Time Series Forecasting with Metric Analysis Approach

Victor V. Ivanov[†*], **Alexander V. Kryanev**[*†],
Leonid A. Sevastianov[‡†], **David K. Udumyan**[*‡§]

[*] *National Research Nuclear University "MEPhI"*
[†] *Joint Institute for Nuclear Research*
[‡] *People' Friendship University of Russia (RUDN University)*
[§] *University of Miam*

Email: `ivanov@jinr.ru, a_v_kryanev@mtu-net.ru, sevastianov_la@rudn.university, mathudum@gmail.com`

Time series forecasting scheme based on the metric analysis approach is presented. This approach provides preliminary filtering of noisy components with the help of singular-spectral analysis.

The scheme uses an auto regression model in which the predicted value is considered as a function of the m previous values of the filtered values of this series. Thus, the auto regression model reduces the prediction problem to the problem of nonlinear interpolation of a function of several variables that are the smoothed (filtered) values of the original chaotic time series. To solve the problem of interpolation of such functions, the article uses metric analysis, which makes it possible to reveal the patterns of auto regression dependence in the deterministic component of the investigated chaotic time process. In classical interpolation schemes, the interpolation function is recovered immediately in the entire region of independent variables under consideration. In the metric analysis, the interpolation values are restored separately at each point, taking into account the location of the point at which the value of the function is restored, with respect to the location of the points at which the values of the function are given. Therefore, metric analysis allows us to take into account individual features of the location of points in which the values of the function are restored directly in the interpolation formula, which makes it possible to obtain a more accurate result in the absence of additional information about the function.

The article reviews time series from various areas, such as stock prices, sales volumes, passenger traffic volumes in the metro, and electricity consumption in the Moscow region. The presented examples demonstrate the effectiveness of the presented scheme. In particular, it is shown that with the help of the presented scheme it is possible to predict the dynamics trend of the time chaotic time series under investigation with acceptable accuracy by several tens of time steps in advance. The accuracy of the forecast largely depends on the choice of the dimension of the forecast model — the number m of previous values of this auto regression series.The scheme allows one to select an optimal value of the dimension of the forecasting model, which, on average, provides the best accuracy for prediction.

Key words and phrases: ime series, auto regression model, interpolation of functions of many variables, metric analysis, prediction scheme, forecasting examples .

1. Introduction

One of the main problems of data processing in many areas is the problem of predicting the values of time processes. To date, many different methods and schemes have been developed which solve various particular problems of forecasting time processes [1, 2]. Below there is a brief description of the metric analysis of predicting the values of time series [3–5] and its application for the prediction of specific time series [6].

2.   Forecasting Scheme

The prediction scheme uses the interpolation method for the functional dependence:

$$Y = F(X_1, \ldots, X_m) = F(\vec{X}),$$

where the function $F(\vec{X})$ is unknown and is the subject for recovery, either at one point $\vec{X}^*$ or in a set of given points on the basis of known values of the function Y_k, $k = 1, \ldots, n$, at fixed points $\vec{X}_k = (X_{k1}, \ldots, X_{km})^T$ [3].

According to the method of interpolation, based on metric analysis, interpolation values are found as solutions of problems to a minimum according to $\vec{z} = (z_1, \ldots, z_n)^T$ the measure of metric uncertainty [3–5]

$$\sigma_{ND}^2(Y^*; \vec{z}) = (W(\vec{X}^*; \vec{X}_1; \ldots; \vec{X}_n)\vec{z}, \vec{z}),$$

and the interpolation value is determined by a linear combination

$$Y^* = \sum_{i=1}^{n} z_i Y_i,$$

and is given by

$$Y^* = \frac{(W^{-1}\vec{Y}, \vec{1})}{(W^{-1}\vec{1}, \vec{1})}.$$

The matrix of the metric W uncertainty is defined by

$$W = \begin{pmatrix} \rho^2\left(\vec{X}_1, \vec{X}^*\right)_{\vec{w}} & \left(\vec{X}_1, \vec{X}_2\right)_{\vec{w}} & \ldots & \left(\vec{X}_1, \vec{X}_n\right)_{\vec{w}} \\ \left(\vec{X}_2, \vec{X}_1\right)_{\vec{w}} & \rho^2\left(\vec{X}_2, \vec{X}^*\right)_{\vec{w}} & \ldots & \left(\vec{X}_2, \vec{X}_n\right)_{\vec{w}} \\ \ldots & \ldots & \ldots & \ldots \\ \left(\vec{X}_n, \vec{X}_1\right)_{\vec{w}} & \left(\vec{X}_n, \vec{X}_2\right)_{\vec{w}} & \ldots & \rho^2\left(\vec{X}_n, \vec{X}^*\right), \end{pmatrix},$$

where $\rho_{\vec{w}}^2\left(\vec{X}_i, \vec{X}^*\right) = \sum\limits_{k=1}^{m} w_k \left(X_{ik} - X_k^*\right)^2$,

$$\left(\vec{X}_i, \vec{X}_j\right)_{\vec{w}} = \sum_{k=1}^{m} w_k \left(X_{ik} - X_k^*\right) \cdot \left(X_{jk} - X_k^*\right), \quad i, j = 1, \ldots, n.$$

Consider the function of time $ty = f(t)$ with values $Y_1 = f(t_1), \ldots, Y_n = f(t_n)$ for $t_1 < \ldots < t_n \in [t_1, t_n]$.

It is required to find the predicted value y_{n+1} for t_{n+1}.

The problem of finding the predicted value y_{n+1} is reduced to the problem of interpolation of functions of several variables by means of a nonlinear autoregressive model [3–5]

$$y(t_{m+1}) = y_{m+1} = F(y_1, \ldots, y_m),$$
$$y(t_{m+2}) = y_{m+2} = F(y_2, \ldots, y_{m+1}),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y(t_N) = y_N = F(y_{N-m}, \ldots, y_{N-1}).$$

Then the prediction of the function $y = f(t)$ is reduced to interpolating the function of m variables $Y = F(y_1, y_2, \ldots, y_m)$ with values in $n - m$ points

$$\vec{X}_1 = (Y_1, \ldots, Y_m)^T,$$
$$\vec{X}_2 = (Y_2, \ldots, Y_{m+1})^T,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\vec{X}_{n-m} = (Y_{n-m}, \ldots, Y_{n-1})^T.$$

The predicted value $y_{\text{for}} = y_{n+1}$ is defined as the interpolation value of the function $Y = F(y_1, y_2, \ldots, y_m)$ at the point $\vec{X}^*$:

$$y_{n+1} = F(\vec{X}^*) = \frac{(W^{-1}\vec{1}, \vec{Y})}{(W^{-1}\vec{1}, \vec{1})},$$

where, $\vec{X}^* = (Y_{n-m+1}, \ldots, Y_n)^T$, W^{-1} is the inverse matrix to the $(n - m) \times (n - m)$ matrix of metric uncertainty, $\vec{Y} = (Y_{m+1}, \ldots, Y_n)^T$ is the $(n - m)$-dimensional vector of the values of the predicted time process.

The natural number m determines the dimensionality of the space of vectors $\vec{X}$ and its value is found as the solution of the extremal problem [3–5]

$$m = \operatorname{argmin}\|\vec{Y} - \vec{Y}_{\text{for}}\|.$$

In the scheme proposed in this paper, preliminary filtering (trend isolation) of the original time series is used with the help of singular-spectral analysis [1].

3. Numerical Results

Fig. 1 shows one example of forecasting the daily closing prices for a company's shares using the metric analysis scheme.

Fig. 2 shows the forecast of the sums of one-time shoe store sales (data provided by A. Khokhlov).

Figs. 3 and 4 show the forecasting of the daily passenger traffic (in thousands of passengers) on the Moscow metro during various periods of 2014 (the source of the data is the Moscow Metro) (data provided by E. Osetrov, see also [6]).

Figs. 5 and 6 show the forecasting the daily electricity consumption (in billion kilowatt-hours) in the Moscow region (Moscow and Moscow region) only on working days during different periods of 2014 (data source — System Operator of the Unified Energy Systems / JSC "SO UES" branch of the Moscow Regional Dispatch Office) (data provided by E. Osetrov, see also [6]).

4. Conclusion

The forecasting scheme presented in this article allows one to predict the trend dynamics of chaotic time series under analysis. The obtained numerical results of prediction with respect to time processes taken from various fields show that the presented scheme yields acceptable results in the accuracy of the forecast.
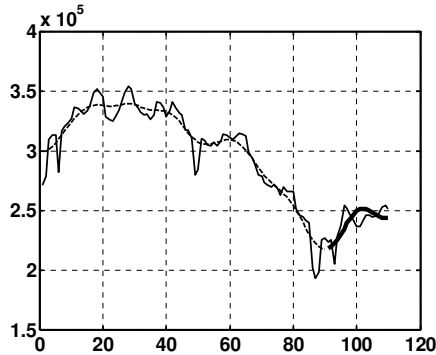
Figure 1. Forecasting 20 steps ahead. The continuous line is original row data, the dashed line is the filtered component, the solid line is the forecast
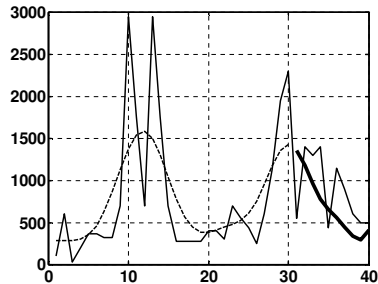(the optimal value of m is 20)

Figure 2. Forecasting 10 steps ahead. The continuous line is the original row data, the dashed line is the filtered component, the solid line is the forecast
(the optimal value of m is 10)

References

1. N. Golyandina, V. Nekrutkin, A. Zhigljavsky. Analysis of Time Series Structure. SSA and Related Techniques. Chapman & Hall / CRS, 2001.
2. A. V. Kryanev, G. V Lukin. Mathematical methods for processing indeterminate data. Leningrad: Nauka, 2006 [in Russian].
3. A. V. Kryanev, D. K. Udumyan, Metric Analysis, Properties and Applications as a Tool for Forecasting. International Journal of Mathematical Analysis (2014), Vol. 8, no. 60, pp. 2971–2978.
4. V. V. Ivanov, A. V. Kryanev, D. K. Udumyan, G. V. Lukin, Metric Analysis Approach for Interpolation and Forecasting of Time, Processes. Applied Mathematical
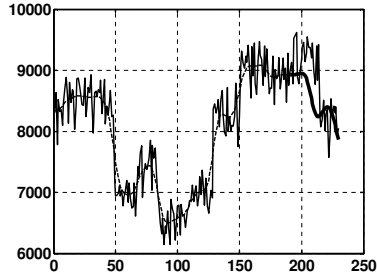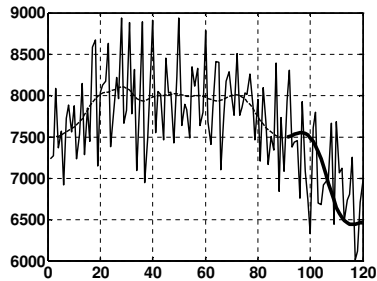
Figure 3. Forecasting 30 steps ahead. The continuous line is the original row data, the dashed line is the filtered component, the solid line is the forecast (the optimal value of m is 20)

Figure 4. Forecasting 40 steps ahead. The continuous line is the original row data, the dashed line is the filtered component, the solid line is the forecast (the optimal value of m is 30)

Sciences, Vol. 8, 2014, no. 22, pp. 1053–1060.

5. A. V. Kryanev, G. V. Lukin, D. K. Udumyan, Metric Analysis and data processing. Leningrad: Nauka, 2012 [in Russian].

6. V. V. Ivanov, A. V. Kryanev, E. S. Osetrov, Forecasting daily electricity consumption in the Moscow region using artificial neural networks, Physics of Particles and Nuclei Letters (2017), Issue 2.
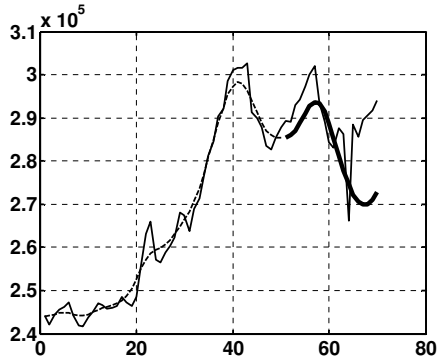
Figure 5. Forecasting 20 steps ahead. The continuous line is the original row data, the
dashed line is the filtered component, the solid line is the forecast
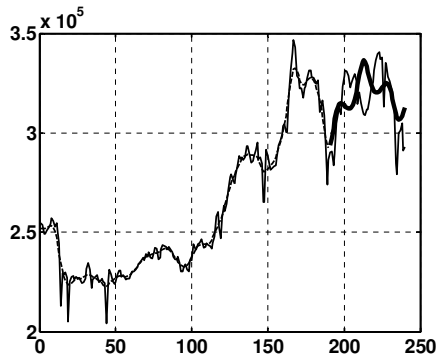(the optimal value of m is 10)

Figure 6. Forecasting 50 steps ahead. The continuous line is the original row data, the
dashed line is the filtered component, the solid line is the forecast
(the optimal value of m is 12)