UDC 004.94
DOI https://doi.org/10.32782/IT/2024-4-8

Vyacheslav GOREV
Candidate of Physical and Mathematical Sciences, Associate Professor, Head of the Department of Physics, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, Gorev.V.M@nmu.one
ORCID: *0000-0002-9528-9497*
Scopus-Author ID: *55047688400*

Yana SHEDLOVSKA
Candidate of Technical Sciences, Associate Professor at the Department of Information Technology and Computer Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, shedlovska.y.i@nmu.one
ORCID: *0000-0003-4931-4070*
Scopus Author ID: *57191853306*

Ivan LAKTIONOV
Doctor of Technical Sciences, Associate Professor, Professor at the Department of Software Engineering, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, laktionov.i.s@nmu.one
ORCID: *0000-0001-7857-6382*
Scopus Author ID: *57194557735*

Grygorii DIACHENKO
Candidate of Technical Sciences, Associate Professor at the Department of Electric Drive, Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, Ukraine, 49005, diachenko.g@nmu.one
ORCID: *0000-0001-9105-1951*
Scopus Author ID: *57201252081*

To cite this article: Gorev, V., Shedlovska, Ya., Laktionov, I., Diachenko, G. (2024). LMS prohnozuvannia ta filtratsiia Kalmana dlia ymovirnosti vynyknennia khvorob kukurudzy [LMS Prediction and Kalman Filtering for the Probability of the Maize Disease Occurrence]. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 61–67, doi: https://doi.org/10.32782/IT/2024-4-8

LMS PREDICTION AND KALMAN FILTERING FOR THE PROBABILITY OF THE MAIZE DISEASE OCCURENCE

The paper is devoted to an urgent agriculture problem such as the maize disease occurrence prediction. The data for the probability of the maise disease occurrence used in the paper are taken from the dataset of professional weather stations from Metos by Pessl Instruments using the FieldClimate IoT platform, access to which is provided by Metos Ukraine LLC. The data are collected from September 2022 to September 2023 with a 1-hour interval for the Dnipropetrovsk region. In our recent papers the data prediction is made on the basis of the neural networks. It should be stressed that the data contain discontinuities, so the question occurs whether the data smoothing may enhance the prediction. In our recent papers we used the smoothing algorithm which is based on an artificial exponential decay which is used in order to smooth the data discontinuities. However, such an approach is rather artificial one. So a question occurs whether a standard smoothing algorithm, for example such as the Kalman one, may enhance the data prediction. For simplicity, in this paper we restrict ourselves to the investigation of the LMS prediction of the smoothed and non-smoothed data. It is shown that the Kalman filtering may slightly enhance some metrics of the LMS prediction. The corresponding investigation for the neural network prediction may be a plan for the future.

The aim of the work *is to investigate the LMS prediction of the probability of the maize disease occurrence in the case where the data are non-smoothed and in the case where the data are smoothed on the basis of the Kalman algorithm.*

The methodology *consists in the Kalman filtering and the LMS prediction of the data for the probability of the maize disease occurrence.*

The scientific novelty *consists in the use of the Kalman filtering of the data for the probability of the maize disease occurrence in order to enhance the data prediction.*

The conclusions *are as follows. The data Kalman filtering may slightly enhance some metrics of the LMS prediction.*

Key words: *Kalman filtering, LMS prediction, probability of the maize disease occurrence.*

В'ячеслав ГОРЄВ
кандидат фізико–математичних наук, доцент, завідувач кафедри фізики, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0002-9528-9497*
Scopus-Author ID: *55047688400*

Яна ШЕДЛОВСЬКА
кандидат технічних наук, доцент кафедри інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0003-4931-4070*
Scopus Author ID: *57191853306*

Іван ЛАКТІОНОВ
доктор технічних наук, доцент, професор кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0001-7857-6382*
Scopus Author ID: *57194557735*

Григорій ДЯЧЕНКО
кандидат технічних наук, доцент кафедри електропривода, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького 19, Дніпро, Україна, 49005
ORCID: *0000-0001-9105-1951*
Scopus Author ID: *57201252081*

Бібліографічний опис статті: Горєв, В, Шедловська, Я., Лактіонов, І., Дяченко, Г. (2024). LMS прогнозування та фільтрація Калмана для ймовірності виникнення хвороб кукурудзи. *Information Technology: Computer Science, Software Engineering and Cyber Security,* 4, 61–67, doi: https://doi.org/10.32782/IT/2024-4-8

LMS ПРОГНОЗУВАННЯ ТА ФІЛЬТРАЦІЯ КАЛМАНА ДЛЯ ЙМОВІРНОСТІ ВИНИКНЕННЯ ХВОРОБ КУКУРУДЗИ

Стаття присвячена такій актуальній проблемі сільського господарства, як прогнозування захворювання кукурудзи. Використані в роботі дані для ймовірності виникнення хвороби кукурудзи взяті з набору даних професійних метеостанцій Metos від Pessl Instruments за допомогою платформи FieldClimate IoT, доступ до якої надає ТОВ «Метос Україна». Дані збирались з вересня 2022 року по вересень 2023 року з інтервалом в 1 годину для Дніпропетровської області. У наших останніх роботах прогнозування даних зроблено на основі нейронних мереж. Слід підкреслити, що дані містять розриви, тому виникає питання, чи може згладжування даних покращити прогнозування. У наших нещодавніх роботах ми використовували алгоритм згладжування, який базується на штучному експоненційному спаді, який використовується для згладжування розривів даних. Однак такий підхід є досить штучним. Тому виникає питання, чи може стандартний алгоритм згладжування, наприклад, такий як алгоритм Калмана, покращити прогнозування даних. Для простоти в цій статті ми обмежимося дослідженням LMS прогнозуванням для згладжених та незгладжених даних. Показано, що фільтрація Калмана може трохи покращити деякі показники LMS прогнозування. Відповідне дослідження для прогнозування на основы нейронних мереж може бути планом на майбутнє.

Метою роботи є *дослідити LMS прогнозування ймовірності виникнення хвороби кукурудзи у випадку незгладжених даних та у випадку згладжених даних на основі алгоритму Калмана.*

Методологія *полягає у фільтрації Калмана та LMS прогнозуванні даних для ймовірності виникнення хвороби кукурудзи.*

Наукова новизна *полягає у використанні фільтрації Калмана даних для ймовірності виникнення хвороби кукурудзи з метою покращення прогнозування даних.*

Висновки *є такими. Фільтрація даних на основі алгоритму Калмана може дещо покращити деякі показники LMS прогнозування.*

Ключові слова: *Фільтрація Калмана, LMS прогнозування, ймовірність виникнення хвороби кукурудзи.*

Introduction. An urgent agriculture problem devoted to the prediction of the probability of the maize disease occurrence is considered in this paper. We use the corresponding data taken from the dataset of professional weather stations from Metos by Pessl Instruments using the FieldClimate IoT platform, access to which is provided by Metos Ukraine LLC, see (Laktionov et al, 2023; Diachenko et al, 2024). The data under investigation contain discontinuities, so the data may be smoothed in order to enhance the prediction. In paper (Diachenko et al, 2024) the use of exponentially decreasing functions is proposed to describe the data smoothing in the vicinity of the discontinuities. However, such a trick is rather artificial one, so in this paper the data smoothing based on the Kalman filtering is considered. Kalman filtering is rather popular for the prediction improvement; see, for example, (Lai et al, 2019). In papers (Laktionov et al, 2023; Diachenko et al, 2024) neural networks were used in order to obtain the data prediction. For simplicity, in this paper we restrict ourselves to the use of the adaptive LMS prediction method, which is rather popular in the literature, see, for example (Tajdari, 2021; Prasetyowati et al, 2021). Both the prediction of smoothed and non-smoothed data are investigated.

Investigation of the LMS prediction of smoothed and non-smoothed data. The data under consideration, see (Laktionov et al, 2023; Diachenko et al, 2024), is shown in Fig. 1 where p is the probability of the maize disease occurrence and n is the number of a point; the data are collected from September 2022 to September 2023 with a 1-hour interval for the Dnipropetrovsk region. The data contains more than 8000 points.

As is known, the algorithm of the Kalman filter is as follows, see (Lai et al, 2019; Gusev et al, 2019). Let us suppose that physically the data obey the equation

$$p_n = ap_{n-1} + bu_{n-1} + \xi_{n-1} \tag{1}$$

where ξ_n is a random process which describes the random character of system evolution, u_n is the known quantity which governs the system evolution, a, b are known constants. Let us suppose that the data are measured by a noisy device which readings are

$$z_n = cp_n + \eta_n \tag{2}$$

where η_n is a random process which describes the device error and c is the gain coefficient. The filter error, ξ_n and η_n are supposed to be independent stationary processes.

Let us denote the smoothed data as p^{opt}. At the first step we put

$$p_1^{\text{opt}} = p_1 , \ \left\langle e_1^2 \right\rangle = \sigma_\eta^2 \tag{3}$$

where $\left\langle e_n^2 \right\rangle$ is the mean square error of the Kalman filter and σ_η^2 is the variance of the process η_n.

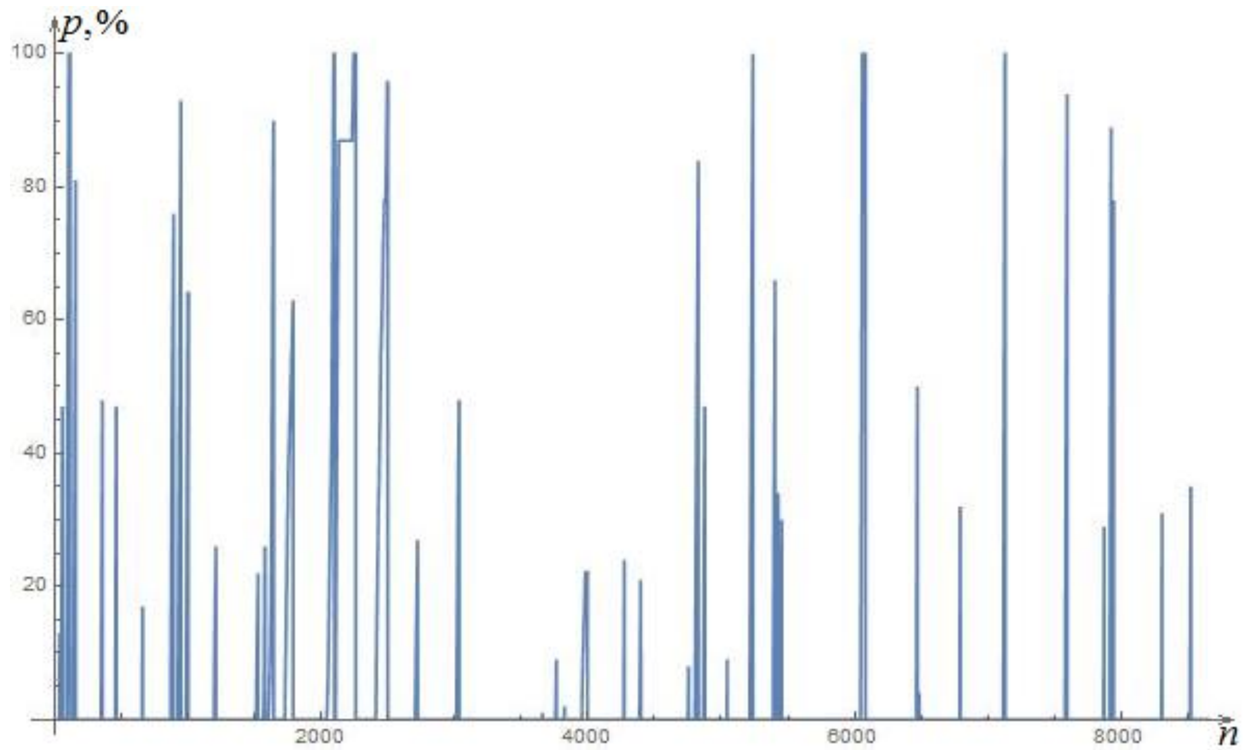

Fig. 1. The data for the probability of the maize disease occurrence

Then at each step the following recurrent formulas are used:

$$p_n^{\text{opt}} = K_n p_n + \left(1 - cK_n\right)\left(a p_{n-1}^{\text{opt}} + b u_{n-1}\right),$$

$$K_n = c\frac{a^2\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2 a^2\left\langle e_{n-1}^2\right\rangle + c^2\sigma_\xi^2},$$

$$\left\langle e_n^2\right\rangle = \sigma_\eta^2\frac{a^2\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + c^2 a^2\left\langle e_{n-1}^2\right\rangle + c^2\sigma_\xi^2} \qquad (4)$$

where σ_ξ^2 is the is the variance of the process ξ_n. According to (Lai et al, 2019), the following parameter choice should be made: the gain coefficient $c = 1$, the physically process does not change its state at each step, so $a = 1$, $b = 0$. In such a case the expressions (4) read as

$$p_n^{\text{opt}} = K_n p_n + \left(1 - K_n\right) p_{n-1}^{\text{opt}},$$

$$K_n = \frac{\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + \left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2},\ \left\langle e_n^2\right\rangle = \sigma_\eta^2\frac{\left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}{\sigma_\eta^2 + \left\langle e_{n-1}^2\right\rangle + \sigma_\xi^2}. \qquad (5)$$

In (Lai et al, 2019) it is stressed that σ_η^2 should be known on the basis of the device characteristics and σ_ξ^2 should be artificially chosen in order to obtain the lowest prediction errors. But in our case the data for the probability of the maize disease occurrence is calculated on the basis of a couple of measurements which are considered to be rather accurate. So, in this paper we artificially put $\sigma_\xi^2 = 1$ and we investigate under which values of σ_η^2 the prediction of the smoother process leads to the most reliable results.

In this paper we investigate a one-point-forward prediction based on the LMS adaptive algorithm. In fact, we provide the prediction algorithm for the smoothed data, but we consider the obtained result as a prediction for the non-smoothed data. The LMS algorithm is as follows, see, for example, (Tajdari, 2021; Prasetyowati et al, 2021). The step size μ and the algorithm depth N are chosen. Let us denote the predicted process as $\bar{p}$. The first N points of the predicted and actual process are considered to be the same and at the zero step the algorithm weights are considered to be zero ones:

$$\bar{p}_n = p_n^{\text{opt}},\ n = \overline{1,N};\ h_j^{(0)} = 0,\ j = \overline{1,N}. \qquad (6)$$

Then a one-point-forward prediction at each step $n > N$ is as follows:

$$\bar{p}_n = \begin{cases} 0, \sum_{j=1}^{N} h_j p_{n-j}^{\text{opt}} < 0 \\ 100, \sum_{j=1}^{N} h_j p_{n-j}^{\text{opt}} > 100 \\ \sum_{j=1}^{N} h_j p_{n-j}^{\text{opt}}, \text{otherwise} \end{cases},$$

$$h_i^{(n)} = h_i^{(n-1)} + 2\mu\left(p_n^{\text{opt}} - \bar{p}_n\right)p_{n-i}^{\text{opt}},\ i = \overline{1,N}, \qquad (7)$$

here we take into account the fact that obviously $\bar{p}_n \in [0,100]$. The following prediction metrics are investigated, see (Laktionov et al, 2023; Diachenko et al, 2024):

$$\text{RMSE} = \sqrt{\frac{1}{M-N}\sum_{j=N+1}^{M}\left(p_j - \bar{p}_j\right)^2},$$

$$\text{MAE} = \frac{1}{M-N}\sum_{j=N+1}^{M}\left|p_j - \bar{p}_j\right|,$$

$$R^2 = 1 - \frac{\sum_{j=N+1}^{M}\left(p_j - \bar{p}_j\right)^2}{\sum_{j=N+1}^{M}\left(p_j - \langle p\rangle\right)^2}, \qquad (8)$$

where $M = 8658$ is the number of data points, and here it is taken into account that only the points with numbers $N+1$, $N+2$,…, M are predicted.

First of all, we investigate the case where $N \ge 2$. An exhaustive search of the parameters σ_η (in the range $\sigma_\eta \in [0,1]$ with the step equal to 10^{-1}), μ (in the range $\mu \in [10^{-7}; 2\cdot 10^{-5}]$ with the step equal to 10^{-7}) and N (in the range $N \in [2;20]$ with the step equal to 1) is provided in order to obtain the minimal MAE value. The corresponding results are given in Table 1.

Table 2

Minimal MAE values in the case N=1

σ_η	**Minimal MAE**	**Corresponding μ value**
0	0.823	$9.1\cdot10^{-6}$
0.1	0.825	$9.2\cdot10^{-6}$
0.2	0.828	$9.5\cdot10^{-6}$
0.3	0.836	$9.9\cdot10^{-6}$
0.4	0.847	$1.06\cdot10^{-5}$
0.5	0.863	$1.12\cdot10^{-5}$
0.6	0.883	$1.23\cdot10^{-5}$
0.7	0.905	$1.38\cdot10^{-5}$
0.8	0.930	$1.58\cdot10^{-5}$
0.9	0.956	$1.89\cdot10^{-5}$
1	0.982	$2.00\cdot10^{-5}$

The value $\sigma_\eta = 0$ describes the non-smoothed data. So, as can be seen, the data smoothing enhances the corresponding prediction, for $\sigma_\eta = 0.7$ we obtain the minimal MAE value during all the above-mentioned exhaustive search, and the corresponding MAE value 0.884 is less than the MAE in the case of the non-smoothed data. In the case of the non-smoothed data ($\sigma_\eta = 0$, $N = 10$, $\mu = 8.8\cdot10^{-6}$) we have $\text{MAE} = 0.919$, $\text{RMSE} = 5.22$ and $R^2 = 0.912$ while in the case where $\sigma_\eta = 0.7$, $N = 9$, $\mu = 1.45\cdot10^{-5}$ we have $\text{MAE} = 0.884$, $\text{RMSE} = 5.14$ and $R^2 = 0.915$, so the MAE and RMSE metrics are better for the considered smoothed data rather than for the non-smoothed data. The corresponding actual and predicted values for the probability of the maize disease occurrence are shown in Fig. 2

Let us investigate the case where $N = 1$. A similar exhaustive search of the parameters σ_η and μ is provided in order to obtain the minimal MAE value. The corresponding results are given in Table 2. As can be seen, the data smoothing does not enhance the MAE value. However, for example, for the case where $\sigma_\eta = 0.1$, $N = 1$, $\mu = 9.2 \cdot 10^{-6}$ we have $\text{MAE} = 0.825$, $\text{RMSE} = 4.995$ and $R^2 = 0.920$ while for the non-smoothed data ($\sigma_\eta = 0$, $N = 1$, $\mu = 9.1 \cdot 10^{-6}$) we have $\text{MAE} = 0.823$, $\text{RMSE} = 4.997$ and $R^2 = 0.920$, so the RMSE parameter is slightly better in the case where the data is smoothed. It should be stressed that in the paper we write the investigated values rounded off up to three significant digits except for the above-mentioned RMSE values. The corresponding values are given rounded off up to four significant digits in order to demonstrate their difference.

The actual and predicted data values for the case $\sigma_\eta = 0.1$, $N = 1$, $\mu = 9.2 \cdot 10^{-6}$ are shown in Fig. 3. As can be seen, the MAE and RMSE metrics are slightly better for the case $N = 1$ rather than for the case where $N \geq 2$. However, some peaks are better predicted in the case where $N \geq 2$ rather than in the case where $N = 1$.

Table 1

Minimal MAE values in the case N≥2

σ_η	**Minimal MAE**	**Corresponding μ value**	**Corresponding N value**
0	0.919	$8.8 \cdot 10^{-6}$	10
0.1	0.918	$8.9 \cdot 10^{-6}$	10
0.2	0.914	$9.3 \cdot 10^{-6}$	10
0.3	0.908	10^{-5}	10
0.4	0.900	$1.14 \cdot 10^{-5}$	9
0.5	0.893	$1.26 \cdot 10^{-5}$	9
0.6	0.886	$1.41 \cdot 10^{-5}$	9
0.7	0.884	$1.45 \cdot 10^{-5}$	9
0.8	0.888	$1.5 \cdot 10^{-5}$	9
0.9	0.899	$1.5 \cdot 10^{-5}$	9
1	0.919	$1.49 \cdot 10^{-5}$	9

Conclusions. The paper is devoted to an urgent agriculture problem of the prediction of the probability of the maize disease occurrence. In this paper for simplicity we restrict ourselves to the LMS prediction algorithm, and it shown that the Kalman smoothing of initial data may lead to the enhancement of some prediction metrics, in particular, the RMSE one. By comparison of the obtained prediction metrics we can conclude that the LMS prediction may work not worse than some

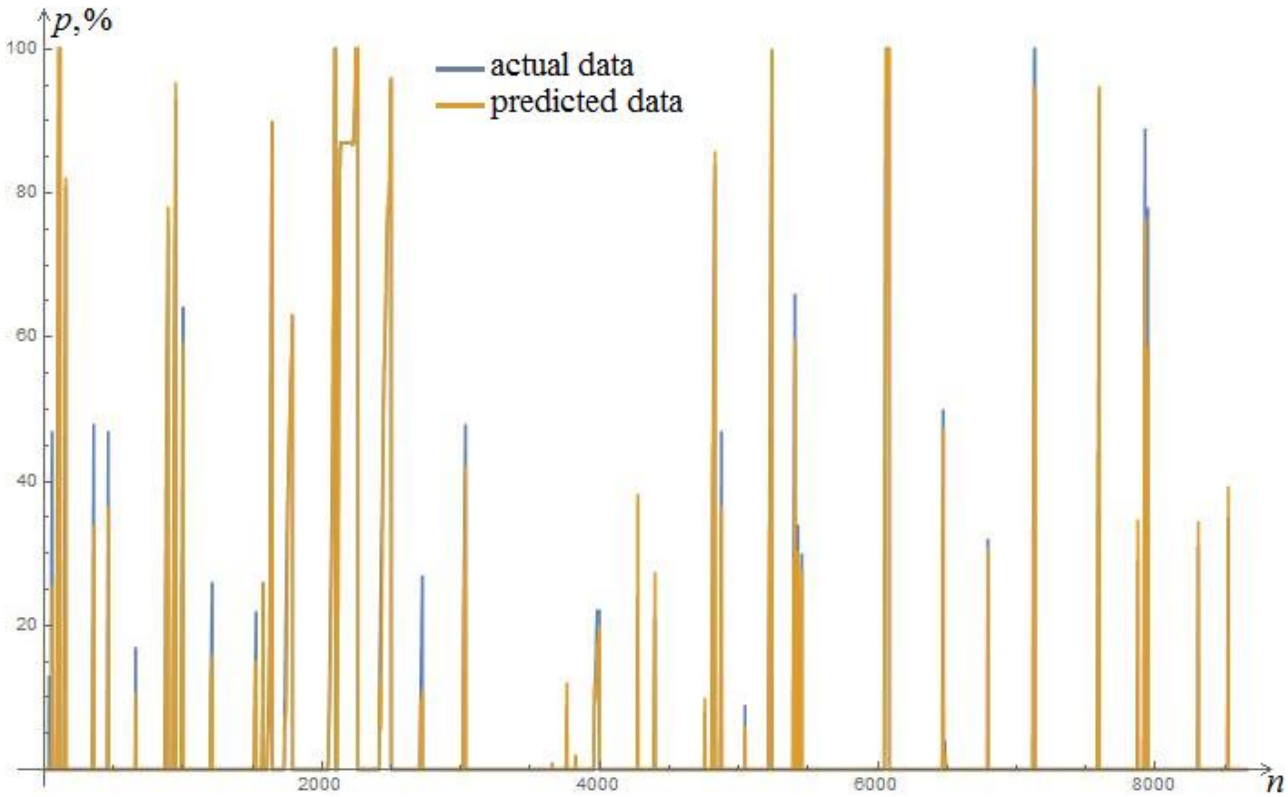

Fig. 2. Actual and predicted data for $\sigma_\eta = 0.7$, $N = 9$, $\mu = 1.45 \cdot 10^{-5}$

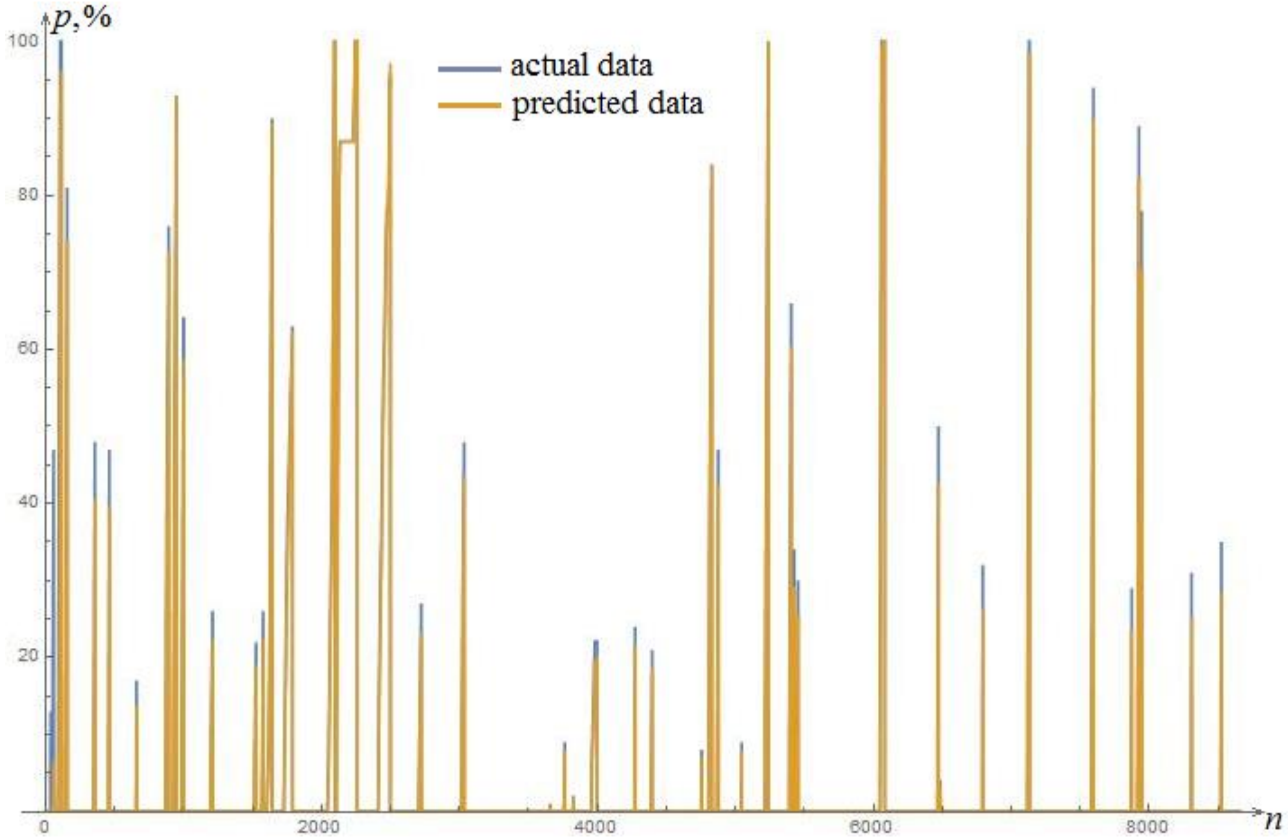

Fig. 3. Actual and predicted data for $\sigma_\eta = 0.1$ **,** $N = 1$ **,** $\mu = 9.2 \cdot 10^{-6}$

neural networks in some cases, see, for example, (Diachenko et al, 2024). However, the Random Forest neural network leads to better prediction results for the data under consideration rather than the LMS prediction, see (Diachenko et al, 2024). The corresponding investigation of the Random Forest prediction on the basis of the Kalman smoothing may be a plan for the future.

Acknowledgements. This research is carried out as part of the scientific project «Development of software and hardware of intelligent technologies for sustainable cultivation of agricultural crops in war and post-war times» funded by the Ministry of Education and Science of Ukraine at the expense of the state budget (State Registration No. 0124U000289).

BIBLIOGRAPHY:

1. Laktionov I., Diachenko G., Rutkowska D., Kisiel-Dorohinick M., An Explainable AI Approach to Agrotechnical Monitoring and Crop Diseases Prediction in Dnipro Region of Ukraine. *Journal of Artificial Intelligence and Soft Computing Research*. 2023. Vol. 13, issue 4, p. 247–272. doi: 10.2478/jaiscr-2023-0018

2. Diachenko G., Laktionov I., Vizniuk A., Gorev V., Kashtan V., Khabarlak K., Shedlovska Y., An Improved Approach to Prediction of Maize Disease Occurrence Based on Weather Monitoring and Machine Learning: a Case of the Forest-Steppe and Northern Steppe of Ukraine. *Engineering Applications of Artificial Intelligence*, 2024, under review.

3. Lai X., Yang T., Wang Z., Peng C. IoT Implementation of Kalman Filter to Improve Accuracy of Air Quality Monitoring and Prediction. *Applied Sciences*. 2019. Vol. 9, issue 9, 1831 (23 pages). doi: 10.3390/app9091831

4. Tajdari T. Adaptive method to predict an track unknown system behaviors using LMS and RLS algorithms. *Facta Universitatis, Series: Electronics and Energetics*. 2021. Vol. 34, issue 1, pp. 133–140. doi: 10.2298/FUEE2101133T

5. Prasetyowati S. A. D., Arifin B., Ardalli J., Ismail M., Subroto I. M. I., Purnomo M. H. A Hybrid FLC-LMS Algorithm for Predicting Sediment Volume in the River. *International Journal of Intelligent Engineering and Systems*. 2021. Vol. 14, issue 2, pp. 395–409. doi: 10.22266/ijies2021.0430.36

6. Gusev O. Yu., Gorev V. M., Korniienko V. I. Theory of adaptive filtration: tutorial. Dnipro: NTU “DP”, 156 p, 2019.

REFERENCES:

1. Laktionov, I., Diachenko, G., Rutkowska, D. & Kisiel-Dorohinick, M. (2023). An Explainable AI Approach to Agrotechnical Monitoring and Crop Diseases Prediction in Dnipro Region of Ukraine. *Journal of Artificial Intelligence and Soft Computing Research*, 13(4), 247–272. doi: 10.2478/jaiscr-2023-0018

2. Diachenko, G., Laktionov, I., Vizniuk, A., Gorev, V., Kashtan, V., Khabarlak, K. & Shedlovska, Y. (2024). An Improved Approach to Prediction of Maize Disease Occurrence Based on Weather Monitoring and Machine Learning: a Case of the Forest-Steppe and Northern Steppe of Ukraine. *Engineering Applications of Artificial Intelligence*, under review.

3. Lai, X., Yang, T., Wang, Z. & Peng, C. (2019). IoT Implementation of Kalman Filter to Improve Accuracy of Air Quality Monitoring and Prediction. *Applied Sciences*, 9(9), 1831 (23 pages). doi: 10.3390/app9091831

4. Tajdari, T. (2021). Adaptive method to predict an track unknown system behaviors using LMS and RLS algorithms. *Facta Universitatis, Series: Electronics and Energetics*, 34(1), 133–140. doi: 10.2298/FUEE2101133T

5. Prasetyowati, S. A. D., Arifin, B., Ardalli, J., Ismail, M., Subroto I. M. I. & Purnomo, M. H. (2021). A Hybrid FLC-LMS Algorithm for Predicting Sediment Volume in the River. *International Journal of Intelligent Engineering and Systems*, 14 (2), 395–409. doi: 10.22266/ijies2021.0430.36

6. Gusev, O. Yu., Gorev, V. M. & Korniienko, V. I. (2019). *Theory of adaptive filtration: tutorial*. Dnipro: NTU “DP”, 156 p.