Optimal Nonlinear Extrapolation of Stationary Continuous Random Processes: Some Examples*

A.M. Yaglom
Institute of Atmospheric Physics
Academy of Sciences of the U.S.S.R., Moscow

A part of the presentation with E.G. Gladyshev and M.I. Fortus

Abstract

This paper presents explicit formulae for the mean square error (MSE) of the best linear predictor and the best nonlinear predictor for thirteen non-Gaussian continuous random processes. First, by introducing three examples of continuous Markov random processes, the author shows that the difference between the MSEs is very small. Next, via introducing 10 examples, the focus is on deriving MSE expressions for the best nonlinear predictor in case the random stochastic process follows a piecewise constant and a piecewise Gaussian random process. In all cases, the improvement of the nonlinear prediction method over the linear prediction method is minimal.

*Translated from Russian by Victor Privalsky. Copyediting of the text was done by Jan G. De Gooijer. The uncorrected proof of the paper is also available in the book by V. Privalsky, titled *Practical Time Series Analysis in Natural Sciences*, Springer-Verlag, 2023, Attachment 2.2, pp. 82–104.

1 Introduction

The major goal of the theory of extrapolation of stationary random processes is to find the solution to the following problem, which is very important for practical purposes. Let $\xi(s)$, $-\infty < s < \infty$, be a stationary random process whose "past" values (upon the semi-axis $\xi(s)$, $-\infty < s \leq t$) are known. The task is to produce the best prediction of the future value $\xi(t+\tau)$, for $\tau > 0$, using these known values. Usually, the "best prediction" is understood at the value of the functional $\widetilde{\xi}_{t,\tau}$, $\{\xi(s), -\infty < s \leq t\} = \xi(t,\tau)$ of all past values of the process for which the mean square error (MSE) of prediction is given by

$$\sigma^2(\tau) = \mathbb{E}\{\xi(t+\tau) - \widetilde{\xi}(t,\tau)\}^2, \tag{1}$$

is the smallest and, additionally, the functionals $\widetilde{\xi}(t,\tau)$ are supposed to be linear. Finding the "optimal" linear functional $\widetilde{\xi}(t,\tau)$ and the corresponding MSE value $\sigma^2(\tau)$ is the content of the profound theory of linear extrapolation of stationary random processes created as long as about 20 years ago by [1] and developed further by [2] and [3]. Currently, the theory has achieved a significant degree of completeness.

Yet, it is clear that the practical task of finding "the best prediction" cannot be solved exhaustively within the framework of the linear extrapolation theory. Actually, the usual limitation of using only linear functionals $\widetilde{\xi}(t,\tau)$ is just a consequence of the fact that a solution becomes much simpler and it is by no means related to the substance of the task. Moreover, the condition of the minimal MSE is just a specific example belonging to a very large set of other permissible (and equally natural from the common-sense point of view) conditions of optimal forecasting. The selection of that specific condition is dictated, first of all, by it being the most convenient for applying analytical methods of solution and for obtaining viewable ultimate results.

In this paper, we will follow the usual way of characterizing the "quality of prediction" $\widetilde{\xi}(t,\tau)$ by the MSE $\sigma^2(\tau)$ but we will remove the requirement of linearity of the functional $\widetilde{\xi}(t,\tau)$. It is well known that for Gaussian processes, this removal cannot result in any improvement in the solution of the prediction task provided by the theory of linear extrapolation – the lowest value of $\sigma^2(s)$ within the class of arbitrary functionals $\widetilde{\xi}(t,\tau) = \widetilde{\xi}_{t,\tau}$, $\{\xi(s), -\infty < s \leq t\}$ is always achieved here upon some linear functional $\widetilde{\xi}_{t,\tau}$.[1]

If, however, not all scalar and multidimensional probability distributions are Gaussian, there are, generally, some reasons to hope that the MSE of linear extrapolation can be made smaller due to the transfer to nonlinear functionals $\widetilde{\xi}(t,\tau)$. In this respect, suggestions have been made multiple times in the engineering literature to look for the prediction $\widetilde{\xi}(t,\tau)$ within the class of some specific nonlinear functionals defined with a finite number of "arbitrary functions" (playing the role of parameters in this task), which are more general than the class of linear functionals (which can formally be presented as $\int_0^\infty \xi(t-\tau)\omega(\tau)\mathrm{d}\tau$ that contains just one "arbitrary function" $\omega(\tau)$). In this case, the minimum MSE condition (1) allows one to obtain a system of equations (usually, integral or integral-differential) with respect to the unknown "arbitrary functions" that define the optimal functional $\widetilde{\xi}(t,\tau)$, and which can possibly be solved, for example, numerically (cf. [6]– [9]).[2]

A different approach to the task of optimal nonlinear prediction is based upon the circum-

[1] Note that in the case of Gaussian processes using that same condition of the MSE also turns out to be immaterial; in this case, all reasonable conditions of optimality lead to one and the same solution of the prediction task (see, e.g., [4], p. 77, or [5]).

[2] In most of those works, however, the subject was not the task of predicting a future value of a random process but rather a more general task of filtering of random processes, which includes the prediction task as a special case.

stance that the conditional mathematical expectation

$$\widetilde{\xi}_0(t,\tau) = \mathbb{E}\{\xi(t+\tau)|\xi(s), s \le t\} \tag{2}$$

of the future values of the process under the condition that all its past values are known, assuming this mathematical expectation exists, will be exactly the same functional of $\xi(s)$, $-\infty < s \le$, for which the MSE (1) turns out to be minimal (see, e.g., [4, p. 77] or [9, Section 140]). Therefore, the conditional mathematical expectation $\xi_0(t,\tau)$ presents the best possible nonlinear prediction of the value $\xi(t+\tau)$ while the MSE

$$\sigma_0^2(\tau) = \mathbb{E}\{\xi(t+\tau) - \widetilde{\xi}_0(t,\tau)\}^2 \tag{3}$$

defines the lower limit of the MSE for all possible prediction methods.

For the stationary random processes $\xi(n)$ with a discrete argument (stationary random sequences), which have moments of all orders and satisfy several more general conditions, it is possible to indicate a general method that, in principle, allows one to calculate, using the values $\xi(m)$ at $m \le n$, on the basis of the functions of statistical moments of different orders, the conditional mathematical expectation $\mathbb{E}\{\xi(n+h)|\xi(m), m \le n\}$ with any given degree of accuracy (see 10]). However, this method requires extremely cumbersome computations even when the required degree of accuracy is quite modest and it is barely effective in practice even when using modern high-speed computers. The methods of nonlinear extrapolation proposed in [6]– [9] also require burdensome calculations, which strongly impede their practical applications. As for specific examples of nonlinear extrapolation, they are practically impossible to find in the literature.[3] Therefore, the answer to the question of what gain can be achieved due to moving from the linear extrapolation of some process that exists in practice to the nonlinear extrapolation remains undiscovered.

It should be stressed that, strictly speaking, this question does not belong to the mathematical probability theory. It is even difficult to formulate the question of possible gains due to the use of the nonlinear methods of extrapolation as compared to the linear methods is given by the following well-known example of a stationary process

$$\xi(\tau) = a\cos(\lambda\tau + \varphi). \tag{4}$$

Here a is a constant, φ is a random variable uniformly distributed over the interval $(0, 2\pi)$, and λ is a random variable independent of φ, and having an arbitrary probability distribution function $F(\lambda)$ (see, for example, [11], Chapter X, example 4 and Chapter XI, Section 3, example 4). Indeed, all sample records of such process will have a strictly sinusoidal shape so that the nonlinear prediction $\widetilde{\xi}_0(t,\tau)$ at any τ will be absolutely accurate, i.e., $\sigma_0^2(\tau) \equiv 0$. On the other hand, the spectral function of the process will be $(1/4)a^2[1 + F(\lambda) - F(-\lambda)]$ meaning that it can be chosen such that for any $\tau > 0$ there is a process described by (4) for which the MSE $\sigma^2(\tau)$ of linear extrapolation can be arbitrarily close to $\mathbb{E}\{\xi(t+\tau)\}^2 = \sigma^2(\tau)$, i.e. the variance of the process (4).

However, (4) is not a real stationary process from an applied point of view. The only random quantities here are the phase and frequency of a harmonic oscillation and only those are required to be estimated instead of applying to $\xi(t)$ the theory of random processes. Note also that the practical application of this theory to $\xi(t)$ turns out to be extremely difficult because the process is not ergodic so that its statistical properties cannot be determined from a one sample record or from a small number of samples. Clearly, Eq. (4) can be easily used to construct an ergodic

[3] The situation seems to be a little bit better in the area of filtering the stationary processes; yet, even in that area the number of published examples of applying the nonlinear methods is still very insignificant.

process with very close properties: it is sufficient to assume, for example, that $\xi(t)$ consists of long slices of sinusoids (4) that end at random "discontinuity points" (distributed, for example, according to the Poisson law with a very small parameter) where the transfer to the slice of a new sinusoid occurs with the values of τ and λ chosen in accordance with their probability distributions given by the process (4). However, from an applier's point of view, this process looks exceedingly exotic; besides, in order for it to be "very close" to the process described by (4), the discontinuity points should be distributed "very rarely" so that the ergodicity of the process will become purely conditional because it will be almost impossible to utilize.

On the other hand, it is not difficult to build a series of not complicated examples of stationary random processes that are ergodic and not too exotic from the point of view of practical application, and for which it can be possible to write the best nonlinear prediction and compare the respective root MSE (RMSE) $\sigma_0(\tau)$ with the RMSE $\sigma(\tau)$ of the best linear prediction. The depiction of several examples of this type constitutes the main content of the present paper. Clearly, no number of such examples will allow us to precisely determine the gain that we obtain by switching to nonlinear extrapolation in any specific and novel case. But some general idea about the degree of magnitude of the expected gain can still be obtained in this manner. In particular, the examples given below show that quite frequently the difference between $\sigma_0^2(\tau)$ and $\sigma^2(\tau)$ turns out to be quite small though the extrapolation formulae defining the linear and nonlinear predictions are very dissimilar. This feature agrees very well with the fact that even in the case of linear extrapolation, a significant change in the extrapolation formulae often leads to very small changes in the prediction error. This shows that in empirical applications, the expediency of converting to a new and more complicated method of prediction requires a thorough and dedicated study in all cases.

2 Continuous Markov Random Processes

The Markov process $\xi(s)$ is characterized by the property that its conditional probability distribution of the variable $\xi(t+\tau)$ with the known past values $\xi(s)$, $-\infty < s \leq t$ depends only upon $\xi(t)$ – the last of the known values. Clearly, this setup makes the task of the optimal linear extrapolation much simpler for this process. Here

$$\widetilde{\xi}_0(t,\tau) = \mathbb{E}\{\xi(t+\tau)|\xi(t)\}, \tag{5}$$

so that the functional $\widetilde{\xi}_0(t,\tau) = \widetilde{\xi}_0^{t,\tau}\ \{\xi(s), -\infty < s \leq t\}$ becomes a function $\widetilde{\xi}_0^{t,\tau}(\xi(t))$ of a single variable and all that is needed in order to find is the bivariate probability distribution of the process.

2.1 Unique functions of the Ornstein–Uhlenbeck process

The simplest stationary Markov process is the so-called Ornstein–Uhlenbeck process – a real-valued Gaussian process $\eta(s)$ with $\mathbb{E}[\eta(s)] = 0$ and correlation function $R_{\eta\eta}(\tau) = \mathbb{E}\{\eta(s)\eta(s+\tau)\} = e^{-|\tau|}$, i.e. with the spectral density $f_{\eta\eta}(\lambda) = 1/\big(\pi(\lambda^2+1)\big)$. As the process is Gaussian, its best linear prediction is also the best prediction in general:

$$\widetilde{\eta}(t,\tau) = \widetilde{\eta}_0(t,\tau) = e^{-|\tau|}\eta(t). \tag{6}$$

Simple examples of non-Gaussian stationary Markov processes can be obtained by assuming $\xi(s) = \varphi\big(\eta(s)\big)$, where $\eta(s)$ is an Ornstein–Uhlenbeck process and $y = \varphi(x)$ is a known nonlinear function that has a unique inverse function $x = \varphi(y)$. In the specific case when $\varphi(x) = ax^{2n+1}$, the correlation function of $\xi(s)$ will be easily calculated following the rule of calculating higher moments of the Gaussian distribution, and will have a rational Fourier transform. Consequently, for such process, the best linear extrapolation function will be easy to obtain as an explicit linear extrapolation formula. Moreover, the conditional distribution for $\xi(t+\tau)$ with the known $\xi(s)$ will be given here with a plain explicit formula that also allows one to easily obtain an explicit expression for the function $\widetilde{\xi}_0^{t,\tau}(\xi(t))$.

Example 1. Let $\xi(t) = a\eta^3(t)$ and $a = 1/\sqrt{15}$, where $\eta(s)$ is the Ornstein–Uhlenbeck process, and where the multiplier $1/\sqrt{15}$ is intentionally selected to ensure that $R_{\xi\xi}(0) = 1$. In this case

$$R_{\xi\xi}(\tau) = \frac{1}{15}\mathbb{E}\{\eta^3(t)\eta^3(t+\tau)\} = \frac{1}{15}[9R^2_{\eta\eta}(0)R_{\eta\eta}(\tau) + 6R^3_{\eta\eta}(\tau)]$$
$$= \frac{3}{5}e^{-\tau} + \frac{2}{5}e^{-3\tau}. \tag{7}$$

The corresponding spectral density is given by

$$f(\lambda) = \frac{3}{5\pi}\Big(\frac{1}{\lambda^2+1} + \frac{2}{\lambda^2+9}\Big) = \frac{9}{5\pi}\Big(\frac{|\lambda + i\sqrt{11/3}|^2}{|(\lambda+i)(\lambda+3i)|^2}\Big). \tag{8}$$

From this it follows that the best linear prediction of the process $\xi(s)$ is given by

$$\widetilde{\xi}(t+\tau) = D_0\xi(t) + D_1\int_0^\infty \exp\Big(-\sqrt{\frac{11}{3}}s\Big)\xi(t-s)\mathrm{d}s, \tag{9}$$

where

$$D_0 = \frac{1}{2}\Big[\Big(\sqrt{\frac{11}{3}}-1\Big)e^{-\tau} + 3\Big(3-\sqrt{\frac{11}{3}}\Big)e^{-3\tau}\Big], \quad D_1 = \frac{2}{3}\Big(3\sqrt{\frac{11}{3}}-5\Big)(e^{-\tau}-e^{-3\tau}). \tag{10}$$

The MSE of this prediction is given by

$$\sigma^2(\tau) = (1-e^{-2t})\Big[1 + \frac{3\sqrt{33}-11}{10}e^{-2\tau} + \frac{19-3\sqrt{33}}{10}e^{-4\tau}\Big]$$
$$\approx (1-e^{-2\tau})[1 + 0.62e^{-2\tau} + 0.18e^{-4\tau}]. \tag{11}$$

Clearly, the process $\xi(s)$ is non-Gaussian. The univariate probability distribution function (pdf) of $\xi(s)$ will obviously be

$$p(x) = \frac{\sqrt[6]{15}}{3\sqrt{2\pi}x^{2/3}}\exp\Big(-\frac{\sqrt[3]{15}x^{2/3}}{2}\Big). \tag{12}$$

This expression becomes infinite at $x = 0$. The conditional pdf of $\xi(t+\tau)$ for a known $\xi(t)$ is given by

$$p_\tau\big(x|\xi(t)\big) = \frac{\sqrt[6]{15}}{3\sqrt{2\pi(1-e^{-2\tau})}x^{2/3}}\exp\Big(-\frac{\sqrt[3]{15}[x^{1/3}-e^{-\tau}\xi^{1/3}(t)]^2}{2(1-e^{-2\tau})}\Big). \tag{13}$$

The functional $\widetilde{\xi}_0(t,\tau)$, which defines the best nonlinear forecast, can be found through calculating the mean value of the distribution given by (13), or by calculating the mean value of the variable $\eta^3/15$ using the known conditional distribution of the variable η. It turns out to be

$$\widetilde{\xi}_0(t,\tau) = \frac{3}{\sqrt[3]{15}}e^{-\tau}(1-e^{-2\tau})\xi^{1/3}(t) + e^{-3\tau}\xi(t). \tag{14}$$

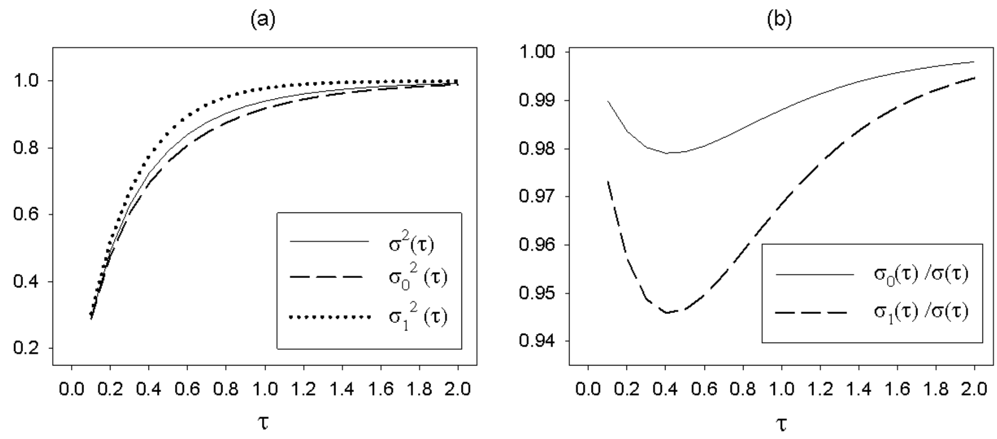

Figure 1: (a) Plots of the MSEs $\sigma^2(\tau)$, $\sigma_0^2(\tau)$, and $\sigma_1^2(\tau)$ for different values of τ; (b) plots of the ratios $\sigma_0(\tau)/\sigma(\tau)$ and $\sigma_1(\tau)/\sigma(\tau)$ as a function of τ (Example 1).

Substituting (14) into (3) will show that the MSE of extrapolation according to this formula equals

$$\sigma_0^2(\tau) = (1 - e^{-2\tau}[1 + 0.4e^{-2\tau} + 0.4e^{-4\tau}]. \tag{15}$$

Also note that we have one more way to build the optimal extrapolation of the process $\xi(s)$, which may well be the most natural: as the best linear prediction of the variable $\eta(t+\tau)$ given the known values $\eta(s), s \le t$, is inarguably, $\widetilde{\eta}(t+\tau) = e^{-\tau}\eta(t)$ [see Eq. (6)]. The best prediction of the variable $\xi(t+\tau) = \widetilde{\eta}^3(t+\tau)/15$ can also be written as

$$\xi_1(t+\tau) = \frac{1}{15}\widetilde{\eta}^3(t,\tau) = e^{-3\tau}\xi(t). \tag{16}$$

However, we would like to stress that the expression "the best" as applied to formula (16) and to formula (14) has a different meaning. The prediction $\widetilde{\eta}(t+\tau)$ will be the best possible solution in the sense that $\mathbb{E}\{\eta(t+\tau) - \widetilde{\eta}(t,\tau)\}^2 =$ minimal. It clearly shows that (16) is the version of the extrapolation formulae for which the quantity that takes the minimal value is given by

$$\widehat{\sigma}^{2/3}(\tau) = \mathbb{E}\{\xi^{1/3}(t+\tau) - \widehat{\xi}^{1/3}(t,\tau)\}^2. \tag{17}$$

The condition of minimizing the value of (17) is also a reasonable way for optimal extrapolation but it has a drawback: it is very inconvenient for calculating in all cases except for the one that we are discussing here. In particular, the MSE of (16) is equal to

$$\sigma_1^2(\tau) = \mathbb{E}\{\xi(t+\tau) - e^{-3\tau}\xi(\tau)\}^2 = (1 - e^{-2\tau})(1 + e^{-2\tau} - 0.2e^{-4\tau}), \tag{18}$$

which is, of course, higher than the MSE (15) of the best linear extrapolation $\widetilde{\xi}_0(t,\tau)$ as well as greater than the MSE (11) of the best (in the same sense) linear extrapolation $\widetilde{\xi}(t,\tau)$, because the extrapolation formula (16) is also linear.

However, it is interesting to compare quantitatively the error values for extrapolations in accordance with formulae (9), (14), and (16). Figure 1(a) shows the values of $\sigma^2(\tau)$, $\sigma_0^2(\tau)$, and $\sigma_1^2(\tau)$ for different values of τ. Figure 1(b) shows the ratios $\sigma_0(\tau)/\sigma(\tau)$ and $\sigma_1(\tau)/\sigma(\tau)$. It is seen that in this case the best linear prediction $\widetilde{\xi}(t,\tau)$ has an MSE which exceeds the MSE of the best nonlinear prediction $\widetilde{\xi}_0(t,\tau)$ by not more than 2%. Moreover, even the linear prediction

$\widetilde{\xi}_1(t,\tau)$, which is optimal in the sense of a rather strange criterion (17), has an MSE that exceeds the MSE of the best linear prediction by just 5%. As for comparison of only the linear prediction methods, switching from the minimum of the MSE to the condition of minimizing (14) results in an increase of the MSE by not more than 3%.

Example 2. Consider the process $\xi(s) = \eta^5(s)/3\sqrt{105}$, where the coefficient $1/3\sqrt{105}$ is selected such the variance of $\xi(s)$ equals 1. One may think that this process is "more non-Gaussian" than the process discussed in Example 1 so that the benefit due to the transfer from the linear method to a nonlinear method will be greater than in the previous case. Actually, this assumption turns out to be incorrect.

The correlation function of the process $\xi(s)$ is given by

$$R_{\xi\xi}(\tau) = \frac{1}{945}\mathbb{E}\{\eta^5(s)\eta^5(s+\tau)\} = \frac{5}{21}e^{-\tau} + \frac{40}{63}e^{-3\tau} + \frac{8}{63}e^{-5\tau}, \tag{19}$$

so that

$$f_{\xi\xi}(\lambda) = \frac{5}{63\pi}\Big[\frac{3}{\lambda^2+1} + \frac{24}{\lambda^2+9} + \frac{8}{\lambda^3+25}\Big] = \frac{5}{63\pi}\,\frac{35\lambda^4+806\lambda^2+1347}{(\lambda^2+1)(\lambda^2+9)(\lambda^2+25)}. \tag{20}$$

Hence, it can be deduced that the best linear prediction of the process $\xi(s)$ is given by

$$\widetilde{\xi}(t+\tau) = D_0\xi(t) + \int_0^\infty \big[D_1e^{-\alpha_1 s} + D_2e^{-\alpha_2 s}\big]\xi(t-s)\mathrm{d}s, \tag{21}$$

where $\alpha_1 = \sqrt{-x_1} \approx 1.34$ and $\alpha_2 = \sqrt{-x_2} \approx 4.61$, with x_1 and x_2 the roots of the quadratic equation $35x^2 + 806x + 1347 = 0$, while D_0, D_1, and D_2 denote some linear combinations of exponential functions $e^{-2\tau}$, $e^{-3\tau}$, and $e^{-5\tau}$ with numerical coefficients. The MSE of the prediction is given by

$$\sigma^2(\tau) = 1 - e^{-2\tau}\big[\beta_0 + \beta_1e^{-2\tau} + \beta_2e^{-4\tau} + \beta_3e^{-6\tau} + \beta_4e^{-8\tau}\big], \tag{22}$$

where $\beta_0, \dots, \beta_4$ are constants, which will not be given here.

In order to determine the best nonlinear prediction $\widetilde{\xi}_0(t,\tau)$ of the process $\xi(s)$, one should calculate the mean value of the variable $\eta^5/945$, where η is distributed in accordance with the Gauss law with mean value $\sqrt[10]{945}e^{-\tau}\xi^{1/5}(t)$ and variance $1-e^{-2\tau}$. It easily leads to the formula

$$\widetilde{\xi}_0(t,\tau) = \frac{15}{945^{2/5}}e^{-\tau}(1-e^{-2\tau})^2\xi^{1/2}(t) + \frac{10}{945^{1/5}}e^{-3\tau}(1-e^{-2\tau})\xi^{3/5}(t) + e^{-5\tau}\xi(t). \tag{23}$$

According to this formula, the MSE of extrapolation is given by

$$\sigma_0^2(\tau) = (1-e^{-2\tau})\Big[1 + \frac{16}{21}e^{-2\tau} + \frac{16}{21}e^{-4\tau} + \frac{8}{63}e^{-6\tau} + \frac{8}{63}e^{-8\tau}\Big]. \tag{24}$$

Finally, we can study a third method of extrapolating the process $\xi(s)$ using the formula

$$\widehat{\xi}_1(t,\tau) = e^{-5\tau}\xi(t). \tag{25}$$

This has the property that the smallest possible value is given by

$$\widetilde{\sigma}^{2/5}(\tau) = \mathbb{E}\{\xi^{1/5}(t+\tau) - \widetilde{\xi}_1^{1/5}(t,\tau)\}^2. \tag{26}$$

In this case the MSE of extrapolation is given by

$$\sigma_1^2(\tau) = \mathbb{E}\{\xi(t+\tau) - \widetilde{\xi}(t,\tau)\}^2 = (1-e^{-2\tau})\Big[1 + e^{-2\tau} + e^{-4\tau} + \frac{11}{21}e^{-6\tau} - \frac{47}{63}e^{-8\tau}\Big]. \tag{27}$$

This will naturally be greater than (24) and even greater than (25). The differences between the three functions $\sigma^2(\tau)$, $\sigma_0^2(\tau)$, and $\sigma_1^2(\tau)$ turn out to be still as small, as in the case of Example 1. In particular, $\sigma(\tau)$ exceeds $\sigma_0(\tau)$ by not more than 2% while the RMSE $\sigma_1(\tau)$ according to (2.23) exceeds $\sigma_0(\tau)$ by not more than 5% and $\sigma(\tau)$ by not more than 3–4%.

2.2 Markov processes of diffusion type

Consider a stationary diffusion Markov process $\xi(s)$ having a stationary distribution $w(x)$ and the transition probabilities $p(\tau, x, y)$ that satisfy the Fokker-Planck equation of the form

$$\frac{\partial p}{\partial \tau} = L_y p,\ L_y p(y) = -\frac{\partial}{\partial y}[A(y)p(y)] + \frac{\partial^2}{\partial y^2}[B(y)p(y)]. \tag{28}$$

If the elliptic differential operator L_y has an infinite sequence of eigenvalues $\lambda_0 = 0, -\lambda_1, -\lambda_2, \ldots$, $(0 < -\lambda_1 \leq -\lambda_2 \leq \cdots)$, which has a full orthonormalized system of eigenfunctions $\ell_0(y) = aw(y), \ell_1(y), \ell_2(y), \ldots$, then $p(\tau, x, y)$ for such process will be given by

$$p(\tau, x, y) = \sum_{k=0}^{\infty} e^{-\lambda_k \tau} \ell_k(x) \overline{\ell_k(y)}. \tag{29}$$

Therefore, the optimal linear forecast $\widetilde{\xi}_0(t, \tau)$ (which coincides with the mean value of probability distribution having a density $p(\tau, \widetilde{\xi}(t), y)$) is given by

$$\widetilde{\xi}_0(t, \tau) = \sum_{k=0}^{\infty} a_k e^{-\lambda_k \tau} \ell_k[\xi(t)], \tag{30}$$

where a_k are constants and the correlation function and the spectral density $f_{\xi\xi}(\lambda)$ are, respectively, equal to

$$R_{\xi\xi}(\tau) = \sum_{k=0}^{\infty} b_k e^{-\lambda_k |\tau|},\ f_{\xi\xi}(\lambda) = \sum_{k=0}^{\infty} \frac{b_k \lambda_k}{\pi[\lambda^2 + \lambda_k^2]}, \tag{31}$$

where b_k are (nonnegative) constants not coinciding with a_k.

The spectral density $f_{\xi\xi}(\lambda)$ is generally a meromorphic function of λ having an infinite number of poles, which means that finding a new formula for the best linear prediction $\widetilde{\xi}(t, \tau)$ will hardly be possible. But we can always cut off the series after a small finite number of its terms and investigate a linear extrapolation formula which would be optimal for this "broken" correlation function. However, we will see that even with the linear extrapolation using this formula (which is not the "best" of the "best" for our process $\xi(s)$ but seemingly leads to a RMSE just slightly higher than $\sigma(\tau)$), the RMSE of extrapolation differs by just a small value from the RMSE of the best nonlinear prediction (30). We show that as compared with the best linear forecast, the best nonlinear forecast cannot give us a noticeable gain. These general considerations will now be illustrated by a simple example.

Example 3. Consider a stationary diffusion process within a finite interval $[-\pi/2, \pi/2]$ described with the common diffusion equation

$$\frac{\partial p}{\partial \tau} = a^2 \frac{\partial^2 p}{\partial y^2}, \tag{32}$$

with the boundary conditions $\partial p/\partial y|_{y=\pi/2} = \partial p/\partial y|_{y=\pi/2} = 0$. Obviously, the density of the stationary distribution here is a constant: $w(x) = 1/\pi$, $-\pi/2 \leq x \leq \pi/2$. The orthonormalized system of eigenfunctions of the operator L_y and the respective system of eigenvalues is given by

$$\ell_0(x) = \frac{1}{\sqrt{\pi}}, \quad \ell_n(x) = \begin{cases} \frac{2}{\sqrt{\pi}} \sin(nx), & n = 1, 3, 5, \ldots, \\ \frac{2}{\sqrt{\pi}} \cos(nx), & n = 2, 4, 6, \ldots, \end{cases} \tag{33}$$

and $\lambda_n = a^2n^2$, $n = 0, 1, 2, \ldots$ Using (33), we have

$$p(\tau, x, y) = \frac{1}{\pi} + \frac{2}{\pi}\sum_{k=0}^{\infty} \exp[-(2k+1)^2a^2\tau]\sin(2k+1)x\sin(2k+1)y$$
$$+ \frac{2}{\pi}\sum_{k=0}^{\infty} \exp[-(2k)^2a^2\tau]\cos(2kx)\cos(2ky), \tag{34}$$

$$\widetilde{\xi}_0(t,\tau) = \widetilde{\xi}_0^{t,\tau}(\xi(t)) = \frac{4}{\pi}\sum_{k=0}^{\infty} \frac{(-1)^k\exp[-(2k+1)^2a^2\tau]}{(2k+1)^2}\sin[(2k+1)\xi(t)], \tag{35}$$

$$R_{\xi\xi}(\tau) = \frac{8}{\pi^2}\sum_{k=0}^{\infty} \frac{\exp[-(2k+1)^2a^2|\tau|]}{(2k+1)^4}, \tag{36}$$

$$f_{\xi\xi}(\lambda) = \frac{8a^2}{\pi^3}\sum_{k=0}^{\infty} \frac{1}{(2k+1)^2[\lambda^2 + (2k+1)^4a^4]}. \tag{37}$$

It is not difficult to show that the MSE of extrapolation of the process $\xi(s)$ in accordance with formula (36) is equal to

$$\sigma_0^2(\tau) = \int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2}[y - \xi_0^{t,\tau}(x)]^2 w(x)p(\tau, x, y)\mathrm{dxdy}$$
$$= \frac{\pi}{12} - \frac{8}{\pi^2}\sum_{k=0}^{\infty} \frac{e^{-2(2k+1)}a^2\tau}{(2k+1)^2}. \tag{38}$$

As for the best linear prediction that corresponds to the correlation function (36), it cannot be presented explicitly. Yet, as the absolute values of the consecutive terms of the series on the right-hand side of (36) diminish very fast, one can hope that the linear extrapolation formula, which will be optimal in the case when the correlation function is given only by the first term of (36), will result in a MSE that is close to the MSE of the best linear extrapolation. Therefore, it will be rather interesting to compare the MSE $\sigma_0^2(\tau)$ with the MSE $\sigma_1^2(\tau)$ of linear extrapolation of $\xi(s)$ according to the following equation

$$\widetilde{\xi}_1(t,\tau) = e^{-a^2\tau}\xi(t), \tag{39}$$

which is equal to

$$\sigma_1^2(\tau) = \mathbb{E}\{\xi(t+\tau) - e^{-a^2\tau}\xi(t)\}^2 = (1 + e^{-2a^2\tau})R_{\xi\xi}(0) - 2e^{-a^2\tau}R_{\xi\xi}(\tau)$$
$$= \frac{\pi}{12}(1 + e^{-a^2\tau}) - \frac{16}{\pi^2}e^{-2a^2\tau}\sum_{k=0}^{\infty} \frac{\exp[-(2k+1)^2a^2\tau]}{(2k+1)^4}. \tag{40}$$

By computing the values on the right-hand side of (38) and (40) for different values of a and τ, it can be easily seen that even at the most "unproductive" value of $a\tau$ (it turns out that $a\tau$ is close to 0.05), $\sigma_1^2(\tau)$ exceeds $\sigma_0^2(\tau)$ by just about 1.5% of the latter value; at most other values of $a\tau$, the deviation of $\sigma_1^2(\tau)$ from $\sigma_0^2(\tau)$ is much smaller than even that small number. It becomes clear now that the MSE $\sigma^2(\tau)$ of the best linear forecast which takes a value somewhere between $\sigma_0^2(\tau)$ and $\sigma^2(\tau)$, will be barely different in this case from $\sigma_0^2(\tau)$.

3 Disconnected random processes

In issues related to applications, there are some processes which, along with continuous variations, may contain suddenly appearing and very fast "jumpy" changes of respective quantities.

The mathematical models of such processes is the disconnected random processes that have discontinuities of the first kind ("jumps") at some random sequence of points and with some dynamic or stochastic law that controls their continuous variability between the points of discontinuity. In this section, we discuss a number of examples of such disconnected processes. These examples allow us to find explicit solutions of the tasks of linear and nonlinear optimal extrapolation.

3.1 Piecewise constant random processes

Consider first some processes belonging to the class of point processes with adjoint random variables (cf. [13]). That is, the processes that include jump points at some random sequence of points $\ldots, t_{-1}, t_0, t_1, \ldots$, that defines a "point random process" along the axis $-\infty < t < \infty$. Within the intervals between the jump points, the values of the process are constant and coincide with the values of some sequence of mutually independent and identically distributed random variables. Moreover, we may assume, without any loss of generality, that the mean value of those variables is zero and the variance equals one, i.e., $\mathbb{E}[\xi(s)] = 0$, $\mathbb{E}[\xi^2(s)] = 1$.

Example 4. Consider a Poisson process with adjoint random variables for which the probability of having exactly m discontinuity points within any interval of length N on the axis s equals $[(\beta T)^m/m!]e^{-\beta T}$, where β is the positive *mean density* of the Poisson point process $\{t_n\}$. In this case the correlation function is given by

$$R_{\xi\xi}(\tau) = \mathbb{E}\{\xi(s)\xi(s+\tau)\} = e^{-\beta\tau} \tag{41}$$

(see, e.g., [13], p. 224), so that

$$\widetilde{\xi}(t,\tau) = e^{-\beta\tau}\xi(t), \quad \sigma^2(\tau) = 1 - e^{-2\beta\tau}. \tag{42}$$

Hence,

$$\begin{aligned} R_{\xi\xi}(\tau) = \mathbb{E}\{\xi(s)\xi(s+\tau)\} &= \mathbb{P}\{\text{in the interval}\,(s, s+\tau)\,\text{without points}\ t_i\} \times \xi(t) \\ &+ \mathbb{P}\{\text{in the interval}\,(s, s+\tau)\,\text{with points}\ t_i\} \times 0 = e^{-\beta\tau}\xi(t), \end{aligned} \tag{43}$$

that is,

$$\widetilde{\xi}_0(t,\tau) = \widetilde{\xi}(t,\tau), \quad \sigma_0^2(\tau) = \sigma^2(\tau). \tag{44}$$

Thus, in spite of the fact that the process under study is non-Gaussian, its best prediction turns out to be linear. Note that for this same process with an unknown mean $\mathbb{E}[\xi(s)]$, the variance of the optimal nonlinear estimate of the mean value obtained from the values of $\xi(s)$ on the interval $0 \le s \le t$. That is, the MSE of the nonlinear filter with a T-long set for separating the constant summand from the stationary "noise" $\xi'(s) = \xi(s) - \mathbb{E}[\xi(s)]$, according to [13], will be, for $T \gg 1/\beta$, almost twice smaller than the error variance of the respective best linear estimate.

Example 5. Assume that the random sequence of points $\ldots, t_{-1}, t_0, t_1, \ldots$ is obtained from some Poisson random sequence $\{t_i'\}$ (with mean density β) by selecting every second point:

$$t_i = t_{2i}', \quad i = \ldots, -1, 0, 1, \ldots \tag{45}$$

For the process $\xi(s)$ that corresponds to the sequence $\{t_i\}$ the adjoint random variables are given by

$$\begin{aligned}\mathbb{E}\{\xi(s)\xi(s+\tau)\} &= \mathbb{P}\{\text{in the interval}\,(s,s+\tau)\,\text{without points}\ t_i\} \times \mathbb{E}[\xi(t)]^2\\ &= \mathbb{P}\{\text{in the interval}\,(s,s+\tau)\,\text{without points}\ t_i'\} \times \mathbb{E}[\xi(t)]^2\\ &\quad + \mathbb{P}\{\text{in the interval}\,(s,s+\tau)\,\text{there is one point}\,t_i\,\text{with an odd}\,i\} \times \mathbb{E}[\xi(t)]^2\\ &= e^{-\beta\tau}\big(1+\frac{\beta\tau}{2}\big). \end{aligned} \tag{46}$$

Consequently,

$$f(\lambda) = \frac{\beta}{2\pi}\frac{\lambda^3+3\beta^2}{(\lambda^2+\beta^2)^2} \tag{47}$$

and

$$\widetilde{\xi}(t,\tau) = [1+(\sqrt{3}-1)\beta\tau]e^{-\beta\tau}\xi(t) + 2(2-\sqrt{3})\beta^2\tau e^{-\beta\tau}\int_0^\infty \exp(-\sqrt{3}\beta s\xi(t-s)\mathrm{ds}, \tag{48}$$

$$\sigma^2(\tau) = 1-e^{-2\beta\tau}(1+\beta\tau+(2-\sqrt{3})\beta^2\tau^2) \approx 1-e^{-2\beta\tau}[1+\beta\tau+0.268\beta^2\tau^2]. \tag{49}$$

In this case, the optimal nonlinear prediction $\widetilde{\xi}_0(t,\tau)$ depends only upon the value of $\xi(t)$ and upon the distance τ_0 from point t to the latest observed jump of the process $\xi(s)$ (at the point $(t-\tau_0,t)=t_i=t'_{2i}$). Having in mind that with probability $1/(1+\beta\tau_0)$ in the interval $(t-\tau_0,t)$ there will be no points t'_j and, with probability $\beta\tau_0/(1+\beta\tau_0)$, there will be one such point t'_{2i+1}, we have

$$\begin{aligned}\mathbb{E}\{\xi(t+\tau)|\xi(s),s\le t\} &= \mathbb{E}\{\xi(t+\tau)|\xi(t),\tau_0\}\\ &= \frac{1}{1+\beta\tau_0}e^{-\beta\tau}(1+\beta\tau)\xi(t) + \frac{\beta\tau_0}{1+\beta\tau_0}e^{-\beta\tau}(1+\beta\tau)\xi(t),\end{aligned}$$

that is

$$\widetilde{\xi}_0(t,\tau) = \frac{1+\beta(\tau+\tau_0)}{1+\beta\tau_0}e^{-\beta\tau}\xi(\tau). \tag{50}$$

To calculate $\sigma_0^2(\tau)$ the pdf of the random variable τ_0 is given by $h(t_0)=(1/2)\beta(1+\beta t_0)e^{-\beta t_0}$. This can be used to derive the following result

$$\begin{aligned}\sigma_0^2(\tau) &= \int_0^\infty \mathbb{E}\{[\xi(t+\tau)-\widetilde{\xi}_0(t,\tau)]^2|\tau_0\}h(\tau_0)\mathrm{d}\tau_0\\ &= 1-e^{-2\beta\tau}\Big[1+\beta\tau-\frac{eEi(-1)}{2}(\beta\tau)^2\Big] \approx 1-e^{-2\beta\tau}[1+\beta\tau+0.298\beta^2\tau^2],\end{aligned} \tag{51}$$

where for real non-zero values of x, $Ei(x)$ is the integral exponential function ([14]), defined as $Ei(x) = -\int_{-x}^{\infty}\frac{e^{-t}}{t}\mathrm{dt} = \int_{-\infty}^{x}\frac{e^t}{t}\mathrm{dt}$.

We see now that the formula for $\sigma^2(\tau)$ differs from the formula for $\sigma_0^2(\tau)$ in the last case due to a slightly bigger coefficient at $\beta^2\tau^2$. Calculations of $\sigma^2(\tau)$ and $\sigma_0^2(\tau)$ according to formulae (49) and (51), respectively, shows that for any value $\beta\tau$ the first quantity does not exceed the second one by more than 0.1–0.2%.

This approach allows one to study the case when $t_i = t'_{ki}$, where $\{t'_i\}$ is a Poisson sequence of points and $k\ge 3$. Yet, in this case the difference between $\sigma^2(\tau)$ and $\sigma_0^2(\tau)$ turns out to be more noticeable. Similar results are also obtained for processes with jumps of only ± 2 at t_i while the values between the jumps are equal to either -1 or +1 consecutively. Such processes are used sometimes in engineering as models of "infinitely clipped white noise" when the points t_i play the role of zeroes in the initial noise that is being clipped; see, e.g., [15]).

Example 6. In the case when $\{t_i\}$ is a Poisson sequence, the above described process $\xi(s)$ is usually called a "random telegraph signal" (see, e.g., [16]). The correlation function of this particular process is given by

$$R_{\xi\xi}(\tau) = e^{-2\beta|\tau|} \tag{52}$$

and

$$\begin{aligned}\mathbb{E}\{\xi(t+\tau)|\xi(s), s \le\ t\} &= \mathbb{E}\{\xi(t+\tau)|\xi(t)\} \\ &= \xi(t)e^{-\beta\tau}\Big(1 - \beta\tau + \frac{(\beta\tau)^2}{2!} - \cdots\Big) = e^{-2\beta\tau}\xi(t).\end{aligned}$$

Note that

$$\widetilde{\xi}_0(t,\tau) = \widetilde{\xi}(t,\tau), \quad \sigma_0^2(\tau) = \sigma^2(\tau). \tag{53}$$

Thus, similar to Example 4, the best prediction is linear.

Example 7. Consider the "infinitely clipped noise", for which the sequence $\{t_i\}$ is obtained from a Poisson sequence $\{t_i'\}$ (with mean density β) in accordance with formula (45). In this case, for $\tau > 0$, we have

$$\begin{aligned}R_{\xi\xi}(\tau) = \mathbb{E}\{\xi(t)\xi(t+\tau)\} &= \frac{1}{2}e^{-\beta\tau}\Big(1 + \beta\tau - \frac{(\beta\tau)^2}{2!} + \frac{(\beta\tau)^3}{3!} + \cdots\Big) \\ &+ \frac{1}{2}e^{-\beta\tau}\Big(1 - \beta\tau - \frac{(\beta\tau)^2}{2!} + \frac{(\beta\tau)^3}{3!} + \cdots\Big) = e^{-\beta\tau}\cos(\beta\tau).\end{aligned} \tag{54}$$

That is,

$$f(\lambda) = \frac{\beta}{\pi}\frac{\lambda^2 + 2\beta^2}{\lambda^2 + 4\beta^4}, \tag{55}$$

$$\begin{aligned}\widetilde{\xi}(t,\tau) &= [\cos(\beta\tau) + (\sqrt{2}-1)\sin(\beta\tau)]e^{-\beta\tau}\xi(t) \\ &\quad - 2(2-\sqrt{2})\beta\sin(\beta\tau)e^{-\beta\tau}\int_0^\infty e^{\sqrt{2}\beta s}\xi(t-s)\mathrm{d}s,\end{aligned} \tag{56}$$

$$\sigma^2(\tau) = e^{-2\beta\tau}[1 - 2(\sqrt{2}-1)\sin^2(\beta\tau) \approx 1 - e^{-2\beta\tau}[1 - 0.828\sin^2(\beta\tau)]. \tag{57}$$

The best nonlinear prediction can be defined in the same manner as has been done in Example 5:

$$\begin{aligned}\widetilde{\xi}_0(t,\tau) = \mathbb{E}\{\xi(t+\tau)|\xi(t),\tau_0\} &= \xi(t)\Big\{\frac{e^{-\beta\tau}}{1+\beta\tau_0}\Big[1 + \beta\tau - \frac{(\beta\tau)^2}{2!} - \frac{(\beta\tau)^3}{3!} + \cdots\Big] \\ &+ \frac{\beta\tau_0 e^{-\beta\tau}}{1+\beta\tau_0}\Big[1 + \beta\tau - \frac{(\beta\tau)^2}{2!} + \frac{(\beta\tau)^3}{3!} + \cdots\Big]\Big\} = e^{-\beta\tau}\Big(\cos(\beta\tau) + \frac{1-\beta\tau_0}{1+\beta\tau_0}\sin(\beta\tau)\Big)\xi(t).\end{aligned} \tag{58}$$

The MSE is given by

$$\begin{aligned}\sigma_0^2(\tau) &= \int_0^\infty \mathbb{E}\Big\{\Big[\xi(t+\tau) - \widetilde{\xi}_0(t,\tau)\Big]^2|\tau_0\Big\}h(\tau_0)\mathrm{d}\tau_0 \\ &= 1 - e^{-2\beta\tau}[1 - 2(1 + eEi(-1)\sin^2(\beta\tau)] \approx 1 - e^{-2\beta\tau}[1 - 0.810\sin^2(\beta\tau)].\end{aligned} \tag{59}$$

Comparison of (59) with (57) shows that for all values $\beta\tau$ the value of $\sigma(\tau)$ does not differ from $\sigma_0(\tau)$ by more than 0.3%.

3.2 Piecewise Gaussian random processes

An extension of the "point process with adjoint random variables" is the "point process with an adjoint continuous random process" – the random process $\xi(t)$ that has discontinuities of the first kind within the random sequence process coincide with pieces of the continuous stationary random process $\xi_i(s)$ with known statistical properties. In what follows, we will discuss only the case when within all intervals $t_i \leq s < t_{i+1}$, the process coincides with a segment of a Gaussian stationary random process $\xi_i(s)$ with zero mean value, and correlation function $R_i(\tau)$; these processes are not correlated with all processes $\xi_i(s)$ at $j \neq i$.

Consider the case when all correlation function are identical (with $R_i(\tau) = R(0) =$ constant and also if $\beta \to \infty$, such a process turns into a process with adjoint random values). The correlation function of the process $\xi(s)$ is given by

$$R_{\xi\xi}(\tau) = p_0(\tau)R_i(\tau), \tag{60}$$

where $p_0(\tau)$ is the probability of having at least one point t_i within an interval of length τ. If $R_i(\tau)$ has a rational Fourier transform and $p_0(\tau)$ is expressed with a combination of power and exponential functions, the spectral density $f(\lambda)$ that corresponds to $R_{\xi\xi}(\tau)$ will be rational so that the quantities $\widetilde{\xi}(t,\tau)$ and $\sigma^2(\tau)$ can be defined with explicit formulae. As for the best nonlinear forecast, it can be described with a formula of the type

$$\widetilde{\xi}_0(t,\tau) = p(t, t+\tau)\widetilde{\xi}_i(t,\tau,\tau_0), \tag{61}$$

where τ_0 has the same meaning as in formula (50). The quantity $\widetilde{\xi}_i(t,\tau,\tau_0)$ is the best linear prediction at lead time τ of the process having the correlation function $R_i(\tau)$ obtained by using the values of the process at $s - \tau_0 \leq s \leq t$, while $p(t, t+\tau)$ is the conditional probability of having at least one point t_i within the interval $t, t+\tau$ on the semi-axis $-\infty < s \leq t$.

Example 8. Let $R_i(\tau) = e^{-|\tau|}$ and $\{t_i\}$ is a Poisson sequence with mean density β. Then $p_0(\tau) = p_0(t, t+\tau) = e^{-\beta\tau}$, and

$$R_{\xi\xi}(\tau) = e^{-(\beta+1)|\tau|}, \quad \widetilde{\xi}_0(t,\tau) = e^{-(\beta+1)|\tau|}\xi(t) = \widetilde{\xi}(t,\tau). \tag{62}$$

Thus, similar to the case discussed in Examples 4 and 6, the best prediction turns out to be linear.

Example 9. Let $R_i(\tau) = e^{-|\tau|}$, as in the previous example, but $\{t_i\}$ is obtained from a Poisson sequence in accordance with (45). Then, obviously,

$$R_{\xi\xi}(\tau) = \Big(1 + \frac{\beta|\tau|}{2}\Big)e^{-(\beta+1)|\tau|}, \quad f(\lambda) = \frac{\beta+2}{\pi}\,\frac{\lambda^2 + \dfrac{(\beta+1)^2(3\beta+2)}{\beta+2}}{[\lambda^2 + (\beta+1)^2]^2}. \tag{63}$$

In this case we have

$$\widetilde{\xi}(t,\tau) = A\xi(t) + B\int_0^\infty e^{-\gamma s}\xi(t-s)\mathrm{d}s, \tag{64}$$

where A and B are functions of β and τ, γ is a function of β, and

$$\sigma^2(\tau) = 1 - e^{-2(\beta+1)\tau}\{1 + \beta\tau + (\beta+1)[2(\beta+1) - \sqrt{3(\beta+2)(\beta+2)}]\tau^2\}. \tag{65}$$

Now, similar to (50) and (51), we get

$$\widetilde{\xi}_0(t,\tau) = \frac{1+\beta(\tau+\tau_0)}{1+\beta\tau_0} e^{-(\beta+1)\tau}\xi(t), \tag{66}$$

$$\begin{aligned}\sigma_0^2(\tau) &= 1 - e^{-2(\beta+1)\tau}\Big\{1+\beta\tau - \frac{1}{2}eEi(-1)\beta^2\tau^2\Big\}\\ &\approx 1 - e^{-2(\beta+1)\tau}\{1+\beta\tau+0.298\beta^2\tau^2\}.\end{aligned} \tag{67}$$

When comparing (65) and (67), it is useful to keep in mind that for all nonnegative values of β,

$$0.250\beta^2 \le \Big[(\beta+1)[2(\beta+1) - \sqrt{3(\beta+2)(\beta+2)}]\Big] < 0.265\beta^2. \tag{68}$$

Comparing the values of $\sigma^2(\tau)$ and $\sigma_0^2(\tau)$ shows that for any value of β and τ, the first value exceeds the second one by not more than 1%.

Example 10. Let $R_i(\tau) = (1+|\tau|)e^{-(\beta+1)|\tau|}$ and $\{t_i\}$ is a Poisson sequence with mean density β. Then $p_0(\tau) = e^{-\beta\tau}$ and

$$R_{\xi\xi}(\tau) = (1+|\tau|)e^{-(\beta+1)|\tau|}, \quad f(\lambda) = \frac{\beta}{\pi}\frac{\lambda^2 + \dfrac{(\beta+1)^2(\beta+2)}{\beta}}{[\lambda^2+(\beta+1)^2]^2}. \tag{69}$$

In this case $\widetilde{\xi}(t,\tau)$ is given by (64), and the corresponding MSE is of the form

$$\sigma^2(\tau) = 1 - e^{-2(\beta+1)\tau}\{1+2\tau+2(\beta+1)\big[(\beta+1)-\sqrt{\beta(\beta+2)}\big]\tau^2\}. \tag{70}$$

On the other hand, according to (61),[4]

$$\widetilde{\xi}_0(t,\tau) = e^{-(\beta+1)\tau}[(1+\tau)\xi(t) + \tau\xi'(t)] \tag{71}$$

so that

$$\sigma_0^2(\tau) = 1 - e^{-2(\beta+1)\tau}\{1+2\tau+2\tau^2\}. \tag{72}$$

In order to make a comparison of the above formula with (70), note that $1 < 2(\beta+1)[(\beta+1) - \sqrt{\beta(\beta+2)}] < 2$ for all positive values of β. Using (70) and (72), one can verify that in this example the ratio $\sigma_0^2(\tau)/\sigma^2(\tau)$ will always stay between 1 and 0.95 for all values of β and τ.

Example 11. Let$\{t_i\}$ be a Poisson sequence with mean density β, and

$$R_i(\tau) = e^{-|\tau|}\cos(\tau), \quad f(\lambda) = \frac{\beta+1}{\pi}\frac{\lambda^2+(\beta^2+2\beta+2)}{\lambda^2+2\beta(\beta+2)\lambda^2+(\beta^2+2\beta+2)}. \tag{73}$$

In this case, the best linear prediction $\widetilde{\xi}(t,\tau)$ is again given by (64), while the best prediction $\widetilde{\xi}_0(t,\tau)$ is given by

$$\widetilde{\xi}_0(t,\tau) = a\xi(t) + b\xi(t-\tau_0) + \int_0^{\tau_0}\Big(ce^{-\sqrt{2s}} + de^{\sqrt{2s}}\Big)\xi(t-s)\mathrm{dt}, \tag{74}$$

[4]Seemingly, formula (71) shows that the best prediction $\widetilde{\xi}_0(t,\tau)$ is linear; however, then it becomes difficult to understand how it can be different from $\widetilde{\xi}(t,\tau)$. Actually, the process $\xi(s)$ that we are discussing now is not differentiable in the MSE sense (which follows immediately from (69)); therefore, the quantity $\xi'(s)$ cannot be obtained with linear operations over the set of random variables $\xi(s)$, $s \le t$, and, consequently, (71) cannot be regarded as a linear prediction. At the same time, the process $\xi(s)$ will be differentiable almost certainly; therefore, formula (71) makes sense here.

where a, b, c, and d are some specific functions of β, τ, and τ_0 (cf. formulae (1.54) in [17]).

The MSEs of extrapolating with formulae (64) and (74) are, respectively, given by

$$\sigma^2(\tau) = 1 - e^{-2(\beta+1)\tau}\Big\{\cos^2(\tau) + (\sqrt{\beta^2+2\beta+2} - \beta - 1)^2 \sin^2(\tau)\Big\} \tag{75}$$

and

$$\sigma_0^2(\tau) = 1 - e^{-2(\beta+1)\tau}\Big\{\cos^2(\tau) + \frac{4F(1, \frac{\beta}{2\sqrt{2}}, \frac{\beta}{2\sqrt{2}} + 2; 17 - 12\sqrt{2})}{(4+3\sqrt{2})(\beta+2\sqrt{2})}\sin^2(\tau)\Big\}, \tag{76}$$

where $F(\alpha, \beta; \gamma; z)$ is a hypergeometric function. These formulae can be used to verify that the ratio $\sigma_0^2(\tau)/\sigma^2(\tau)$ in this case is not smaller than 0.99 for all values of β and τ.

Now assume that the sequence $\{t_i\}$ is again a Poisson sequence with mean density β, but the correlation function $R_i(\tau)$ alternatively takes two different values: $R_0(\tau) = R_{2k}(\tau)$ (within the intervals $t_i \le s < t_{i+1}$, where $i = 0, \pm 2, \pm 4, \ldots$) and $R_1(\tau) = R_{2k+1}(\tau)$, where $i = \pm 1, \pm 3, \ldots$). Then, the correlation function of the process $\xi(s)$ is given by

$$R_{\xi\xi}(\tau) = \frac{1}{2} e^{-\beta\tau}[R_0(\tau) + R_1(\tau)], \tag{77}$$

and if the Fourier transforms $f_0(\lambda)$ and $f_1(\lambda)$ of both $R_0(\tau)$ and $R_1(\tau)$ are rational, then $R_{\xi\xi}(\tau)$ will also have a rational Fourier transform $f(\lambda)$ that can be used to obtain explicit formulae for, respectively, the best linear prediction $\widetilde{\xi}(t,\tau)$ and for the MSE $\sigma^2(\tau)$.

In the case when the functions $f_0(\lambda)$ and $f_1(\lambda)$ are such that $\lim(f_1(\lambda)/f_0(\lambda)) \ne 1$ for $\lambda \to \infty$, then, according to the well-known result obtained by [18], for any (arbitrarily small) continuous piece of the sample record of the process $\xi(s)$ can be exactly established whether the piece belongs to a process with correlation function $R_0(\tau)$ or $R_1(\tau)$. Therefore, the best nonlinear prediction will still be given by formula (61) where $p(t, t+\tau) = e^{-\beta\tau}$, while $\widetilde{\xi}_i(t, \tau, \tau_0)$ should now be understood as the best linear prediction of the process with the correlation function $R_i(\tau)$ that can be $R_0(\tau)$ or $R_1(\tau)$ depending upon which correlation function corresponds to the last observed continuous piece of the process $\xi(s)$. Also, it is easy to show that if the values of $\xi(s)$ are known over the entire semi-axis $-\infty < s \le t$; this last result will stay valid even in the case when $\lim(f_1(\lambda/f_0(\lambda)) = 1$ for $\lambda \to \infty$. Note that in the limit, when $\beta \to 0$, the process $\xi(s)$ becomes a non-Gaussian and non-ergodic process which presents, with probability 1/2, either a Gaussian process with correlation function $R_0(\tau)$ or, with the same probability, a Gaussian process with correlation function $R_1(\tau)$. However, the process $\xi(s)$ will be ergodic for any positive value of β though this ergodicity will be very difficult to make useful if β is small.

Example 12. Let $R_0(\tau) = e^{-\beta|\tau|}$ and $R_1(\tau) = e^{-\beta|\tau|}(1 + |\tau|)$. Then

$$R_{\xi\xi}(\tau) = \Big(1 + \frac{|\tau|}{2}\Big)e^{-(\beta+1)|\tau|}, \quad f(\lambda = \frac{2\beta+1}{\pi}\,\frac{\lambda^2 + \dfrac{(\beta+1)^2(2\beta+3)}{2\beta+1}}{[\lambda^2 + (\beta+1)^2]^2}. \tag{78}$$

Therefore, the best linear prediction $\widetilde{\xi}(t,\tau)$ is again expressed by (64) while the respective MSE is given by

$$\sigma^2(\tau) = 1 - e^{-2(\beta+1)\tau}\Big\{1 + \tau + (\beta+1)[(2\beta+1) - \sqrt{(2\beta+1)(2\beta+3)}]\tau^2\Big\}. \tag{79}$$

Then, in this case,

$$\widetilde{\xi}(t,\tau) = \begin{cases} e^{-(\beta+1)|\tau|}\xi(t), & \text{if } t_{2i} < t < t_{2i+1}, \\ e^{-(\beta+1)|\tau|}[(1+\tau)\xi(t) + \tau\xi'(t)], & \text{if } t_{2i-1} < t < t_{2i}, \end{cases} \tag{80}$$

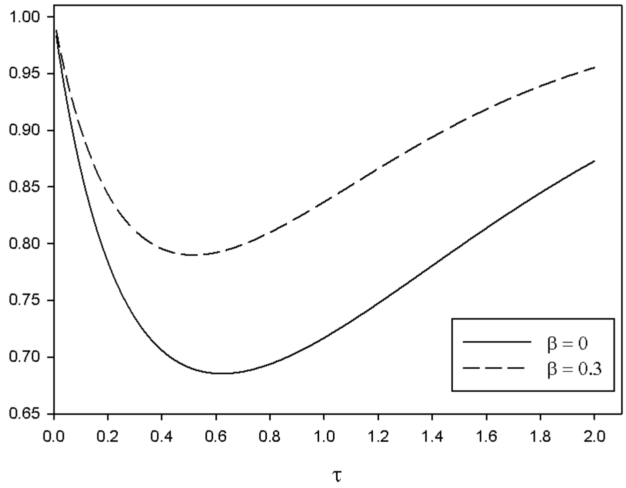

Figure 2: Plot of the ratio $\sigma_0^2(\tau)/\sigma^2(\tau)$ as a function of τ for $\beta = 0$ and $\beta = 0.3$ (Example 12).

and

$$\begin{aligned}\sigma_0^2(\tau) &= \frac{1}{2}\left[1 - e^{-2(\beta+1)|\tau|}\right] + \frac{1}{2}\left[1 - e^{-2(\beta+1)|\tau|}(1 + 2\tau + 2\tau^2)\right] \\ &= 1 - e^{-2(\beta+1)|\tau|}[1 + \tau + \tau^2]. \end{aligned} \tag{81}$$

According to formulae (79) and (81), if $\beta = 0$ (that is, in the case of a non-ergodic process $\xi(s)$), the ratio $\sigma_0^2(\tau)/\sigma^2(\tau)$ changes as a function of τ between 1 and 0.686. However, even for $\beta = 0.3$ (which should be regarded as a small value in this case), these limits become 1 and 0.790; see Figure 2.

Example 13. Assume that the correlation functions $R_0(\tau)$ and $R_1(\tau)$ decrease at different rates and have different initial variances $R(0)$. In particular, let

$$R_0(\tau) = \frac{2}{C+1} e^{-|\tau|} \text{ and } R_1(\tau) = \frac{2C}{C+1} e^{-\alpha|\tau|}. \tag{82}$$

Then, obviously,

$$R(\tau) = \frac{e^{-(\beta+1)|\tau|} + Ce^{-(\beta+\alpha)|\tau|}}{C+1} \tag{83}$$

and

$$f(\lambda) = \frac{1 + c\alpha + (C+1)\beta}{(C+1)\alpha} \frac{\lambda^2 + \dfrac{(1+\beta)(\alpha+\beta)[C + \alpha + (C+1)\beta]}{1 + C\alpha + (C+1)\beta}}{[\lambda^2 + (1+\beta^2)^2][\lambda^2 + (\alpha+\beta)^2]}. \tag{84}$$

The best linear extrapolation formula corresponding to the spectral density (84) is a formula similar to (64), but with A and B being functions of α, β, C, and τ, while γ is a function of α, β, and C, and the MSE is given by

$$\begin{aligned}\sigma^2(\tau) = 1 - \frac{e^{-2\beta\tau}}{C+1}\Big\{ e^{-2\tau} + Ce^{-2\alpha\tau} - \frac{2\sqrt{(1+\beta)(\alpha+\beta)}}{(\alpha-1)^2} \\ \times\Big[\sqrt{(1 + C\alpha + \alpha(C+1)\beta)(C + \alpha + (C+1)\beta)} - (C+1)\sqrt{(1+\beta)(\alpha+\beta)}\Big](e^{-\tau} - e^{-\alpha\tau})^2\Big\}.\end{aligned} \tag{85}$$

Further, we have

$$\widetilde{\xi}(t,\tau) = \begin{cases} e^{-\tau}\xi(t), & \text{if } t_{2i} < t < t_{2i+1}, \\ e^{-\alpha\tau}\xi(\tau), & \text{if } t_{2i-1} < t < t_{2i}, \end{cases} \tag{86}$$

and

$$\sigma_0^2(\tau) = 1 - \frac{e^{-2\beta\tau}(e^{-2\tau} + Ce^{-2\alpha\tau})}{C+1}. \tag{87}$$

It is straightforward to verify that expression (85) is always not smaller than (87); however, the difference between them will usually be very small. In fact, it easily follows from (86) (and also from (85) and (87)) that when $\alpha = 1$ the best prediction turns out to be linear and therefore $\sigma^2(\tau) = \sigma_0^2(\tau)$ (If $C = 1$, the case that we are discussing now obviously coincides with the case discussed in Example 8). Now, when α and β are fixed and $C \to \infty$ or $C \to 0$, $(\sigma_0^2(\tau)/\sigma^2(\tau)) \to 1$ for all τ. The situation will be the same for fixed values of C, α and β, for $\tau \to 1$ and $\tau \to 0$. Finally, if C and α are fixed but $\beta \to \infty$, the ratio $\sigma_0^2(\tau)/\sigma^2(\tau)$ will anyway tend to 1.

All these properties of (85) and (87) make natural the fact that even at intermediate parameter values the ratio $\sigma_0^2(\tau)/\sigma^2(\tau)$ will be close to 1. Indeed, direct calculations show that if, for example, $C = 1$ and $\alpha = 2$ (or $\alpha = 1/2$) the ratio $\sigma_0^2(\tau)/\sigma^2(\tau)$ will stay between 1 and 0.97 for all values of β and τ. It is only when α differs from 1 very much (that is, when the damping rates of $R_0(\tau)$ and $R_1(\tau)$ are very different) and if "the most unproductive" values of β and τ are intentionally selected, it becomes possible to get smaller ratios of $\sigma_0^2(\tau)/\sigma^2(\tau)$ tending to 0.9 which is not insignificant for practical applications. Thus, if $C = 1$ and $\alpha = 9$ (or $\alpha = 1/9$), the smallest value of this ratio (that can be reached at $\beta \approx 0.03$ and $\tau \approx 3$) turns out to be just a little bit smaller than 0.93.

4 Summary

We saw (Examples 1–3) that for a number of non-Gaussian continuous processes the difference between the MSE of the best linear and nonlinear predictions turns out to be very small; therefore, the results obtained in Examples 8–13 make it natural to assume that for disturbance processes consisting of fragments of non-Gaussian continuous processes $\xi_i(s)$, the improvement obtained due to the switching from the optimal linear extrapolation to the nonlinear extrapolation will be very meager in many cases. Of course, we can try to validate this assumption by constructing specific examples with processes consisting of fragments of the Markov processes discussed in Section 2; but we will not explore this task here.

Acknowledgement The author is grateful to R.L. Dobrushin, A.N. Kolmogorov, and A.M. Obukhov for their useful comments on the issues discussed here. He is also grateful N.D. Morozova, E.Ya. Svirina, and G.I Simonova for conducting the numerical calculations that served as a part of the results given above.

References

[1] Kolmogorov AN (1941) Interpolation and extrapolation of stationary random sequences. Izv. AN SSSR, Mathematics, 5, 1, 3–14 SSSR, 48, 8, 339–342.

[2] Krein M (1945) On an A.N. Kolmogorov's extrapolation problem. Dokl. AN SSSR, 48, 8, 339–342.

[3] Wiener N (1949) Extrapolation, interpolation, and smoothing of the stationary time series. New York.

[4] Yaglom AM (1952) Introduction in theory stationary random function. Adv. Math, Sci. 7, 5 (51), 3–168.

[5] Sherman S (1958) Non-mean-square error criteria. IRE Trans. on Inform. Theory, I, T-4, 3, 125–126.

[6] Zade LA (1953) Optimum nonlinear filters. Jour. Appl. Physics, 24, 4, 396–404.

[7] Kuznetsov P, Stratonovich R, Tikhonov V. (1954) Transmission of random functions through nonlinear systems. Automatics and Telemechanics, 15, 3, 200–205.

[8] Lubbok I (1959) The optimization of a class of non-linear filters. Proc. Instn. Electr. Engrs, C, 344 E, 1–15.

[9] Pugatchev V (1960) Theory of random functions and its applications to the problems of automatic control. 2nd edn., Moscow, 1960. English translation by Pergamon Press, Oxford, 1965.

[10] Masani N, Wiener N (1959) Non-linear prediction. Probability and statistics (the Harald Cramer Volume), Stockholm–New York, 190–212.

[11] Doob JL (1953) Stochastic processes. John Wiley and Sons, New York.

[12] Doob JL (1942) The Brownian movement and stochastic equations. Ann. Math., 43, 2, 361–369.

[13] Grenander U (1950) Stochastic processes and statistical inference. Ark. f. Math, 1, 3, 195–227.

[14] Gradsteyn I, Ryzhik I (1980) Table of integrals, series, and products. Academic Press, San Diego.

[15] McFadden J (1958) The fourth product moment of the infinitely clipped noise. IRE Trans of Inform. Theory, I T-4, 4, 159–162.

[16] Rice S (1944) Theory of fluctuation noise. The Bell System Technical Journal, 24, 1, 282–332.

[17] Yaglom AM (1955) Extrapolation, interpolation and filtering of stationary processes with rational spectral density. Tr. Mosk. Math. Obs, 4, 333–374.

[18] Stepian D (1958) Some comments on the detection of Gaussian signals in Gaussian noise. IRE Trans. on Inform Theory, I T-4, 2, 65–68.