ON THE DEVELOPMENT OF AN INFORMATION TECHNOLOGY FOR PLASMA TURBULENCE RESEARCH

Andrey Gorshenin
Institute of Informatics Problems, Russian Academy of Sciences
Vavilova str., 44/2, Moscow, Russia
MIREA, Faculty of Information Technology
Email: agorshenin@ipiran.ru

Victor Korolev
Lomonosov Moscow State University
Leninskie Gory, Moscow, Russia
Institute of Informatics Problems, Russian Academy of Sciences
Email: victoryukorolev@yandex.ru

Dmitry Malakhov
Nina Skvortsova
Prokhorov General Physics Institute,
Russian Academy of Sciences
Vavilova str., 38, Moscow, Russia
Email: mukudori@mail.ru

Sergey Shorgin
Institute of Informatics Problems, Russian Academy of Sciences
Vavilova str., 44/2, Moscow, Russia
Email: sshorgin@ipiran.ru

Victor Kuzmin
"Wi2Geo LLC", Russia
Email: shadesilent@gmail.com

KEYWORDS

Information technology, Turbulent plasma, Simulation, Probabilistic models, Spectrum.

ABSTRACT

The paper deals with an example of real information technology developed for the examination of specific structures of plasma turbulence by the spectral analysis. The mathematical basis is a probabilistic approach using a special simulation algorithm to construct sample for the probabilistic modeling. To describe the fine structure of stochastic processes, finite mixtures of various probability distributions are used.

In the paper, the general structure of the developed information technology including mathematical models, algorithms and software realization is considered. Procedures for finding statistical estimators of the unknown parameters of model are based on modifications of EM algorithm. Examples of application of the developed technology in an important area of modern physics are presented.

INTRODUCTION

To create effective and safety sizeable scientific plants using concept of tokamaks (e.g., International Thermonuclear Experimental Reactor, ITER) adequate models of plasma functioning have to be developed. And the first step for it is the construction of mathematical models, algorithms and software for special laboratory plants.

To estimate process parameters, spectrum measured by spectrometer, spectrograph or analog-to-digital converter must be split into components (continuous spectrum, bands, single components). Spectral analysis is one of the traditional tools of a signal processing in the plasma turbulence. But the decomposition of a plasma turbulence spectrum is an ill-posed problem. The unique solution exists only under additional assumptions about the object's structure (Lochte-Holtgreven, 1968; Akhmanov, 2004). So, it is impossible to obtain important spectral information about the functioning of plasma turbulence by the traditional approach implying spectrum's approximation by Kolmogorov-Obukhov model or shot (fluctuation) noise model.

However, the spectra must be analyzed since it makes possible to reveal the set of important physical parameters: the type of instability, the mechanism of turbulence formation, the proportion between plasma fluctuations and plasma structures, e.g., ion-acoustic solitons and drift vortices.

To overcome difficulties, special probabilistic procedure for plasma spectrum decomposition was implemented in the paper (Gorshenin et al., 2011). The spectrum is interpreted as density of unknown distribution. Then a test sample with pre-specified size is simulated using special bootstrap-type procedure. To this sample we apply an approach based on compound Cox process model and thus an approximation to the spectrum can be obtained. The choice of Cox processes for turbulence modelling is based on known empirical and theoretical results (see, e.g., the book (Korolev and Skvortsova, 2006)). Surely, one of the most important problems is that of the choice and optimization of computational methods for the

Proceedings 28th European Conference on Modelling and Simulation ©ECMS Flaminio Squazzoni, Fabio Baronio, Claudia Archetti, Marco Castellani (Editors)
ISBN: 978-0-9564944-8-1 / ISBN: 978-0-9564944-9-8 (CD)

estimation of unknown parameters.

Under some experimental conditions, the efficiency of the methodology was demonstrated. The spectra were successfully decomposed into the components. Moreover, new results confirmed earlier models. For example, only few components are significant for plasma turbulence functioning despite of huge overall number of processes in plasma.

The paper represents the information technology created for automating of experimental data processing using the ideas mentioned above. In the further sections we discuss the models, algorithms and improvement techniques included in the information technology. Experimental samples for various plasma conditions from L-2M stellarator (Pshenichnikov et al., 2005) were used as data sets for probabilistic modelling.

MATHEMATICAL MODELS

The design of methods for the analysis of stochastic processes is very important for the evaluation of turbulence characteristics when conditions of the plasma confinement are changed. In this section we discuss mathematical approach for probabilistic modeling in the implemented information technology.

The spectrum is interpreted as probability density function of unknown probability distribution $\mathcal{F}$. We can obtain any α-quantile (values of α are chosen pseudo-random from uniform distribution on the interval $[0,1]$) by solving the equation

$$\mathcal{F}(x_\alpha)=\alpha.$$

A sample $\mathcal{X}$ from unknown distribution $\mathcal{F}$ could be simulated for any predetermined size. Investigating low-frequency fluctuations of the plasma it was found that the processes in plasma turbulence could be adequately described by compound Cox processes (Korolev, 2011). So using this model we can assume that $\mathcal{F}(x)$ is a finite mixture of probability distributions (e.g., normal, gamma) with unknown parameters, i.e.:

$$\mathcal{F}(x)=\sum_{i=1}^{k}p_iF(x,a_i,b_i), \qquad (1)$$

$$\sum_{i=1}^{k}p_i=1,\ p_i\geq 0, \qquad (2)$$

where $F(\cdot)$ denotes some type of cumulative distribution functions; $k\geqslant 1$ is a known natural number; $a_i\ \ b_i$, $i=1,\dots,k$, are parameters of distribution under correct conditions (e.g., standard deviations for normal mixtures are strictly positive). Quantities $F(x,a_i,b_i)$, $i=1,\dots,k$, are components of mixture; $p_1,\dots,p_k$ are weights; k is a number of components in mixture. Values p_i, $a_i\ \ b_i$, $i=1,\dots,k$, are usually unknown and should be estimated by sample. Resulting mixture approximates the spectrum.

For the normal mixtures $\mathcal{F}(x)$ in equality (1) has the following form:

$$\mathcal{F}(x)=\sum_{i=1}^{k}p_i\Phi\left(\frac{x-a_i}{\sigma_i}\right),$$

where

$$a_i\in\mathbf{R},\ \sigma_i>0,\ i=1,\dots,k,$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left\{-\frac{t^2}{2}\right\}dt,\quad x\in\mathbf{R},$$

and conditions (2) hold.

For the gamma mixtures $\mathcal{F}(x)$ in equality (1) has the following form:

$$\mathcal{F}(x)=\sum_{i=1}^{k}p_iF_{r_i,\theta_i}(x),$$

where

$$r_i>0,\ \theta_i>0,\ i=1,\dots,k,$$

$$F_{r,\theta}(x)=\int_{0}^{x}\frac{\theta^r t^{r-1}e^{-\theta t}}{\Gamma(r)}dt,\quad x>0,$$

and conditions (2) hold.

To find unknown parameters, various modifications of EM algorithm (Dempster et al., 2011) are used in the information technology. For the case of normal mixtures, the estimators can be found in (Korolev, 2011), gamma-mixtures estimators were described, e.g., in (Gorshenin and Korolev, 2013). Characteristics of grid modifications are considered in (Gorshenin et al., 2013). Combination of different methods and special statistical procedures make it possible to realize an acceptable computation accuracy and to obtain a reasonable time for different types of spectra.

The initial data are the one- and two-sided Fourier spectra. The size of the simulated samples is 100000 elements. Maximum number of components for EM algorithm equals 6. As a stopping criterion we use the relation

$$\max\left|\theta^{(m)}-\theta^{(m-1)}\right|<\varepsilon,$$

where $\theta^{(m)}$ is a vector of estimated parameters at m-th iteration step, and the ε is an accuracy. The computing accuracy ε equals 10^{-8} (multiple experiments with the EM algorithm demonstrate the fact that this value provides both the correct results and a reasonable time). Decomposition of the spectrum into components gives opportunity to understand the behavior of different types of fluctuations and structures in the plasma.

ALGORITHMS AND SOFTWARE

In this section we discuss issues of functioning and interactions of software modules representing the computational part of the information technology.

In modern researches with huge sizes of experimental data and critical required processing time it is impossible to imagine an analysis without a creation of specialized information technologies. To decompose plasma spectra into the components the corresponding solution based on the mathematical model of compound Cox processes was suggested.

Having developed mathematical tools, new results for adequacy and velocity were obtained. It predetermined set of software's algorithms. With built-in MATLAB fourth-generation programming language the computational module for spectrum's analysis was created and optimized for plasma physics research.

The information technology consists of mathematical, computational and visualization tools and the user interface for the researchers usability. Let us consider the structure of the developed software (see Fig. 1).

Experimental data are external files with access via software interface. The program can be run with a help both the standard MATLAB console (there is the main processing function) and special interface to specify parameters of methods. The first way is suitable for "developer mode", while the most of users definitely prefer interface which hides the details of implementation and simplifies working with a program.

There are three logical modules in the software's functional:

1. the simulation module (we simulate a sample for modelling);
2. the estimation module (we estimate unknown parameters by the sample using different methods and obtain approximating mixture);
3. the vizualization module (we plot various figures).

Obviously, each of these steps must be performed consecutively (in Fig. 1 it is demonstrated by the dashed arrows), but you can start at any point (in Fig. 1 possible execution paths are shown by solid arrows from the module "Interface"). It was realized to process data previously saved on the disk without re-simulation, re-estimating of parameters or both of them.

The test sample is formed as described in the previous section. To estimate unknown parameters the following methods can be used by user's choice in the estimation module:

1. EM algorithm for normal mixtures;
2. EM algorithm for gamma mixtures;
3. grid method for normal mixtures;
4. grid method for gamma mixtures.

The methods for the normal mixtures can be used for one- and two-sided spectra, the methods for the gamma mixtures should be applied in case of one-sided spectra.

EM algorithm is a quite universal method for finding maximum likelihood estimates. It allows to achieve a balance between velocity and quality of approximation. However, if we have some additional information about parameters, the grid methods can be more efficient. So, both types of methods were included in the final software.

The problem of correct determination of number of components arises during an approximation by finite distributions mixtures. Therefore, most powerful tests (see, e.g., (Gorshenin, 2011)) and values of the Lo statistic (Lo et al., 2001) should be used for these purposes. The program checks significance of components with low weights by different criteria and returns the number of components in the mixture. The significance level of criterion is a parameter of the method.

The vizualization module forms graphical output. According to parameter estimates the approximating curve and the constituent components are plotted to provide an intuitive interpretation of the results. The graphs allow to determine a number of components (i.e. processes in turbulent plasma) and specify some physical characteristics.

Moreover, analyzed experimental data are the series of spectra obtained under various conditions. To demonstrate evolution of spectra in time you can plot three-dimensional surface in the vizualization module. A profile in concrete time moment represents approximating mixture for the corresponding spectrum. Three-dimensional graphics can be rotated and zoomed by mouse. The images could be saved in the graphic (JPEG, PNG, EPS) and MATLAB's formats (FIG).

Mathematical algorithms were approved on the test samples with known characteristics. As part of the development of information technology the special procedures to improve efficiency and accuracy were created. Using different algorithms and special techniques in the united software we made a detailed analysis of data for sustainable series of spectra obtained in low-frequency plasma turbulence at the edge and in the center of the plasma filament in the L-2M stellarator.

The software can be run on computers with an installed MATLAB as well as without it (but in this situation freeware MATLAB Compiler Runtime have to be installed).

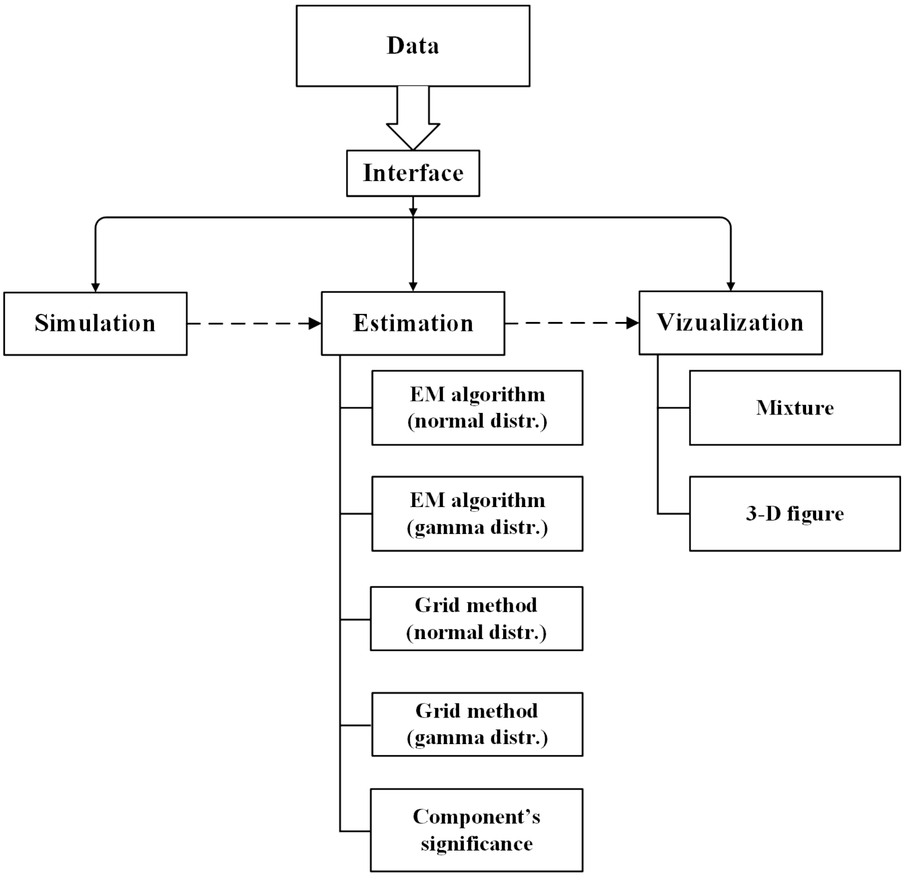

Figure 1: Structure of the information technology

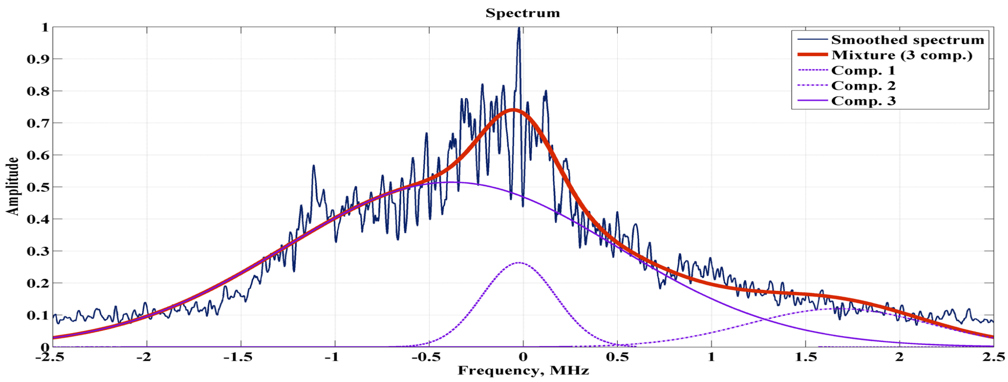

Figure 2: Two-sided spectrum

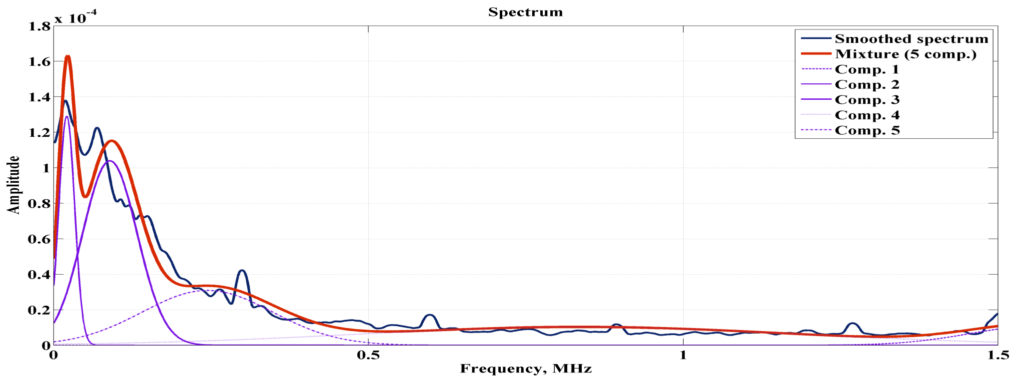

Figure 3: One-sided spectrum

APPLICATION

In this section, we demonstrate examples of applying the implemented information technology to the real data which represent measurements in the laboratory stellarator L-2M.

Figures 2-3 demonstrate examples of the software's output. We can see initial spectrum (thin solid line), approximating mixture (bold line) and the components. The x-coordinate corresponds to frequencies, the y-coordinate shows the amplitude of spectrum. There are three dominant components with different non-zero weights. It corresponds to the energy transfer between different types of turbulence. Most of the trends in the original spectrum are approximated by modelled mixture quite good.

Figure 2 demonstrates example of processing for two-sided spectrum in the certain time moment of plasma discharge in the stellarator. There are 3 forming components with non-zero weights.

Figure 3 demonstrates example of processing for one-sided spectrum in the L-2M. There are 5 forming components with non-zero weights. As seen on figures the number of forming components is finite and moreover it is a modest as compared with total number of stochastic processes in plasma turbulence. Both figures demonstrates spectrum's decomposition into the normal components.

The main present reached physical results by the applying of information technology to the real data are following:

1. the decomposition of the spectra: the form of harmonics in amplitude spectra was found out;
2. the specification of correct number of the forming processes in plasma turbulence;
3. the reveal of recurrence of stochastic processes with typical mean frequencies of spectrum's half-width;
4. the determination values of the radial electric field fluctuations, phase velocities, etc.

POSSIBLE IMPROVEMENTS

In this section we discuss possible ways for the information technology to improve a productivity. Undoubtedly, it can be embodied by special programming solutions, e.g., we can realize computational modules using any low-level programming language. The source code might be very effective and fast but too difficult for programming and especially for debugging. Moreover, it needs much time and in fact you will create a new information technology. So in this section we consider approaches of possible optimizations of the existent technology.

The first way is simply to use more actual and efficient hardware. For example, we have used the Intel Core i7 CPU and obtained up to 3 times acceleration of working with one spectrum comparing our previous hardware. Moreover, modern CPUs have more than one logical core and so you can process multiple data sets simultaneously. The ratio between velocity and time is not linear, nevertheless you obtain significant acceleration without special techniques!

The second way logically follows the first one in terms of parallelism. Modern integrated development environments (e.g., built-in MATLAB IDE) support mechanisms for automatization of parallel computing for a source code. Using special directives, program can work faster without wide modifications of the code.

The third way is based on new hardware ideas and creating special source code for these purposes. It leads to computing on GPUs, clusters, etc. At that, GPU solutions are not so expensive as clusters and supercomputers. The fact that nowadays CPUs are created with integrated graphics chips (e.g., AMD Fusion, Intel Ivy Bridge, etc.) demonstrates extent of perspectiveness of GPU computing. The world leading GPU producers offer special solutions for researchers in different ares (CUDA technology by NVIDIA, ATI Stream Technology by AMD). It should be noted that in modern GPUs the number of cores equals from several hundred to thousands ones. Obviously, their performance may be extremely high for various complex computational problems in the areas with the critical requirements for accuracy and processing time. Surely, one of the most important problems is a creation of an effective software that would be able to use the full power of the hardware solutions. In fact the optimal application performance on multi-core systems can be achieved through rational use of program threads for the correct allocation of subproblems. Threads execution can be optimized for running on a different physical cores.

CONCLUSION

As mentioned above the proposed methodology is quit successful for problems of approximation of a spectrum and its decomposition into the components. The number of components and the forms are invariable in specific regions in the series. The resulting number of significant components in the decomposition of the spectrum adjusts with the physical essence of the processes under study because the amount of significant processes in real data is a modest.

The use of non-parametric estimation procedures in plasma physics is very prospective for the analysis of short-wavelength plasma turbulence. With the help of a physical interpretation of the component it is possible to create more precise models of the turbulent plasma functioning.

Thus, the created information technology presents an effective software tool of plasma processes research. Moreover, just now we obtained informal outcomes for the real data sets.

Due to the informational properties of probability distributions, some known physical aspects the problem of modelling of spectra by mixtures of another distributions can be very interesting and perspective. Although it seems to be an object of special independent research.

The problem of spectrum components decomposition is a relevant in other physical spheres, e.g., quantum decoding of the harmonic spectra of diatomic molecules like a titanium (II) oxide TiO (Hermann et al., 2001).

The implemented methods contributes to the application of mathematical methods and computational techniques in the physics of plasma turbulence. The created information technology can be used as software basis of processing experimental data in the data centers. To use the solution in real plants like ITER some software and hardware improvements including mentioned in previous section are needed. But success methodology for laboratory plant allows to expect potentially interesting results in this case too.

ACKNOWLEDGEMENTS

The research is supported by the Grant of the President of the Russian Federation (project MK-4103.2014.9) and by the Russian Foundation for Basic Research (projects 12-07-00115a, 12-07-00109a, 14-07-00041a).

REFERENCES

Akhmanov, S.A.; and S.Yu. Nikitin. 2004. *Physical Optics.* Nauka, Moscow.

Dempster, A., N. Laird and D. Rubin. 1977. "Maximum likelihood estimation from incompleted data." *Journal of the Royal Statistical Society. Series B* 39, No.1, 1-38.

Gorshenin, A.K. 2011. "Testing of statistical hypotheses in the splitting component model." *Moscow University Computational Mathematics and Cybernetics* 35, No.4, 176-183.

Gorshenin, A.K., V.Yu. Korolev, D.V. Malakhov and N.N. Skvortsova. 2011. "Analysis of fine stochastic structure of chaotic processes by kernel estimators." *Matematicheskoe modelirovanie* 23, No.4, 83-89.

Gorshenin, A., V. Korolev. 2013. "Modeling of statistical fluctuations of information flows by mixtures of gamma distributions". In *Proceedings of 27^{th} European Conference on Modelling and Simulation* (May 27-30, 2013, Alesund, Norway). Digitaldruck Pirrot GmbHP, Dudweiler, Germany, 569-572.

Gorshenin, A., V. Korolev, V. Kuzmin and A. Zeifman. 2013. "Coordinate-wise versions of the grid method for the analysis of intensities of non-stationary information flows by moving separation of mixtures of gamma-distribution". In *Proceedings of 27^{th} European Conference on Modelling and Simulation* (May 27-30, 2013, Alesund, Norway). Digitaldruck Pirrot GmbHP, Dudweiler, Germany, 565-568.

Hermann, J., A. Perrone and C. Dutouquet. 2001. "Analyses of the TiO-γ system for temperature measurements in a laser-induced plasma." *J. Phys. B: At. Mol. Opt. Phys.* 34, 153-164.

Korolev, V. Yu. 2011. *Probabilistic and Statistical Methods of Decomposition of Volatility of Chaotic Processes.* Moscow University Publishing House, Moscow.

Korolev, V.Yu.; and N.N. Skvortsova (Eds). 2006. *Stochastic Models of Structural Plasma Turbulence.* VSP, Utrecht.

Lo, Y., N.R. Mendell and D.B. Rubin. 2001. "Testing the number of components in a normal mixture." *Biometrika* 88, No.3, 767-778.

Lochte-Holtgreven, W. 1968. *Plasma Diagnostics.* North-Holland, Amsterdam; Interscience (Wiley), New York.

Pshenichnikov, A.A., L.V. Kolik, N.I. Malykh, A.E. Petrov et al. 2005. "The use of Doppler reflectometry in the L-2M stellarator." *Plasma Phys. Rep.* 31, No.7, 554-561.

AUTHOR BIOGRAPHIES

ANDREY GORSHENIN is Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Russian Academy of Sciences; associate professor, Faculty of Information Technology, Moscow State Institute of Radio Engineering, Electronics and Automation. His e-mail address is: `agorshenin@ipiran.ru`

VICTOR KOROLEV is Doctor of Science in physics and mathematics, professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M.V. Lomonosov Moscow State University; leading scientist, Institute of Informatics Problems, Russian Academy of Sciences. His e-mail address is: `victoryukorolev@yandex.ru`

DMITRY MALAKHOV is Candidate of Science (PhD) in physics and mathematics, Prokhorov General Physics Institute, Russian Academy of Sciences. His e-mail address is: 89199945791@mail.ru

NINA SKVORTSOVA is Doctor of Science in physics and mathematics, Prokhorov General Physics Institute, Russian Academy of Sciences. Her e-mail address is: mukudori@mail.ru

SERGEY SHORGIN is Doctor of Science in physics and mathematics, professor, deputy director, Institute of Informatics Problems, Russian Academy of Sciences. His e-mail address is: sshorgin@ipiran.ru

VICTOR KUZMIN is Head of Development, "Wi2Geo LLC", Russia. His e-mail address is: shadesilent@gmail.com