93.94	89.09
hies4	92.12	56.36	83.64	81.21	90.30	87.27	82.42	87.88	92.12	90.91	92.12	89.70	90.91	89.70
hies5	90.91	56.36	82.42	76.97	85.45	90.91	84.24	90.30	90.91	86.67	90.91	84.24	91.52	91.52
hies6	90.30	56.36	82.42	85.45	84.85	89.70	81.21	86.06	90.30	90.30	90.30	89.70	89.09	83.64
hies7	91.52	56.36	85.45	80.61	89.09	88.48	85.45	87.27	91.52	90.91	91.52	90.30	91.52	91.52
r	90.30	56.36	78.79	77.58	83.03	89.09	81.21	89.09	90.30	84.85	90.30	84.24	90.30	90.30
	90.30	56.36	81.21	81.21	89.70	84.24	88.48	85.45	90.30	89.09	90.30	87.88	89.70	84.85
2	88.48	56.36	78.79	81.82	83.64	87.88	80.00	81.82	88.48	89.09	87.88	88.48	88.48	88.48
	93.94	56.36	87.27	79.39	91.52	92.12	83.64	92.12	93.94	93.94	80.00	92.73	93.33	84.85
	90.74	56.36	81.82	80.56	86.80	88.09	83.85	87.23	90.74	89.57	89.61	88.83	90.39	88.49
ank	9	0	0	0	0	0	0	1	9	5	8	1	6	5

o base learners which classified the given transformed signal with the highest accuracy rate were selected. The wrapper method was used for feature selection. The base r with the highest accuracy rate for the wrapper method was used as a base learner.

REFERENCES

Adams, R. D., Victor, M., & Ropper, A. H., (1997). *Principles of Neurology* (6th ed.). New York, NY: McGraw-Hill.

Aydın, F., (2017). *Analyzing Parkinson's Diseases Data by Machine Learning Methods (Doctoral dissertation)*. Istanbul Aydin University, Istanbul, Turkey.

Bermingham, M. L., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., … Haley, C. S., (2015). Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Scientific Reports*, *5*(1), 10312. https://doi.org/10.1038/srep10312.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems*, *34*(3), 483–519. https://doi.org/10.1007/s10115-012-0487-8.

Brown, J. W., & Churchill, R. V., (1993). *Fourier Series and Boundary Value Problems* (50th ed.). Mc Graw-Hill.

Colominas, M. A., Schlotthauer, G., & Torres, M. E., (2014). Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control*, *14*, 19–29. https://doi.org/10.1016/j.bspc.2014.06.009.

Daliri, M. R., (2013). Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease. *Biomedical Signal Processing and Control*, *8*(1), 66–70. https://doi.org/10.1016/j.bspc.2012.04.007.

Deo, R. C., (2015). Machine learning in medicine. *Circulation*, *132*(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593.

Ertuğrul, Ö. F., Kaya, Y., Tekin, R., & Almalı, M. N., (2016). Detection of Parkinson's disease by Shifted One Dimensional Local Binary Patterns from gait. *Expert Systems with Applications*, *56*, 156–163. https://doi.org/10.1016/j.eswa.2016.03.018.

Flandrin, P., (1999). *Time-Frequency/Time-Scale Analysis*. Academic Press.

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., & Hausdorff, J. M., (2005a). Effect of gait speed on gait rhythmicity in Parkinson's disease: variability of stride time and swing time respond differently. *Journal of NeuroEngineering and Rehabilitation*, *2*(1), 23. https://doi.org/10.1186/1743-0003-2-23.

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., & Hausdorff, J. M., (2005b). Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. *Movement Disorders*, *20*(9), 1109–1114. https://doi.org/10.1002/mds.20507.

Gibberd, F. B., (1986). The management of Parkinson's disease. *The Practitioner*, *230*(1412), 139–146. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3703811.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., … Stanley, H. E., (2000). PhysioBank, PhysioToolkit, and PhysioNet :

Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, *101*(23), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215.

Gröchenig, K., (2001). *Foundations of Time-Frequency Analysis*. Boston, MA: Birkhäuser Boston. https://doi.org/10.1007/978-1-4612-0003-1hausdorff, J. M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., & Giladi, N., (2007). Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *European Journal of Neuroscience*, *26*(8), 2369–2375.https://doi.org/10.1111/j.1460-9568.2007.05810.xhuang,

Huang, N. E., (2005). Introduction to the Hilbert-Huang Transform and Its Related Mathematical Problems. In N. E. Huang & S. S. P. Shen (Eds.), *Hilbert-Huang Transform and Its Applications* (pp. 1–26). Singapore: World Scientific Publishing.

Huang, N. E., Long, S. R., & Shen, Z., (1996). The Mechanism for Frequency Downshift in Nonlinear Wave Evolution. *Advances in Applied Mechanics*, *32*, 59–117C. https://doi.org/10.1016/S0065-2156(08)70076-0huang, N. E., Shen, Z., & Long, S. R., (1999). A New View of Nonlinear Water Waves: The Hilbert Spectrum. *Annual Review of Fluid Mechanics*, *31*(1), 417–457. https://doi.org/10.1146/annurev.fluid.31.1.417huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Liu, H. H., (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193huang, N. E., & Wu, Z., (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, *46*(2), RG2006. https://doi.org/10.1029/2007RG000228.

James, G., Witten, D., Hastie, T., & Tibshirani, R., (2013). *An Introduction to Statistical Learning* (Vol. 103). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7.

Jane, Y. N., Nehemiah, H. K., & Arputharaj, K., (2016). A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease. *Journal of Biomedical Informatics*, *60*, 169–176. https://doi.org/10.1016/j.jbi.2016.01.014.

Kantz, H., & Schreiber, T., (2004). *Nonlinear Time Series Analysis* (2nd ed.). Cambridge University Press.

Kohavi, R., & John, G. H., (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324. https://doi.org/10.1016/S0004-3702 (97)00043-X.

Lee, S.-H., & Lim, J. S., (2012). Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. *Expert Systems with Applications*, *39*(8), 7338–7344. https://doi.org/10.1016/j.eswa.2012.01.084.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C., (2009). *Machine Learning, Neural and Statistical Classification* (2009th ed.). Overseas Press.

Morlet, J., (1981). Sampling Theory and Wave Propagation. In *Proc. 51st Annual International Meeting of the Society of Exploration Geophysi-cists*. Los Angeles.

Obermeyer, Z., & Emanuel, E. J., (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *New England Journal of Medicine*, *375*(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181.

Özaras, N., & Yalçın, S., (2002). Normal Yürüme ve Yürüme Analizi. *Turkish Journal of Physical Medicine and Rehabilitation*, *48*(3). Retrieved from http://www.ftrdergisi.com/content.php?id=262.

Prochazka, A., & Storek, M., (1995). Wavelet transform use for signal classification by self-organizing neural networks. In (*1995) Fourth International Conference on Artificial Neural Networks* (pp. 209–213). Cambridge, UK: The Institution of Engineering and Technology.

Shrivastava, P., Shukla, A., Vepakomma, P., Bhansali, N., & Verma, K., (2017). A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease. *Computer Methods and Programs in Biomedicine*, *139*, 171–179. https://doi.org/10.1016/j.cmpb.2016.07.029.

Tong, H., (2002). Nonlinear Time Series Analysis Since (1990): Some Personal Reflections. *Acta Mathematicae Applicatae Sinica, English Series*, *18*(2), 177–184. https://doi.org/10.1007/s102550200017.

Yogev, G., Giladi, N., Peretz, C., Springer, S., Simon, E. S., & Hausdorff, J. M., (2005). Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding? *European Journal of Neuroscience*, *22*(5), 1248–1256. https://doi.org/10.1111/j.1460-9568.2005.04298.x.

Zeng, W., Liu, F., Wang, Q., Wang, Y., Ma, L., & Zhang, Y., (2016). Parkinson's disease classification using gait analysis via deterministic learning. *Neuroscience Letters*, *633*, 268–278. https://doi.org/10.1016/j.neulet.2016.09.043.

ABOUT THE EDITORS

Professor Dr. Zafer Aslan
Istanbul Ayın University, Halit Aydın Campus, Istanbul, Turkey

She is a faculty member of Department of Computer Engineering in Istanbul Aydın University (IAU). She was Director of International Relations and Vice Rector of IAU. Her recent studies are on data mining, atmospheric physics, climate changing and renewable energy. She is the senior associate member of ICTP, the Board Member of OSTIV (International Scientific and Technical Soaring Organization), and Chairs of OSTIV Scientific Section and MET Panel. She is Executive Committee member of OWSD NC-Turkey and Advisor to the Rector of IAU. She is Member of Turkish Meteorological Engineering Chamber. Her awards are Prize of OWSD 25th Anniversary Video Competition, Federation Aeronautic International (FAI) The Paul Tissandier Diploma Award Lausanne, Switzerland; ICTP TRIL Award, Regular Associateship Award of ICTP, EGS Travel Award, Edinburgh. Grant of Scientific Committee on Oceanic Research (SQR), Lisbon, İTÜ Research Foundation and TUBITAK (The Scientific and Technical Research Council of Turkey), Borlange, Sweden, ICTP Grant for Training in the Italian Laboratories CNR, Rome, NATO Grant, University of Washington, USA, ICTP - scholarships, Trieste, Italy. She has 164 research and review articles published in international journals and proceedings, Articles in national journals and proceedings, three books, seven book chapters, 24 project reports.

Professor Dr. Funda Dökmen
Kocaeli University, Food and Agricultural Vocational School,
Campus of Arslanbey, Kartepe, Kocaeli-Turkey

She had her undergraduate degree at the Faculty of University of Uludağ in Bursa, Turkey. Her PhD is in the same field at the Faculty of Agricultural, Department of

Irrigation of Trakya University in Turkey. She studied at The University of London in England as Young Scientist. She is one of the academic members at Food and Agricultural Vocational School in Kocaeli University (KOÜ) and Head of Department of Plant and Animal Production. Her specialization is in the field of irrigation water quality, water management, irrigation and modelling, and agricultural meteorology. She published more than 100 scientific papers, articles and review papers in international and national journals and proceedings. She has published four different book chapters (Nova Publishing Company and Springer Publishing) in international relation to her topics. She has chaired and carried on committee memberships of some international and national scientific meetings in her field.

Professor Dr. Enrico Feoli

University of Trieste, Department of Life Sciences, Trieste, Italy

His recent research topics are mainly concerned with "Decision Support Systems for Sustainable Development Integrating Ecology and Economy". He is particularly interested in vegetation analysis in vegetation-climate interaction and in evaluating the environmental impacts of industrial development. He introduced in ecology new methods of data analysis for measuring hierarchical diversity and hierarchical similarity. He developed a series of original computer programs for information analysis and simulation of ecological gradients, and established vegetation databases of Italy. He worked to introduce the fuzzy set theory in vegetation sciences and ecology and developed a new method for analyzing the vegetation variation based on the association concept and fuzzy set theory. In these last years, he worked on similarity theory and on the use of probabilistic similarity functions to validate vegetation classifications. He was principal or local coordinator of several international projects in Africa, China, Europe, India, and South America, funded by European Commission and cooperated with UNIDO (United Nations Industrial Development Organization) in several projects**.** In 1998, he got the Distinguished Statistical Ecologist Award by INTECOL (International Ecological Society). He is one of the academic members University of Trieste, Italy.

Professor Dr. Abul Hasan Siddiqi

Senior Teaching Member, Sharda University, Delhi, Greater Noida, NCR, India

He is a distinguished scientist and professor emeritus at School of Basic Sciences and Research at Sharda University, Greater Noida, India. He is also a visiting consultant at the International Centre for Theoretical Physics (ICTP) (Trieste, Italy), Sultan Qaboos University (Muscat, Oman), King Fahd University of Petroleum and Minerals (KFUPM)

Dhahran, Saudi Arabia) and several other well-known universities of the world. He has a long association with ICTP (a UNESCO institution) in capacities of short – time visitor, long – duration visitor, senior associate, guests of the director, senior associate and ICTP visiting consultant to Turkey. He was awarded the German Academic Exchange Fellowship thrice to carry oft mathematical research in Germany. He has published more than 100 research papers jointly with his research collaborator, 6 books and edited proceeding of 10 international conferences, as well as supervised more than 30 PhD scholars. He is the founder secretary of the Indian Society of Industrial and Applied Mathematics (ISIAM), which was celebrated its Silver Jubilee in January 2016. He is editor – in- chief of Indian Journal of Industrial and Applied Mathematics, published by ISIAM and Industrial and Applied Mathematics, a book series with Springer.

INDEX

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

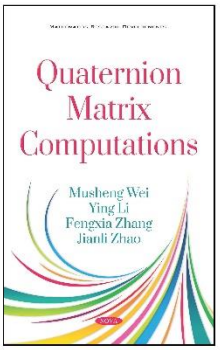
V

W

Y

Related Nova Publications

Quaternion Matrix Computations



Authors: Musheng Wei, Ying Li, Fengxia Zhang and Jianli Zhao

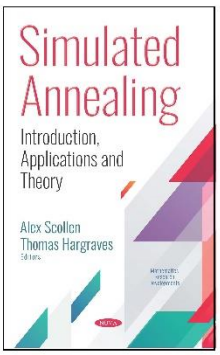
Series: Mathematics Research Developments

Book Description: In this monograph, the authors describe state-of-the-art real structure-preserving algorithms for quaternion matrix computations, especially the LU, the Cholesky, the QR and the singular value decomposition of quaternion matrices, direct and iterative methods for solving quaternion linear systems, generalized least squares problems, and quaternion right eigenvalue problems.

Hardcover ISBN: 978-1-53614-121-4
Retail Price: $160

Simulated Annealing: Introduction, Applications and Theory

Editor: Alex Scollen and Thomas Hargraves



Series: Mathematics Research Developments

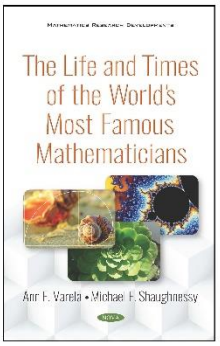
Book Description: The opening chapter of this book aims to present and analyze the application of the simulated annealing algorithm in solving parameter optimization problems of various manufacturing processes.

Hardcover ISBN: 978-1-53613-674-6
Retail Price: $195

To see a complete list of Nova publications, please visit our website at www.novapublishers.com

Related Nova Publications

The Life and Times of the World's Most Famous Mathematicians



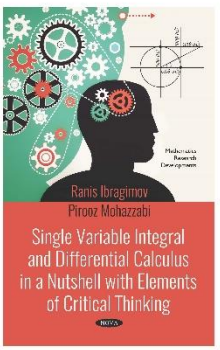
Authors: Ann F. Varela and Michael F. Shaughnessy

Series: Mathematics Research Developments

Book Description: Arithmetic, math, algebra, geometry, trigonometry, calculus and statistics are all part of the modern world, yet so few of us know where these fields of study originated. In this book, the lives and discoveries of various mathematicians will be explored and examined, so as to provide some insight into various mathematical functions and operations.

Hardcover ISBN: 978-1-53613-975-4
Retail Price: $195

Single Variable Integral and Differential Calculus in a Nutshell with Elements of Critical Thinking



Authors: Ranis Ibragimov and Pirooz Mohazzabi

Series: Mathematics Research Developments

Book Description: This book presents a variety of calculus problems concerning different levels of difficulty with technically correct solutions and methodological steps that look also correct, but that have obviously wrong results (like 0 = 1).

Hardcover ISBN: 978-1-53614-047-7
Retail Price: $195

To see a complete list of Nova publications, please visit our website at www.novapublishers.com