UDC 519.246.8+519.216.3+550.385.1 DOi 10.30837/ bi.2021.1(96).01

Vadim Romanuke
Polish Naval Academy, Gdynia, Poland
v.romanuke@amw.gdynia.pl

MAXIMUM-VERSUS-MEAN ABSOLUTE ERROR IN SELECTING CRITERIA OF TIME SERIES FORECASTING QUALITY

In time series forecasting, a commonly accepted criterion of the forecasting quality is the root-mean-square error (RMSE). Sometimes only RMSE is used. In other cases, another measure of forecasting accuracy is used along with RMSE. It is the mean absolute error (MAE). Although RMSE and MAE are the common criteria of time series forecasting quality, they both register information about averaged errors. However, averaging may remove information about volatility, which is typical for time series, in a few points (outliers) or narrow intervals. Information about outliers in time series forecasts (with respect to test data) can be registered by the maximum absolute error (MaxAE). The MaxAE criterion does not have any relation to averaging. It registers information about the worst outlier instead. Therefore, the goal is to ascertain the best criteria of time series forecasting quality, wherein the RMSE criterion is always present. First, 12 types of benchmark time series are defined to test and select criteria. The time series is of 168 points, whereas the last third of the series is forecasted. After having generated 200 times series for each of those 12 types, ARIMA forecasts are made at 56 points of every series. All the 2400 RMSEs are sorted in ascending order, whereupon the respective MAEs and MaxAEs are re-arranged as well. The interrelation between the RMSE and MAE/MaxAE is studied by their intercorrelation function. RMSEs and MaxAEs are "more different" than RMSEs and MAEs, because the correlation between the RMSE and MAE is stronger. Consequently, the MAE criterion is useless as it just nearly replicates information about the forecasting quality from the RMSE criterion. Inasmuch as the MaxAE criterion can import additional information about the forecasting quality, the best criteria are RMSE and MaxAE.

TIME SERIES FORECASTING, FORECASTING QUALITY, ROOT-MEAN-SQUARE ERROR, MEAN ABSOLUTE ERROR, MAXIMUM ABSOLUTE ERROR, OUTLIERS, ARIMA FORECASTING, INTERCORRELATION FUNCTION

Романюк В. В. Максимальна проти середньої абсолютної похибки при виборі критеріїв якості прогнозування часових рядів. При прогнозуванні часових рядів якість прогнозів загальноприйнято оцінювати за критерієм середньоквадратичної помилки (RMSE). Іноді тільки RMSE й використовують. В інших випадках, разом з RMSE використовується ще одна міра точності прогнозування. Такою мірою є середня абсолютна помилка (MAE). Хоча RMSE й MAE є загальноприйнятими критеріями якості прогнозування часового ряду, вони обидва фіксують інформацію про усереднені помилки. Однак усереднення може стирати інформацію про волатильність, яка є типовою для часових рядів у точках викидів або на вузьких інтервалах. Інформація про викиди у прогнозах часового ряду (відносно тестових даних) може бути зафіксована за допомогою максимуму абсолютної помилки (MaxAE). MaxAE-критерій не має жодного стосунку до усереднення. Натомість він фіксує інформацію про найгірший викид. Тому мета полягає у встановленні найкращих критеріїв якості прогнозування часових рядів, де, щоправда, RMSE-критерій завжди присутній. Спочатку визначаються 12 типів контрольних часових рядів для тестування і вибору критеріїв. Часовий ряд складається зі 168 точок, причому прогнозується остання третина цього ряду. Згенерувавши 200 часових рядів для кожного з 12 типів, виконуються ARIMA-прогнози у 56 точках кожного ряду. Всі 2400 значень RMSE сортуються у порядку зростання, після чого відповідні значення MAE й MaxAE також упорядковуються наново. Взаємоспіввідношення між RMSE й MAE/MaxAE вивчається за їх взаємокореляційною функцією. Значення RMSE й MaxAE є "більш різними", ніж значення RMSE й MAE, оскільки кореляція між RMSE й MAE є сильнішою. Отже, MAE-критерій не має сенсу, тому що він практично повторює інформацію про якість прогнозування з RMSE-критерію. Оскільки MaxAE-критерій може вносити додаткову інформацію про якість прогнозування, найкращими критеріями є RMSE й MaxAE.

ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ, ЯКІСТЬ ПРОГНОЗУВАННЯ, СЕРЕДНЬОКВАДРАТИЧНА ПОМИЛКА, СЕРЕДНЯ АБСОЛЮТНА ПОМИЛКА, МАКСИМУМ АБСОЛЮТНОЇ ПОМИЛКИ, ВИКИДИ, ARIMA-ПРОГНОЗУВАННЯ, ВЗАЄМОКОРЕЛЯЦІЙНА ФУНКЦІЯ

Романюк В. В. Максимальная против средней абсолютной ошибки при выборе критериев качества прогнозирования временных рядов. При прогнозировании временных рядов качество прогнозов общепринято оценивать по критерию среднеквадратической ошибки (RMSE). Иногда только RMSE и используют. В других случаях, вместе с RMSE используется ещё одна мера точности прогнозирования. Такой мерой является средняя абсолютная ошибка (MAE). Хотя RMSE и MAE являются общепринятыми критериями качества прогнозирования временного ряда, они оба фиксируют информацию об усреднённых ошибках. Однако усреднение может стирать информацию о волатильности, которая типична для временных рядов в точках выбросов или на узких интервалах. Информация о выбросах в прогнозах временного ряда (относительно тестовых данных) может быть зафиксирована при помощи максимума абсолютной ошибки (MaxAE). MaxAE-критерий не имеет никакого отношения к усреднению. Вместо этого он фиксирует информацию о наихудшем выбросе. Поэтому цель состоит в установлении наилучших критериев качества прогнозирования временных рядов, где, впрочем, RMSE-критерий присутствует всегда. Сначала определяются 12 типов контрольных временных рядов для тестирования и выбора критериев. Временной ряд состоит из 168 точек, причём прогнозируется последняя треть этого ряда.

Сгенерировав 200 временных рядов для каждого из 12 типов, осуществляются ARIMA-прогнозы в 56 точках каждого ряда. Все 2400 значений RMSE сортируются в порядке возрастания, после чего соответствующие значения MAE и MaxAE также упорядочиваются заново. Взаимосоотношение между RMSE и MAE/MaxAE изучается по их взаимокореляционной функции. Значения RMSE и MaxAE являются “более разными”, чем значения RMSE и MAE, поскольку корреляция между RMSE и MAE сильнее. Итак, MAE-критерий бесполезен, потому что он практически дублирует информацию о качестве прогнозирования из RMSE-критерия. Поскольку MaxAE-критерий может привносить дополнительную информацию о качестве прогнозирования, наилучшими критериями являются RMSE и MaxAE.

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ, КАЧЕСТВО ПРОГНОЗИРОВАНИЯ, СРЕДНЕКВАДРАТИЧЕСКАЯ ОШИБКА, СРЕДНЯЯ АБСОЛЮТНАЯ ОШИБКА, МАКСИМУМ АБСОЛЮТНОЙ ОШИБКИ, ВЫБРОСЫ, ARIMA-ПРОГНОЗИРОВАНИЕ, ВЗАИМОКОРЕЛЯЦИОННАЯ ФУНКЦИЯ

1. Common criteria of time series forecasting quality

Time series forecasting is applied to control and predict processes comprising sequences of data. The quality (accuracy) of forecasting depends on which approaches are used to forecast, length of forecast, and which criteria are used to estimate the quality. A commonly accepted criterion is the root-mean-square error (RMSE) [1, 2]. Sometimes only RMSE is used. In other cases, another measure of forecasting accuracy is used along with RMSE. It is the mean absolute error (MAE) [3]. Although RMSE and MAE are the common criteria of time series forecasting quality [4], they both register information about averaged errors. Meanwhile, time series forecasts are never guaranteed to be similarly scattered around test data points (with respect to which the forecasting accuracy is estimated). Volatility is typical for times series, and it becomes more intense for farther points forecasted. However, averaging may remove information about volatility in a few points (spikes) or narrow intervals.

Information about spikes (which also may be referred to as outliers) in time series forecasts (with respect to test data) can be registered by the maximum absolute error (MaxAE). The MaxAE criterion does not have any relation to averaging [5]. An open question is whether it is better to estimate forecasting accuracy by using only RMSE, or the pair of RMSE and MAE, or the pair of RMSE and MaxAE.

2. The goal and tasks to achieve it

The goal is to ascertain the best criteria of time series forecasting quality (accuracy). To achieve the goal, the following three tasks are to be completed. First, benchmark time series will be defined to test and select criteria. The four options are the single RMSE criterion, RMSE and MAE, RMSE and MaxAE, or RMSE by MAE and MaxAE. Second, an analysis of MaxAE versus MAE will be carried out after forecasts are made by the ARIMA approach [1, 2, 6, 7]. Finally, the selection of criteria should be justified followed by an appropriate conclusion.

3. MaxAE criterion

Consider a time series defined on a sequence of time points $t=\overline{1,T}$, where T is an available amount of data (not to be confused with the availability of data in real-world practice, where forecasting is “blind” and the factual accuracy of forecasts is principally indeterminable until time point T is reached). Hereinafter, it is easier to presume that, without losing generality, $t_i = i$. Let $\{y(t_i)\}_{i=1}^{T_0}$ by $T_0 < T$ be the data by which forecasts are made at $t_i = \overline{T_0+1, T}$. So, set $\{y(t_i)\}_{i=1}^{T_0}$ is the time series to be forecasted for $T - T_0$ time points ahead.

Data $\{y(t_i)\}_{i=T_0+1}^{T}$ are used for testing the forecasting accuracy. These data are normalized (standardized to the range from 0 to 1) as follows [8, 9]:

$$u(t_i) = \frac{y(t_i) - \min\limits_{k=\overline{T_0+1,\,T}} y(t_k)}{\max\limits_{k=\overline{T_0+1,\,T}} y(t_k) - \min\limits_{k=\overline{T_0+1,\,T}} y(t_k)}. \quad (1)$$

The standardization by (1) allows comparing the forecasting quality for different time series defined along the same number of points to be forecasted. Similarly to (1), if $\{\tilde{y}(t_i)\}_{i=T_0+1}^{T}$ are forecasted data (regardless of approaches used to forecast), they are normalized also (with respect to the initial data):

$$\tilde{u}(t_i) = \frac{\tilde{y}(t_i) - \min\limits_{k=\overline{T_0+1,\,T}} y(t_k)}{\max\limits_{k=\overline{T_0+1,\,T}} y(t_k) - \min\limits_{k=\overline{T_0+1,\,T}} y(t_k)}. \quad (2)$$

Then the RMSE is [1, 2]

$$\rho_{RMSE} = \sqrt{\frac{1}{T-T_0} \sum_{i=T_0+1}^{T} \left[u(t_i) - \tilde{u}(t_i) \right]^2} \quad (3)$$

and the MAE [3] is

$$\rho_{MAE} = \frac{1}{T-T_0} \sum_{i=T_0+1}^{T} \left| u(t_i) - \tilde{u}(t_i) \right|. \quad (4)$$

The difference between RMSE (3) and MAE (4) is not that much. Obviously, owing to the square, the RMSE criterion intensifies greater errors. This is why it is unconditionally used almost everywhere to compare data, functions, surfaces, etc. [5, 10]. The MAE criterion is sometimes claimed to be more suitable but reasons behind this are quite unclear. A very serious drawback of both RMSE and MAE is that they do not show outliers. This is so because outliers, if any, are lost due to averaging. On the contrary, the MaxAE calculated as

$$\rho_{MaxAE} = \max_{i=\overline{T_0+1,\,T}} \left| u(t_i) - \tilde{u}(t_i) \right| \quad (5)$$

registers information about the worst outlier [11, 12]. Therefore, whereas RMSE (3) is a compulsory criterion, using MAE (4) or/and MaxAE (5) requires a thorough research and subsequent justification.

4. Benchmark time series

The benchmark time series are based on 12 random-like sequences with repeatability. Every sequence is generated by using pseudorandom numbers drawn from the standard normal distribution (with zero mean and unit variance) [8, 13, 14]. Apart from the repeating random "pure" sequence, a trend, seasonality, and extinction properties are embedded into the sequences by the following patterns [15]:

$$y_1(t)=\left[a_1+0.25\Theta_1(T)\right]r_1(t)+a_2\Theta_2(T), \quad (6)$$

$$y_2(t)=\left[a_1+0.25\Theta_3(T)\right]r_2(t)+a_2\Theta_4(T)+a_3t, \quad (7)$$

$$y_3(t)=\left[a_1+0.25\Theta_5(T)\right]r_3(t)+$$
$$+a_2\Theta_6(T)+\left[a_4+0.25\Theta_7(T)\right]a_5\cos(\upsilon t), \quad (8)$$

$$y_4(t)=\left[a_1+0.25\Theta_8(T)\right]r_4(t)+a_2\Theta_9(T)+$$
$$+a_3t+\left[a_4+0.25\Theta_{10}(T)\right]a_5\cos(\upsilon t), \quad (9)$$

$$y_5(t)=\left[a_1+0.25\Theta_{11}(T)\right]r_5(t)e^{-a_6t}+a_2\Theta_{12}(T), \quad (10)$$

$$y_6(t)=\left[a_1+0.25\Theta_{13}(T)\right]r_6(t)e^{a_6t}+a_2\Theta_{14}(T), \quad (11)$$

$$y_7(t)=\left[a_1+0.25\Theta_{15}(T)\right]r_7(t)e^{-a_6t}+$$
$$+a_2\Theta_{16}(T)+a_3t, \quad (12)$$

$$y_8(t)=\left[a_1+0.25\Theta_{17}(T)\right]r_8(t)e^{-a_6t}+$$
$$+a_2\Theta_{18}(T)+\left[a_4+0.25\Theta_{19}(T)\right]a_5\cos(\upsilon t)e^{-a_6t}, \quad (13)$$

$$y_9(t)=\left[a_1+0.25\Theta_{20}(T)\right]r_9(t)e^{-a_6t}+a_2\Theta_{21}(T)+$$
$$+a_3t+\left[a_4+0.25\Theta_{22}(T)\right]a_5\cos(\upsilon t)e^{-a_6t}, \quad (14)$$

$$y_{10}(t)=\left[a_1+0.25\Theta_{23}(T)\right]r_{10}(t)e^{a_6t}+$$
$$+a_2\Theta_{24}(T)+a_3t, \quad (15)$$

$$y_{11}(t)=\left[a_1+0.25\Theta_{25}(T)\right]r_{11}(t)e^{a_6t}+$$
$$+a_2\Theta_{26}(T)+\left[a_4+0.25\Theta_{27}(T)\right]a_5\cos(\upsilon t)e^{a_6t}, \quad (16)$$

$$y_{12}(t)=\left[a_1+0.25\Theta_{28}(T)\right]r_{12}(t)e^{a_6t}+a_2\Theta_{29}(T)+$$
$$+a_3t+\left[a_4+0.25\Theta_{30}(T)\right]a_5\cos(\upsilon t)e^{a_6t}, \quad (17)$$

where $r_g(t)$ is a sequence of identical randomly-structured subsequences (whose shape is not that random and it still may have some roughly-regular convexities/concavities) of type g, $\left\{\Theta_l(T)\right\}_{l=1}^{30}$ are vectors of T pseudorandom numbers (these vectors are used to simulate noise and volatility), $\left\{a_h>0\right\}_{h=1}^{6}$ is a set of adjustable coefficients, and factor $\upsilon>0$ indicates an oscillation frequency. Initially, a time series is generated by

$$a_1=2,\ a_2=0.175,\ a_3=0.01,\ a_4=5,\ a_5=0.18,$$
$$\upsilon=0.02,\ a_6=0.0005,\ T=1680,$$

where every sequence $r_g(t)$ is of 6, 7, or 8 subsequences, $g=\overline{1,12}$. Then the time series is equidistantly downsampled so that 168 time points remain. These points are smoothed producing thus the benchmark time series. Graphical examples of benchmark time series generated by (6) — (17) are presented in Figure 1.

It is worth noting that the benchmark series are intentionally generated in a way preventing from forecasting trivial time series (being "easy-to-forecast-with-high-accuracy" time series). Although the downsampled time series is smoothed, fluctuations in it are still present (due to every initial sequence of 1680 points has severe spikes which cannot be literally smoothed). Moreover, the smoothing itself may produce outliers at the starting and ending time points, i. e. at $t_1=1$ and $t_{168}=168$. For example, an outlier is seen in the top left subplot in Figure 1. Thus, despite this subplot represents the simplest case without trend and seasonality, the forecasts are not likely to be accurate. Another outlier example is in the second row middle subplot, where the ending time point has unexpectedly dropped down. Similar cases with outliers are seen in Figure 1 as well. Such benchmarking is made to obtain more significant differences in accuracy. This subsequently will allow making more effective decisions on the best criteria of time series forecasting quality.

5. MaxAE versus MAE

After having generated 200 times series for each of those 12 types by $T=168$, ARIMA forecasts are made at $t_i=\overline{113,168}$ (i. e, the forecast length is one third of the available data). The worst and best forecasts (whose RMSEs are the highest and the least for the given type, respectively) are presented in Figure 2. All the 2400 RMSEs are sorted in ascending order, whereupon the respective MAEs and MaxAEs are re-arranged as well. They are shown in Figure 3, where 10 worst RMSEs along with the respective 10 MAEs and MaxAEs are cut off due to they correspond to unacceptable forecasts (see Figure 2).

The interrelation between the RMSE and MAE/MaxAE can be studied by their intercorrelation function [16, 17]. If $\left\{c_j\right\}_{j=1}^{n}$ and $\left\{d_j\right\}_{j=1}^{n}$ are some real-valued data, where $c_j=0$ and $d_j=0$ by $j<1$ or $j>n$, their intercorrelation function is a sequence calculated as follows:

$$b(z)=\sum_{j=1}^{n}c_j\cdot d_{j-z} \quad \text{for} \quad z=\overline{-n+1,n-1}. \quad (18)$$

Inasmuch as the RMSEs, MAEs, and MaxAEs are differently scattered from minimum to maximum values, it is better to normalize them before calculating intercorrelation functions by (18). For this, every value of the respective criterion is divided by the maximum (corresponding to the worst forecast). The intercorrelation functions calculated in this way are presented in Figure 4. The normalized intercorrelation functions are shown in Figure 5 allowing to see more distinctly that RMSEs and MaxAEs are less correlating than RMSEs and MAEs. This is so because the intercorrelation function of RMSEs and MaxAEs is closer to the intercorrelation function of RMSEs and noise, whereas correlation with noise is always the least.

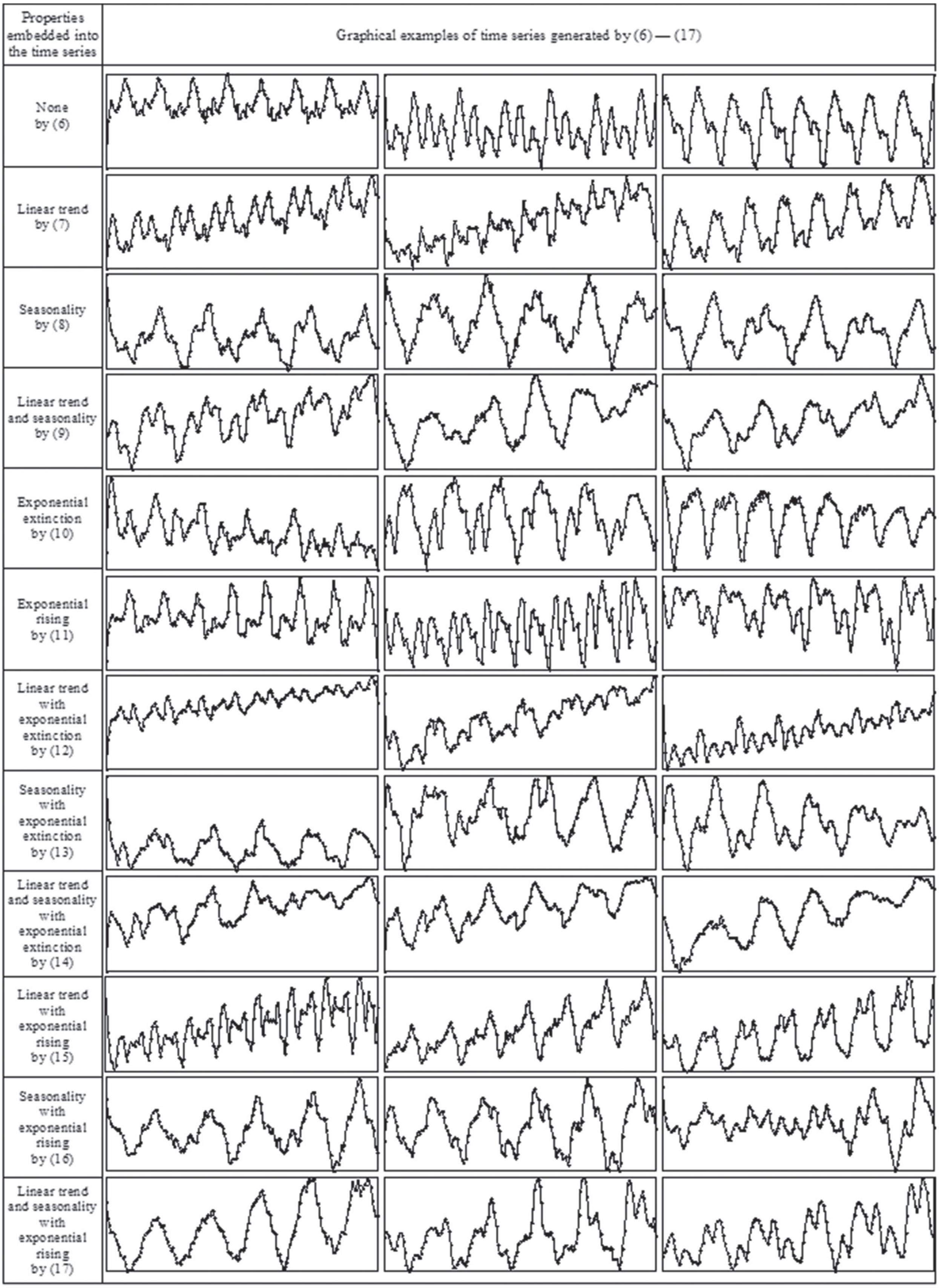

Fig. 1. Graphical examples of the 12 types of benchmark time series

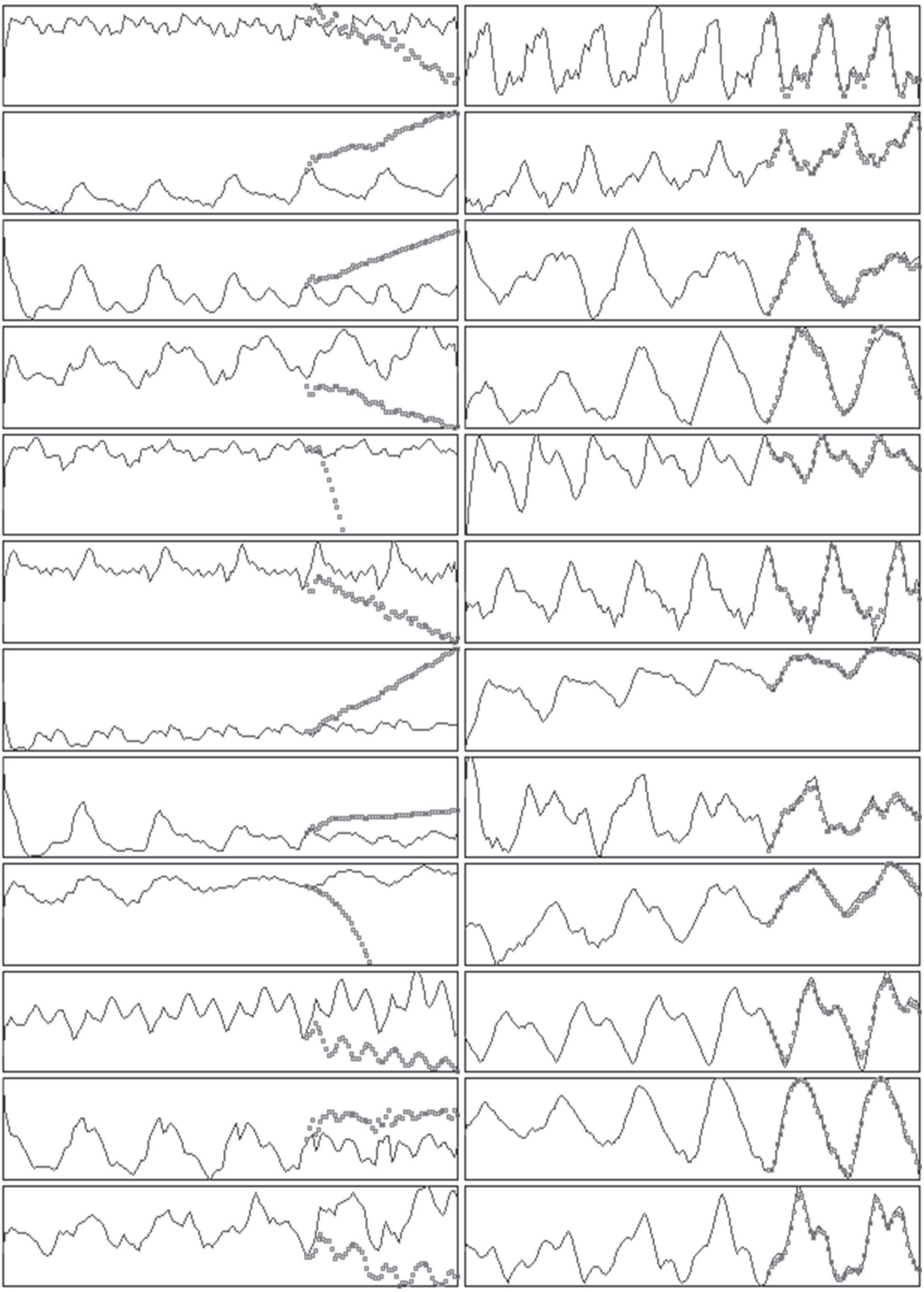

Fig. 2. The worst and best forecasts for each of the 12 types of benchmark time series

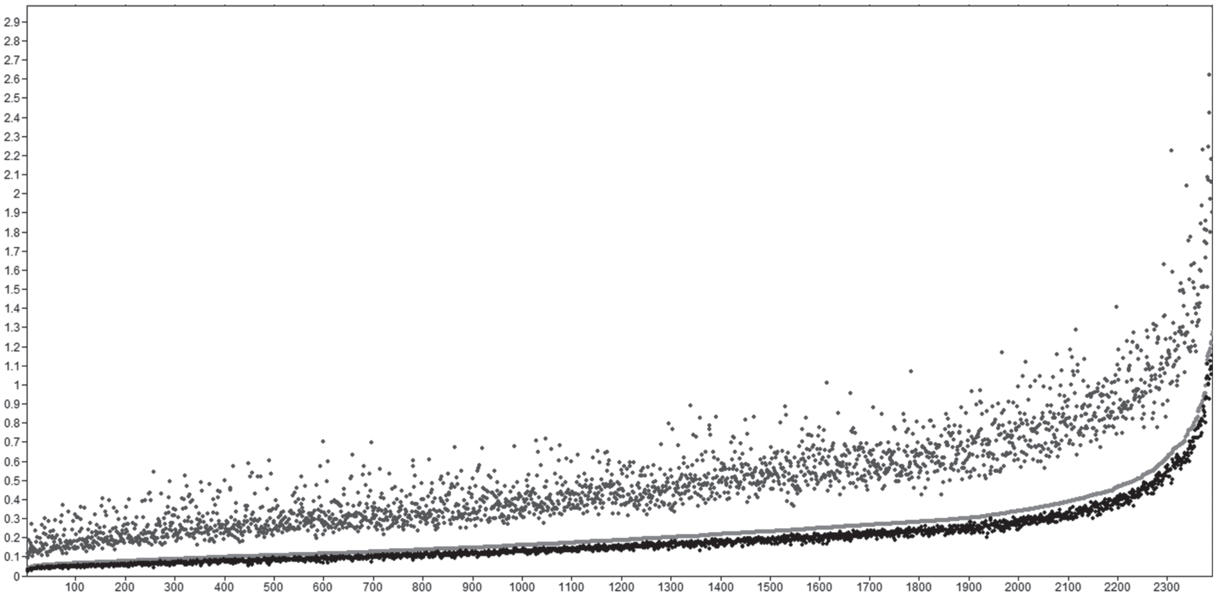

Fig. 3. The distribution of the sorted RMSEs (seen as a line in the middle) of forecasts along with the re-arranged MAEs (points below) and MaxAEs (points above) for the 2390 time series

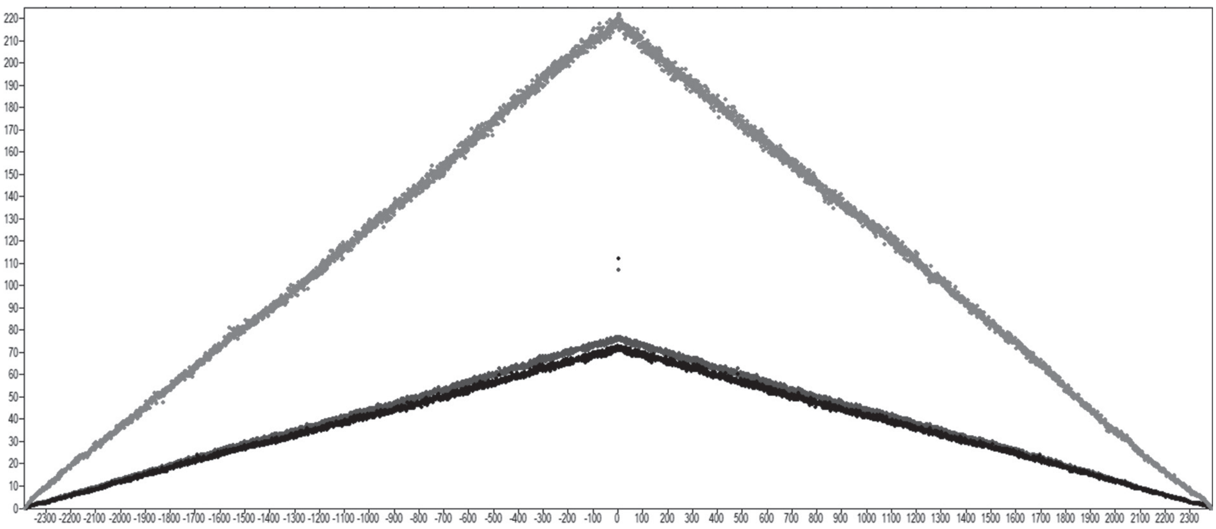

Fig. 4. The three intercorrelation functions of 2390 RMSEs and MAEs, MaxAEs, and a pseudorandom sequence of values (noise) from interval $[0; 1]$

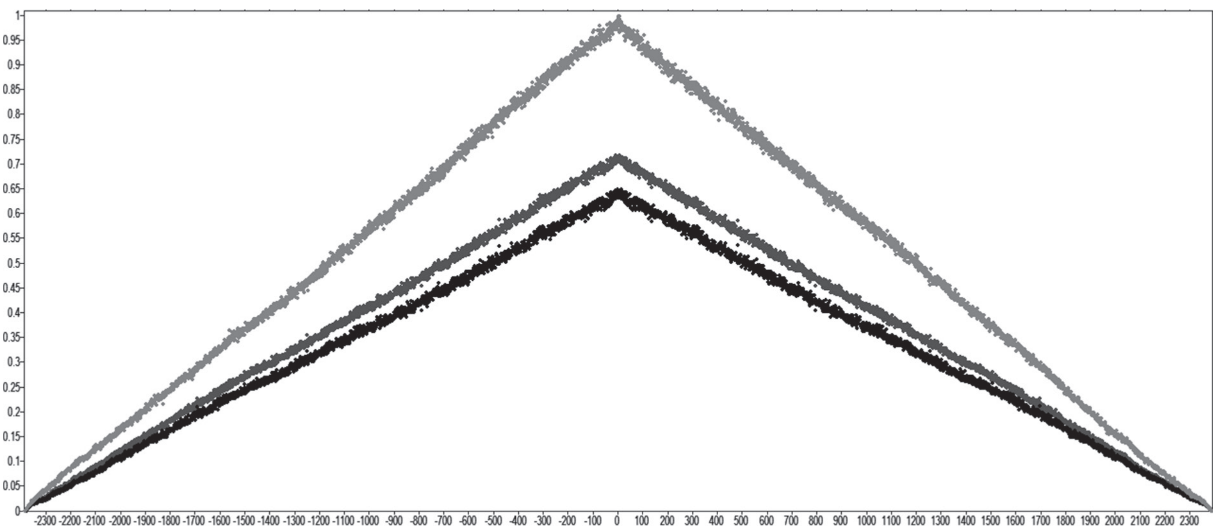

Fig. 5. The normalized intercorrelation functions from Fig. 4

In fact, Figures 3 and 5 are an experimental proof of that using the RMSE and MaxAE criteria is better than using the RMSE and MAE criteria to estimate the time series forecasting quality. When RMSEs are sorted in either ascending or descending order, MaxAEs are more scattered than MAEs (see Figure 3). RMSEs and MaxAEs are “more different” than RMSEs and MAEs (see Figure 5, although Figure 4 may serve herein also), because the correlation between the RMSE and MAE is stronger. This implies that using the MAE criterion along with the RMSE criterion is redundant, whereas the MaxAE criterion can import additional information about forecasts.

Conclusion

Based on the analysis of forecasts for 2400 time series, it is ascertained that the MAE criterion nearly replicates information about the forecasting quality, which is directly drawn from the RMSE criterion. Therefore, the MAE criterion is useless, whichever a group of criteria is, unless the MAE criterion just substitutes the RMSE criterion. Inasmuch as the MaxAE criterion can import additional information about the forecasting quality, the best criteria are RMSE and MaxAE.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

This work was technically supported by the Faculty of Mechanical and Electrical Engineering, Polish Naval Academy, Poland

References

[1] *Schelter B., Winterhalder M., Timmer J.* Handbook of Time Series Analysis: Recent Theoretical Developments and Applications. — Wiley, 2006.

[2] *Kotu V., Deshpande B.* Data Science (Second Edition). — Morgan Kaufmann, 2019.

[3] *Hyndman R., Koehler A.* Another look at measures of forecast accuracy // International Journal of Forecasting. — 2006. — Vol. 22, Iss. 4. — P. 679 — 688.

[4] *De Gooijer J. G., Hyndman R. J.* 25 years of time series forecasting // International Journal of Forecasting. — 2006. — Vol. 22, Iss. 3. — P. 443 — 473.

[5] *Edwards R. E.* Functional Analysis. Theory and Applications. — Hold, Rinehart and Winston, 1965.

[6] *Pankratz A.* Forecasting with Univariate Box — Jenkins Models: Concepts and Cases. — John Wiley & Sons, 1983.

[7] *Box G., Jenkins G., Reinsel G.* Time Series Analysis: Forecasting and Control. — Prentice Hall, Englewood Cliffs, NJ, 1994.

[8] *Romanuke V. V.* Decision making criteria hybridization for finding optimal decisions’ subset regarding changes of the decision function // Journal of Uncertain Systems. — 2018. — Vol. 12, No. 4. — P. 279 — 291.

[9] *Brockwell P. J., Davis R. A.* Introduction to Time Series and Forecasting. — Springer, Cham, 2016.

[10] *Romanuke V. V.* On optimizing WLF equation over experimental viscosity measurements by finite set of fixed temperatures with running metric // Optoelectronic Information-Power Technologies. — 2011. — No. 1 (21). — P. 24 — 43.

[11] *Romanuke V. V.* A minimax approach to mapping partial interval uncertainties into point estimates // Journal of Mathematics and Applications. — 2019. — Vol. 42. — P. 147 — 185.

[12] *Romanuke V. V.* Theoretic-game methods of identification of models for multistage technical control and run-in under multivariate uncertainties. — Mathematical Modeling and Computational Methods. — Vinnytsia National Technical University, Vinnytsia, Ukraine, 2014.

[13] *Kneusel R.* Random Numbers and Computers. — Springer International Publishing, 2018.

[14] *Romanuke V. V.* Multiple state problem reduction and decision making criteria hybridization // Research Bulletin of NTUU “Kyiv Polytechnic Institute”. — 2016. — No. 2. — P. 51 — 59.

[15] *Romanuke V. V.* Time series smoothing and downsampling for improving forecasting accuracy // Applied Computer Systems. — 2021. — Vol. 26, No. 1. — P. 60 — 70.

[16] *Gubner J.* Probability and Random Processes for Electrical and Computer Engineers. — Cambridge University Press, 2006.

[17] *Romanuke V. V.* Computational method of building orthogonal binary functions bases for multichannel communication systems with code channels division. — Mathematical Modeling and Computational Methods. — Ternopil State Technical University, Ternopil, Ukraine, 2006.

The article was delivered to editorial stuff on the 19.02.2021