UDC 004.942: 551.509.32

Kim Thanh Tran[1], Doctor of Philosophy, Senior lecture, University of Finance-Marketing,
E-mail: tkthanh2011@gmail.com, ORCID: 0000-0001-9601-6880
The Vinh Tran[2,3], Doctor of Philosophy, Senior lecture, Researcher, E-mail: ttvinhcntt@gmail.com, ORCID: 0000-0002-4241-1065
[1]University of Finance-Marketing, 2/4 Tran Xuan Soan Street, District 7, Ho Chi Minh City, Vietnam
[2]Ho Chi Minh City University of Transport (UT_HCMC) No. 2, D3 Street, Ward 2, Ward 25, Binh Thanh District, Ho Chi Minh city, Vietnam
[3]Center of Ukrainian-Vietnamese Cooperation, Odessa National Polytechnic University, Shevchenko Avenue, 1, Odessa, Ukraine, 65044

THE APPLICATION OF CORRELATION FUNCTION IN FORECASTING STOCHASTIC PROCESSES

Annotation. *One of the most important applications of the correlation function is establishing a prediction model for stochastic process. Stationary property makes predicting the stochastic process entirely possible based on the correlation function. This predictive model is interested in cases, where the observation data are assumed to have no measurement errors. We provided some processing to make the forecasting model usable. It is proposed to calculate the value of the standardized correlation function in accordance with the actual observed sample and to estimate the necessary values of averaged correlation function that they cannot be calculated from the sample. We replaced the unknown values by their estimates, which we found using one of the predictive tools suitable for the time series. Theoretically, for the stationary stochastic processes, the correlation function and the standardized correlation function depend only on the time distance between two sections, without depending on the specific time value of each section. However, in this application, when we consider an observation process to be a stationary stochastic process, it means that we have approximated this observation process with a stationary stochastic process. Therefore, when calculating for a specific observation sample, the values of the sample correlation function and the sample standardized correlation function between two sections can fluctuate according to time values of each section, although time distance between two sections unchanged. The sample standardized correlation function of a section has been computed as the arithmetic mean of all values of the sample standardized correlation function between two sections. In this article, the prediction model is linear interpolation and extrapolation model and it is obtained by least squares method. The task for application of this model is to give the highest indexes of daily average temperature in July during last three years 2017-2019 in some localities in northern Vietnam using this forecasting model. The data has been compiled from the data source of the General Department of Meteorology and Hydrology of Vietnam. For processes occurring in the atmosphere and hydrosphere, their hypothesis of stationarity is relatively well satisfied in a time and distance that is not very large. Because of that, we selected the aforementioned data set to apply to the forecasting model. The calculation results are obtained by Matlab software.*

Keyword*: stationary stochastic process; correlation function; forecasting model; temperature*

.

I. Introduction

Forecasting issues are posed in all areas of science and practice [1-5]. The predicted phenomenon usually involves a system that evolves over time and space, which at each time and place it is a random variable. Therefore, the forecasting problem is placed in the random function prediction in general and in particular, the stochastic process. In the theory of the stochastic processes, the correlation function plays a very important role, especially in the problem of the stochastic process prediction ([6-9]).

Given a stochastic process $X = \{X_t, t \in T\}$, for $t \neq s$, the correlation function $R_X(t,s)$ or the standardized correlation function $r_X(t,s)$ characterizes the linear dependence level between two sections X_t, X_s. Therefore, in many application problems, it is sufficient to know the mean function and the correlation function of the stochastic processes especially for stochastic processes with normal distribution that are common in practice, the average function and the correlation function are general features of the stochastic processes [9-11].

In chapter "Stochastic Modeling (Time Series Analysis) and Forecasting" [12], prediction of future observations is done by constructing relevant models based on stochastic process concepts. Stochastic processes can be classified as stationary and non-stationary. Special classes of linear models of stationary stochastic processes are: Auto-regressive processes (AR), Moving-average and Auto-regressive and moving average processes (ARMA). Stochastic process may be described as a phenomenon unfolding in 'time' according to certain probability laws. Here, the word 'time' is used as a real variable which may not always stand for time. A geological process may be viewed as a stochastic process because it is associated with different geochemical elements-each of which can be treated as Random Variable having a probability distribution. Here the observations are not in time,

DOI: 10.15276/hait 04.2019.3
ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

but in space. Even then, we can apply the time domain models of stochastic processes (time series analysis) to geological processes. The applicability of a stochastic process and in particular, the subclasses, viz. AR, MA and ARMA, depends on the behavior of the relevant autocorrelation function (act) and the partial autocorrelation function (pact)

The approach of modeling the correlation as a hyperbolic function of a stochastic process has been recently proposed, reviewed this novel concept and generalize this approach to derive stochastic correlation processes (SCP) from a hyperbolic transformation of the modified Ornstein-Uhlenbeck process [13]. In the article [13] they determined a transition density function of this SCP in closed form which could be used easily to calibrate SCP models to historical data. They computed the price of a quantity adjusting option (Quanta) and discussed concisely the effect of considering stochastic correlation on pricing the Quanta. In [13] they had revised concisely some stochastic correlation models. Market observations give strong evidence that financial quantities are correlated in a strongly nonlinear, non-deterministic way. Instead of assuming a constant correlation, correlation has to be modelled as a stochastic process. They discussed first the general stochastic correlation model proposed in [14] and proved that the stochastic correlation process in [15] can be obtained by applying this general approach. They generalized the approach [14] to derive a stochastic correlation model from a hyperbolic transformation of the modified Ornstein-Uhlenbeck process allowing for a transition density function in a closed form and an easy-to-handle calibration to historical data. As an example, they computed the fair price of a Quanto Put-option with stochastic correlation. The numerical results showed that the correlation risk caused by using a wrong correlation model cannot be neglected.

In this paper presented an application of the correlation function for predicting stationary stochastic processes and illustrate by forecasting the highest daily average air temperature. For processes occurring in the atmosphere and hydrosphere, their hypothesis of stationarity is relatively well satisfied in a time and distance that is not very large ([6]). Because of that, we selected the aforementioned data set to apply to the forecasting model.

II. Stochastic processes analysis and forecasting model

1. The correlation function and the standardized correlation function of the stochastic process

Let $X = \{X_t, t \in T\}$ be a stochastic process. For each t ∈ T, the random variable X_t is called the section at t,

$$m(t) = m_X(t) = m_1(t) = E(X_t),$$

is called the average function of the stochastic process. Denote:

$$\sigma_X(t) = \sqrt{D_X(t)},$$

where: $D_X(t) = E\big(X_t - m_X(t)\big)^2$ is the variance function.

The correlation function of the stochastic process X is:

$$R_X(t,s) = cov(X_t, X_s)$$
$$= E\big\{\big(X_t - m_X(t)\big)\big(X_s - m_X(s)\big)\big\}, t, s \in T.$$

The standardized correlation function of the random process X is:

$$r_X(t,s) = \frac{R_X(t,s)}{\sigma_X(t).\sigma_X(s)}.$$

2. The sample correlation function and the sample standard correlation function of the stochastic process

For the random process $X = \{X_t, t \in T\}$, at different times

$$t_1, t_2, \dots, t_m \ (t_1 < t_2 < \cdots < t_m),$$

we have sections $X_{t_1}, X_{t_2}, \dots, X_{t_m}$ which are m random variables.

Suppose at every section X_{t_j} has n observations (n shown):

$$X_{t_j}(1), X_{t_j}(2), \dots, X_{t_j}(n).$$

We then consider the following important statistical characteristics of the stochastic process ([6].

– Sample mean function:

$$\overline{m}_X(t) = \overline{X}_t = \frac{1}{n}\sum_{i=1}^{n} X_t(i).$$

– Sample correlation function:

$$\overline{R}_X(t,s) = \frac{1}{n-1}\sum_{i=1}^{n}\big[X_t(i) - \overline{X}_t\big]\big[X_s(i) - \overline{X}_s\big].$$

– The sample variance for a section is

$$S^2(X_t)\ \Big(or\ \overline{D}(X_t)\Big) = \overline{R}_X(t,t)$$
$$= \frac{1}{n-1}\sum_{i=1}^{n}\big[X_t(i) - \overline{X}_t\big]^2.$$

– Sample deviation:

$$S(X_t)\ \big(or\ \overline{\sigma}_X(t)\big) = \sqrt{S^2(X_t)}.$$

– Standardized correlation function:

$$\overline{r}_X(t,s) = \frac{\overline{R}_X(t,s)}{S(X_t).S(X_s)}.$$

3. Stationary stochastic process

The stochastic process $X = \{X_t, t \in T\}$ is called a stationary stochastic process (in the narrow sense), if its dimensional finite distribution does not change through the steady translation of time [7; 8; 16], i.e.

$$F(x_1, \dots, x_m;\ t_1 + h, \dots, t_m + h) = \\ = F(x_1, \dots, x_m;\ t_1, \dots, t_m), \\ \forall h: t_j + h \in T, \forall j = \overline{1.m},$$

for every finite distribution function

$$F(x_1, x_2, \dots, x_m;\ t_1, t_2, \dots, t_m).$$

The nature of stationaryness can be stated as follows: observe at any point of the system (vector $(X_{t_1}, X_{t_2}, \dots, X_{t_m})$) and observe at m time after which any time h (vector $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_m+h})$) have the same distribution.

The stationary stochastic process

$$X = \{X_t, t \in T\},$$

has the following remarkable properties:

$$m_X(t) = m_X = const; \\ R_X(t,s) = R_X(\tau)\ (\tau = t - s), \\ \forall t, s \in T$$

so:

$$R_X(-\tau) = R_X(\tau); \\ D_X(\tau) = D_X(0); \\ r_X(\tau) = \frac{R_X(\tau)}{R_X(0)}; \\ r_X(0) = 1.$$

These are very important properties of the stationary stochastic process. In terms of application, it is simplified to investigate and calculate the values of the mean and the correlation functions. Therefore, these properties are used to define a random process that stationary in a broad sense.

It is convention that when talking a process $X = \{X_t, t \in T\}$ it is understood that a stationary stochastic process in a broad sense ([7; 8]), i.e. for this process:

$$m_X(t) = m_X = const; \\ R_X(t,s) = R_X(\tau)\ (\tau = t - s), \\ \forall t, s \in T.$$

For t ≠ s, the correlation function $R_X(t,s)$ or the standardized correlation function $r_X(t,s)$ characterizes the linear dependence level between two sections X_t, X_s. Therefore, in many application problems, it is sufficient to know the mean function and the correlation function of the random process. Especially for random processes with normal distributions common in practice, the mean function and the correlation function are general features of the random process.

4. The forecast model without the measurement errors

In the theory of stochastic processes [6; 7; 11], the correlation function plays a very important role.

Let $X = \{X_t, t \in T\}$ be a stationary stochastic process, with the average m_X and the correlation function $R_X(\tau)$ are known. Suppose at m times $t_1, t_2, \dots, t_m$ $(t_1 < t_2 < \dots < t_m)$, get m expressed values (m observed values): $x_{t_1}, x_{t_2}, \dots, x_{t_m}$. They are assumed to have no measurement errors.

In order to predict the expressed value x_{t_0} of the process at a time t_0 (interpolation, extrapolation predictions without observed errors), we use a linear prediction model [1]:

$$\hat{x}_{t_0} = \sum_{k=1}^{m} \alpha_k x_{t_k}, \qquad (1)$$

in which the coefficients α_k $\left(k = \overline{1,m}\right)$ are found by the least square method.

Thereby leading to the system of equations determining the coefficients α_k is:

$$\sum_{j=1}^{m} \alpha_j . r_X(t_j - t_k) = r_X(t_0 - t_k). \qquad (2)$$

Average square error of interpolation, extrapolation above is:

$$\sigma_m^2(\alpha_1, \dots, \alpha_m) = \\ = R_X(0) - \sum_{k=1}^{m} \alpha_k . R_X(t_0 - t_k). \qquad (3)$$

For the forecasting problem, when m_X average and $R_X(\tau)$ correlation are not known, in practice, they are replaced by the corresponding statistical characteristics of the sample mean $\overline{m}_X$ and the sample correlation function $\overline{R}_X(\tau)$. Therefore (1) is a predictive model with the coefficients $\alpha_k (k = \overline{1,m})$ found from the system of equations:

$$\sum_{j=1}^{m} \alpha_j . \overline{r}_X(t_j - t_k) = \overline{r}_X(t_0 - t_k), \\ \left(k = \overline{1,m}\right), \qquad (2a)$$

and the formula (3) for the average squared error of internal, extrapolation in actual calculation is replaced with:

$$\sigma_m^2(\alpha_1, \dots, \alpha_m) = \\ = \overline{R}_X(0) - \sum_{k=1}^{m} \alpha_k . \overline{R}_X(t_0 - t_k). \qquad (3a)$$

In fact, the following index (3b) is often used to measure interpolation, extrapolation errors:

$$\varepsilon_m = \frac{\sigma_m^2(\alpha_1, \dots, \alpha_m)}{\overline{R}_X(0)} = 1 - \sum_{k=1}^{m} \alpha_k . \overline{r}_X(t_0 - t_k). \qquad (3b)$$

5. Some issues need to be addressed in the forecasting model

When using the forecasting model (1), there are some issues that need to be addressed:

a) In fact the function $r_X(\tau)$ is unknown and when $t_0 \neq t_k, \forall k = 1,2, \dots, m$, the right side of the equation system (2a) will have the values $\overline{r}_X(t_0 -$

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

t_k) undetermined, because the stochastic processes $X = \{X_t, t \in T\}$ has not been observed at t_0 yet. Therefore, the system of equations (2a) has not been solved. To fix this problem, we replace the unknown values by their estimates, which we find using one of the predictive tools suitable for the time series.

b) Theoretically, for the stationary stochastic processes $\{X_t\}$, the correlation function $R_X(t,s)$ and the standardized correlation function $r_X(t,s)$ depend only on the distance $\tau = |t-s|$ without depending on the specific value of t and s. However, in the application, when we consider an observation process to be a stationary stochastic process, it means that we have approximated this observation process with a stationary stochastic process. Therefore, for a specific observation sample, when calculating, the values of the sample correlation function $\overline{R}_X(t,s)$ and the sample standardized correlation function $\overline{r}_X(t,s)$ can fluctuate according to t and s, although $\tau = |t-s|$ unchanged. Therefore in (2a) and (3a), we compute $\overline{r}_X(\tau)$ as the arithmetic mean of all values of $\overline{r}_X(t,s)$ where: $|t-s| = \tau$.

III. Implementation and experiental result

1. Pre-processing data

In the data source of the General Department of Meteorology and Hydrology of Vietnam on average daily air temperature in the months of years from 2008 to 2017 at 11 measurement stations of the northern provinces of Vietnam [17], we are interested in the figures in July every year (the month is considered to be the hottest of the year in Vietnam). The problem is based on data from 2008 to 2016, to forecast the highest index of daily average air temperature in July of 2017, 2018, 2019 in the above 11 stations.

In order to solve this problem, from the data source [17], we synthesize the following data table on the average daily temperature index of the highest days in July from 2008 to 2017 at 11 measurement stations above (Table 1, 2). In this table, the measurement stations are numbered: Bac Ninh (1), Cao Bang (2), Cua Ong (3), Ha Dong (4), Hai Duong (5), Lao Cai (6), Lang Son (7), Luc Ngan (8), Mong Cai (9), Van Ly (10), Vinh Yen (11). The data of the years from 2008 to 2016 is used to forecast the highest daily average temperature index for July from 2008 to 2019 at each measurement station (2017 Figures used to control forecasts).

Table 1. Highest daily average temperatures (t,°C) in July of 2008-2017

t (°C) / Year	The measurement stations					
	(1)	(2)	(3)	(4)	(5)	(6)
2008	31.6	28.9	30.2	31.5	30.9	29.9
2009	32.0	29.8	31.0	31.9	31.8	31.2
2010	33.8	31.4	31.9	33.7	33.4	32.3
2011	32.4	30.2	30.4	32.9	32.3	32.3
2012	31.9	29.7	30.8	33.5	32.7	31.1
2013	30.5	29.7	29.5	31.2	31.0	31.4
2014	31.6	30.3	30.4	32.3	31.9	31.2
2015	32.9	30.4	31.8	34.7	34.4	33.1
2016	31.5	31.1	30.9	32.9	33.0	32.8
2017	30.5	29.9	30.1	32.9	33.0	32.6

Table 2. Highest daily average temperatures (t,°C) in July of 2008 – 2017

t (°C) / Year	The measurement stations					
	(7)	(8)	(9)	(10)	(11)	$\overline{x}_{t_k}$
2008	29.2	31.2	30.4	30.0	31.6	30.5
2009	29.7	31.7	30.3	31.3	31.5	31.1
2010	30.8	33.1	31.3	31.3	33.6	32.4
2011	30.1	32.3	30.4	31.0	32.5	31.5
2012	29.7	31.3	30.0	32.4	32.5	31.4
2013	28.7	30.4	28.6	30.4	31.1	30.2
2014	29.5	31.4	29.2	31.7	31.7	31.0
2015	31.4	33.0	32.6	34.1	32.9	32.8
2016	30.4	31.8	31.8	32.6	32.6	31.9
2017	29.7	31.7	30.5	32.6	31.8	31.4

2. Setting up forecasting model

Let X_t is denoted as the highest index of daily average air temperature in July of year t in the region of 11 observation stations in northern Vietnam. The Table 1 is the table of expressed values for 10 sections

$$X_{t_1}, X_{t_2}, \dots, X_{t_{10}},$$
$$(t_1 = 2008,\ t_2 = 2009, \dots, t_{10} = 2017).$$

Of the random process $X = \{X_t, t \in N\}$ that it is considered to be a stationary stochastic process. Each section X_{t_k} has 11 expressed values:

$$x_{t_k}(1), x_{t_k}(2), \dots, x_{t_k}. \quad (11)$$

Observations are assumed to have no measurement errors. Forecasting model for the highest index of daily average temperature of July at an observation station of year t_0 is:

$$\hat{x}_{t_0} = \sum_{k=1}^{9} \alpha_k x_{t_k}. \quad (4)$$

From (2a), we get the system of equations to determine the coefficients $\alpha_1, \alpha_2, \dots, \alpha_9$, in the case of this problem is:

$$\sum_{j=1}^{9} \alpha_j . \overline{r}_X(t_j - t_k) = \overline{r}_X(t_0 - t_k), \quad (k = \overline{1,9}). \qquad (5)$$

The error of the internal and extrapolation forecast (4) is assessed by the index:

$$\varepsilon_9 = \frac{\sigma_9^2(\alpha_1,\dots,\alpha_9)}{\overline{R}_X(0)} = 1 - \sum_{k=1}^{9} \alpha_k . \overline{r}_X(t_0 - t_k). \qquad (6)$$

2. Proceeding with the forecast and experiental result

From Table 1, the sample mean values for the sections X_{t_k} $(k = 1,2,\dots,9)$ are calculated using the following formula and shown in Table 3

$$\overline{x}_{t_k} = \frac{1}{11}\sum_{i=1}^{11} x_{t_k}(i) \; (k = 1,2,\dots,9).$$

In order to find the predictions by the formula (5), it is need to calculate the values of the sample standardized correlation function $\overline{r}_X(t,s)$.

These values are calculated using the following formula:

$$\overline{r}_X(t,s) = \frac{\overline{R}_X(t,s)}{S(X_t).S(X_s)} =$$

$$= \frac{\sum_{i=1}^{11}[X_t(i)-\overline{X}_t][X_s(i)-\overline{X}_s]}{\left[\left(\sum_{i=1}^{11}[X_t(i)-\overline{X}_t]^2\right).\left(\sum_{i=1}^{n}[X_s(i)-\overline{X}_s]^2\right)\right]^{\frac{1}{2}}}.$$

Table 3. The sample mean values

$\overline{x}_{t_1}$	30.5
$\overline{x}_{t_2}$	31.1
$\overline{x}_{t_3}$	32.4
$\overline{x}_{t_4}$	31.5
$\overline{x}_{t_5}$	31.4
$\overline{x}_{t_6}$	30.2
$\overline{x}_{t_7}$	31.0
$\overline{x}_{t_8}$	32.8
$\overline{x}_{t_9}$	31.9

Table 4. Matrix of sample standardized correlation coefficients

$\overline{r}_X(t,s)$	**2008**	**2009**	**2010**
2008	1	0.864811	0.892584
2009	0.864811	1	0.871165
2010	0.892584	0.871165	1
2011	0.814873	0.883549	0.919543
2012	0.751356	0.876107	0.761973
2013	0.545881	0.787467	0.757626
2014	0.65416	0.880152	0.792261
2015	0.766173	0.752474	0.7802
2016	0.510854	0.672146	0.546585

Table 5. Matrix of sample standardized correlation coefficients

$\overline{r}_X(t,s)$	**2011**	**2012**	**2013**
2008	0.814873	0.751356	0.545881
2009	0.883549	0.876107	0.787467
2010	0.919543	0.761973	0.757626
2011	1	0.810547	0.885129
2012	0.810547	1	0.80056
2013	0.885129	0.80056	1
2014	0.856178	0.924191	0.920149
2015	0.828203	0.725507	0.617823
2016	0.737348	0.790627	0.806162

Table 6. Matrix of sample standardized correlation coefficients

$\overline{r}_X(t,s)$	**2014**	**2015**	**2016**
2008	0.65416	0.766173	0.510854
2009	0.880152	0.752474	0.672146
2010	0.792261	0.7802	0.546585
2011	0.856178	0.828203	0.737348
2012	0.924191	0.725507	0.790627
2013	0.920149	0.617823	0.806162
2014	1	0.608032	0.73878
2015	0.608032	1	0.70082
2016	0.73878	0.70082	1

By Matlab software, the values of $\overline{r}_X(t,s)$ are shown in the Tables 4; 5 and 6.

Determining the values of $\overline{r}_X(\tau)$: As mentioned above, we consider the value of $\overline{r}_X(\tau)$ to be the arithmetic mean of those values of $\overline{r}_X(t,s)$ that $|t-s| = \tau$, $\tau = 0,1,2,\dots,8$.. Because the standardized correlation matrix $(\overline{r}_X(t,s))_{9\times 9}$ is a

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

symmetric matrix, $\overline{r}_X(\tau)$ is determined by the following formula

$$\overline{r}_X(\tau) = \frac{1}{9-\tau}\sum_{t=1}^{9-\tau}\overline{r}_X(t, t+\tau),$$
$$\tau = 0, 1, 2, \dots, 8.$$

Thus each value of

$\overline{r}_X(0), \overline{r}_X(1)$,..., $\overline{r}_X(8)$ is the average of the numbers on the same diagonal line that is parallel to the main diagonal of the matrix $(\overline{r}_X(t,s))_{9\times 9}$.

Such as

$$\overline{r}_X(3) = \frac{1}{6}(0.814873 + 0.876107 + 0.757626 + +0.856178 + 0.725507 + 0.806162 = = 0.806075,$$

$$\overline{r}_X(5) = \tfrac{1}{4}(0.545881 + 0.880152 + +0.7802 + 0.737348) = = 0.735895.$$

In order to estimate the highest index of daily average temperature in July of 2017; 2018; 2019, when solving the system of equations (5), we still need the values $\overline{r}_X(9), \overline{r}_X(10), \overline{r}_X(11)$.

However, these values are related to unobserved years (2017; 2018; 2019), so they cannot be calculated from the sample. After selecting the estimation methods, we replace $\overline{r}_X(9), \overline{r}_X(10)$, $\overline{r}_X(11)$ with their estimates that are found using the 3-level moving average method ([9; 11]), i.e.:

$$\overline{r}_X(9) := \frac{1}{3}\left(\overline{r}_X(6) + \overline{r}_X(7) + \overline{r}_X(8)\right),$$
$$\overline{r}_X(10) := \tfrac{1}{3}\left(\overline{r}_X(7) + \overline{r}_X(8) + \overline{r}_X(9)\right),$$
$$\overline{r}_X(11) := \tfrac{1}{3}\left(\overline{r}_X(8) + \overline{r}_X(9) + \overline{r}_X(10)\right).$$

Obtain from there the required values of $\overline{r}_X(\tau)$ in the Table 7.

Table 7. Values of the arithmetic mean $\overline{r}_X(\tau)$

$\overline{r}_X(0)$	1
$\overline{r}_X(1)$	0.811953
$\overline{r}_X(2)$	0.814861
$\overline{r}_X(3)$	0.806075
$\overline{r}_X(4)$	0.789983
$\overline{r}_X(5)$	0.735895
$\overline{r}_X(6)$	0.651073
$\overline{r}_X(7)$	0.71916
$\overline{r}_X(8)$	0.510854
$\overline{r}_X(9)$	0.627029
$\overline{r}_X(10)$	0.619014
$\overline{r}_X(11)$	0.585632

Replace the values of $\overline{r}(\tau)$ found in Table 7 into the system of equations (5), with Matlab software, we find the values of α_j $(j = 1,2, \dots, 9)$ corresponding to the years in the Tables 8; 9 and 10.

Table 8. Values of the coefficients α_j

α	α_1	α_2	α_3
2008	1	0	0
2009	8.10E-16	1	-1.01E-16
2010	-2.81E-16	-3.45E-17	1.00E+00
2011	-3.33E-16	1.90E-16	1.07E-16
2012	-4.03E-16	4.45E-16	-3.03E-16
2013	-3.05E-16	9.33E-17	1.98E-16
2014	3.71E-16	-4.28E-16	1.07E-16
2015	3.53E-16	-1.61E-16	3.25E-17
2016	-2.72E-16	1.77E-16	-8.83E-17
2017	1.123788	-1.24839	0.385304
2018	0.872549	-0.1746	-0.70935
2019	0.739827	-0.03077	0.138435

Table 9. Values of the coefficients α_j

α	α_4	α_5	α_6
2008	0	0	0
2009	-7.57E-17	-3.79E-16	3.50E-16
2010	7.33E-17	1.05E-16	4.62E-17
2011	1	1.05E-16	4.62E-17
2012	-5.57E-17	1	5.13E-16
2013	2.23E-16	1.13E-16	1
2014	0	1.21E-16	1.70E-16
2015	-4.46E-16	-1.44E-16	3.85E-16
2016	1.11E-16	3.12E-16	1.52E-16
2017	-0.39124	-0.24863	0.198257
2018	0.006413	-0.53602	0.056921
2019	-0.98551	-0.14532	-0.36797

Table 10. Values of the coefficients α_j

α	α_7	α_8	α_9
2008	0	0	0
2009	4.48E-16	-7.41E-16	6.97E-16
2010	-5.86E-16	1.40E-16	9.18E-17
2011	-5.86E-16	1.40E-16	9.18E-17
2012	-1.04E-16	7.61E-16	-6.55E-16
2013	-5.50E-16	1.32E-16	8.61E-17
2014	1	-6.50E-16	3.26E-16
2015	-1.79E-16	1	3.75E-16
2016	7.74E-17	1.24E-16	1
2017	0.383919	-0.45711	1.267619
2018	0.6005	0.074048	0.779919
2019	0.33545	0.31072	0.971142

The errors ε_m of internal, extrapolation according to the years are shown in the Table 11.

The obtained forecasting results, which are the highest index of average daily temperature in July (the hottest month of the year) at the 11 observation

stations in Vietnam from year 2008 to year 2019, are shown in the Tables 12 and 13. The obtained forecasting results at the observation station 2 are illustrated by Fig. 1.

Table 11. Errors ε_m of interpolation, extrapolation

Year	ε_m
2008	0
2009	0
2010	0
2011	0
2012	0
2013	1.11E-16
2014	0
2015	1.11E-16
2016	1.11E-16
2017	0.063994
2018	0.143273
2019	0.140577

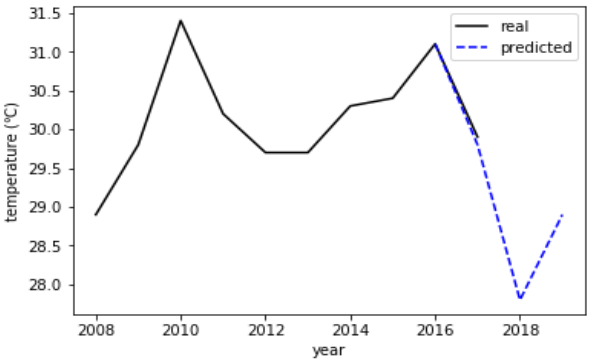

Fig. 1. Chart of the highest index of average daily temperature in July at the observation station 2 in Vietnam

Table 12. The forecasting results at the observation stations 1-5 in Vietnam

Year	Observation stations				
	(1)	(2)	(3)	(4)	(5)
2008	31.6	28.9	30.2	31.5	30.9
2009	32	29.8	31	31.9	31.8
2010	33.8	31.4	31.9	33.7	33.4
2011	32.4	30.2	30.4	32.9	32.3
2012	31.9	29.7	30.8	33.5	32.7
2013	30.5	29.7	29.5	31.2	31.0
2014	31.6	30.3	30.4	32.3	31.9
2015	32.9	30.4	31.8	34.7	34.4
2016	31.5	31.1	30.9	32.9	33.0
2017	30.5	29.8	29.6	31.6	31.3
2018	28.9	27.8	28.3	29.9	29.9
2019	30.1	28.9	29.7	31.2	31.3

Table 13. The forecasting results at the observation stations 6-11 in Vietnam

Year	Observation stations					
	(6)	(7)	(8)	(9)	(10)	(11)
2008	29.9	29.2	31.2	30.4	30	31.6
2009	31.2	29.7	31.7	30.3	31.3	31.5
2010	32.3	30.8	33.1	31.3	31.3	33.6
2011	32.3	30.1	32.3	30.4	31	32.5
2012	31.1	29.7	31.3	30.0	32.4	32.5
2013	31.4	28.7	30.4	28.6	30.4	31.1
2014	31.2	29.5	31.4	29.2	31.7	31.7
2015	33.1	31.4	33.0	32.6	34.1	32.9
2016	32.8	30.4	31.8	31.8	32.6	32.6
2017	30.8	29.2	30.6	30.6	30.6	31.8
2018	29.7	28.0	29.3	29.4	30.3	29.3
2019	29.8	28.8	30.1	30.9	30.9	30.8

Conclusions

The article has mentioned an important application of the correlation function of the stochastic process that is the problem of prediction of stochastic process. In which, we provided the solutions to ensure the system of equations (2a) be solved:

– propose how to calculate the values of sample standardized correlation function $\bar{r}(\tau)$;

– replace the unknown right sides $\bar{r}_X(t_0 - tk$with their estimates by selecting the appropriate estimation method.

A forecasting model has been established to predict the highest index of the daily average temperature in July every year of the observation stations in northern region of Vietnam using the aggregate data from the General Department of Meteorology and Hydrology of Vietnam. The actual observation process cannot perfectly satisfy the theoretical conditions set for the model, such as the stationarity. Therefore, we propose a way to calculate the values of $\bar{r}(\tau)$of the sample standardized correlation function as in the above presented content of the article. As the references have pointed out, for processes occurring in the atmosphere and hydrosphere, their hypothesis of stationarity is relatively well satisfied in a time and distance that is not very large. By examining the sample standardized correlation matrix, it can be found that the more stationarity of the observation process is satisfied, if on each diagonal line that is parallel to the main diagonal, the less the elements oscillate. Then the estimates that are received from the forecasting model (1) are even more accurate.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

References

1. Kannan D., & Lakshmikantham V. (2001). "Handbook of stochastic analysis and application". *CRC Press*. (Published October 23, 2001). Reference – 808 Pages. ISBN 9780824706609 - CAT# DK1885.

2. Gusti Ngurah Agung. (2018). "Advanced Time Series Data Analysis". *Publ. Wiley*, (December 2018). ISBN: 9781119504733.

3. Peter Michael Inness & Steve Dorling. (2012). "Operational Weather Forecasting. Series: Advancing Weather and Climate Science". Wiley-Blackwell. (November 2012, 2018). ISBN: 9781118447642.

4. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm & Kipp Martin. (2013). "Quantitative Methods for Business". *South-Western; 12th International Edition.* Library of Congress Control Number: 2011936338. Book only ISBN-13: 978-0-8400-6234-5. Book only ISBN-10: 0-8400-6234-6.

5. Laurence D. Hoffmann & Gerald L. (2012). "Bradley. Applied Calculus for Business, Economics, and the Social and Life Sciences". *McGraw-Hill Education; 11th edition (January 6, 2012).* ISBN-10: 0073532371. ISBN-13: 978-0073532370.

6. Kazakevits, D. I. (1971). "Fundamentals of the Random Function Theory and their use in Hydrometeorology", *In Russian Gidrometeoizdat*. Leningragad: Russian Federation.

7. Lindgren, Georg. (2006). "Lectures on Stationary Stochastic Processes". Lund University, Lund, Sweden. DOI: https://doi.org/10.1002/9780470012505.tas030.

8. Lindgren, Georg Rootzen, Holger & Sandsten, Maria. (2013). "Stationary Stochastic Processes for Scientists and Engineers". *CRC Press*, 1st Edition. Reference, 1060 Pages. ISBN 9781466586185 - CAT# K20279.

9. Nguyen Duy Tien & Dang Hung Thang. (2005). "Probability Models and Applications. Part II - Stationary Processes and Applications". Hanoi National University; 2th edition (2005). GT.005745. Press.Reference – 126 Pages (in Vietnamese).

10. Ross, S. M. (2009). "Introduction to Probability Models". *Academic Press*; 10th edition. (Published December 17, 2009). Reference – 808 Pages. ISBN-10: 0123756863. ISBN-13: 978-0123756862.

11. Cramér H. & Leadbetter M. R. (1967). "Stationary and Related Stochastic Processes". John Wiley and Sons, Inc., New York: *United States*. (Published January 01, 1967). Reference – 360 Pages. Document ID: 19680032160 (Acquired December 04, 1995). Accession Number:68A13132.

12. Sarma, D. D. (2009). "Stochastic Modeling (Time Series Analysis) and Forecasting". In: *Geostatistics with Applications in Earth Sciences*, pp 62-77. *Publ. Springer*, Dordrecht. Print ISBN978-1-4020-9379-1. Online ISBN978-1-4020-938. DOI https://doi.org/10.1007/978-1-4020-9380-7_4.

13. Long Teng Matthias Ehrhardt & Michael Günther. (2016). "Modelling Stochastic Correlation". *Journal of Mathematics in Industry*, Volume 6, Article number: 2. DOI https://doi.org/10.1186/s13362-016-0018-4.

14. Teng, L, Van Emmerich C, Ehrhardt M. & Günther, M. A. (2016). "Versatile Approach for Stochastic Correlation using Hyperbolic Functions". Int J. Comput Math. 2016; 93(3):524-39. DOI: https://doi.org/10.1080/00207160.2014.1002779.

15. Van Emmerich C. (June 2006). "Modeling Correlation as a Stochastic Process". Preprint 06/03, University of Wuppertal. [Electronic resource]. Access mode: https://pdfs.semanticscholar.org/fdb9/1b0a9fc8b2bbe92888c020df3f291b605b7b.pdf. – Active Link – (June 2006).

16. Dang Hung Thang. (2009). "Stochastic Process and Stochastic Calculus". Lesson of Hanoi National University (Vietnam). [Electronic resource] Access mode: http://www.ebook.edu.vn/?page=1.14&view=17636 – Active Link 2009 (in Vietnamese).

17. General Department of Meteorology and Hydrology of Vietnam. (2019). "Average Daily Temperature in the years from 2008 to 2017 at the Measurement Stations of the Northern Region Vietnam: Bac Ninh, Cao Bang, Cua Ong, Ha Dong, Hai Duong, Lao Cai, Lang Son, Luc Ngan, Mong Cai, Van Ly, Vinh Yen". In statistical data of General Department of Meteorology and Hydrology of Vietnam. Hanoi, Vietnam.

Received 21.10.2019
Received after revision 13.11.2019
Accepted 29.11.2019

[1]**Чан, Кім Тхань,** доктор філософії, старший викладач, E-mail: tkthanh2011@gmail.com, ORCID: 0000-0001-9601-6880
[2,3]**Чан, Тхе Вінь,** доктор філософії, старший викладач, науковий співробітник, E-mail: ttvinhcntt@gmail.com, ORCID: 0000-0002-4241-1065
[1]Університет фінансів та маркетингу, вулиця Чанхуа Шоан 2/4, Район 7, місто Хошимін, В'єтнам.
[2] Університет транспорту міста Хошимін (UT_HCMC), вулиця Д 3, район Бінь Тхань, місто Хошимін, В'єтнам
[3]Центр українсько-в'єтнамського співробітництва, Одеський національний політехнічний університет, проспект Шевченка 1, Одеса, Україна, 65044

ЗАСТОСУВАННЯ ФУНКЦІЇ КОРЕЛЯЦІЇ У ПЕРЕДБАЧЕННІ ВИПАДКОВИХ ПРОЦЕСІВ

Анотація. *Одним з найважливіших застосувань кореляційної функції є встановлення моделі прогнозування стохастичного процесу. Стаціонарна властивість робить прогнозування стохастичного процесу цілком можливим на основі функції кореляції. Ця модель прогнозування зацікавлена у тих випадках, коли в даних спостережень не існує помилок вимірювань. Ми надали деяку обробку, щоб зробити модель прогнозування корисною. Пропонується обчислити значення стандартизованої функції кореляції відповідно до фактично спостережуваної вибірки та оцінити необхідні значення, які вони не можуть бути обчислені з вибірки. Ми замінили невідомі значення їх оцінками, які ми знайшли за допомогою одного з інструментів прогнозування, придатних для часових рядів. Теоретично, для стаціонарних випадкових процесів кореляційна функція і стандартизована кореляційна функція залежать тільки від часової відстані між двома секціями, не залежати від конкретного значення часу кожної секції. Однак в цьому додатку, коли ми розглядаємо процес спостереження як стаціонарний стохастичний процес, це означає, що ми апроксимували цей процес спостереження стаціонарним стохастичним процесом. Отже, при розрахунку для конкретної вибірки спостережень значення кореляційної функції вибірки і стандартизованої кореляційної функції вибірки між двома секціями можуть коливатися відповідно до значень часу кожної секції, хоча часова відстань між двома секціями залишається незмінним. Вибіркова стандартизована кореляційна функція розділу була обчислена як середнє арифметичне всіх значень вибіркової стандартизованої кореляційної функції між двома розділами. У цій статті модель прогнозування - це модель лінійної інтерполяції і екстраполяції, отримана методом найменших квадратів. Завдання застосування цієї моделі полягає в тому, щоб за допомогою цієї моделі прогнозування дати найвищі показники середньодобової температури в липні за останні три роки 2017-2019 в деяких населених пунктах північного В'єтнаму. Дані були зібрані з джерела даних Генерального департаменту метеорології та гідрології В'єтнаму. Для процесів, що відбуваються в атмосфері та гідросфері, їх гіпотеза про стаціонарність порівняно добре задоволена за час та відстань, які не дуже великі. Через це ми вибрали вищезгаданий набір даних для застосування до моделі прогнозування. Результати розрахунків отримують програмне забезпечення Matlab.*

Ключові слова: *стохастичний процес; кореляційна функція; зупиняючі властивості; ергодичні властивості; модель прогнозування; температура*

[1]**Чан, Ким Тхань,** доктор философии, старший преподаватель, E-mail: tkthanh2011@gmail.com, ORCID: 0000-0001-9601-6880
[2,3]**Чан, Тхе Винь,** доктор философии, старший преподаватель, научный сотрудник, E-mail: ttvinhcntt@gmail.com, ORCID: 0000-0002-4241-1065
[1]Университет финансов-маркетинга, улица Чанхуан Шоан 2/4, Район 7, город Хошимин, Вьетнам.
[2]Университет транспорта города Хошимин (UT_HCMC), улица Д3, район Бинь Тхань, г. Хошимин, Вьетнам
[3]Центр украинско-вьетнамского сотрудничества, Одесский национальный политехнический университет, проспект Шевченко 1, Одесса, Украина, 65044

ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННОЙ ФУНКЦИИ В ПРОГНОЗИРОВАНИИ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ

Аннотация. *Одним из наиболее важных применений корреляционной функции является создание модели прогнозирования для случайного процесса. Стационарное свойство делает предсказание стохастического процесса полностью возможным на основе корреляционной функции. Эта прогностическая модель интересна в тех случаях, когда предполагается, что данные наблюдений не имеют ошибок измерения. Мы предоставили некоторую обработку, чтобы сделать модель прогнозирования пригодной для использования. Предлагается рассчитать значение стандартизированной корреляционной функции в соответствии с фактической наблюдаемой выборкой и оценить необходимые значения, чтобы их нельзя было рассчитать по выборке. Мы заменили неизвестные значения их оценками, которые мы нашли с помощью одного из инструментов прогнозирования, подходящих для временных рядов. Теоретически, для стационарных*

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

стохастических процессов корреляционная функция и стандартизированная корреляционная функция зависят только от временного расстояния между двумя секциями, не завися от конкретного значения времени каждой секции. Однако в этом приложении, когда мы рассматриваем процесс наблюдения как стационарный стохастический процесс, это означает, что мы аппроксимировали этот процесс наблюдения стационарным стохастическим процессом. Следовательно, при расчете для конкретной выборки наблюдений значения корреляционной функции выборки и стандартизированной корреляционной функции выборки между двумя секциями могут колебаться в соответствии со значениями времени каждой секции, хотя временное расстояние между двумя секциями остается неизменным. Выборочная стандартизированная корреляционная функция раздела была вычислена как среднее арифметическое всех значений выборочной стандартизированной корреляционной функции между двумя разделами. В этой статье модель прогнозирования - это модель линейной интерполяции и экстраполяции, полученная методом наименьших квадратов. Задача применения этой модели состоит в том, чтобы с помощью этой модели прогнозирования дать самые высокие показатели среднесуточной температуры в июле за последние три года 2017–2019 в некоторых населенных пунктах северного Вьетнама. Данные были собраны из источника данных Главного управления метеорологии и гидрологии Вьетнама. Для процессов, происходящих в атмосфере и гидросфере, их гипотеза о стационарности относительно хорошо выполняется во времени и на расстоянии, которое не очень велико. Из-за этого мы выбрали вышеупомянутый набор данных для применения к модели прогнозирования. Результаты расчетов получены с помощью программного обеспечения Matlab.

Ключевые слова: *стохастический процесс; корреляционная функция; останавливающие свойства; эргодические свойства; модель прогнозирования; температура*

Tran Kim Thanh , Doctor of Philosophy, Senior lecture
Research field: Statistical probability

Tran The Vinh, Doctor of Philosophy, Senior lecture, Researcher
Research field: Number theory, Data analysis