11. Recurrent method of optimum forecasting the Gaussian random process

Introduction

We consider the vector Gaussian random process *y(t)* with a zero a priori mean and the autocorrelation function

$$M\left\{y(t)\cdot y^T(t)\right\}=K_y(t,\tau)_0 \,. \tag{11.1}$$

The problem of determining the optimum estimate of the process $y(t),\quad t \ge t_k$ is solved on the basis of measurements

$$z_i = y(t_i)+v(t_i), \quad i=1,\ldots,k \,. \tag{11.2}$$

Here $v(t_i),\quad i=1,2,\ldots,k$ is the Gaussian random process with discrete time and independent values with a zero mean value and specified covariation matrix R_i.

The processes *y(t)* и $v(t_i)$ are assumed to be mutually uncorrelated. The estimate, the errors of which have a minimum variance, is considered to be optimum.

This problem (for a scalar stationary process) was first solved by A.N. Kolmogorov [1]. In subsequent works the researchers obtained the generalizations associated with considering the non-stationary and vector random processes [2], [3], The inconvenience in applying the developed techniques was caused by the necessity of solving the system of linear equations, the order of which grows with increasing *k*.

A considerable step on the way of expanding the field of applying the techniques of filtering and forecasting the random processes was done in the works by Kalman, Bucy and their followers [4]. The problem was reduced to the necessity of applying the forming filter and standard procedures of solving the difference equations, which can easily be implemented on computers. However, the application and this approach meet some difficulties as well. A complicated task is the construction of a forming filter, for the non-stationary processes especially. The difficulties are aggravated when the correlation function (11.1) is poorly known a priori and has to be updated in the process of filtering based on obtained estimates $\hat{y}(t)$, which necessitates the reconstruction of the filtering and forecasting algorithm.

The presented brief review indicates the urgency of constructing such a filtering and forecasting technique, which could be applicable to the work with the Gaussian random process of general form. In this case the correlation function (11.1) can be specified discretely over some grid of arguments. The algorithm of such type was first described, apparently, in papers [5] and

[6]. Its basic statements were presented above in section 7. The justification of the technique stated below was published in [7].

Derivation of recurrence functional relationships

For constructing the algorithm we apply the criterion of maximum of the a posteriori probability

$$p[y(t)|Z_i] = \frac{p[y(t), z_i|Z_{i-1}]}{p(z_i|Z_{i-1})} \to \max\ , \tag{11.3}$$

where the sequence $z_1, z_2, ..., z_i$ is designated as Z_i for brevity.

We assume that the estimates

$$M\{y(t)|Z_{i-1}\} = \hat{y}(t)_{i-1}\ , \tag{11.4}$$

$$M\left\{[y(t)-\hat{y}(t)]\cdot[y(\tau)-\hat{y}(\tau)]^T|Z_{i-1}\right\} = K_y(t,\tau)_{i-1}, \quad t,\tau \ge t_{i-1} \tag{11.5}$$

are constructed based on measurements Z_{i-1}.

For determining the value of $\hat{y}(t)$, which provides a maximum to criterion (11.3), it is sufficient to consider the numerator only, because function *y(t)* does not appear in a denominator. We write the expression for density in a numerator

$$p[y(t), z_k|Z_{i-1}] = C\cdot\exp\left\{-\frac{1}{2}\begin{Vmatrix} y(t)-\hat{y}(t)_{i-1} \\ z_i-\hat{y}(t)_{i-1}\end{Vmatrix}^T\cdot\begin{Vmatrix} K & L \\ L^T & M\end{Vmatrix}\cdot\begin{Vmatrix} y(t)-\hat{y}(t)_{i-1} \\ z_i-\hat{y}(t)_{i-1}\end{Vmatrix}\right\}. \tag{11.6}$$

Here

$$M\{z_i|Z_{i-1}\} = \hat{y}(t_i)_{i-1}\ ,$$

$$M\left\{\left[z_i-\hat{y}(t_i)_{i-1}\right]\cdot\left[z_i-\hat{y}(t_i)_{i-1}\right]^T|Z_{i-1}\right\} = K_y(t_i,t_i)_{i-1}+R_i\ ,$$

$$M\left\{\left[y(t)-\hat{y}(t)_{i-1}\right]\cdot\left[z_i-\hat{y}(t_i)_{i-1}\right]^T|Z_{i-1}\right\} = K_y(t,t_i)_{i-1}\ ,$$

$$\begin{Vmatrix} K & L \\ L^T & M\end{Vmatrix} = \begin{Vmatrix} K_y(t,t)_{i-1} & K_y(t,t_i)_{i-1} \\ K_y^T(t,t_i)_{i-1} & K_y(t_i,t_i)_{i-1}+R_i\end{Vmatrix}^{-1},$$

where *K, L, M* are some square block matrixes of size ($n\times n$), n is the dimension of vector *y*.

One can easily show [5] that the maximum of density (11.6) is achieved for

$$\hat{y}(t)_i = \hat{y}(t)_{i-1} + W(t)\cdot\left[z_i-\hat{y}(t_i)_{i-1}\right],$$

where

$$W(t) = -K^{-1}\cdot L = K_y(t,t_i)_{i-1}\cdot\left[K_y(t_i,t_i)_{i-1}+R_i\right]^{-1}. \tag{11.7}$$

So,

$$\boxed{\hat{y}(t)_i = \hat{y}(t)_{i-1} + K_y(t,t_i)_{i-1}\cdot\left[K_y(t_i,t_i)_{i-1}+R_i\right]^{-1}\cdot\left[z_i-\hat{y}(t_i)_{i-1}\right], \quad t\ge t_i\ .} \tag{11.8}$$

Now, it remains to construct the a posteriori correlation function $K_y(t,\tau)_i$. We express the error in forecasting $\Delta y(t)_i$ of the $y(t)$ process at the *i*-th step in terms of a corresponding error at the (*i*-*1*)-th step. For this purpose we make use of expressions (11.8) and (11.2). We obtain

$$\Delta y(t)_i = y(t) - \hat{y}(t)_i = \left[y(t) - \hat{y}(t)_{i-1}\right] - W(t)\cdot\left[z_i - \hat{y}(t_i)_{i-1}\right] = \Delta y(t)_{i-1} - W(t)\cdot\left[\Delta y(t_i)_{i-1} + v(t_i)\right].$$

Further,

$$\begin{aligned} K_y(t,\tau)_i &= M\left\{\Delta y(t)_i \cdot \Delta y^T(\tau)_i \middle| Z_i\right\} = \\ &= M\left\{\left[\Delta y(t)_{i-1} - W(t)\cdot\left(\Delta y(t_i)_{i-1} + v(t_i)\right)\right]\cdot\left[\Delta y(\tau)_{i-1} - W(\tau)\cdot\left(\Delta y(t_i)_{i-1} + v(t_i)\right)\right]^T \middle| Z_{i-1}\right\}\end{aligned} \qquad (11.9)$$

When multiplying the terms in the right-hand part of (11.8) we take into consideration that all mean values of type

$M\left\{\Delta y(t)_{i-1}\cdot v^T(t_i)\right\} = 0$, and that

$M\left\{\left[\Delta y(t_i)_{i-1} + v(t_i)\right]\cdot\left[\Delta y(t_i)_{i-1} + v(t_i)\right]^T \middle| Z_{i-1}\right\} = K_y(t_i,t_i)_{i-1} + R_i$.

With regard to designation (11.7), the expression (11.9) takes the form

$$\boxed{K_y(t,\tau)_i = K_y(t,\tau)_{i-1} - K_y(t,t_i)_{i-1}\cdot\left[K_y(t_i,t_i)_{i-1} + R_i\right]^{-1}\cdot K_y^T(\tau,t_i)_{i-1}, \quad t,\tau \ge t_i. \qquad (11.10)}$$

Thus, the stated problem is solved. Two recurrence relations (11.8) and (11.10) are developed, which allow one, on the basis of initial data (11.4) and (11.6), to process the last measurement and to prepare the necessary initial data for performing the next step in processing the measurements. In so doing, the maximum possible accuracy of obtained estimates is provided. The constructed recurrence relations represent a basis of the filtering and forecasting algorithm for measurements in a discrete time. Unlike the recurrence relationships of the Kalman-Bucy filter, these relations are *functional*. It is this feature that made it possible to construct the filtering and forecasting algorithm for the Gaussian random process.

<u>Example 1.</u>

Determination of correlation function $K_y(t,\tau)_i$ in the stable filtering mode was carried out for the scalar stationary process with the correlation function

$$K_y(t,\tau)_0 = \begin{cases} 1 - \dfrac{|t-\tau|}{\Lambda}, & |t-\tau| \le \Lambda, \\ 0, & |t-\tau| > \Lambda. \end{cases}$$

Such a choice was caused by the absence of analytical solution for the function of indicated form. In performing calculations it was accepted that:

- the number of grid steps on the correlation interval is (Λ) $p = \Lambda/\Delta = 50$;
- the number of grid steps on the time interval between the measurements (Δt) is $m = \Delta t/\Delta = 1, 5, 10, 15, 25, 35, 45$;

- the root-mean-square error of measurements is ($\sigma = \sqrt{R}$) $\sigma = 0$, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0.

Table 11.1 and figure 11.1 below present the matrix $K_y(t,\tau)_{end}$ for one of initial data versions.

Table 11.1. Correlation matrix $K_y(t,\tau)_{end}$, m=5, $\sigma = 0.1$

τ/Λ	Forecast interval t/Λ										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.009	0.009	0.009	0.008	0.008	0.008	0.008	0.007	0.007	0.0064	0.000
0.1	0.009	0.146	0.138	0.134	0.130	0.126	0.122	0.119	0.115	0.1115	0.100
0.2	0.009	0.138	0.269	0.256	0.248	0.241	0.233	0.227	0.221	0.2142	0.200
0.3	0.008	0.134	0.256	0.383	0.367	0.355	0.344	0.335	0.326	0.3168	0.300
0.4	0.008	0.130	0.248	0.367	0.490	0.470	0.455	0.443	0.431	0.4193	0.400
0.5	0.008	0.126	0.241	0.355	0.470	0.589	0.566	0.551	0.537	0.5218	0.500
0.6	0.008	0.122	0.233	0.344	0.455	0.566	0.682	0.659	0.642	0.6242	0.600
0.7	0.007	0.119	0.227	0.335	0.443	0.551	0.659	0.772	0.747	0.7263	0.700
0.8	0.007	0.115	0.221	0.326	0.431	0.537	0.642	0.747	0.857	0.8288	0.800
0.9	0.006	0.112	0.214	0.317	0.419	0.522	0.624	0.726	0.829	0.9361	0.900
1.0	0.000	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.9000	1.000

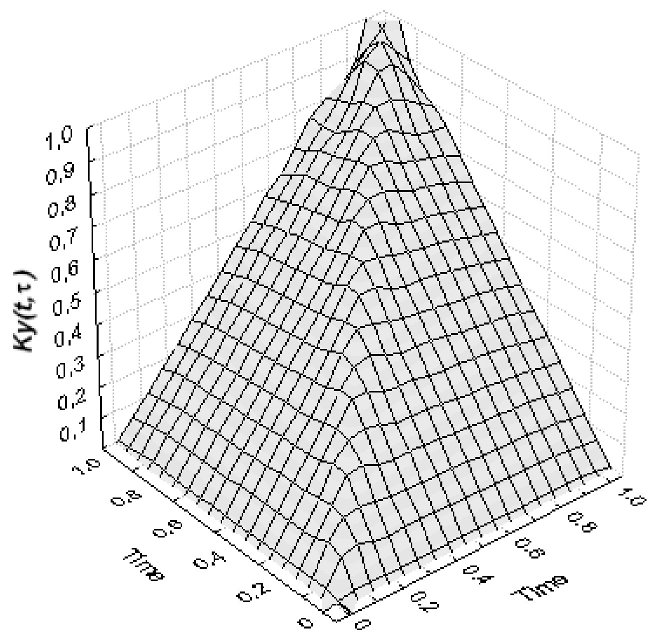

Figure 11.1. Correlation matrix $K_y(t,\tau)_{end}$, m=5, $\sigma = 0.1$

Table 11.1 and figure 11 present the correlation matrix $K_y(t,\tau)_{end}$ for the steady filtering mode. The highest accuracy is achieved at the last measurement instant (the forecast time = 0). The values of matrix components monotonously grow with increasing forecast interval. When the forecast interval is equal to the correlation interval Λ, the variance of errors is equal to 1, i.e. to the value of a priori correlation function $K_y(t,t)_0$. Here the values of correlation moments $K_y(t,\Lambda)_{end}$ и $K_y(\Lambda,\tau)_{end}$ coincide with the values of a priori correlation function (11.11). Table 11.2 presents the values of RMS errors as a function of forecast interval.

Table 11.2. RMS of forecast errors $\sqrt{K_y(t,t)_{end}}$, m=5, $\sigma = 0.1$

Forecast interval t/Λ										
0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.097	0.38	0.52	0.62	0.70	0.77	0.83	0.88	0.92	0.97	1.00

The RMS value for t=0 virtually coincides with the value $\sigma = 0.1$.

Table 11.3 gives the values of variances of estimates for the time instant of the last measurement ($K_y(0,0)_{end}$), i.e. the results of solution of the filtering problem.

Table 11.3. Values of $K_y(0,0)_{end}$ for various m and σ

m	Values of $K_y(0,0)_{end}$						
	0	0.1	0.2	0.3	0.5	0.7	1.0
1	0	0.0080	0.023	0.040	0.076	0.112	0.165
5	0	0.0093	0.032	0.063	0.133	0.206	0.309
15	0	0.0097	0.037	0.076	0.175	0.279	0.423
25	0	0.0098	0.037	0.079	0.187	0.305	0.464
35	0	0.0098	0.038	0.082	0.196	0.321	0.488
45	0	0.0098	0.038	0.082	0.199	0.328	0.498

The data of this table testify to a rather weak influence of the value of interval m between the measurements on the accuracy of measurements filtering results.

Comment. The random process, considered in this example, which has correlation function (11.11), possesses interesting properties. Namely, this process can be constructed in the following manner on the basis of a sequence of independent random numbers $x_j,\ j = 1,\ldots$, distributed according to the normal law $norm(0,1)$:

$$\begin{aligned} y_i &= \left(\sum_{j=1}^{\Lambda} p_j \cdot (x_{i+j})\right) \Big/ \sqrt{\Lambda}, \\ y_{i+k} &= \left(\sum_{j=1}^{\Lambda} p_j \cdot (x_{i+k+j})\right) \Big/ \sqrt{\Lambda}. \end{aligned}, \qquad (11.12)$$

For the value of weighting coefficients p_j=1 the random process with the correlation function (11.11) takes place. Varying the values of weighting coefficients makes it possible to construct random processes with various correlation functions. Here the important question arises, whether there exists the correlation function $K_y(t,\tau)_0$, in which the corresponding correlation coefficients are greater, than those in function (11.11)? It is intuitively clear that in this case the forecasting errors are lower, than those presented above. To answer this question we consider the simplest case with $\Lambda = 2$, $K_y(0,0.5\Lambda)_0 = k_{0.5}$.

This value in function (11.11) equals 0.5. Is it possible to increase it? The application of model (11.12) allows one to construct the equations for determining coefficients p_1 and p_2. We get:

$$p_1^2 + p_2^2 = 1,$$
$$p_1 \cdot p_2 = k_{0.5}.$$

From these relations it follows that the condition

$$p_1^2 + p_2^2 - 2 \cdot p_1 \cdot p_2 = (p_1 - p_2)^2 = 1 - 2 \cdot k_{0.5} \geq 0 \qquad (11.13)$$

should be satisfied.

It is obvious that the random process with the value $K_y(0{,}0.5\Lambda)_0 = k_{0.5} > 0.5$ doesn't exist!!!

Thus, one can draw the conclusion that in forecasting random processes it is impossible to improve the forecasting accuracy as compared to the estimates presented in table 11.2.

Example 2.

We will consider the application of the technique described above for forecasting the intensity of radio emission of the Sun at the wavelength of 10.7 cm (index $F_{10.7}$)). This index is widely used in studying physical processes in the near-earth space. In particular, it represents one of arguments of modern atmospheric density models. Based on the data of site [8], figures 11.2 and 11.3 present the average daily values of index $F_{10.7}$) over the time interval from May, 2002 to December, 2015 (for time periods with the heightened level of solar activity). The figures present also the averaged estimates (on the previous 81-day time interval).

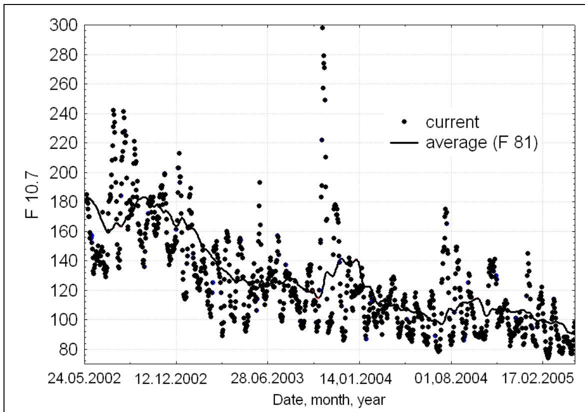

Fig. 11.2. Data for time interval from June, 2002 to May, 2005

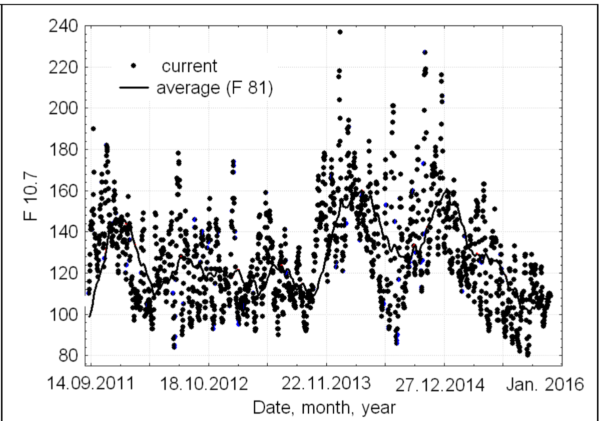

Fig. 11.3. Data for time interval from September, 2011 to December, 2015

The changes in the index values are typical. They reflect the effect of well-known 28-day and 11-year cycles of solar activity. The plots clearly indicate also the irregular (random) deviations, whose prediction is a problematic issue now.

Below we use the assumption that deviations of current estimates of the index from the average values are random quantities. For each time instant the normalized deviations were calculated

$$y(t) = \left[F_{10.7}(t) - F81\right] / F81 . \tag{11.14}$$

Figures 11.4 and 11.5 present statistical distributions of estimates (11.14).

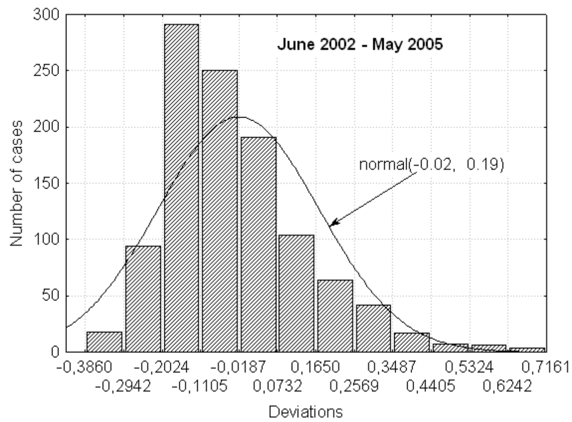

Figure 11.4. Histogram for the time interval from June, 2002 to May, 2005

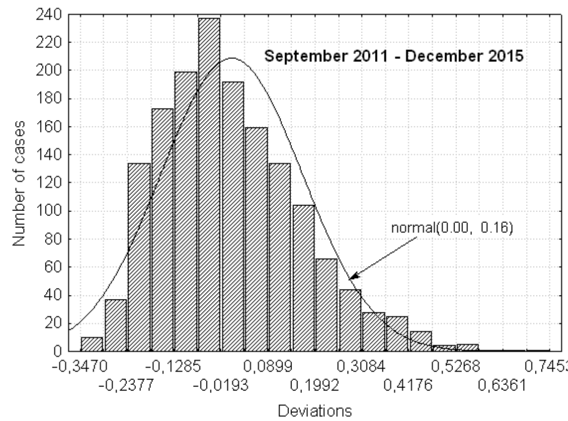

Figure 11.5. Histogram for time interval from September, 2011 to December, 2015

The constructed distributions are similar. In both cases there exists some asymmetry. The amplitude of positive deviations exceeds the amplitude of negative deviations from the average one. This is associated, apparently, with the features of physical processes on the Sun. Nevertheless, the histograms not too highly differ from corresponding normal distributions. Therefore, the application of the considered technique of Gaussian random process forecasting is acceptable.

Figure 11.6 presents the autocorrelation function of random deviations (11.14) constructed according to the data of figures 11.2 and 11.3.

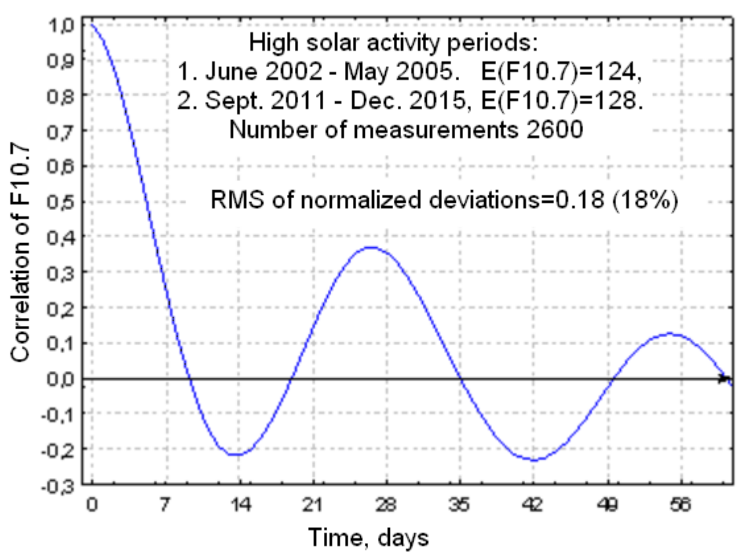

Figure 11.6. Autocorrelation function of normalized deviations of index $F_{10.7}$ from average values

The form of the constructed autocorrelation function is expected. The manifestation of the well-known 28-day period of solar activity variations is clearly seen. The correlation sharply decreases from 1.0 to 0 in 9-10 days. The subsequent correlation maxima don't exceed the value of 0.4. One can also see essential decrease of correlation with time: it becomes less than 0.1 in 2 months.

Figure 11.7 presents the correlation matrix $K_y(t,\tau)_{end}$, calculated by formula (11.10) for the steady filtering mode with the values of parameters m=1, $\sigma = 0.1$. This matrix differs from the similar matrix in figure 11.1 in the existence of a periodic component. This difference is a natural consequence of features of the autocorrelation function presented in figure 11.6.

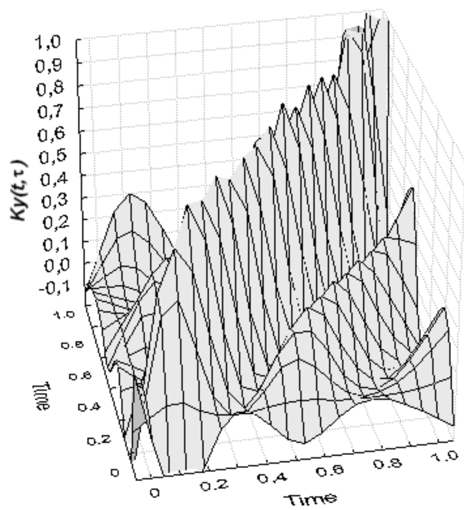

Figure 11.7. Correlation matrix $K_y(t,\tau)_{end}$, m=1, $\sigma = 0.1$, « Time» = t/Λ

When the forecast interval is equal to the correlation interval Λ ("Time" =1), the variance of errors is equal to 1, i.e. to the value of a priori correlation function $K_y(t,t)_0$. Here the values of correlation moments $K_y(t,\Lambda)_{end}$ and $K_y(\Lambda,\tau)_{end}$ coincide with the values of a priori correlation function presented in figure (11.6).

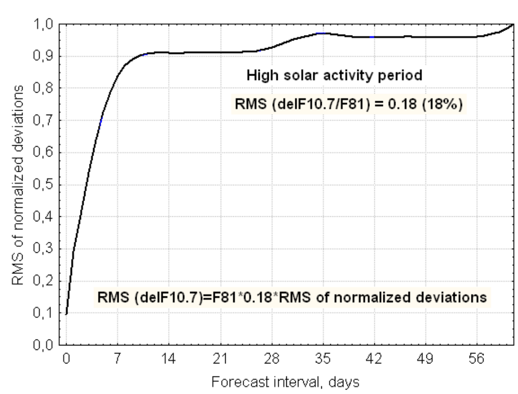

Figure 11.8. RMS of forecast errors $\sqrt{K_y(t,t)_{end}}$, m=1, $\sigma = 0.1$

Figure 11.8 presents the RMS of forecast errors. They were calculated based on the estimates of diagonal terms of matrix $K_y(t,\tau)_{end}$. The data of axis x represent the forecast interval in days. The maximum value of x (61) is equal to the correlation interval Λ .

The data of this figure essentially differ from the similar data presented in table 11.2. On the time interval up to 10 days (10/61=0/16) the RMS errors rapidly increase up to the value of 0.9. For the forecast interval > 0 the RMS values exceed the corresponding data of table 11.2. This fact agrees with the above statement that in forecasting random processes it is impossible to improve the forecasting accuracy as compared to the estimates presented in table 11.2. The accounting for the periodic nature of a priori correlation function (11.1) in this example didn't lead to increasing the accuracy. This accounting was manifested only in a slow increase of RMS forecast errors from 0.9 to 1.0 for the forecast interval larger than 10 days. This result is explained by the fact that on the time interval larger than 10 days the correlation coefficients (figure 11.6) don't exceed the value of 0.4 and decrease down to 0.

The data presented in figure 11.8 can easily be applied to estimate the RMS errors in forecasting the index $F_{10.7}$. For this purpose one should perform multiplication of 3 quantities: 1) the mean value of index (F81), 2) the RMS of relative deviations of current index values from the mean value (σ), 3) the RMS of normalized errors according to the data of figure 11.8. Table 11.4 gives the example of such a calculation.

Table 11.4. RMS errors of forecasting the index $F_{10.7}$

Source	F81	σ	Forecast interval, days				
			1	2	3	4	5
According to the data of figure 11.8	126	0.18	**6.7**	**9.5**	**12.2**	**14.5**	**16.4**

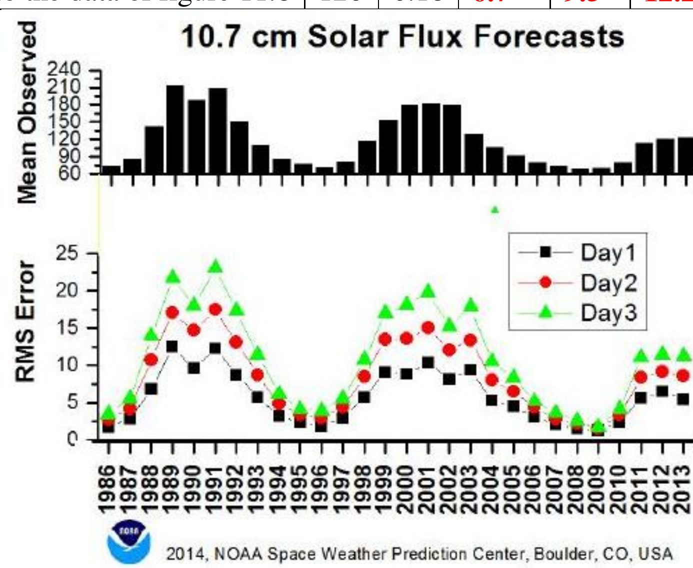

Figure 11.9. NOAA[1] data on forecast errors

[1] National Oceanic and Atmosphere Administration

It was mentioned above that the $F_{10.7}$ estimates were downloaded from the site [8]. This site contains many other materials on the solar activity. In particular, the report "10.7 cm Solar Flux Forecast" [9] presents the detailed statistical data on the index forecasting errors over the time interval up to 5 days. Some of these materials (for the forecast interval up to 3 days) are given in figure 11.9. It is seen that the "RMS Error" estimates for 2013 agree with the estimates of table 11.4.

Table 11.5 gives the more detailed data from the report [9] as well as the results of calculation of errors according to the data of figure 11.8 obtained with using the same estimates of parameters F81 and σ.

Table 11.5. Comparison of RMS of errors of forecasting the $F_{10.7}$ index

Source	F81	σ	Forecast interval, days				
			1	2	3	4	5
According to the NOAA data, 2013	122	0.167	5.4	8.6	11.2	13.7	15.6
According to the data of figure 11.8	122	0.167	5.9	8.6	10.9	13.0	14.8

The table data demonstrate very good compliance of forecast errors estimates obtained by various techniques. This testifies, apparently, to impossibility of increasing the accuracy of solar activity forecasting with the modern level of knowledge of its nature.

Conclusions

1. The technique of optimum forecasting the Gaussian random process, based on the measurements in a discrete time, is substantiated. This technique differs from known approaches in the possibility of specifying a priori autocorrelation function of the process in arbitrary form. The problem solution is reduced to successive application of two functional relations.

2. The best forecasting accuracy is shown to be achieved for the process with a linear autocorrelation function.

3. The application of the developed technique for forecasting the Sun radio emission index $F_{10.7}$ in the periods of high solar activity level is considered. The comparison of obtained results with the corresponding NOAA data demonstrated very good compliance of forecast errors estimates. This testifies, apparently, to impossibility of further increasing the accuracy of solar activity forecasting with the modern level of knowledge of its nature.

References

1. A.N. Kolmogorov. Interpolation and extrapolation of stationary random sequences. Vestnik MGU (the MSU bulletin), ser. mat., 1941, No. 5.

2. A.M. Yaglom. Introduction into the theory of stochastic functions. Uspechi matem. nauk (Progress in math. sciences), 1955, vol. 7, No. 5.

3. Yu.A. Rozanov. The spectral theory of multidimensional stationary processes with discrete time. Uspechi matem. nauk (Progress in math. sciences), 1958, vol. 13, No. 2.

4. R.E. Kalman, R.S. Bucy. New results in a linear filtering and forecasting theory / English Translation. Technical mechanics, ser. D. 1, 1961, 83.

5. A.I. Nazarenko, L. G. Markova. Methods of determination and forecasting the AES orbits in the presence of errors in the mathematical description of motion //Applied problems of space ballistics: collection. M.: Nauka, 1973. pp. 36-67.

6. V.D. Anisimov, et al. Forecasting and determination of AES orbits with regard to the estimate of atmospheric drag fluctuations. In the book Determination of motion of spacecraft. M: Nauka, 1975.

7. A.I. Nazarenko. Recurrent method of optimum forecasting the Gaussian random process. Measurements in a discrete time. Izvestiya AN SSSR (News of the USSR Academy of Sciences). Technical cybernetics, No. 2, 1985.

8. Site NOAA http://www.swpc.noaa.gov.

9. http://www.swpc.noaa.gov/sites/default/files/images/u30/F10.7%20Solar%20Flux.pdf