Российская Академия Наук

РОССИЙСКАЯ АКАДЕМИЯ НАУК

ИНСТИТУТ ПРОБЛЕМ БЕЗОПАСНОГО РАЗВИТИЯ АТОМНОЙ ЭНЕРГЕТИКИ

RUSSIAN ACADEMY OF SCIENCES

NUCLEAR SAFETY INSTITUTE

Препринт ИБРАЭ № IBRAE-2000-09

Preprint IBRAE-2000-09

E. Savelieva, M. Kanevski, A. Kravetski, V. Ogar

GEOSTATISTICALLY BASED METHODS FOR ELECTRICITY LOAD ANALYSIS AND FORECASTING

Москва 2000

Moscow 2000

УДК 502.3
Савельева Е.А., Каневский М.Ф., Кравецкий А.С., Огарь В.П. ГЕОСТАТИСТИЧЕСКИЕ МЕТОДЫ ДЛЯ АНАЛИЗА И ПРОГНОЗА ЭЛЕКТРОПОТРЕБЛЕНИЯ (на англ. яз.). Препринт № IBRAE-2000-09. Москва: Институт проблем безопасного развития атомной энергетики РАН. Декабрь 2000. 18 с. — Библиогр.: 15 назв.

Аннотация

Работа посвящена геостатистическим методам краткосрочного и среднесрочного прогноза электропотребления. Рассматриваются такие методы, как обычный кригинг, частичные Гауссовы симуляции, отжиг, прогноз с помощью многослойного персептрона. Анализируются достоинства и недостатки этих методов в применении к одномерным (временны*м*) данным.

Savelieva E., Kanevski M., Kravetski A., Ogar V.. GEOSTATISTICALLY BASED METHODS FOR ELECTRICITY LOAD ANALYSIS AND FORECASTING. Preprint IBRAE-2000-09. Moscow: Nuclear Safety Institute. December 2000. 26 p. — Refs.: 15 items.

Abstract

In this work some geostatistically based methods are applied to short term and middle term electricity load forecasting. Application of such methods as ordinary kriging, sequential Gaussian simulations, simulated annealing, MLP prediction can be found there. Advantages and disadvantages of application of these methods for 1-dimensional time data are analyzed.

Geostatistically based methods for electricity load analysis and forecasting

E. Savelieva, M. Kanevski, A. Kravetski, V. Ogar

THE INSTITUTE OF NUCLEAR SAFETY
113191, Moscow, B. Tulskaya, 52
tel.: (095) 955-22-31, fax: (095) 958-11-51, E-mail: esav@ibrae.ac.ru

Contents

Introduction

Future forecasting is significantly important problem in many different domains of science. Ideal situation is the whole knowledge of physical laws present as a system of differential equations together with it's initial conditions describing the system under study. But even in those rare cases when system of equations is known absolutely precise forecasting is hindered because of complexity of real dynamic systems (the problem of dynamic chaos). The usual situation is when our knowledge about system is presented as a consequence of observations of some inputs (parameters which depend on system's behaviour) and an output which is of special interest for us. Such observations are present for some period in the past (historical data). In this case we deal with a time series. Sometimes it is possible to construct a mathematical model able to reproduce the relation between known inputs and outputs. Afterwards this relation can be used for forecasting of a future behavior of a system. It is important to state here, that uncertainty and some noise in input information make analysis more complex.

It is impossible to construct a correct model without correctly organized analysis of data. Data analysis have to include several steps [1]:

- System understanding – investigation of historical data presenting a process to detect main relations of data (trends, seasonality, correlations, linearity, stationarity), noise reduction, determining of a set of parameters depending on forecasting one and analysis of such relations (classifications, mutual correlations etc.);
- System modelling – selection of a mathematical model able to reproduce relations between known inputs and output, analysis of residuals of a model and an attempt to model them, validation on different data presenting the system.
- System forecasting – using of a constructed and validated model for future forecasting.

Each step is rather difficult and requires development and adaptation of modern methods.

The main characteristics of a model are:

- The ability to adopt to changes and system evolution;
- The absence of sensibility to measurements' uncertainties (or errors while input of information);
- The simplicity for user;
- The accuracy – the most important characteristic. More accurate model will be preferable even in spite of all previous characteristics are not the best.

Problems concerning electrical loads and electro consumption can be divided into three main groups:

- Very short term forecasting – from several seconds up to an hour;
- Short term forecasting – from an hour up to 1-2 weeks;
- Middle term forecasting – from a week up to a year;
- Long term forecasting – from a year up to 20 years.

Each group has its own peculiarities and domains of application[2]. The results of very short term forecasting are very much dependant by the possibility of fast reaction. For example, the weather changes such as passing of a cloud over the region (or its part) will depend on it. The long term forecasting is strongly dependent by economic and demographic factors. While short term and middle term forecasting can be considered as mathematical tasks on analysis and forecasting of time series. It makes them more interesting for mathematicians.

Short term forecasting is very much important for electricity production and electricity selling companies. It is used for planning of functioning of production capacities and planning of regular repair measures. For short term forecasting hourly (or half-hourly) measurements are usually used, which have got rather high frequent component. The value of an amplitude is dependent by large number of parameters – weather conditions, type of a day, time, season and some others. For short term forecasting adaptive and hybrid methods can be used. They have to be learned on great number of different historical data. The problem of optimal model and number of input parameters is not solved yet [3,4].

Middle term forecasting is usually used to establish electricity tarif and, thus, is directly related with the forecast of companies' benefit. Usually for middle term forecasting one deals with integrated (or averaged) value for some definite period. So these data have no high frequency component and clearly seen autocorrelation. Such data also need detailed analysis.

This work is devoted to show the application of some geostatistically based methods for both short term and middle term forecasting. Problems, advantages and disadvantages of such approach and the direction of following analysis will be indicated.

1. Classical geostatistics for electricity load short term forecasting

Geostatistics is a set of methods for data analysis based on stochastic treatment of data and using data spatial correlation structure [5]. Usually *Intrinsik hypothesis* is proposed: the mean exists and is equal for all sub-regions of region under study, variogram exists and depends only on distance between points. Variogram is

$$\gamma(x,h) = E\{(L(x) - L(x+h))^2\} = \gamma(h), \quad (1)$$

where E means averaging for all pairs separated by distance h. Variogram is used in geostatistics to describe the spatial correlation structure of data.

1.1. Ordinary kriging for direct short term electricity load forecasting

One of the most widely used geostatistical methods is ordinary kriging.

Ordinary kriging is a best linear unbiased estimation (BLUE) method. We have a set of points (x_i) of some value L. The theoretical linear estimation of value ($\hat{L}$) in some point (x) can be written as follows:

$$\hat{L}(x) = \sum_{i=1}^{\infty} w_i L(x_i), \quad (2)$$

where w_i are weight coefficients. In real analysis the summation is made till some number N, describing the number of points influencing on the point under estimation. Ordinary kriging is best interpolation in sense of minimizing the square of estimation error in each point. Unbiasedness means that mean value of estimations is close to mean value of known measurements. For ordinary kriging weight coefficients can be found by solving following system of linear equations:

$$\begin{cases} \sum_{j=1}^{N} w_j \gamma_{ij} + \mu = \gamma_{i0} \\ \sum_{i=1}^{N} w_i = 1 \end{cases},$$

where μ is a Lagrange multiplier and γ_{ij} is the value of variogram for distance between points *i* and *j*.

Together with kriging estimation the variance of estimation is calculated. Formula is the following:

$$\sigma_k^2 = \sum_{i=1}^{N} w_i \gamma_{i0} + \mu .$$

Kriging variance is dependent by density of measurement points in the region where estimation has been done.

Detailed information about geostatistics can be found in [6,7].

Geostatistics was used for electricity load forecasting for data described in details in [3]. Separate analyses have been performed for different seasons. Variograms obtained for summer and winter load data are presented in figure 1. Variogram models for following forecasting are also presented in this figure. One can see that summer variogram model is closer to experimental one.

It is known from geostatistical theory, that the value of kriging variance is significantly dependent by distribution of sample points in the neighbourhood of the point under estimation. The growth of kriging means the growth of estimation's uncertainty. In our case we want to forecast the future values and the space we are working in is 1-dimensional, so we have:

1. The case of extrapolation – estimation out of the region where a model has been constructed.
2. The number of neighbours decreases while going forward in time.

There are several artificial ways how to overcome this difficulty. First is to make consequent estimation – to use once estimated value in following estimations as it is used in sequential approach to simulation [6]. But in this work we use another approach: the known data for the previous year take part in both variogram analysis and in kriging estimation. Only the period of a week under forecasting is eliminated from previous year data, for otherwise we shall obtain exactly the load from the previous year. The results of kriging estimation for 5 working days in winter and 3 working days in summer are presented in figures 2 and 3. The number of forecasted days is different because of the value of radius of correlation radius. One can see from variograms (fig. 1), that in winter it is larger then in summer.

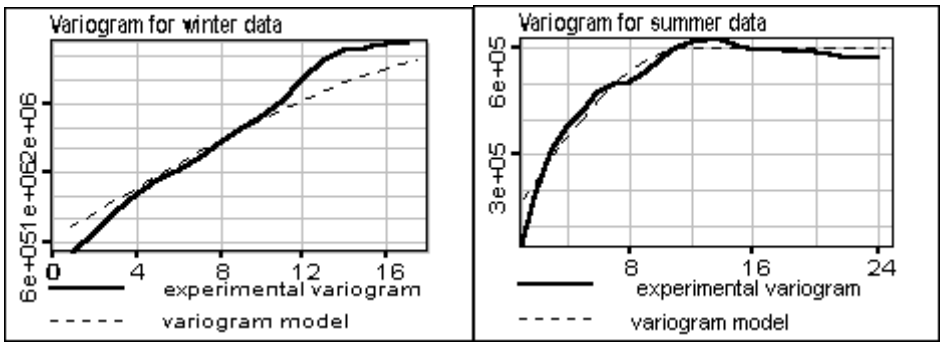

Fig. 1. Variograms for electricity load. Left – for winter data, right – for summer.

In figures 4 and 5 relative errors of ordinary kriging electricity load forecasting are presented. One can see that their absolute values never overcome 15% and the most part of them is whithin 5% error. It can be considered as very good result. But the error is growing with the distance from known values due to this year. Maybe previous year values in some way spoil the estimation for weather this can be different.

The above example illustrates the possibility to use linear estimator – ordinary kriging for electricity load forecasting. But in spite of such promissing results this method has got a lot of problems and difficulties. Both examples were made for stable periods – without any jumps of weather and without holidays whithin the week. But how to be sure future week will have stable weather. Also the artificial overcoming of the extrapolation problem can bring additional distortions.

The examples of possible distortions are presented in figure 6. There are presented ordinary kriging estimations for working days and holidays when they appear in data together. There is no need to perform any analysis of errors, for it is seen that types of days are perplexed.

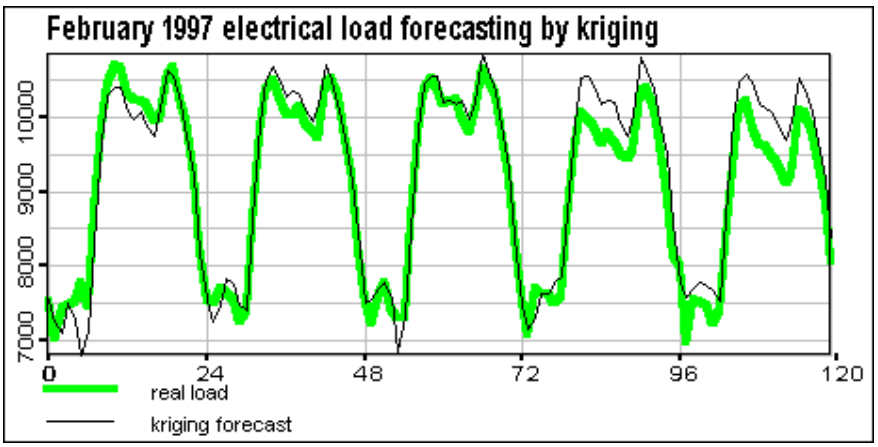

Fig. 2. Electricity load forecasting by ordinary kriging (February 1997)

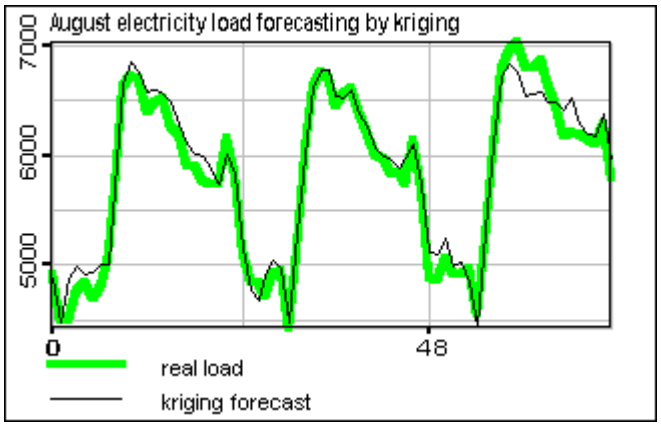

Fig. 3. Electricity load forecasting by ordinary kriging (august 1997)

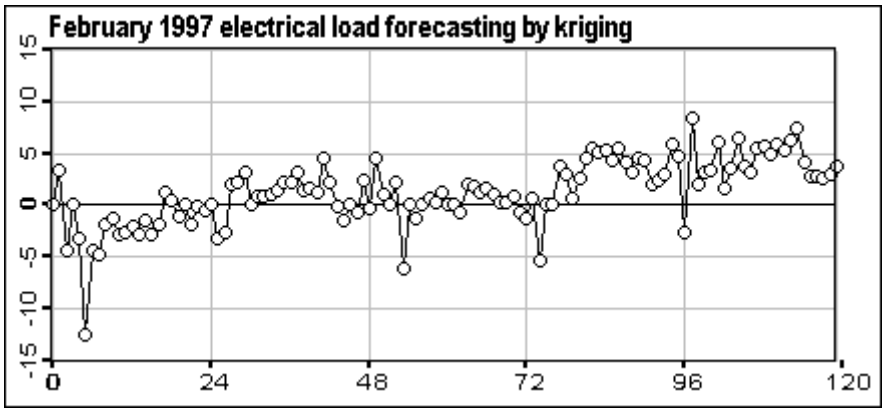

Fig. 4. Relative errors of electricity load forecasting by ordinary kriging (february 1997)

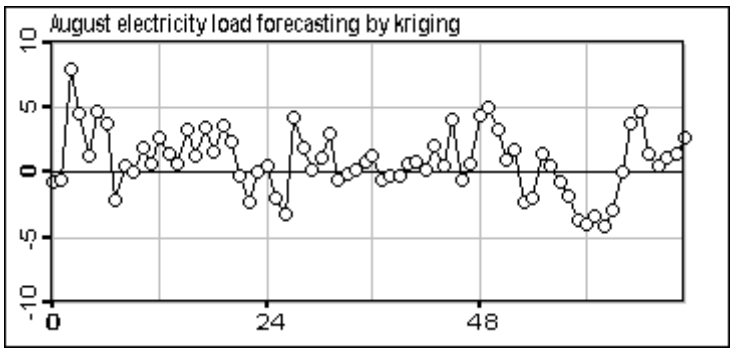

Fig. 5. Relative errors of electricity load forecasting by ordinary kriging (august 1997)

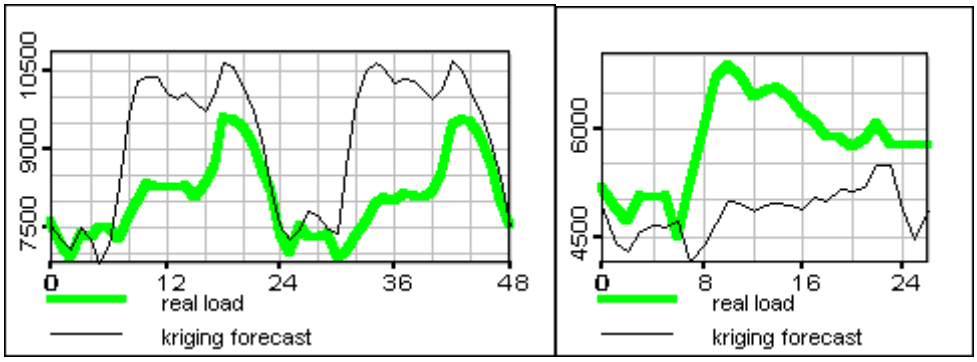

Fig. 6. Ordinary kriging estimations for weekends (left) and Monday (right) when using wokring days and holidays together

The above examples show the general possibility to use geostatistics for 1-dimensional data. Ordinary kriging can be used for electricity load forecasting for stable periods.

1.2. Geostatistical stochastic simulations for ANN residuals

Stochastic simulation approach provides many equally probable realizations with the same global statistical peculiarities (mean, correlation structure, etc). Each point realisation is modeled as a stochastic selection from conditional cumulative distribution function (ccdf). Conditioning is to the known initial data. The whole (*N* point) realisation is modeled as a selection from *N*-variable ccdf, presenting joint distribution *N*-variable distribution.

Number of realisations allows for each time point to estimate probability distribution. If the hypotheses on Gaussian distribution can be accepted then a mean and a variance can be estimated. The mean is considered as E-Type estimation of a residual and the variance as a variability of residual's distribution. 95% confidence level can be estimated using 0.95 quintile. It can be used for estimation of uncertainty of obtained estimate.

For stochastic simulations there is no problem of extrapolation and a number of points to simulate forward as it simply reproduce the statistical and correlation structure of known time series (residuals in our case).

There exist several approaches to stochastic simulations: sequential Gaussian simulations, simulation with constraints, indicator simulations and some others. In this work sequential Gaussian simulation approach and simulation with constraints (simulated annealing) are applied to residuals obtained after artificial neural network (ANN) multi layer perceptron electricity load forecasting. The description of ANN used for electricity load forecasting can be find in [4].

1.2.1. Sequential Gaussian simulation of ANN residuals

Sequential approach is an attempt to change *N*-variable joint conditional distribution function by *N* 1variable conditional distribution functions. It can be done under the proposition of independence of values in separated points.

Gaussian simulations are made under the proposition of multi-Gaussian distribution of initial data. If such proposition can't be made Nscore transform is first performed on data. Nscore doesn't guarantee multi-Gaussian distribution, it need to be checked.

Detailed description of theory on sequential Gaussian simulations can be find, for example, in [6]. Here only the sequence of operations is presented below.

1. Nscore transform of initial data if they don't obey normal distribution.
2. Check at least bi-Gaussianity. There are several empirical methods. For example, the correspondence of following equation can be checked

$$\frac{\sqrt{\gamma(h)}}{M(h)} = \sqrt{\pi},$$

where $\gamma(h)$ is a variogram and $M(h)=E\{|Z(x)-Z(x+h)|\}$ is a madogram.

3. Selection of sequence of points for simulations.
4. Perform simulation in one point. After simulation this point with obtained result is included into the set of data using for following simulations during this realisation. For simulation of value in the point two steps are performed: kriging (ordinary or simple) [6] is used to estimate parameters of Gaussian distribution – kriging

estimate is treated as mean and kriging variance as a variance of normal distribution; random value according to known distribution is generated.

ANN residuals don't obey normal distribution, so Nscore transform has been performed on them. Correlation structure for transformed data was analysed. The experimental variogram and variogram model for nscore transformed residuals are presented in Figure 7.

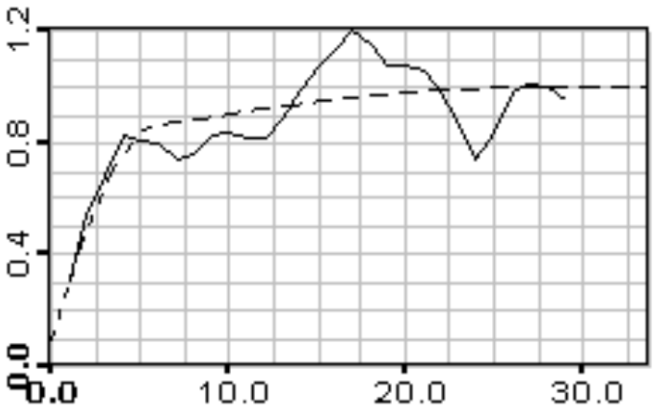

Fig. 7. Experimental variogram (solid) and variogram model (dashed) for ANN residuals after Nscore transform

100 realisations of sequential Gaussian simulations have been performed for ANN residuals. E-type estimate is treated as a result of residuals estimation. Final electricity load forecasting was calculated as a sum of ANN estimate and estimate of a residual. 95% confidence levels were estimated too. They can be considered as uncertainty of estimate. The example of a result of ANN + sequential Gaussian simulations for short term electricity load forecasting is presented in Figure 8

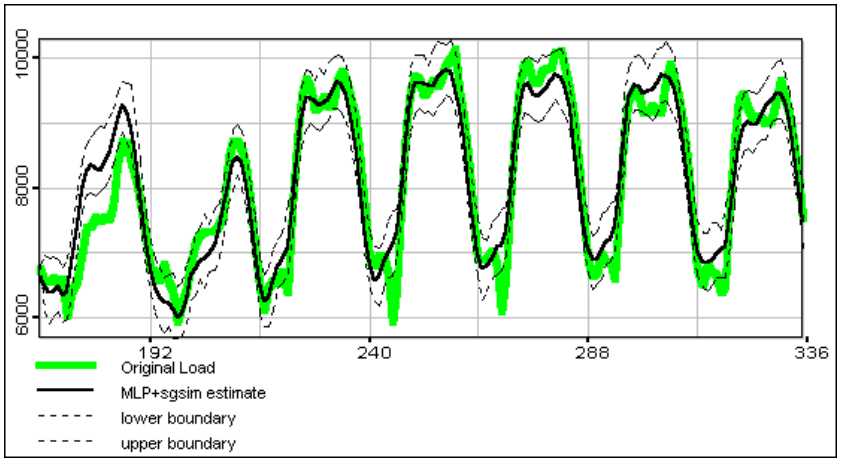

Fig. 8. Example of final presentation of a result. ANN + sequential Gaussian simulations

1.2.2. Simulated annealing for ANN residuals

Another approach to generate simulations was general constrained randomization [8,9]. For this approach all constraints which have to be reproduced in simulated time series are presented in the form:

$$F_i(\{\hat{s}_n\}) = 0, i = 1,...., L,$$

where $\{\hat{s}_n\}$ denotes a simulated time series, L – a number of constraints. Different types of constraints have a different form. In this work histogram levels and correlation structure model were used as constraints. For histogram a constraint connected with *k*th quantile is presented in form:

$$F_i(\{\hat{s}_n\}) = \left(Q_k(\{s_n\}) - Q_k(\{\hat{s}_n\})\right)^2,$$

where Q_k is a value of *k*th quantile.

Constraint connected with *l*th variogram lag is presented in form:

$$F_i(\{\hat{s}_n\}) = \frac{\left(\gamma_{h_i}(\{s_n\}) - \gamma_{h_i}(\{\hat{s}_n\})\right)^2}{\left(\gamma_{h_i}(\{s_n\})\right)^2}.$$

Cost function is constructed to include all constraints in simulation process:

$$C(\{\hat{s}_n\}) = \sum_{i=1}^{L} w_i F_i(\{\hat{s}_n\}).$$

The constraints are fulfilled when cost function has got its global minimum.

Minimization of the cost function was made by method of simulated annealing [10]. Permutations of pairs original time series (residuals) are made. Annealing scheme chooses which changes to accept and which to reject according to the following condition:

$$\begin{cases} p = 1 & \Delta C < 0 \\ p = e^{-\Delta C / T} & \Delta C \geq 0 \end{cases}.$$

Modeled correlation structure of initial data (ANN residuals) is presented in Figure9.

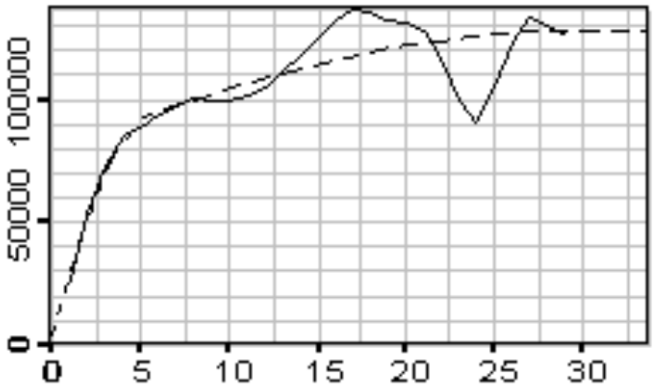

Fig.9. Experimental variogram (solid) and variogram model (dashed) for ANN residuals

100 realisations of general constrained randomisation simulations have been performed for ANN residuals. E-type estimate is treated as a result of residuals estimation. Final electricity load forecasting was calculated as a sum of ANN estimate and estimate of a residual. 95% confidence levels were estimated too. They can be considered as uncertainty of estimate. The example of a result of ANN + simulated annealing for short term electricity load forecasting is presented in Figure 10.

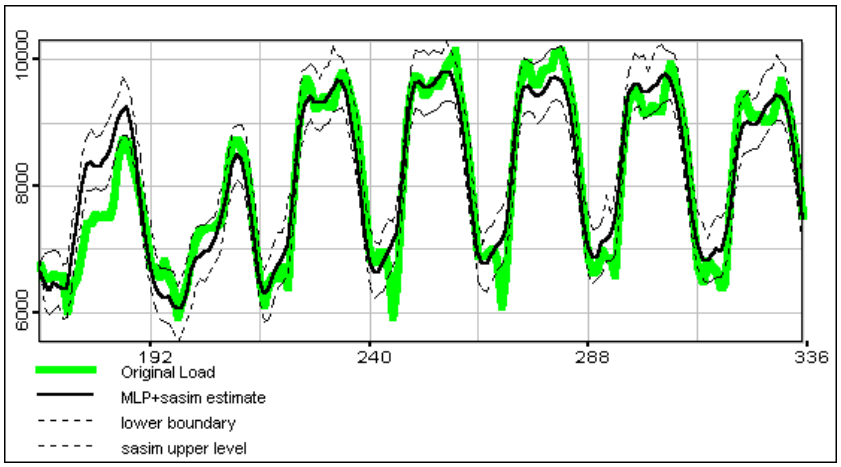

Fig. 10. Example of final presentation of a result. ANN + simualted annealing

2. Analysis and middle term forecasting of integrated data

2.1 Description of data

As an example of integrated data to perform analysis and middle term forecasting data from [11] were used. These data present monthly electricity production in Australia for the period from January 1956 till April 1990 (see figure 11). It is known that an amount of produced electricity is closely connected with electricity consumption and such time series present the same properties.

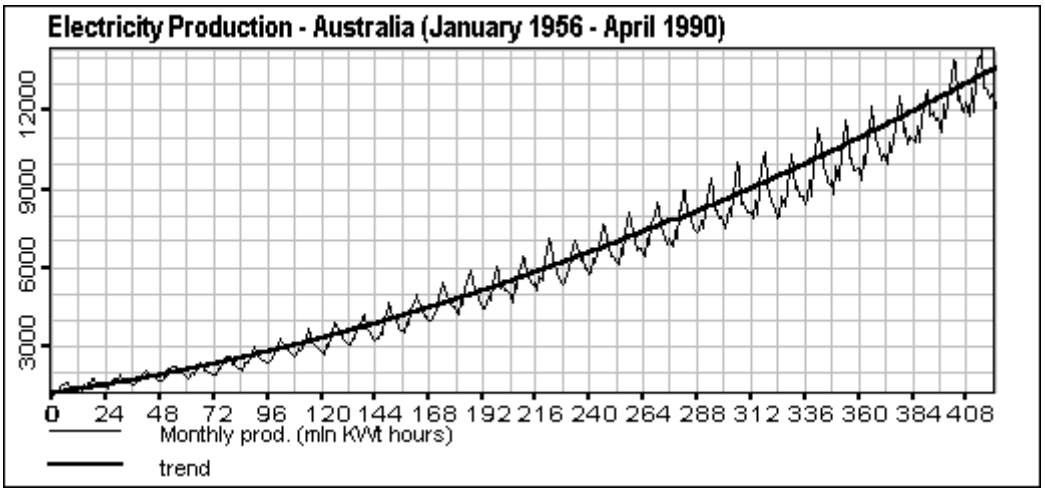

Fig. 11. An example of monthly integrated data – electricity production in Australia

Some trend and year periodicity can be clearly seen in data. There exist several approaches how to eliminate trend and periodical component. Some of them are the following:

1. Polynomial trend modelling;
2. Differential transformation;
3. Deseasonal transformation;
4. Construction of a low frequency filter;
5. Application of artificial neural networks.

The possible model of polynomial trend of data is presented in figure 11. It is a quadratic model:

$$trend(i) = 1240 + 12.58i + 0.04i^2,$$

where i is a number of points in time series which emulates time. Data after subtraction of this trend are presented in figure 12. The periodical component is still in data.

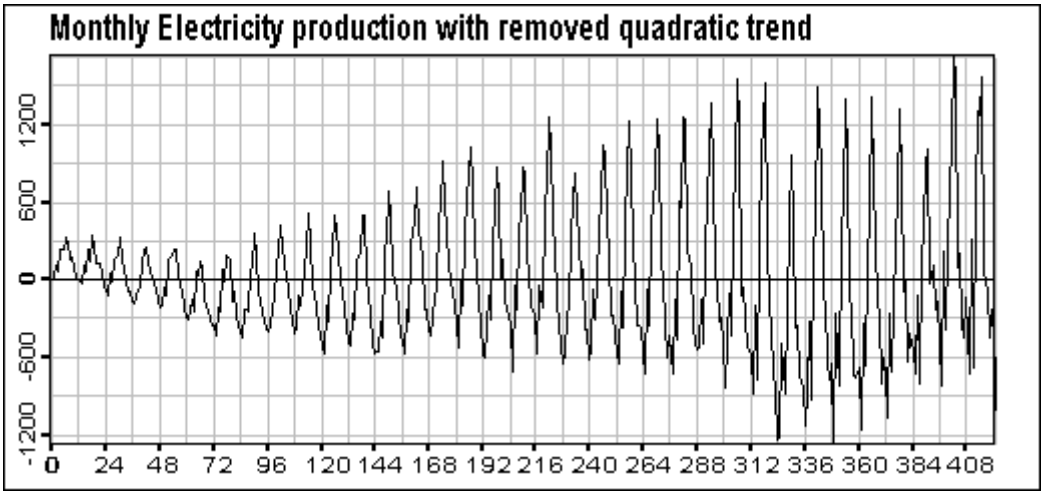

Fig. 12. Integrated data with removed quadratic trend

The periodical component can be modeled as low frequency component. Such model and the obtained result of periodical structure removing are presented in figures 13.

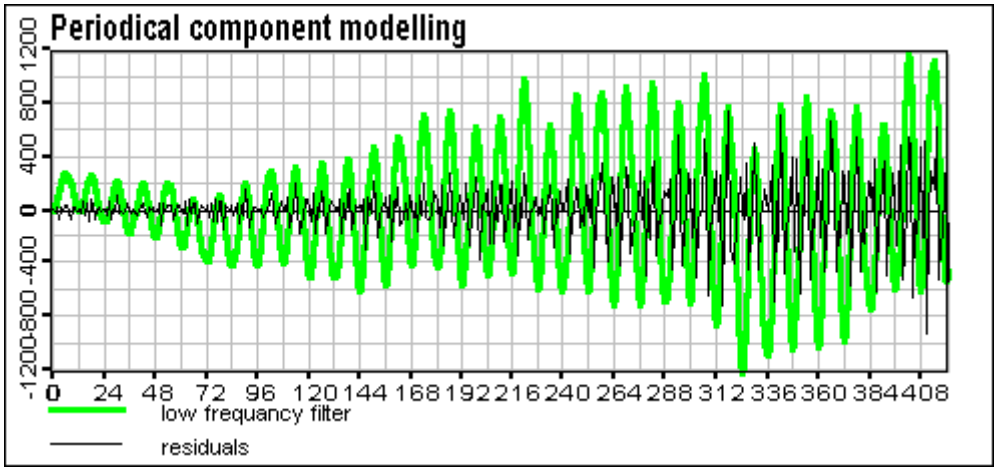

Fig. 13. Removing of periodical component by low frequancy filter

Both trend and periodicity can be removed by using different kinds of differential transformations [11]. The simplest way is to construct new time series $\{y_i\}$ from the old one ($\{x_i\}$) $i=1…n$ as follows:

$$y_i = x_{i+k} - x_i,$$

where k is the lag of obtained differences. Usually k is selected to be equal to *1*, but other values are also possible and sometimes can be preferable. The differences are calculating while $i+k \le n$. So the new time series will be on k elements shorter then the old one.

For periodic structures with known period (d) special type of transformation called deseasonality can be done. To simplify the following formulas we consider d to be even ($d=2q$). In our case we have an a-priori information about the period – 12 months. It is caused by seasonal changes in electricity consumption. No objections exist for odd case, there are some minor changes in formulas. This transformation is made in several steps:

1. Calculating mean values for period. For all i such that $q<i\le n$ m_i are calculated as follows:

$$m_i = (0.5x_{i-q} + x_{i-q+1} + ... + x_{i+q-1} + 0.5x_{i+q}).$$

2. For each step of a period the average of differences of a value and a corresponding mean is computed. The formula for calculations is the following:

$$w_k = E\{(x_{k+jd} - m_{k+jd})\}.$$

   Here j is selected so that $k+jd \le n$.

3. The final estimation of a seasonal component s_k is the following:

$$s_k = w_k - d^{-1}\sum_{j=1}^{d} w_j, k = 1,...., d.$$

   It is obvious that $s_k=s_{k+jd}$ for all j such that $k+jd \le n$.

The estimation of seasonal component is subtracted from each point of initial time series.

Some examples of performing differential transformations to our data are presented in figure 14. The result of deseasonal transformation with lag equal to 12 (1 year) is presented in figure 15.

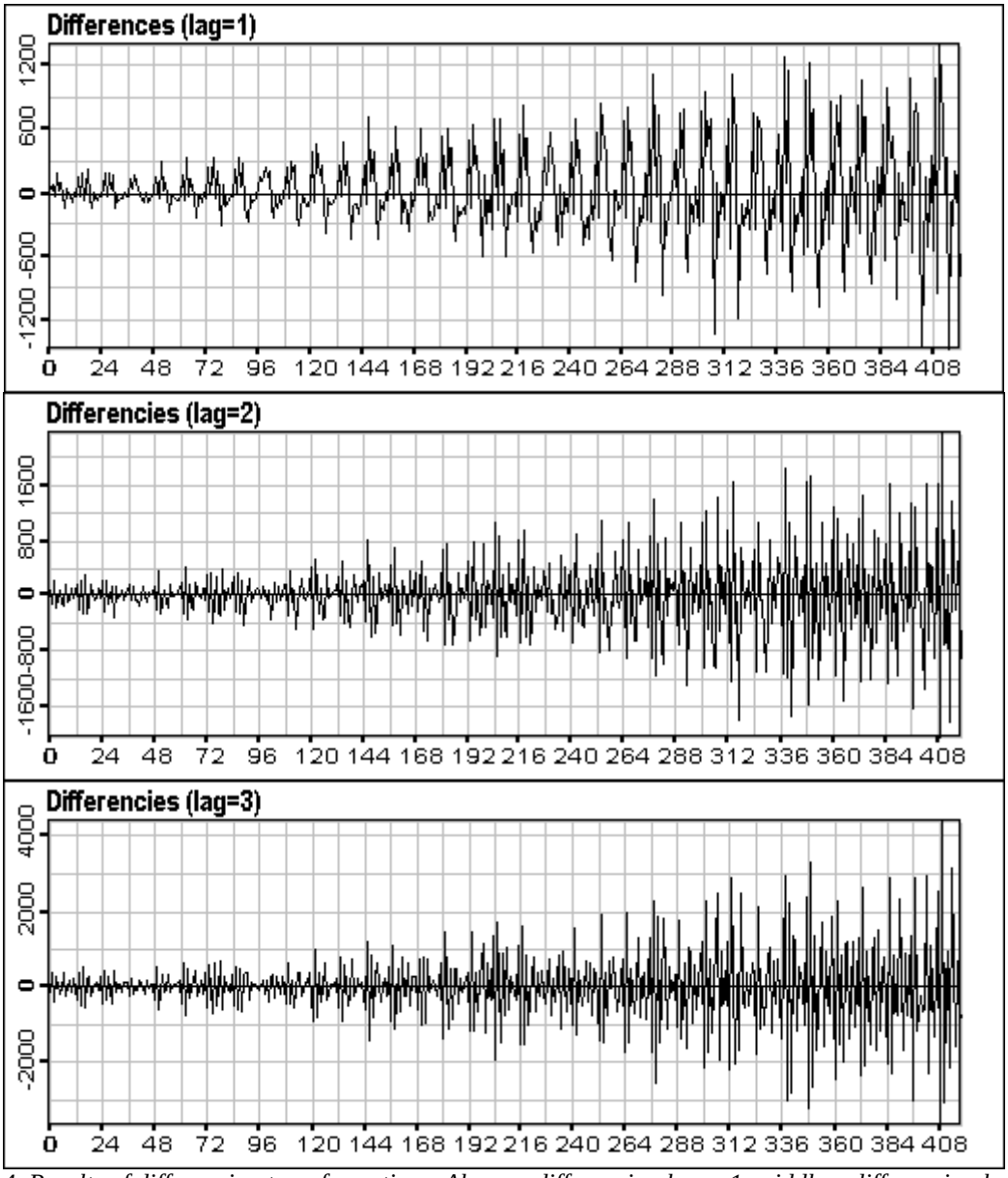

Fig. 14. Results of differencing transformations. Above – differencing lag = 1, middle – differencing lag = 2, below – differencing lag = 3

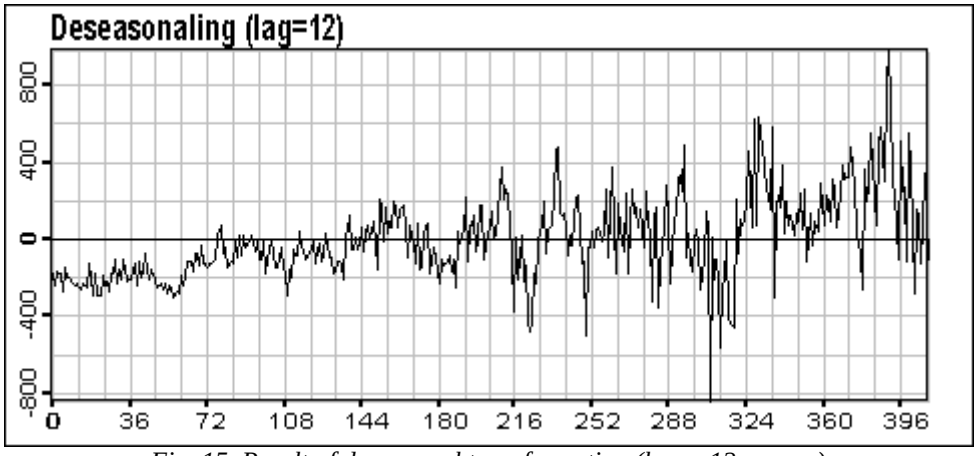

Fig. 15. Result of deseasonal transformation (lag = 12, a year)

2.2 Data understanding

To perform any time series forecasting the detailed analysis of data have to be done. The historical data (and maybe some a-priori knowledge) are the only thing we have for model construction. This stage is called system understanding [1].

The main features that have to be detected about system are the following:

1. Is there any correlation between data? If not data can be considered as white noise and only mean value and standard deviation can be estimated. The future of such process can be forecasted by a random generator. Only the presence of correlations between data allows to construct a model.
2. Is there any additional noise in data? The presence of noise in data complicates the construction of model. The noise brings in data perturbations which can give visual illusion of more complex model of data.
3. Is the process under study linear or not? The solution of this problem helps to choose the type of model able to reproduce the process.

To solve each of the problems stated above serious analysis of data have to be performed. There exist several developments in the directions of solving all these problems. Some of them (data correlation analysis and some non-linearity tests) will be discussed below and applied for analysis of real data described in section 2.1. Noise reduction is not important for such type of data. The integration over some period reduces the noise in the high frequency component.

2.2.1 Data correlation analysis

Data correlation structure of a process $\{X_t\}$ is traditionally described by covariance function:

$$C_X(r,s) = Cov(X_r, X_s) = E\{(X_r - \mu_X(r))(X_s - \mu_X(s))\},$$

where $\mu_X(t)=E\{X_t\}$ means the mean function for process $\{X_t\}$.

For (weakly) stationary process autocovariance and autocorrelation functions can be introduced. The weak stationarity of time series $\{X_t\}$ means that the mean function is independent of t and covariance functions is independent of t and depends only of h – the distance between points. Autocovariance function of stationary process $\{X_t\}$ is

$$C_X(h) = Cov(X_{t+h}, X_t).$$

Autocorrelation function (ACF) of $\{X_t\}$ is

$$\rho_X(h) \equiv \frac{C_X(h)}{C_X(0)} = Cor(X_{t+h}, X_t).$$

Using autocorrelation function one can introduce partial autocorrelation function (PACF) $\alpha(\cdot)$ by the following equations:

$$\alpha(0) = 1$$
$$\alpha(h) = \varphi_{hh}, h \geq 1.$$

where ϕ_{hh} is the last component of $\varphi_h = \Gamma_h^{-1} C_h$, $\Gamma_h = [C(i-j)]_{i,j=1}^{h}$ and $C_h = [C(1), C(2), \ldots, C(h)]'$. It can be seen that PACF is special function used especially for time series, where as h an integer number is used – the number of time delays between time series points.

Another way to estimate data correlation structure is variogram (see formula (1)). Usually it is used in geostatistics. The advantage of variogram is that we work with differences, so less strong assumption then stationarity can be made – Intrinsyk hypothesis.

For analysis of real data sample analogs of described above functions are used. The theoretical mathematical expectation is replaced by sample averaging. Averaging is made on number of data (mean) or on number of pairs of samples separated by distance h (autocorrelation function, variogram and others).

Correlation analysis of real data was performed by using both autocorrelation function and variogram. Correlation analysis were made for all kinds of data – initial data, with removed quadratic trend, after deseasonal transformation, after difference transformations with different lags. The obtained results are presented in figures 16 – 20.

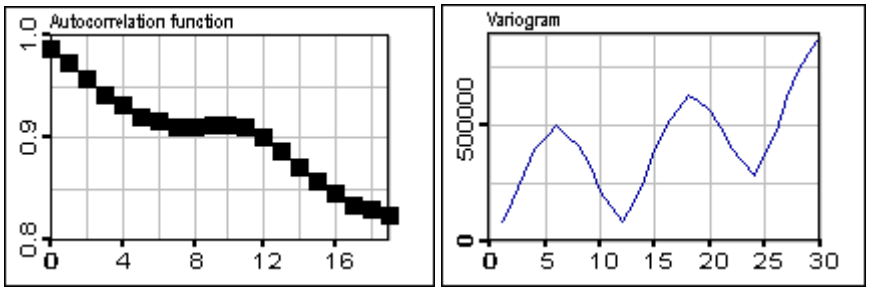

Fig. 16. Correlation analysis of initial data: autocorrelation function (left), variogram (right)

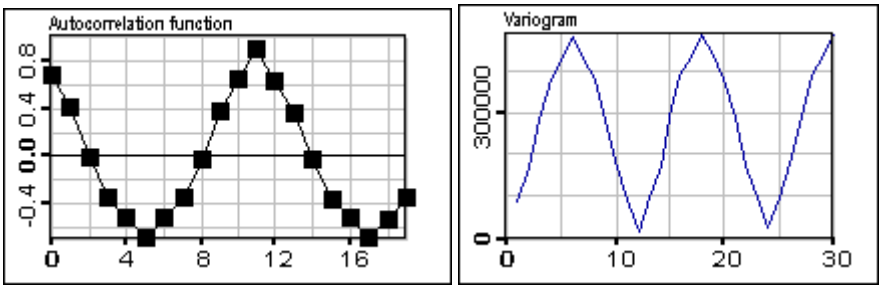

Fig. 17. Correlation analysis of data with removed quadratic trend: autocorrelation function (left), variogram (right)

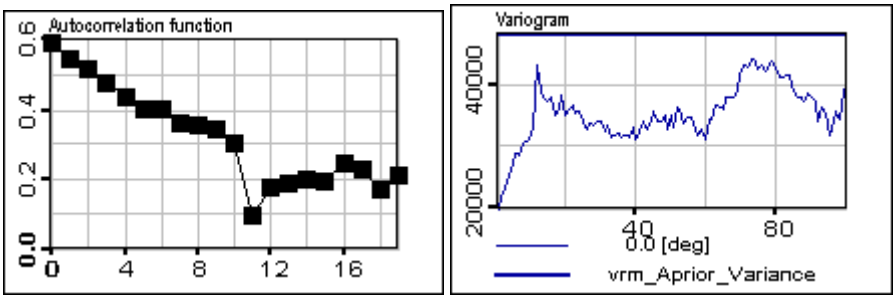

Fig. 18. Correlation analysis for data after deseasonal transformation (lag 12): autocorrelation function (left), variogram (right)

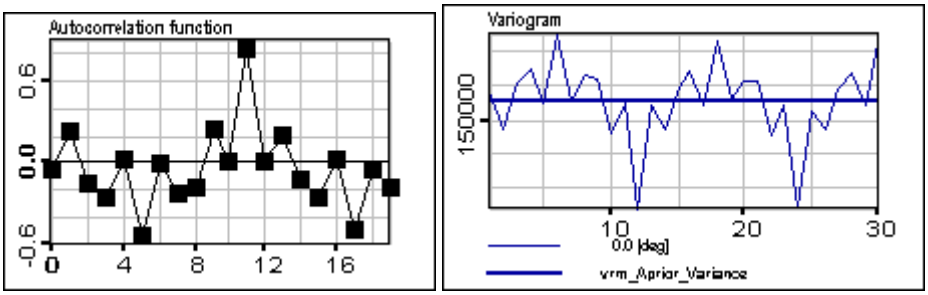

Fig. 19. Correlation analysis of data after differential transformation (lag 1): autocorrelation structure (left), variogram (right)

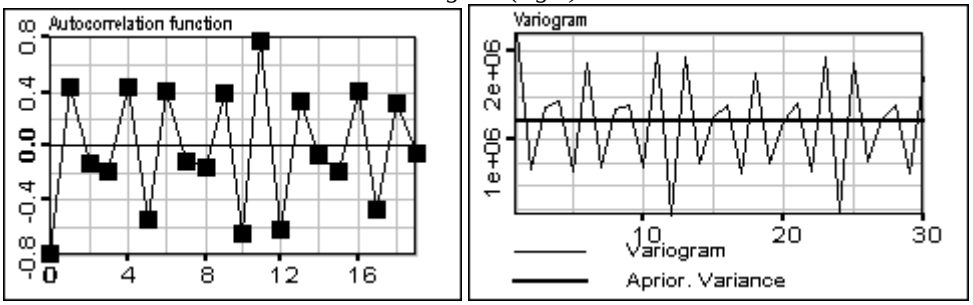

Fig. 20. Correlation analysis of data after differential transformation (lag 3): autocorrelation function (left), variogram (right)

Performed correlation analysis shows the following properties of analyzed data:

1. Periodical structure of correlations is present for nearly all data. It is absent only for data after deseasonal and differential (lag 3) transformations. For data without any transformation non-stationarity of data is also seen – trend in correlation structure.
2. The correlations between data are present in all time series except data after differential transformation with lag 3.
3. It seems that correlation structure for data with removed trend can be modeled within one period (a year). The modelling of correlation structure is also possible for deseasoned data and data after differential transformation with lag 1.
4. For following analysis and forecasting we select data where correlation structure can be modeled: data with removed quadratic trend, deseasoned data and after differential with lag 1.

2.2.2 Nonlinearity analysis

Nowadays it became very popular to use non-linear methods for all kinds of time series analysis. There are different motivations of such behaviour: linear methods have already been used and couldn't produce the desired result, the system producing time series is supposed to be non-linear. In the first case the use of non-linear methods is proved, what can't be said about the second case. The measured output of a non-linear system can appear to be linear, or it can become linear after some simple procedures. So it is proposed to perform a test on non-linearity before application of non-linear methods.

The main task of test on non-linearity is to determine the presence of low non-linearity, for strong non-linear features can be distinguished rather simple [8]. The sketch of test on non-linearity is the following:

1. Formulate the null hypothesis, for example – correspondence of a time series under study to a Gaussian linear process.
2. Generate a number of time series (surrogates) corresponding to a null hypothesis.
3. Calculate some parameter possible to characterize non-linear features of a time series for all generated surrogates and a given time series.
4. Make attempts to reject the null hypothesis by comparing non-linear parameters of a given time series with their distribution for surrogate time series.

From the above sketch two problems can be formulated:

1. What parameter to select to characterize non-linear features?
2. How to generate surrogate time series?

There are 2 possibilities to be non-correct: to reject a null hypothesis when it is in fact true and accept a null hypothesis when it is in fact false. The first one is specified by a *level of significance* (*1-p)* of the test amounts so that the probability to reject a null hypothesis when it is in fact true don't exceed *p*. It depends on the size of statistical test. The other possibility is the main characteristic of criterion to be selected for such type of tests. This criterion can be applicable for detecting of weak non-linearity features. To characterize this quality of criterion the discrimination power (β) has been introduced. It is defined as:

$$\beta = P\{\text{to reject a null hypothesis when it is indeed false}\}.$$

In the work [12] 5 different measures of nonlinearity (of different order and underground nature) were used to detect features of non-linearity in different data. According to criterion β it was found out that the best criteria for non-linearity test are: higher order autocorrelations (time reversibility – autocorrelation of 3d order) and nonlinear prediction errors.

In general time reversibility is the measure of symmetry under the time rotation, linear process is symmetrical according to the time. Time reversibility is defined and can be calculated as follows:

$$\phi^{REV}(\tau) = E\{(s_n - s_{n-\tau})^3\} = \frac{\sum_{n=\tau+1}^{N}(s_n - s_{n-\tau})^3}{\left[\sum_{n=\tau+1}^{N}(s_n - s_{n-\tau})^2\right]^{3/2}},$$

where τ is the time delay and *N* is the length of a time series.

The other proposed measure of nonlinearity is based on the usage of locally linear predictor $F(s_n)$. The prediction over one time step is made as a linear combination of all future values of vectors closer than ε to a vector $\mathbf{s}_n=\{s_{n-1},\ldots,s_{n-m}\}$ in *m* dimensional embedded space. To simplify the situation (not to adjust any weights) simple averaging is used. For nonlinear time series the error will grow together with the size of local area. Such relationships also can be used as nonlinear test (without surrogates). A nonlinear prediction error as a measure of nonlinearity is defined as follows:

$$e^{PE}(m,\tau,\varepsilon) = \left(\sum [s_{n+1} - F(\mathbf{s}_n)]^2\right)^{1/2},$$

where m is the embedded dimension, τ is a time delay and ε is a size of local area.

Now let us consider the problem of surrogate time series construction. A general Gaussian linear process can be written as:

$$x_n = \sum_{i=1}^{M} a_i x_{n-i} + \sum_{i=0}^{N} b_i \eta_{n-i}\ , \qquad (3)$$

where a_i and b_i are weight coefficients and $\{\eta_i\}$ are Gaussian uncorrelated random increments. The statistical test we want to perform is not to check correspondence to a particular linear process, but to a whole class of such processes (complete hypothesis). For this purpose we need to generate surrogate time series which are linear and have the same first and second order quantities as our process. The simplest way to get such surrogates is to generate, so called, Fourier based surrogates [8]. They are generated as follows:

1. The squared amplitude of Fourier transform, which is the periodogram estimator of the power spectrum, is calculated:

$$|X_k|^2 = \left|\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} x_n e^{i2\pi kn/N}\right|^2 .$$

2. Surrogates are created by multiplying the Fourier transform of data by random phases and back transform to time domain. They are the following:

$$\bar{x}_n = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} e^{ia_k}\,|X_k|\,e^{-i2\pi kn/N}\ ,$$

where a_k are independent uniform random numbers from interval $[0,2\pi)$.

The correspondence of real data to Gaussian distribution is rather rare event in nature. The most often data can be transformed to Gaussian distribution by some invertable function $s_n=s(x_n)$, where $\{x_n\}$ is the process presented by formula (3). To test such type of null hypothesis amplitude adjusted Fourier transform (AAFT) method can be used for surrogate series generation [8]. This method is good for test if N is rather large, and correlations are not too strong. Otherwise, there is a bias towards too flat spectrum. In our case we can use such method for our tests.

The main part of this method is the invertion to Gaussian distribution. After that described above Fourier based surrogates are generated and result is rescaled back to empirical distribution. The rescaling procedure is the following:

1. Generate the Gaussian process (g_n) with the same range of values, mean and variance as real data.
2. Order both series in increasing order.
3. For each element of initial time series construct the rescaled one:

$$r_n = g_{rank(s_n)}\ ,$$

where $rank(s_n)$ means the number of element s_n in ordered version of initial series.

Test on nonlinearity with surrogate series has been performed to real data. Surrogates were generated by AAFT method. As a measure of nonlinearity time reversibility was used.

To perform test on nonlinearity first the size of test was determined. We want to have significance level 96%. So the probability to reject the true hypothesis (α) in this case is – 0.04. We are going to use two sided test (time asymmetry can go in both two ways), so we need to generate 50 surrogate realisations.

Several surrogates generated for deseasoned data are presented in figure 21.

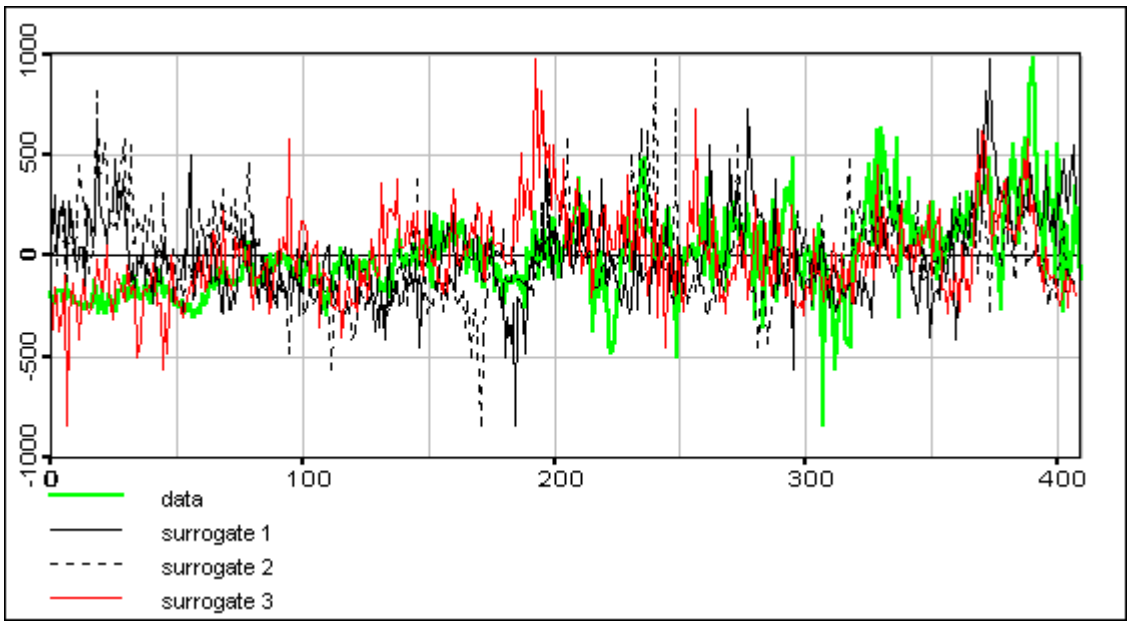

Fig. 21. Examples of surrogate series generated by AAFT method for data after deseasonalized transformation

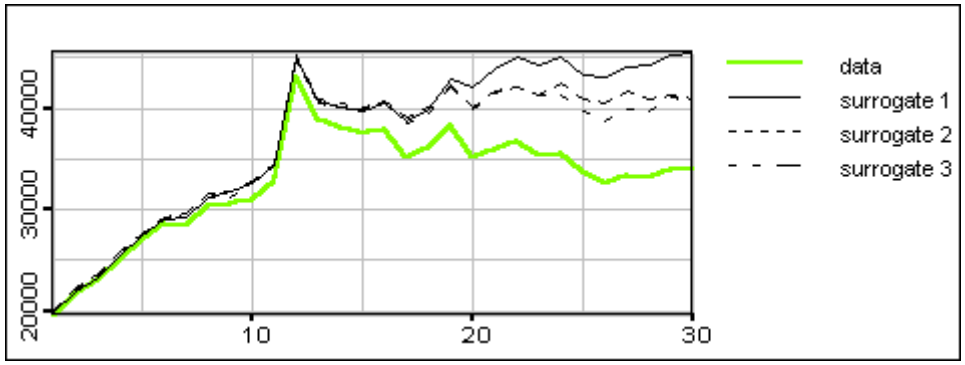

Fig. 22. Variograms for surrogate series generated by AAFT method for data after deseasonalized transformation

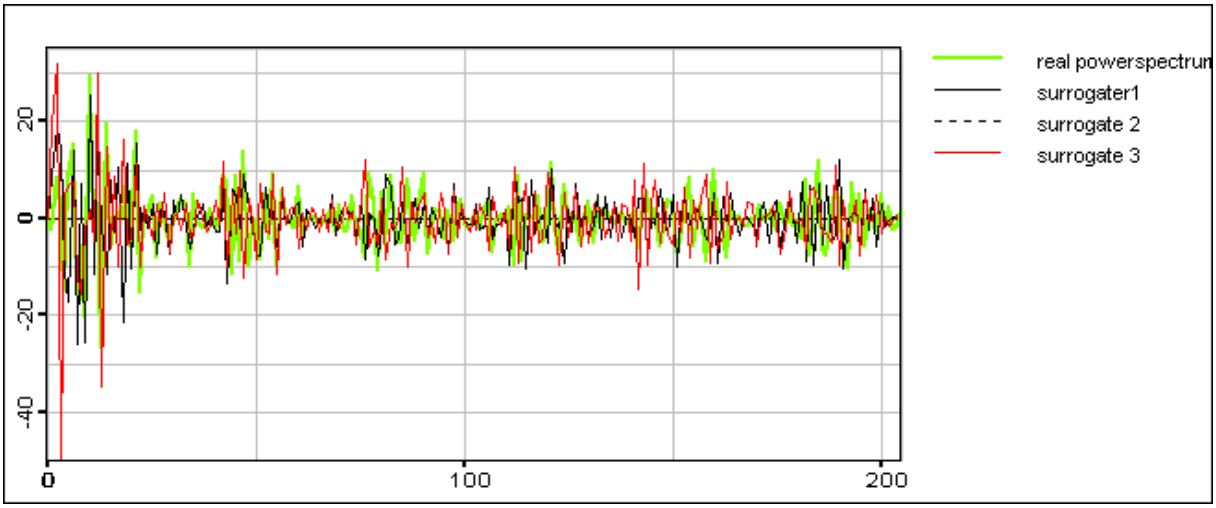

Fig. 23. Power spectrum of surrogate series generated by AAFT method for data after deseasonalized transformation

In figures 22 and 23 one can see how generated surrogates reproduce some characteristics of data – correlation structure and power spectrum. I'd like to note here, that AAFT method is based on power spectrum of

data, correlation structure has been reproduced by the way. So these surrogates can represent data. The obtained result of nonlinearity test is presented in figure 24. The value of data time reversibility is nearly equal to the averaged value of surrogates' time reversibilities. So we can consider deseasoned data as linear process.

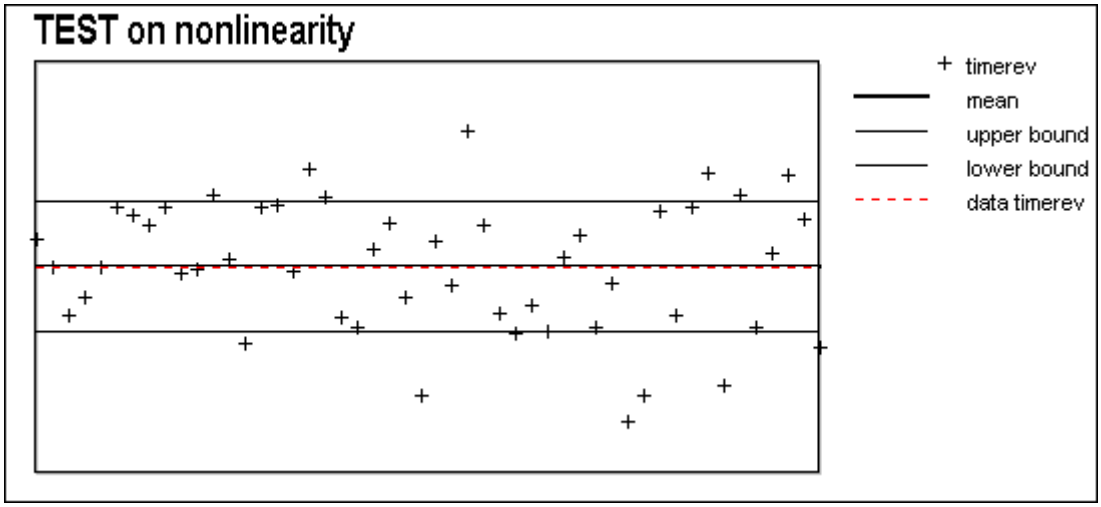

Fig. 24. Results of test on non-linearity of deseasonalised data

Several surrogate series generated for data after differencing with lag 1 are presented in figure 25. Analysis on their reproduction of correlation structure and power spectrum are presented in figures 26 and 27. The reproduction of correlation structure is much better then for previous surrogates (for deseasonalized data). It means that this surrogates are more similar to the data and test is more reliable.

The result of test is presented in figure 28. One can see that data time reversibility is in the range of values considered as linear, but rather close to the upper boundary. So the process can be considered as linear one or more surrogates have to be generated to improve the quality of statistics.

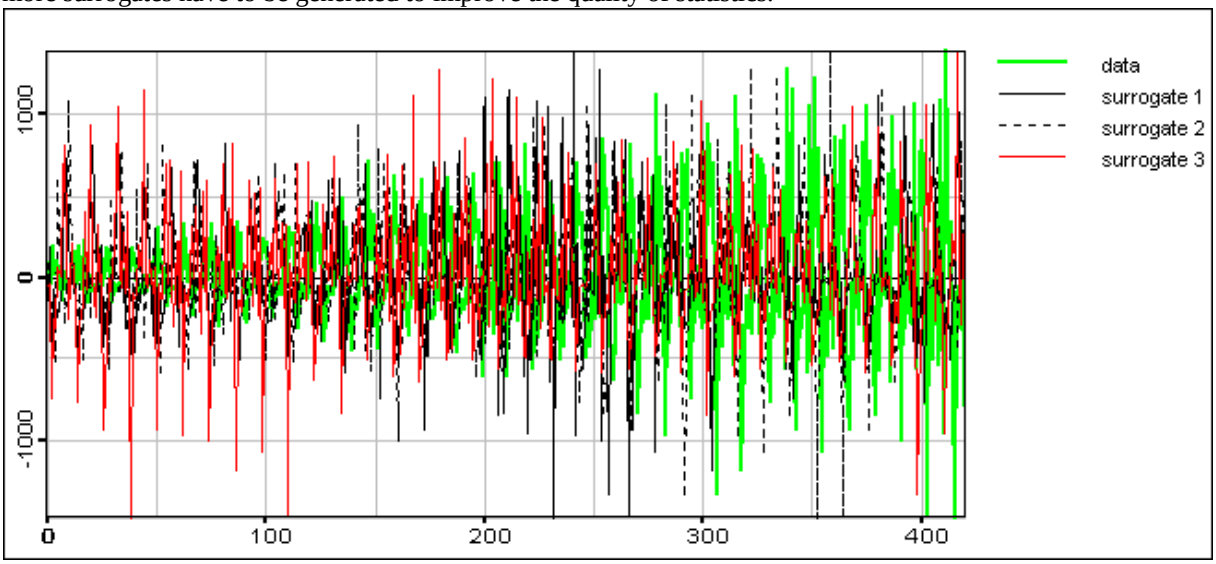

Fig. 25. Examples of surrogate series generated by AAFT method for data after differencing transformation with lag 1

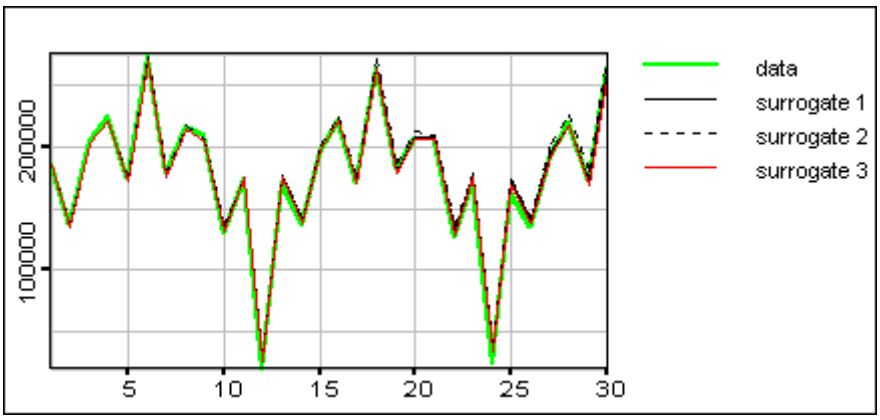

Fig. 26. Variograms for surrogate series generated by AAFT method for data after differencing transformation with lag 1

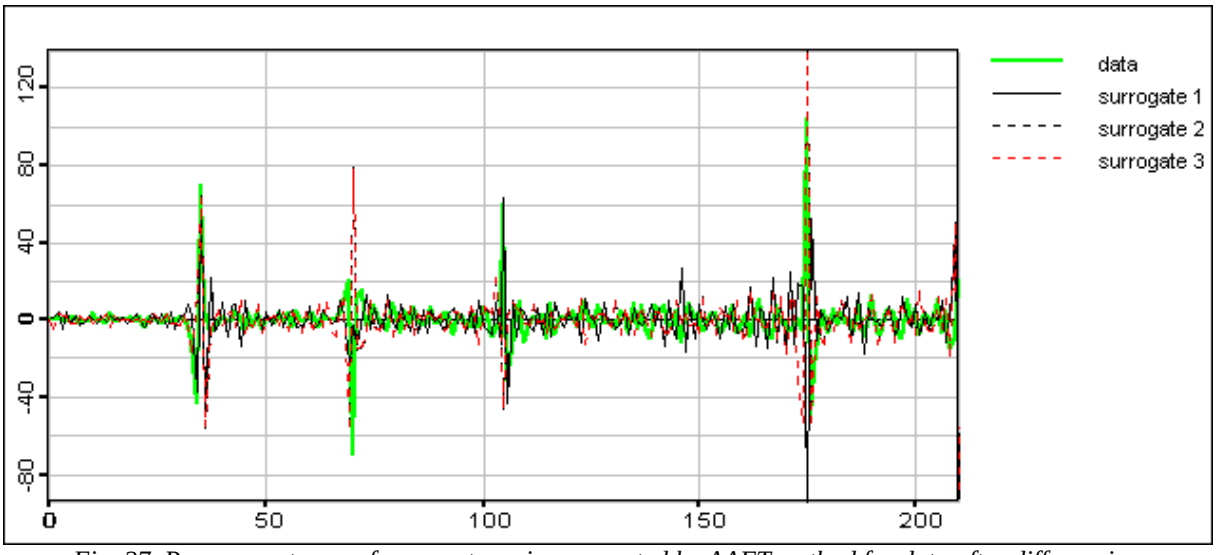

Fig. 27. Power spectrums of surrogate series generated by AAFT method for data after differencing transformation with lag 1

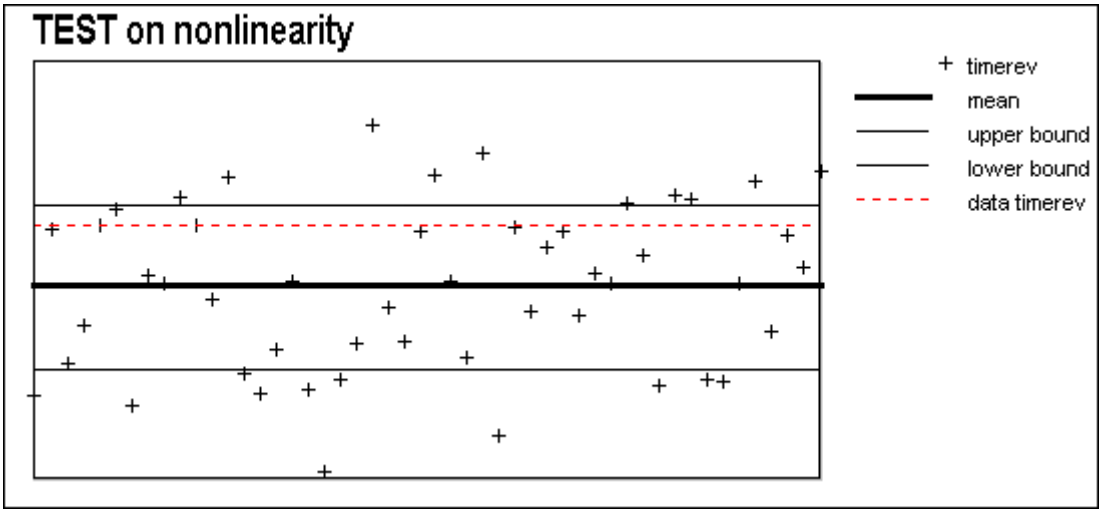

Fig. 28. Results of test on non-linearity for data differenced with lag 1

2.3 Data modelling and forecasting

The second step of data analysis is data modelling: the construction of a model describing the historical behaviour of the system. The main idea is to establish a relationship between some number of inputs and outputs

so to make possible the following forecasting. The model itself, its input parameters, architecture and internal model parameters have to be adjusted using as criteria the error of the following forecasting.

There exist the great amount of different methods for modelling and forecasting time series [13]. All methods can be divided into linear and non-linear models. Linear methods describe the output as a linear combination of inputs or any other known variables (for example realisations of totally known process). Application of linear methods supposes linearity of system or measured signal.

The examples of linear methods are kriging (see section 1) and ***ARMA(p,q)*** process.

The time series $\{X_t\}$ is an ***ARMA(p,q)*** process if $\{X_t\}$ is stationary and if for every t

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},$$

where $\{Z_t\}$ – is white noise process with zero mean and σ^2 variance and polynomials $(1-\phi_1 z - \ldots \phi_p z^p)$ and $(1+\theta_1 z + \ldots + \theta_q z^q)$ have no common factors. It has been proven that if polynomial $(1-\phi_1 z - \ldots \phi_p z^p)$ has no solutions in the circle with radius 1 ($|z| \le 1$), such ***ARMA(p,q)*** can be rewritten as a linear autoregressive process:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \text{with} \quad \sum_{j=0}^{\infty} \left|\psi_j\right| < \infty .$$

It is known about ***ARMA(p,q),*** that all its coefficients $\phi_1 - \phi_p$, $\theta_1 - \theta_p$ can be defined through partial autocorrelation function of a process. All methods of ARMA coefficients' estimation are based on this knowledge. Details about ***ARMA(p,q)*** process and its application to time series forecasting can be found in [3, 11].

Non-linear methods include any non-linear function to model non-linearity of a process. The most distributed non-linear methods are polynomials of order 2 and higher, radial basis functions, multi layer perceptrons and others. Sometimes such systems can be considered as non-linear dynamic system. There some special approaches for their modeling in embedded space (details see for example in [1, 14]).

The answer on the question what type of model to chose has to be based on the previously made analysis. Some parameters of model also can be taken from knowledge obtained during that analysis.

For modelling and analysis of obtained residuals data was divided into two parts – training data and test data. Training data were used to train a model: adjust its parameters. Test data were not used for parameters' fitting, only to test obtained forecasting.

2.3.1 Linear modeling and forecasting

In the previous section we found out that the process under study after some transformations (deseasoning and differential with lag 1) behaves as linear process. So we shall try some linear models for their modelling and forecasting. As test data the last 20 values of time initial time series have been selected.

ARMA(p,q) model of a process was used for modeling and forecasting of deseasoned data. ***ARMA(1,26)*** was chosen by using minimal AICC criterion. The results of applying such model are presented in figures 29 and 30. Results are presented after backward to deseasoning transformation.

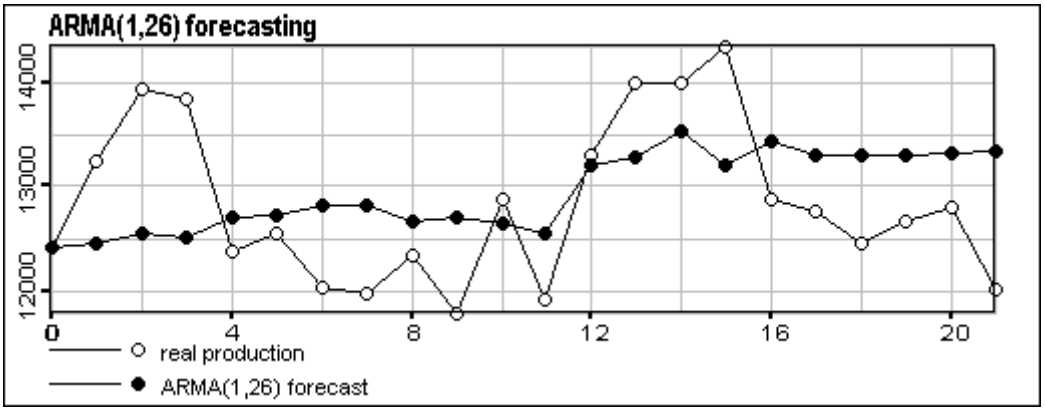

Fig. 29. The result of ARMA(1, 26) forecasting for deseasonalised data

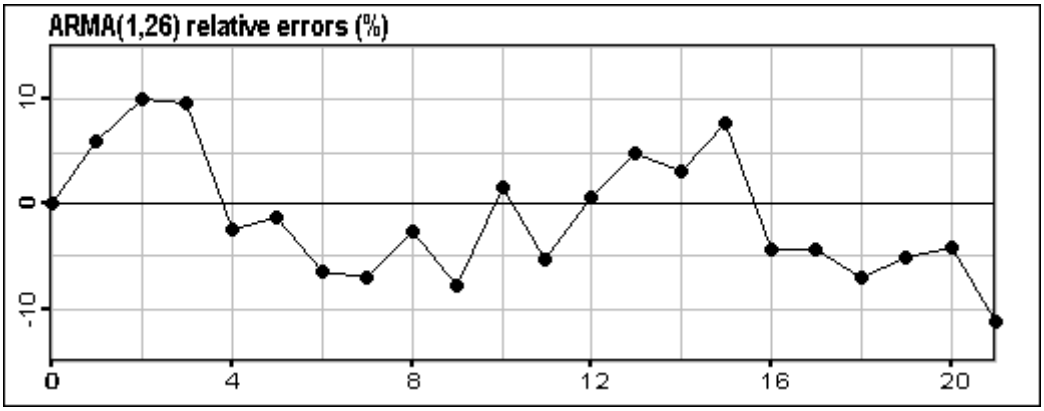

Fig. 30. Relative errors of ARMA(1, 26) forecasting for deseasonalised data

Correlation analysis of residuals of ***ARMA(1,26)*** forecasts were performed. They have been done with the help of variogram (figure 31). One can see the presence of correlation structure in residuals.

To perform geostatistics test data have been divided into two parts: one to model the residuals and the second to test the final result. The result of variogram fitting by a spherical model can be seen in figure 31. Ordinary kriging for residuals have been performed. The result of applying kriging to residuals is presented in figures 32 and 33.

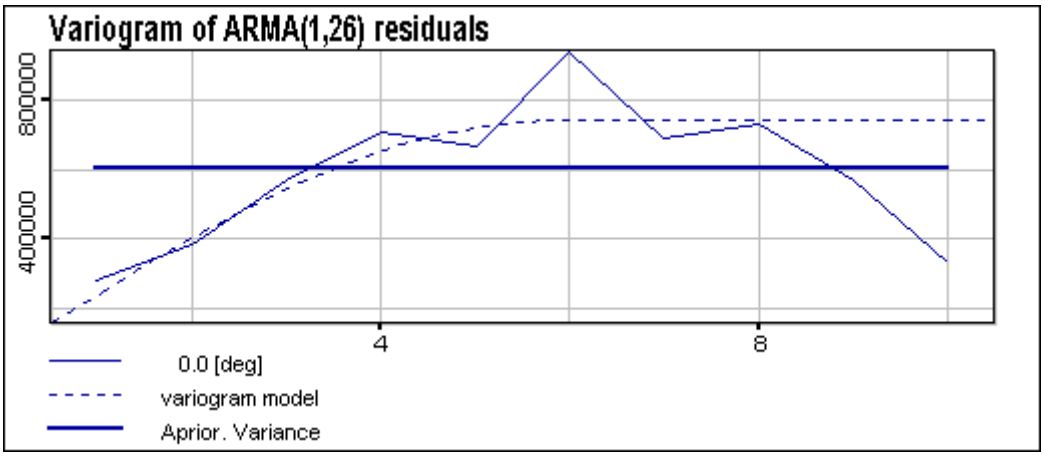

Fig. 31. Correlation structure of residuals of model ARMA(1,26) applied to deseasoned data

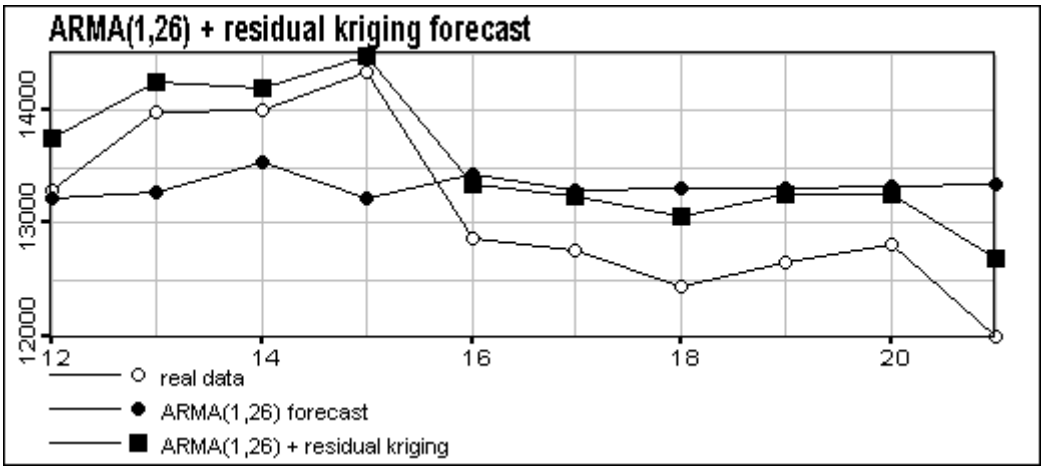

Fig. 32. The result of ARMA(1, 26) forecasting for deseasonalised data plus residual kriging

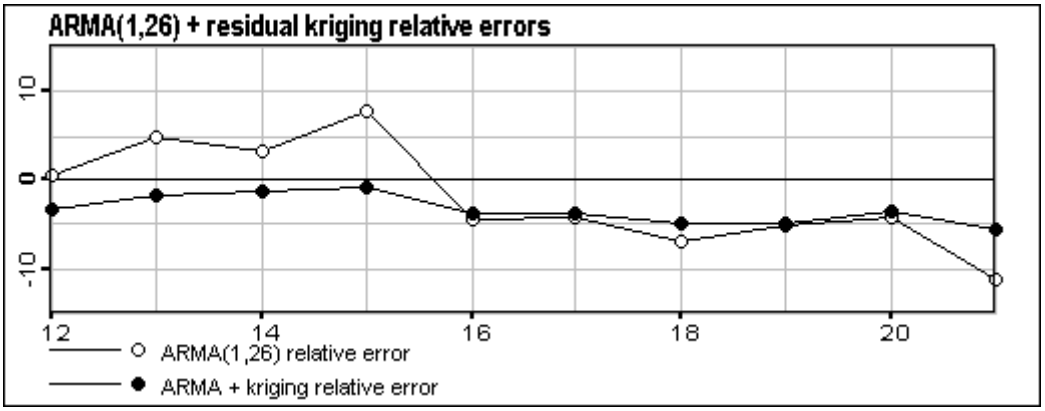

Fig. 33. Relative errors of ARMA(1, 26) forecasting for deseasonalised data plus residual kriging

Correlation analysis for residuals of the whole method (ARMA(1,26) for deseasoned data + ordinary kriging of residuals) has been performed. In spite of some improvement of forecasting results the final residuals still have some correlation structure (see figure 34). Even the radius where correlation between data is visible remained the same (compare figures 31 and 34). It can be explained by difference of kriging application for extrapolation zone. The correlation between errors more distant from the beginning point. Other methods (for example simulations) may be better, for they reproduce the whole process at once.

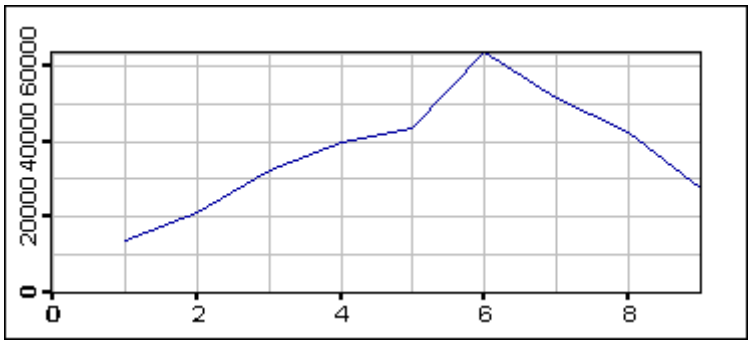

Fig. 34. Correlation structure of residuals of ARMA(1,26) for deseasonalised data plus residual kriging

Analogous modelling has been done for data inflicted by differential transformation with lag = 1. In this case using minimum AICC criterion *AR(24)* model has been chosen. The results of applying this model are presented in figures 35 – 38. One can see that the quality of forecasting is rather good – relative errors never overcome 10% and the most part of them is within 5% border. But the amplitude of waving of residuals around zero is growing.

The correlation analysis of residuals has been performed. The variograms constructed for all (22) residuals and for first part (first 12) are presented in figure 38. One can see the presence of correlation structure. But it is impossible to model it and perform kriging using first part of residuals. So we can't verify the final forecasting result. Any model without testing can't be proposed for further usage.

So in this case we have correlation structure but can do nothing with it to improve the model. The presence of correlation in residuals shows that linear model doesn't reproduce all correlation structure of differences. But as there is no structure in first part the correlation can be caused by errors correlations between errors in more distant from the beginning of forecasting points.

We have seen that linear models can be used for modelling and forecasting of the process, but there are some problems concerning forecasting of distant times.

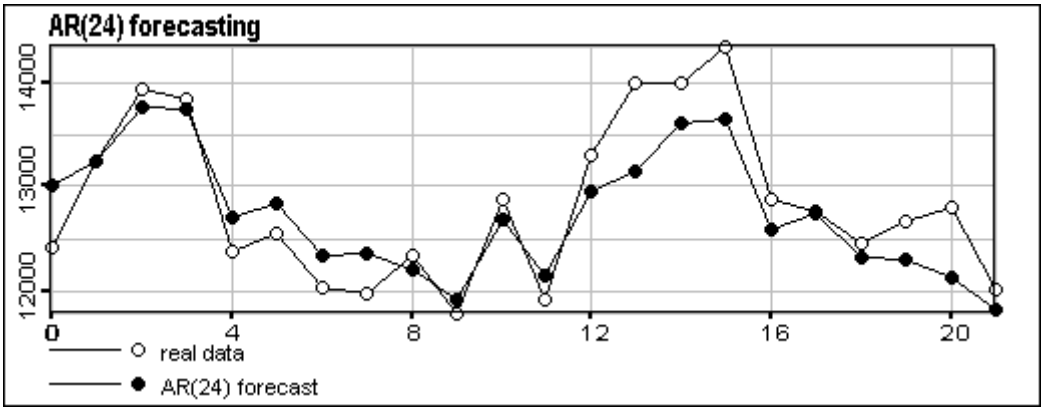

Fig. 35. The result of AR(24) forecasting for data differenced with lag 1

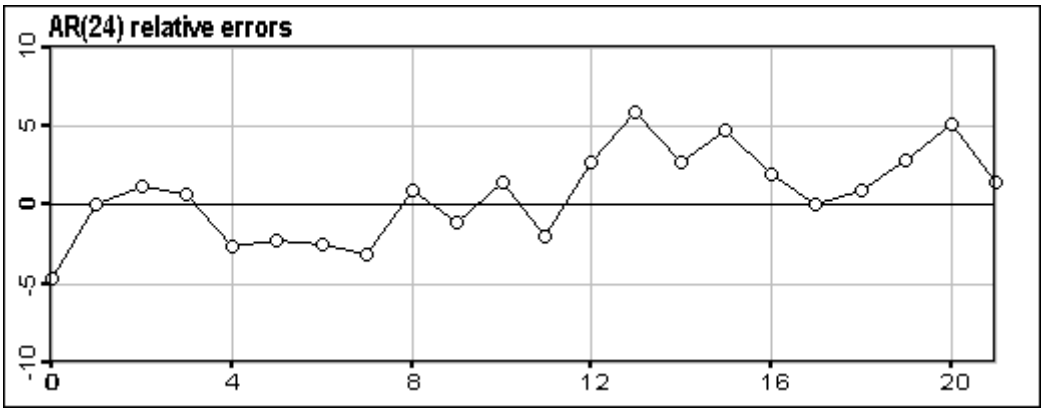

Fig. 36. Relative errors of AR(24) forecasting for data differenced with lag 1

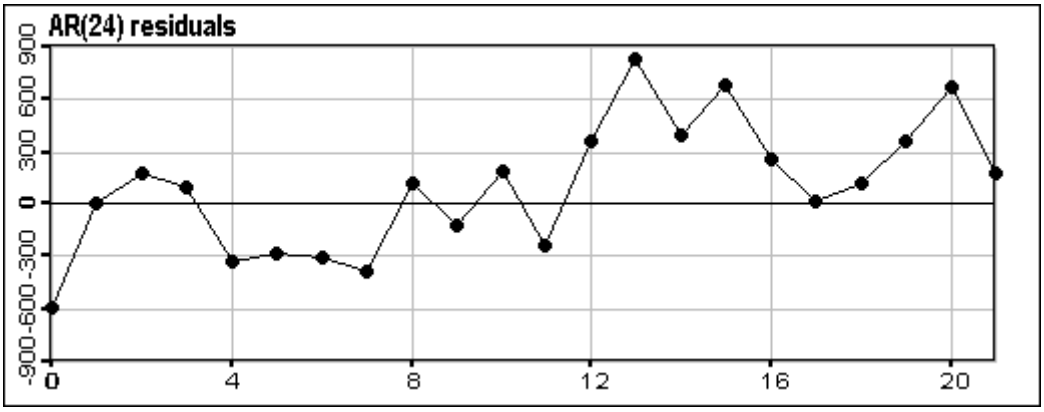

Fig. 37. Residuals of AR(24) forecasting for data differenced with lag 1

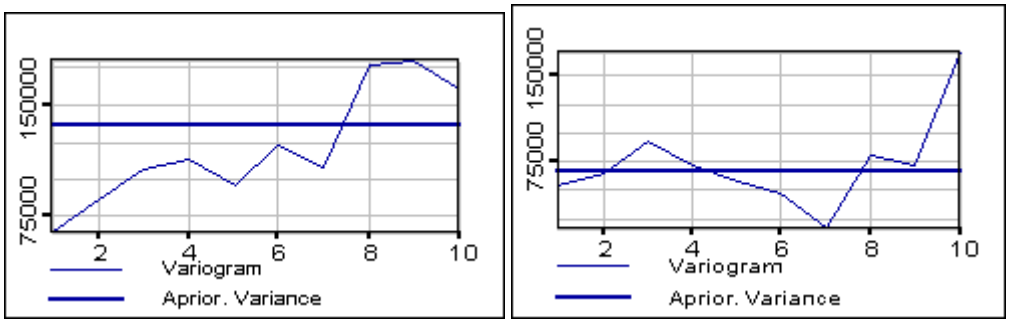

Fig. 38. Variograms for residuals of AR(24) applied to data differenced with lag 1: variogram constructed on all residuals - left, on first 12 residuals – right

2.3.2 Non-linear modelling and forecasting

As a tool for nonlinear modelling and forecasting artificial neural network multi layer perceptron (ANN MP) was selected. Theory on multi layer perceptron – its structure, history and principles of learning can be found in great amount of literature, for example in [15].

For modelling of an electricity monthly production in Australia we constructed a perceptron with 3 inputs, 2 hidden layers with 2 neurons in each and 1 linear output. Inputs were the following:

1. the value of electricity production this month;
2. the value of electricity production in the previous month;
3. the value of electricity production in the same month (as forecasting) a year before.

Such selection of inputs was based on a simple common sense: the periodical waving is taken into account by the value one a year ago and the trend – by 2 previous months. The forecasting was made in recursive way – the forecasted value was instantaneously included in forecasting procedure for the following one.

The results obtained by applying of described above ANN MP are presented in figures 39 and 40. One can see that relative errors of the forecast never overcome 5%. It seems to be rather good result. It was not obtained by using linear methods. Nevertheless the correlation structure analysis of residuals has been performed. The variogram of residuals is presented in figure 41. The presence of correlation structure is seen. It can be modeled by spherical model and kriging can be performed to model residuals.

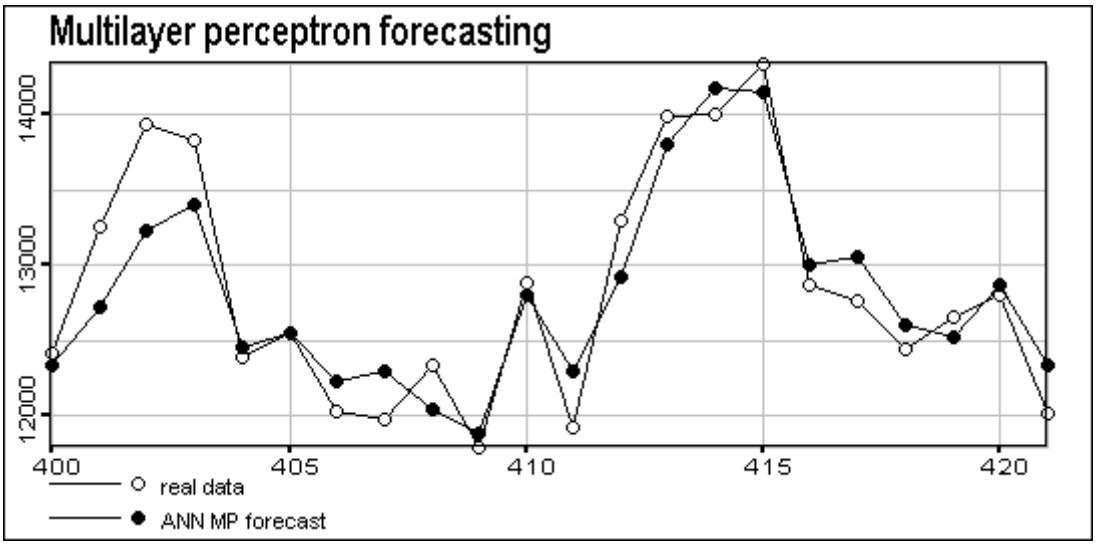

Fig. 39. Monthly electricity production forecasting by ANN multilayer perceptron

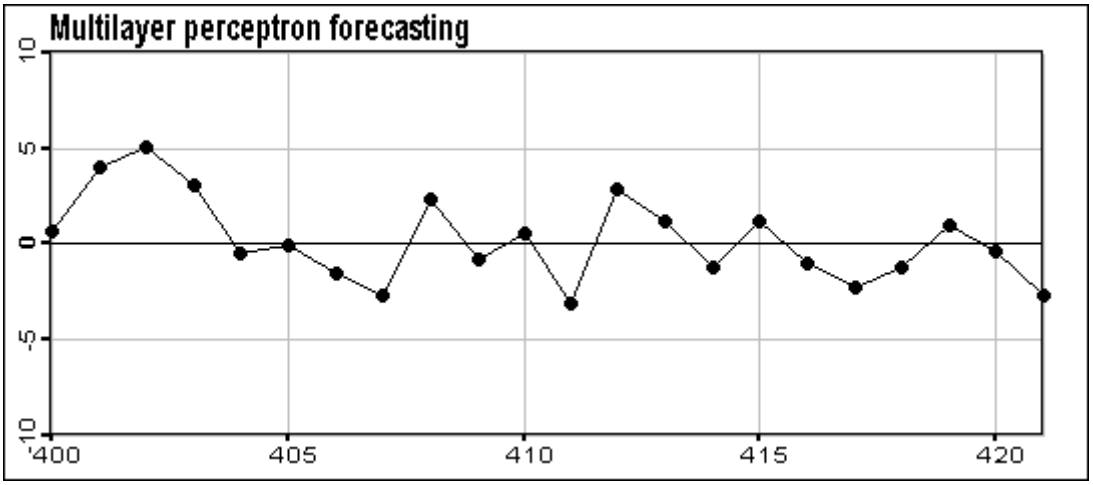

Fig. 40. Relative errors of monthly electricity production forecasting by ANN multilayer perceptron

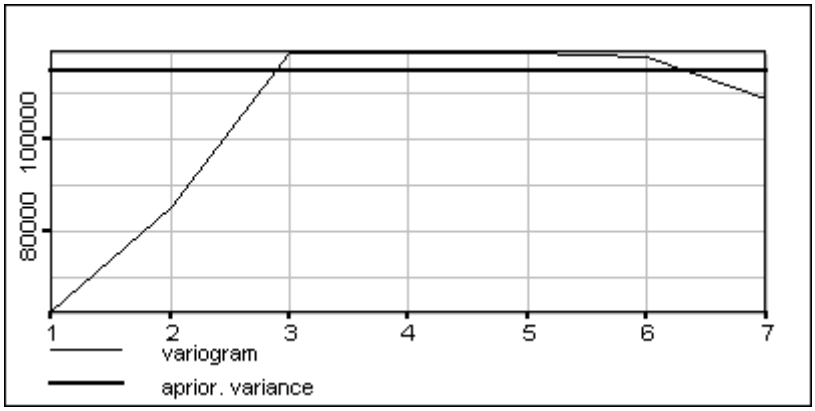

Fig. 41. Variogram of residuals of ANN multilayer perceptron forecasting

The results obtained after application ordinary kriging to ANN residuals are presented in figures 42 and 43. One can see that somewhere improvement has been achieved. But in general the improvement by ordinary kriging is under question. The presence of correlation in residuals indicates the possibility to use geostatistics, maybe any other method will work better.

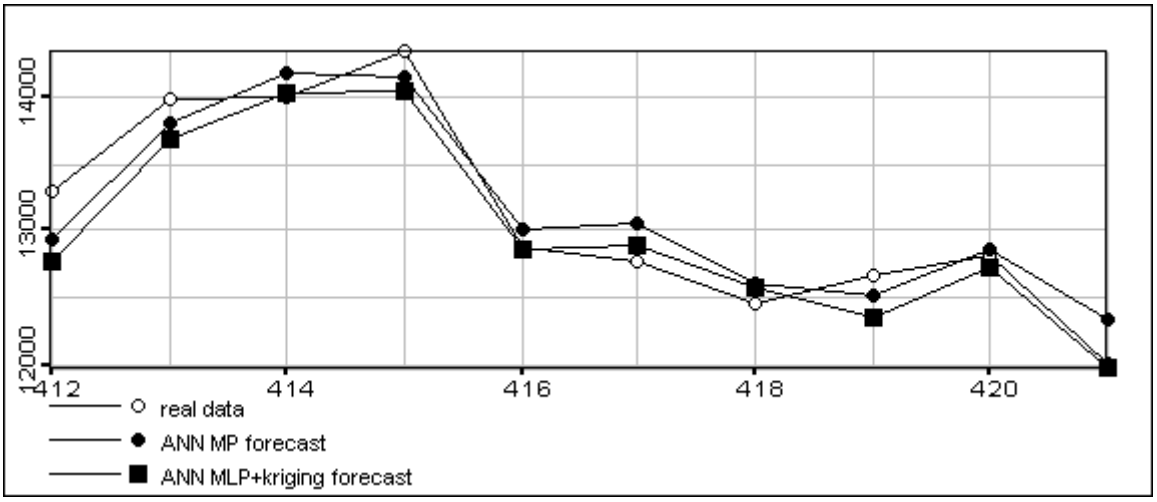

Fig. 42. Monthly electricity production forecasting by ANN multilayer perceptron and ordinary kriging

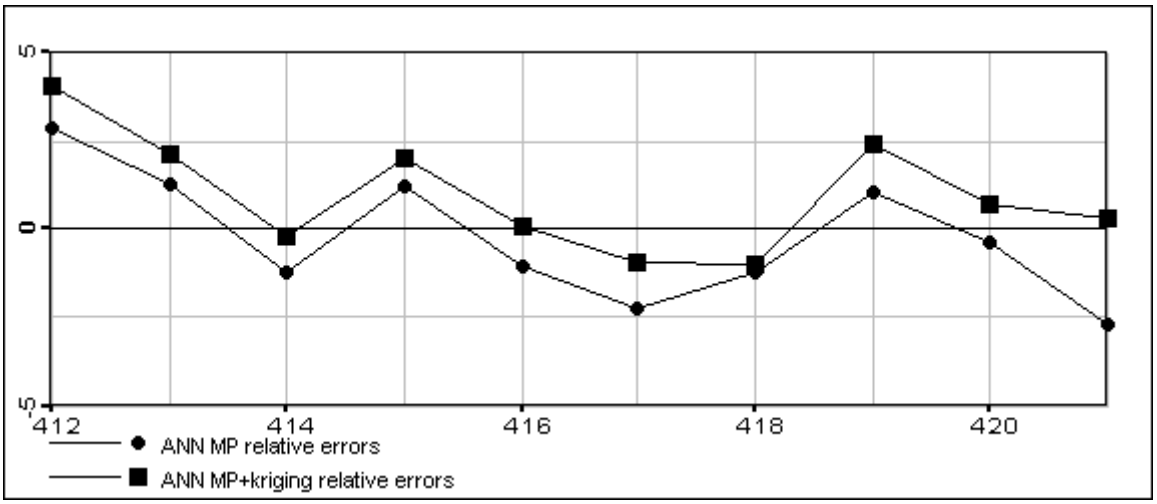

Fig. 43. Relative errors of monthly electricity production forecasting by ANN multilayer perceptron and kriging

We have seen that linear models can be used for modelling and forecasting of the process, but there are some problems concerning forecasting of distant times.

Conclusions

In this work it was shown that geostatisticaly based methods can be applied to 1D problem of electrical load forecasting. Their application can be different:

1. Ordinary kriging can be used for electrical load forecasing for stable periods – without unexpected changes in the weather and for days of one type (working days). Perturbations of data types lead to distortion in forecast.
2. Stochastic simulation methods can be applied to residuals of other (for example MLP) residuals if those residuals are correlated. Stochastic simulations provide description of uncertainty of obtained result.
3. Ordinary kriging also can be applied to residuals of both linear and non-linear methods. It was demonstrated on middle term forecasting.

It was also shown in this work that preliminary data analysis is of a great importance for further system modeling and forecasting. Different methods of preliminary time seriees analysis have been demonstrated.

Acknowlegments

The work was partly supported by INTAS grant 97-31726.

References

1. Weigend A.S., Gershenfeld N.A., eds., *Time Series Prediction: Forecasting the future and understanding the past*, Santa Fe Institute Studies in Science of Complexity, Proc. Vol. XII, Addison-Wesley, Reading, MA, 1993.
2. Piras. A. *A multiresponse structural connectionist model for short term electrical load forecasting*, These 1546, Ecole Polytechnique federal de Lausanne, 1996.
3. Arutyunyan R.V., Bogdanov V.I., Bolshov L.A., Demyanov V.V., Kanevski M.F., Lasarev A.L., Ogar V.P., Savelieva E.A., Timonin V.A., Chernov S.Yu., Jushin V.N. Forecasting of electricity comsuumption: Time series analysis, geostatistics, artificial neural networks (in Russian). Preprint IBRAE-99-05, Moscow, 1999, 45 p.
4. Kravetski A.S., Kanevski M.F., Savelieva E.A., Timonin V.A., Chernov S.Yu., Ogar V.P., Prediction of electricity load by multi layer perceptron (in Russian), Preprint IBRAE-2000-07, Moscow, 2000, 22 p.
5. Mathron G., Principles of geostatistics, Economic Geology, V. 58, pp. 1246-1266, 1963.
6. Goovaerts P., *Geostatistics for Natural resources*, Oxoford University Press, 1997.
7. Problems of Environmet and Nature resources (in Russian) N11, VINITI, Moscow, 1999
8. Schreiber T., Schmidt A., Surrogate time series, http://www.mpipks-dresden.mpg.de/…/docs/surropaper/Surrogates.html, submitted to Physica D, 1999.
9. Savelieva E., Kravetski A., Chernov S., Demyanov V., Timonin V., Arutyunyan R., Bolshov L., Kanevski M., Application of MLP and stochastic simulations for electricity load forecasting in Russia, Procceedings of ESANN2000, Bruges April 26-28 2000, pp.413-418, 2000.
10. Kirkpatric S., Gelatt C.D., Vecci M.P., Optimization of simulated annealing, Science, V.220, N. 4598, pp. 671-680, 1983.
11. Brockwell P.J., Davis R.A., *Introduction to time series and forecasting*, Springer-Verlag New-York Berlin Heidelberg, 1995, 420 p.
12. Schreiber T., Discrimination power of measures for nonlinearity in time series, Physical Review E, V. 55, N. 5, pp. 5443-5447, 1997.
13. Hamilton J.D., Time series analysis, Princeton University Press, Princeton, New Jersey, 1994, 798 p.
14. Schreiber T., Interdisciplinary application of nonlinear time series methods, Physics Reports, V. 308, N. 1, pp. 1-64, 1999.
15. Hyakin S., Neural networks. A comprehensive foundation, Macmillan College Publishing Company, 1994, 490 p.