Modeling from Time Series of Complex Brain Signals

Galina S. Panayotova
Department Computer Science, University of Library Studies and Informational Technologies, Sofia, Bulgaria
Email: panayotovag@gmail.com

Dimitar A. Dimitrov
University "Prof. Dr. As. Zlatarov"- Burgas, Bulgaria
Email: dimitrovda@abv.bg

***Abstract*—Signals obtained from most of real-world systems, especially from living organisms, are irregular, often chaotic, non-stationary, and noise-corrupted. Since modern measuring devices usually realize digital processing of information, recordings of the signals take the form of a discrete sequence of samples (a time series). In the paper given a brief overview of the possibilities of such experimental data processing based on reconstruction and usage of a predictive empirical model of a time series. Brain signals can be recorded by brainwave controlled applications, such as EMotiv Epoc +14. The paper investigates the models of the observed brain signals using time series, analyzes their applicability and develops new statistical models for their study.**

***Index Terms*—time series, signal processing, reconstruction of signals, empirical modelling**

I. INTRODUCTION

One of the main problems of modern statistics is to study the process of change and development of the studied phenomena over time. This problem is solved with the help of time series analysis.

A time series is a numerical sequence of observations that characterize the change in the phenomenon under study over time.

In mathematical statistics, time series are considered as random processes, i. e. are considered as arbitrary functions of a variable, as a rule this variable is time.

The output signals of almost all modern measuring devices use digital processing, including electroencephalographs, cardiographs and many other medical devices. Usually the processed signals are recorded as statistics. They are recorded in the form of time series, where information about the properties of the objects distributed in space is presented. Rows can be vector and scalar depending on the type of their elements [1].

All available approaches for processing such rows can be divided into direct processing, batch and indirect with the help of an additional mathematical model.

The first approach is more traditional - only with the help of appropriate software can be calculated: averages, spectra, correlations or to build phase portraits, etc. The second approach for time series analysis uses models that describe the dynamics - the change of variables over time.

Many studies in the research field of cognitive neuroscience rely on EEG, since EEG hardware is available at relatively low cost and EEG signals enable the capture of the neural correlates of mental acts such as attention, speech, or memory operations with millisecond precision [1]. Analysis of this data could possibly enable our researches gives us an opportunity to analyze signals, received by BCI.

The paper provides a brief overview of the possibilities for processing experimental EEG data through reconstruction and the use of time series models.

Analysis of this data could possibly enable our researches gives us an opportunity to analyze signals, received by BCI.

II. RECONSTRUCTION OF COMPLEX SIGNALS FROM TIME SERIES

A standard way to reconstruct signals is by approximating the data obtained. The purpose of the approximation is to reduce the study of various (unknown or extremely complex) numerical characteristics and qualitative properties of the original objects to work with other objects whose characteristics and properties are already known or more convenient to work with.

Approximation problems related to the analysis of complex signals have been known since the mid-18th century. A real dependence between time and quantity $\eta = f(t, c)$ describing the signal was observed, i. e. there is a function $f(t, c)$, where c is a p-dimensional vector of the signal parameters. If the relationship between the values of t and η - the function $f(t, c)$ - is known to the values of the components of c, the task is to estimate the dependence parameters as accurately as possible [1].

In another, more complex formulation of the problem, it is required to find a function f that predicts the ehavior with minimal error.

In the stochastic case, the model is sought in the form

$$\eta = f(t, c_0) + \xi(t), \tag{1}$$

Manuscript received January 14, 2021; revised March 2, 2021.

where η is a random process with zero mean. Here, when calculating the parameters, we are not talking about their exact value, but about obtaining statistical estimates $\hat{c}$ that are as close as possible to c_0.

As the number of data increases, the model becomes more complicated. High-order rows (2) are used, but this is not always possible.

$$\eta_k = c_0 + c_1 t + \cdots + c_k t^k \quad (2)$$

The problem is simplified if instead of approximating the function we look for an equation whose solution is the order under consideration. Thus the complex time series (Fig. 1) is a solution of the nonlinear recurrence equation

$$x_{n+1} = r x_n (1 - x_n) \quad (3)$$

where r is a parameter and n = 1, 2, 3, ... is a discrete time.

Thus, the problem of reconstructing an equation of the form $x(t_{n+1}) = f(x_n, c)$ is no more complicated than the problem of approximating real time dependences [2].

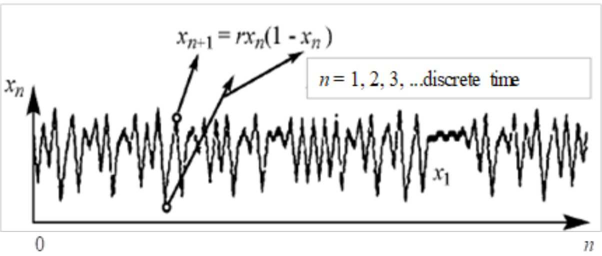

Figure 1. Complex chaotic dependence.

The reconstruction of the equation follows the scheme (Fig. 2) and depends on the type of time series, the initial information for choosing an equation, the choice of function, the number of variables x_i, the way of choosing the variables x_i from the time series of the observed quantities η. The more we know about the model. The more we want to do, the greater the probability of success.

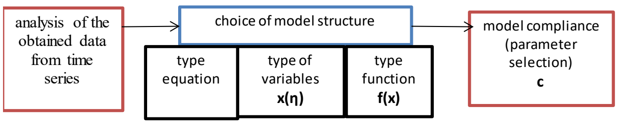

Figure 2. Scheme of the process of reconstruction of the model in time series.

A. *Steps of the Method for Presenting Data Streams by Prototype Encounter Frequency*

Step 1. Segmentation of data streams;

Step 2. Expressing the segments by attributes and determining the similarity between them;

Step 3. Defining templates from similar segments;

Step 4. Represent the data streams through the pattern frequencies [3].

It is essential for the method that the calculation of similarity between segments does not work directly with the original (raw) values of the segments, but uses smoothed data.

Finding a limited number of templates on similar segments is a major step in the process. Each segment is represented by its own template, the encounters of the various templates are calculated, and finally the time series or data streams are represented by the frequency of occurrence of the individual templates.

III. Analysis of Brain Signals

The brain signal is a multidimensional time series having a patio-temporal structure. The time series is characterized at each moment by a time vector of observation, which corresponds to a point in space. The analysis of the characteristics of the current state, their duration and sequence during the transition from one state to another provides important information in the processing of EEG signals.

We consider a multidimensional EEG signal $X = (x\ (1)\ x\ (2),\ ...,\ x\ (T))$ characterized at any time t by an observation vector $x\ (t) = (x_1(t), x_2(t), ..., x_L\ (t))$, where T is the number of time moments and L is the number of channels of the EEG device. Each observation $x\ (t)$ corresponds to a point with coordinates $P_i = (x_i, h_i)$, i = 1 ... L, where L is the number of electrodes projected in an arbitrary plane. Line X is obtained after breaking the continuous EEG signal, i. e the result is a discrete time series.

In the absence of a priori information about the structure of the equations, K. Granger proposes the following approach for identifying causal relationships.

The main idea is that based on the available time series $x\ (t) = (x_1(t), x_2(t), ..., x_L\ (t))$, predictive models are built - "individual" and "joint". Another effective strategy to process time series is to break the rows of segments, build a model for each of the segments and combine the results, in which case instead of building a complex model the task is to build a few simple models for each of the segments. to increase the speed of the analysis and to reduce the volume of pa In addition, in this case the analytical processing can be performed in parallel, which reduces the processing time. A separate analysis can be applied to the model of each segment [4], [5].

IV. Experimental Research

A. *Description of the Output Data*

We have a licensed copy of the EMotiv Cortex SDK software that provides access to the raw EEG data. Cortex is a API powerhouse for creating BCI applications and integrating data streams from headsets with third party software. Built on JSON and WebSockets, Cortex makes it easy to record data for experiments. Cortex is a wrapper around Software Development Kit (SDK) and houses all the tools required to develop with EMotiv [6]. It provides API access to different EMotiv data streams.

The software is developed on Microsoft Visual Studio platform – ASP C# Dot Net. Database is MySQL. The data from EMotiv can be acquired via wireless interface.

The application features include the possibilities for the simultaneous recording and processing of big data. This means different types of data are recorded at the same time, such as: raw data from BCI, what the subject is shown at that moment, the video recording of the experiment and other. The system allows interoperability - receiving the data from the BCI system, processing, exporting of the data to other systems and communication with other apps, which can use these data.

Visually stimulated bio-signals recorded when listening to words with different meanings were studied. When a certain word is heard, the electrical activity of the brain is recorded by 14 sensors for 25 milliseconds by the 14-channel EMotiv Epoc [6], [7].

The signals received from Epoc are transferred via Bluetooth to a computer in real time. The data is obtained with an SDK tool provided by EMotiv (github.com/Emotiv/ community-sdk). Fig. 3 shows a map of the channels on the cerebral cortex.

8 main commands have been determined for the study: START, STOP, LEFT, RIGHT, FORWARD, BACKWARD, FASTER, SLOWER.

Each experiment consisted of recordings of 14 EEG data streams corresponding to the monitored channels. This allows the data from the recorded brain signals to be combined in different ways, in order to look for different dependencies in them.

B. Selection of Characteristics

The study applied 4 methods for selection of characteristics on the formed databases - three filter (Pearson product-moment correlation coefficient, Gain Ratio and Relief) and one covering method (Feature selection with Support Vector Machines). Experimental studies also included an evaluation of the results without performing the feature selection step with the intention of revealing the benefits of including this step in the overall procedure.

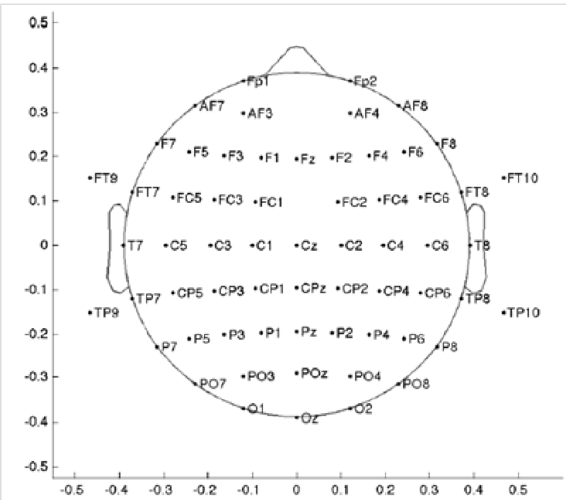

Figure 3. Map of the channels on the cerebral cortex.

In terms of the type of feature selection algorithms used, filter methods are predominant, as they rely solely on data properties and are not dependent on subsequent models and algorithms to derive dependencies on them [8].

C. Structuring Data from Brain Signals

The task of structuring data from brain signals is to create groups of similar objects through grouping and segments.

Each segment is represented by its own template, the number of different templates is calculated and the data streams are represented by the frequency of occurrence of the individual templates. The found prototypes describe the time series and can logically be used in the subsequent analysis of the time series and derivation of dependencies on them. There are many statistical methods for grouping and segmentation.

In the paper, each segment is represented by 12 characteristics obtained using an algorithm of successive extremums [9]. For each segment are determined: 3 minima, 3 maxima, average and the amplitudes between them. The amplitudes of these initial extremes, the average and the time of their occurrence (latency) are considered to be characteristics of the current data flow. The result is an array of data that can be considered in three dimensions:

$$8 \text{ words} \times 14 \text{ Channels} \times 12 \text{ Features}.$$

In the report we consider segments of 25 signals each. We will build a model on the successive segments of the time line with a length of W = 25 signals.

$$W: \{\eta_{(k-1)}W+1, \ldots, \eta_{(k-1)}W+W\}, k= 1, 2, \ldots, L.$$

D. Structuring the Empirical Data of the Signals from the Word STOP

We segment the data from the channel F7. Finding a limited number of prototypes of similar segments is a major step in the process [10]. This allows the application of different variations on how to compare segments and the approach to prototyping.

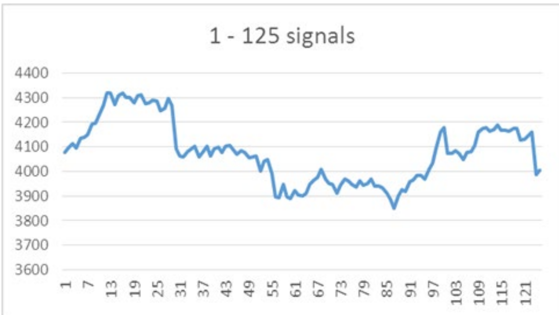

Figure 4. Graph of the obtained empirical data from the first 125 observations on channel F7 of the word STOP.

We consider the first 5 segments of the obtained data of the word STOP (Fig. 4). Table I shows the empirical data of the first 25 observations of the received brain signals. The first 3 minima and 3 maxima and the amplitudes between tach are determined. The mean and standard deviation were found. On the basis of the determined extremes, a polynomial of degree 7 can be reconstructed, choosing the average of the sample as a free member. The middle elements are used primarily due to the simple processing and uniqueness of the solution. In this way we have approximated the first segment [11].

TABLE I. THE FIRST 25 OBSERVATIONS OF THE RECEIVED BRAIN SIGNALS FROM CHANNEL F7

4075,76 MIN1	4095,385	4113,333	4094,359	4136,282
4139,615	4149,359	4192,051	4196,538	4228,974
4269,359	4318,846 MAX1	4318,59	4270,256 MIN2	4306,41
4317,949 MAX2	4298,974	4299,103	4279,744 MIN3	4306,026
4309,231 MAX3	4273,974	4279,359	4290,897	4285,256

For the first segment we found the following characteristics:

mean 4233,83
max 4318,85
min 4075,76
stdev 83,47968286

Fig. 5 shows the graph of the first segment and the approximation of the first segment with a polynomial of degree 7.

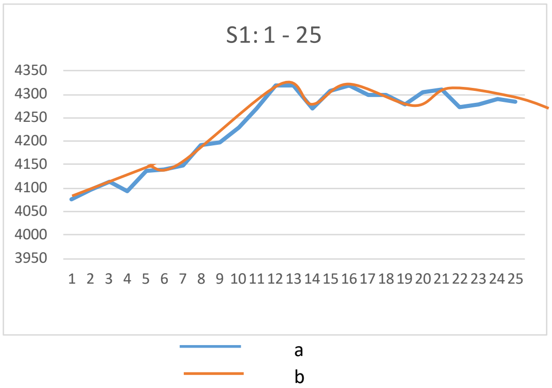

Figure 5. a- shows the graph of the first segment; b - shows the approximation of the first segment with a polynomial of degree 7.

Similarly, we analyze the following 4 segments.

Segment 2 includes the empirical data from 26 to 50 observations obtained from channel F7 of the word STOP. The mean of segment 2 is:

S2 mean = 4108,77

For segment 2 we have (Fig. 6). Fig. 6 shows the graph of the second segment and the approximation of the second segment with a polynomial of degree 7.

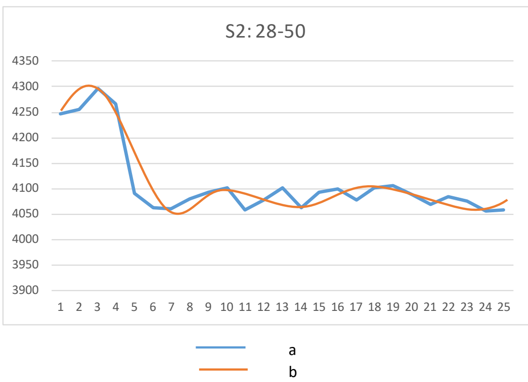

Figure 6. a- shows the graph of the second segment; b - shows the approximation of the second segment with a polynomial of degree 7.

Segment 3 includes the empirical data from 51 to 75 observations obtained from channel F7 of the word STOP. The mean of segment 2 is:

S3 mean = 3954,37

For segment 3 we obtain (Fig. 7). Fig. 7 shows the graph of a segment 3 and the approximation of a segment 3 with a polynomial of degree 7.

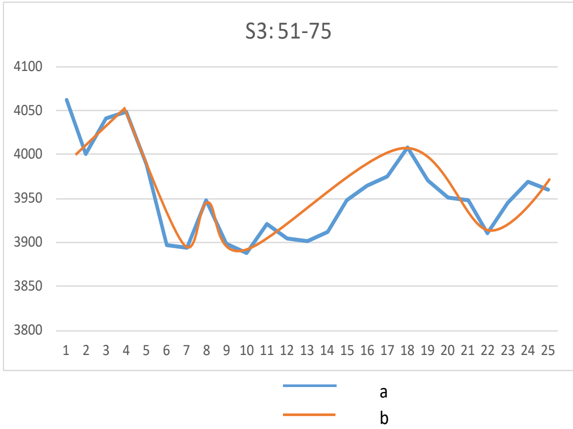

Figure 7. a- shows the graph of a segment S3; b - shows the approximation of a segment S3 with a polynomial of degree 7.

Segment 4 includes the empirical data from 76 to 100 observations obtained from channel F7 of the word STOP. The mean of segment 2 is:

S4 mean = 3970,03

For segment 4 we have (Fig. 8). Fig. 8 shows the graph of a segment 4 and the approximation of a segment 4 with a polynomial of degree 7.

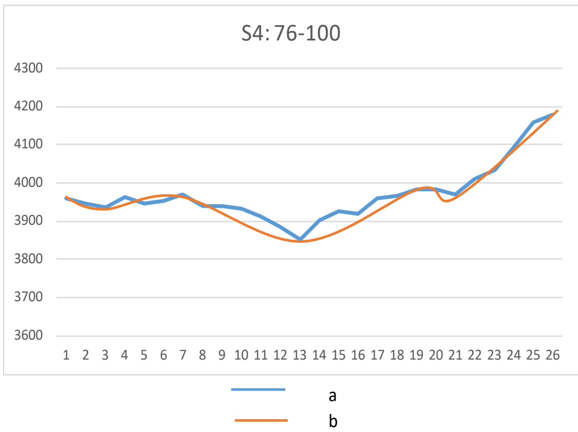

Figure 8. a- shows the graph of a segment S4; b - shows the approximation of a segment S4 with a polynomial of degree 7.

Segment 5 includes the empirical data from 101 to 125 observations obtained from channel F7 of the word STOP. The mean of segment 2 is:

S5 mean = 4121,97

For segment 5 we have (Fig. 9). Fig. 9 shows the graph of a segment 5 and the approximation of a segment 5 with a polynomial of degree 7.

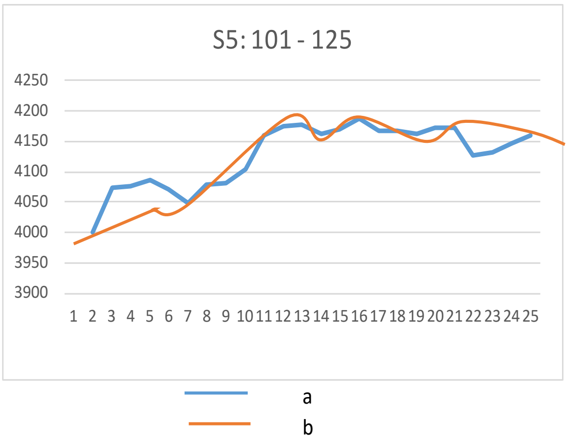

Figure 9. a- shows the graph of a segment S5; b - shows the approximation of a segment S5 with a polynomial of degree 7.

E. Evaluation of the Obtained Results

Fig. 5 and Fig. 9 show similarities in the respective prototypes. The first segment coincides with the fifth segment. This can be used to classify and cluster the data obtained.

The model obtained is: S1, S2, S3, S4, S1,….

We segment the data from the four channels F7, O1, T8, AF4 of the word STOP, because the Euclidean distances between them are relatively close and have strong positive correlations. [12]

The research shows that the model we offer for the obtained data of the word STOP from channel F7 can be used for segmentation of other channels.

Evaluation of the model from the data on the theoretical basis for automatic verification and prediction of the human condition on the basis of the data of the contained experimental evidence. The study can be considered as an initial study to present common representative models of many data. Such developed models can be used as software sensors to detect the desires of subjects of participants in experiments.

V. Conclusions

In recent years, the interaction between man and computer is constantly increasing. Many practical applications are focused mainly on two areas - medicine and robotics.

The identification of individual brain signals and their use for computer-controlled devices as well as the interaction between them will increase in the coming years, but the first steps have already been taken [13].

The immediate approach to signal processing is to create a mathematical models. Modeling remains largely an art, but some general principles and private recipes can be given to increase the chance of a good model.

The proposed models for extracting time dependences on the data need appropriate tools for conducting research. It is necessary to create a software solution that implements the various stages of the model.

The analysis of brain signals is a challenging task due to the great variability and difference between the subjects. Different people generate different brain activity under the same type of external conditions. Even the same individuals respond with different biosignals to the same stimuli depending on their current emotional, physical and mental state. The described data streams from brain signals need proper analysis and interpretation to solve the above problems.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Galina S. Panayotova: II. Rreconstruction of complex signals from time series; III. Analysist of brain signals; IV. Experimental research.

Dimitar Dimitrov: Receive and Select Research Data.

Acknowledgment

This work was supported in part by a grant from PPNP-2020-16 “Analysis and research of non-parametric data systems from the Internet of Things ecosystems”.

References

[1] B. P. Bezruchko, “Modeling from time series and applications to processing of complex signals,” *Processing of University “PND”*, vol. 17, no. 4, pp. 193-207, 2009.

[2] A. Y. Kaplan, *et al.*, “Nonstationary nature of the brain activity as revealed by EEG/MEG: Methodological, practical and conceptual challenges,” *Signal Processing*, vol. 85, no. 11, pp. 2190-2212, 2005.

[3] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, “Querying and mining of time series data: Experimental comparison of representations and distance measures,” *VLDB Endow.*, vol. 1, no. 2, pp. 1542-1552, 2008.

[4] F. Plesinger, J. Jurco, J. Halamek, and P. Jurak, “SignalPlant: An open signal processing software platform,” *Physiol Meas.*, vol. 37, no. 7, pp. 38-48, 2016.

[5] G. Schalk, P. Brunner, L. A. Gerhardt, H. Bischof, and J. R. Wolpawa, “Brain–Computer Interfaces (BCIs): Detection instead of classification brain-computer interface research and development program,” *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 51-62, 2008.

[6] G. Panayotova and D. A. Dimitrov, “Segmentations and models of statistical row data of BCI signals,” presented at the 12th International Conference on Computer Science and Information Technology, 2019.

[7] G. Panayotova and G. P. Dimitrov, “Infographics as a stage of statistical analysis of big data,” presented at Eurasia Research, Budapest, Hungary Conference, September 2018.

[8] G. Panayotova, “Functions of difference and similarity in the study of BCI signals,” in *Proc. International Workshop by ULSIT*, Viena, Austria, may 2019.

[9] R. Nikolov, B. Jekov, and P. Mihaylova, “Big data in a smart city ecosystem: Models, challenges and trends,” in *Proc. International Conference on Big Data, Knowledge and Control Systems Engineering*, Sofia, Bulgaria, 2015, pp. 101-118.

[10] S. Milanov, O. Georgieva, and P. Georgieva, “Comparative analysis of brain data clustering,” in *Proc. Doctoral Conference in Mathematics, Informatics and Education*, 2013, pp. 94-101.

[11] G. Dimitrov, G. Panayotova, K. Aleksiev, I. Kostadinova, P. Petrova, O. Bychkov, and V. Shevchenko, “Using mixed linear programming to classify BCI signals submission,” in *Proc. IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University*, Riga, Latvia, 2019, pp. 1-5.

[12] G. S. Panayotova and G. P. Dimitrov, “Analysis and processing of brain signals,” *International Journal of Signal Processing Systems*, vol. 7, no. 3, pp. 92-95, September 2019.

[13] R. Nikolov, E. Shoikova, M. Krumova, E. Kovatcheva, V. Dimitrov, and A. Shikalanov, "Learning in a smart city environment," *Journal of Communication and Computer*, vol. 13, no. 7, pp. 338-350, 2016.

Galina Panayotova is a Doctor of Mathematics and Professor of Mathematical modeling in the State University of Library Studies and Informational technologies, Head of Department Computer Science, Sofia, Bulgaria and University "Prof. Dr. As. Zlatarov"- Burgas, Bulgaria.
Her research interests are mathematical and computer modeling, big data, use of information technologies and applications in education. Prof. Panayotova is the author of more 110 scientific publications, books and textbooks.

Dimitar Dimitrov is a Doctor of Informatics and Computer science. He works at University "Prof. Dr. As. Zlatarov", Department Electronic, Burgas, Bulgaria. His research interests are computer modeling, big data, digital signal processing systems. Dr. Dimitrov is the author of more 20 scientific publications, books and textbooks.