Methods for statistical analysis of the results of monitoring the avifauna on the territory of wind farms

V V Osadchyi and V S Yeremieiev

1 Bogdan Khmelnitsky Melitopol State Pedagogical University, 20 Hetmanska Str., Melitopol, 72300, Ukraine
E-mail: osadchyi@mdpu.org.ua, eremeev_v@mdpu.org.ua

Abstract. An algorithm for processing the results of monitoring the dynamics of ornithocomplexes on the territory of wind farms using statistical methods is proposed. The solution of the most frequently encountered problems in the analysis of bird migration in the wind farm zone is considered: 1) An algorithm for the primary statistical processing of information on the number of birds of various species, flight altitude and the time of their stay in the zone of interaction with turbines during monitoring has been developed in two ways: a method of route census and observations in accordance with the recommendations of the Scottish Natural Heritage Foundation. 2) The features of the application of correlation and regression analysis have been considered, which allow determining the dependence of the number of birds on a number of factors using the Student, Pearson and Fisher criteria in the presence of strong noise interference. 3) An algorithm of statistical analysis is proposed using a trend approach based on the Student, Irwin, Durbin - Watson, Pearson and Fisher criteria. The considered statistical methods were tested on the results of migratory bird census on the territory of the Prymorsk-1 wind farm located on the coast of the Sea of Azov, which were obtained by a group of researchers led by V Siokhin and P Gorlov.

1. Introduction

The intensive development of wind energy has a significant impact on the avifauna. Numerous observations of researches testify to the need to expand research on the dynamics of ornithocomplexes on the territory of wind farms with the involvement of information technologies and statistical methods of information processing [1–3]. A large number of studies have been devoted to the study of the wind energy impact on ornithological complexes [4–7]. One of the main tasks of monitoring birds on the territory of the wind farms is to obtain data on the dynamics of their quantitative and species characteristics, as well as the height and direction of migration in different seasons [8, 9].

The results of long-term observations over several years make it possible to predict the impact of wind power on the ecological situation associated with changes in the avifauna. Currently, there is an extensive information related to the interaction of birds with wind turbines in various regions, which has been accumulated during the processing of the results of monitoring the territory of wind farms and adjacent regions for several decades [1–3, 9, 10]. The use of various methods of observation expands the possibility of predicting undesirable impact on the avifauna and taking effective measures to reduce this impact.
When organizing observations, the recommendations of the Scottish Natural Heritage Foundation (SNH) [5, 11–13] are often used, which provide for the fulfillment of a number of requirements. In particular, it is assumed that monitoring should be carried out in several areas, which in terms of their landscape and biotope characteristics adequately reflect the entire territory of the wind farm. A sufficient amount of time is allocated for observations, covering the main periods of bird migration in the region. Generally, research should be carried out in spring and autumn during the bird migration, as well as during the breeding period and in winter. Observation is carried out simultaneously by several researches [8, 11, 12]. During monitoring, the following parameters are recorded:

(i) Date, time of the day, species of counted birds and their number at the sites.
(ii) Direction of flight, flight altitude and type of flight (transit, forage, demonstration).
(iii) The time spent by each bird flying below the edge of the wind wheel $H_1$, above its edge $H_2$ and in the risk zone of collision with the rotor blades at heights between $H_1$ and $H_2$.

The results of observations in accordance with the recommendations of the Scottish Natural Heritage Foundation make it possible to determine the activity coefficient of birds of the $j$ species in the risk zone of collision with turbines at the $k$ site $K^{Risk}(k)(j)$, which is necessary to predict the interaction of birds with rotor blades. Its value is [13]

$$K^{Risk}(k)(j) = \sum_i n_i^{Risk(k)(j)} t_i^{Risk(k)(j)},$$

(1)

where $n_i^{Risk(k)(j)}$ - the number of birds of the $j$ species in the $i$ group recorded in the $k$ site in the RZ during the time interval $t_i^{Risk(k)(j)}$.

In a similar way, it is possible to characterize the activity of birds for other heights, as well as in various monitoring areas and the wind farm territory as a whole. For example, the activity coefficient of birds in the $k$ site is determined by the formula

$$K^{(k)} = \sum_i n_i^{(k)(j)} t_i^{(k)(j)},$$

(2)

An alternative way to obtain information about the behaviour of birds on the territory of the wind farm is the method of route census (RAM) [11, 14]. In this case, all birds sitting on the ground and in flight are taken into account. For the last group of birds, the number of individuals of each species, the direction of migration, the height and the type of flight (transit, forage, breeding) are determined. Generally, there are recorded those birds that are located at a distance of no more than 500 m to the left and to the right of the direction of the observer movement along the selected route.

The effectiveness of the analysis of the recorded data largely depends on the adequacy of statistical samples and the correctness of the application of statistical methods. The peculiarity of the processing of primary information in ornithology is associated with the presence of a number of objective reasons, which often make it difficult to obtain reliable conclusions. Therefore, the creation of algorithms that ensure the correct processing of observation results is of great practical importance. This research paper is devoted to the description of the methods of mathematical statistics adapted to the study of the dynamics of ornithocomplexes on the territory of the wind farms.

2. The purpose of the work and setting the task
The purpose of the paper is to develop an algorithm for the use of statistical methods in the analysis of the results of monitoring the avifauna on the territory of the wind farms in solving the following problems:
(i) Primary statistical processing of information on the number of birds of various species, flight altitude and time of their stay at the observation sites, obtained by the SNH [8,12,13] and RAM [8,9,14] methods.

(ii) Correlation and regression analysis to determine the dependence of the number of birds on various factors using the Student, Pearson, Fisher, Irwin criteria.

(iii) Trend analysis of the results of monitoring the wind farm territory based on the Student, Irwin, Durbin - Watson, Pearson and Fisher criteria.

3. Work results

3.1. Primary statistical analysis of the observation results

The number of birds that are recorded in the process of monitoring on the territory of the wind farms varies from several dozens to several thousand individuals and more. Therefore, information processing is carried out with the involvement of computer technologies based on mathematical statistics [1, 9, 11]. In this section, we will focus on the traditional methods of primary statistical processing of observational results, which are usually used in the study of avifauna on the wind farm territory.

The following parameters serve as objects of statistical analysis on the wind farm territory: the number of birds, the number of species, the time spent at the registration site, and the activity coefficients of individuals of various species at different heights, determined by formulas (1), (2). Let $X = x_1, x_2, ..., x_n$ be the number of values that are obtained during the registration of individuals at one of the registration sites during $n$ days. The degree of scattering of each $x_i$ relative to the average value $x_a = \frac{\sum x_i}{n_x}$ is estimated using the empirical variance $s^2_x = \frac{\sum (x_i - x_a)^2}{n_x}$ or the corrected variance $s'^2_x = \frac{\sum (x_i - x_a)^2}{(n_x - 1)}$. The average statistical deviation is taken equal to $s_x = s^2_x$. Statistical data of ornithological research are distinguished by large values of $s_x$, therefore, in some cases (for example, in the graphical interpretation of the results), a smoothing operation is used. The easiest way to smooth is to convert the original data to new values using the formula

$$x_k = (x_{k-1} + x_k + x_{k+1})/3, \quad k = 2, 3, ..., n - 1.,$$  \hspace{1cm} (3)

The weak point of smoothing is the decrease in the number of elements in the original row statistics, which is equal to $n-2$ with a single smoothing. Along with the simple operation (1.1), the averaging is performed over three elements, other methods of linear and nonlinear transformation are used, which are widely used in economic and sociological research [15].

In many cases, there is a problem of comparing monitoring results obtained in different situations, as well as when comparing data at two sites or for different seasons of the year. Let $X = x_1, x_2, ..., x_n$ and $Y = y_1, y_2, ..., y_n$ are two samples, each consisting of $n_x$ and $n_y$ elements. Let us denote the mean sample values as $x_a, y_a$. The verification of the hypothesis about the equality of mathematical expectations is carried out in the case of a known random variable distribution law at a given significance level $q$. Various methods are used to identify the distribution law [15]. As an example, let’s consider the use of the Pearson’s criterion [16] when establishing the normal law of data distribution for the sample $X = x_1, x_2, ..., x_n$. Let’s divide the sample into $m$ equal intervals $h = (x_n - x_1)/m$. Let us denote the number of elements in the $k$ interval as $n_{xk}$. The Pearson criterion is calculated by the formula

$$\chi^2 = \sum_{k=1}^{k=m} \frac{(n_{xk} - n_{xk})^2}{n_{xk}},$$  \hspace{1cm} (4)

where $n'^k_x = hn_x p_k / s_x$ - a theoretical value of the number of elements in the $k$ interval, corresponding to the normal random variable distribution law for the analyzed sample, $p_k$ - the probability that the random variable is in the $k$ interval.
\[ p_k = \exp\left(-\frac{z_k^2}{2}\right)/\sqrt{2\pi}, z_l = \frac{(x_k - x_a)}{s_x}, \]  

(5)

\( x_k \) - the average value of the elements in the k interval. If the value of the \( \chi^2 \)-criterion calculated by the formula (4) is less than the critical value \( \chi^2_{kr} \) for m-3 degrees of freedom at a given significance level \( q \), then the sample obeys the chosen law. For \( \chi^2 > \chi^2_{kr} \), the hypothesis about the possible use of the normal distribution law is rejected. The normal distribution law is characterized by two parameters - the mathematical expectation a and the general variance \( \delta^2 \).

In practice, instead of a, the average value of \( x_a \), is used, and instead of the general variance, the empirical variance \( s_x^2 \) is used. During the primary processing of row statistics, it is recommended to analyze the possibility of the appearance of anomalous values associated with technical errors in obtaining data or processing the original information. The rejection of erroneous data can be carried out by the Irwin’s method. The method is based on determining the coefficients \( \lambda_i = \left|\frac{x_i - x_{i-1}}{s_x}\right|, i = 2, 3, ..., n_x \). The values of \( \lambda_i \) are compared with the critical parameter \( \lambda_a \). If \( \lambda_i > \lambda_a \), then the i element of the row statistics is rejected. Parameters \( \lambda_a \) depending on the sample size for the 5% significance level are presented in table 1.

**Table 1.** Critical Irwin values \( \lambda_a \) depending on the sample size \( n \).

| \( \lambda_a \) | 2.8 | 2.3 | 1.5 | 1.3 | 1.2 | 1.1 | 1.0 |
|-----------------|-----|-----|-----|-----|-----|-----|-----|
| \( n \)         | 2   | 8   | 10  | 20  | 30  | 50  | 100 |

It should be noted that the rejected element may not be the result of a measurement error, but reflect the influence of an unknown factor. But here mathematics is powerless. The decision on rejection is made by the researcher. Testing the hypothesis about the equality of two mathematical expectations of the average value is performed using the Student’s T-test [17]

\[ T = \frac{|x_a - y_a|}{s\sqrt{(n_x n_y)/(n_x + n_y)}}, \]

(6)

where \( s = \sqrt{((n_x - 1)s_x^2 + (n_y - a)s_y^2)/(n_x + n_y - 2)} \)

If \( T \) is less than the critical value \( T_{kr} \) for the significance level \( q \) for \( \nu = n_x + n_y - 2 \) degrees of freedom, then the hypothesis of equality of means is true. For \( |T| > T_{kr} \), the hypothesis is rejected. Sample values of variances \( s_x^2, s_y^2 \) characterizing the scatter of the results of specific measurements generally differ, therefore, before using the formula (6), one should make sure that the general variances \( \delta_x^2 = \delta_y^2 \) are equal. Testing the hypothesis about the equality of general variances can be carried out using the Fisher criterion

\[ F = \left\{ \begin{array}{ll}
\frac{s_x^2}{s_y^2}, & \text{if } s_x^2 > s_y^2 \\
\frac{s_y^2}{s_x^2}, & \text{if } s_y^2 > s_x^2
\end{array} \right. \]

(7)

If the value of \( F \) is less than the critical value \( F_{kr} \) with a given significance level, then the general variances are equal. Otherwise, the hypothesis of equality of dispersions is rejected. Checking the homogeneity of variances with a large number of samples is carried out using the Duncan method and other methods. As an example, we will make a primary statistical analysis of the monitoring results at three vantage points (VP) belonging to the territory of the Prymorsk-1 wind farm. The site sizes of VP1, VP2 and VP3 were respectively 0.50 km\(^2\), 0.86 km\(^2\) and 1.23 km\(^2\), figure 1. Observations were carried out during the spring period of the bird migration in 2017 [11].
Figure 1. Scheme of the location of bird vantage points on the territory of the wind farm.

The results of daily registration of the total number of birds $n^{(k)}$ at the $k$ site per $1 \text{ km}^2$, which flew in transit during the spring migration period, are presented in table 2.

Table 2. The total number of birds flying in transit through the first $n^{(1)}$, second $n^{(2)}$ and third $n^{(3)}$ observation sites on different days.

| Days   | 13.03.17 | 14.03.17 | 25.03.17 | 02.04.17 | 03.04.17 | 04.04.17 | 28.04.17 |
|--------|----------|----------|----------|----------|----------|----------|----------|
| $i$    | 1        | 2        | 3        | 4        | 5        | 6        | 7        |
| $n^{(1)}$ | 74       | 106      | 264      | 102      | 70       | 62       | 238      |
| $n^{(2)}$ | 43.0     | 76.7     | 30.23    | 25.58    | 1.26     | 90.7     | 59.3     |
| $n^{(3)}$ | 32.5     | 39.0     | 108.9    | 85.4     | 23.6     | 52.0     | 63.4     |

At the first stage, we reject possible erroneous measurements using the Irwin method. According to table 1, Irwin’s critical value $\lambda_a$ for the 5% significance level with the number of measurements $n=7$ is about 2.4. Coefficients $\lambda_i$ for all row statistics in accordance with the data in table 2 is much less than the critical value. The exception is the measurement $n^{(2)}=1.26$, which was carried out on April 3, 2017, when $\lambda_i$ turned out to be 3.1.

The results of variance calculations of $s^2_x$, standard deviations $s_x$ and the Pearson’s criteria $\chi^2$ according to the data of table 2, from which the measurement $n^{(2)}=1.26$ obtained on April 3, 2017 is excluded, are presented in table 3. The values of $x^2$ were calculated for four intervals: $m=4$. The Pearson’s criterion is recommended to apply for large samples. Therefore, the obtained data should be treated with caution.

The critical value of the Pearson criterion for the 5% significance level at $m-3=1$ ($m=4$) degrees of freedom is 3.8. The calculated value $x^2=6.7$ for the first area is greater than the critical value, so the first row statistics do not obey the normal random variable distribution law.

Checking the homogeneity of variances in the number of birds in different sites using the Fisher’s formulas (7) gave the following results for a 5% significance level: $F_{23} = 1.4$ at $F_{kr} = 5.0$, $F_{13} = 7.6$ at $F_{kr} = 4.3$, $F_{12} = 10.4$ at $F_{kr} = 5.0$. Here, the numerical indices in the notation of
Table 3. Values of variances $s^2_x$, standard deviations $s_x$ from the mean value $x_a$ and the Pearson’s criteria $\chi^2$ after rejection of erroneous measurements.

| Sites  | OP1 | OP2 | OP3 |
|--------|-----|-----|-----|
| $n$    | 7   | 6   | 7   |
| $x_a$  | 131 | 54.3| 57.8|
| $s^2_x$| 6048| 564 | 800 |
| $s_x$  | 77.8| 23.7| 28.3|
| $\chi^2$ | 6.7 | 2.8 | 3.1 |

the $F_{ij}$ criteria refer to the numbers of the monitoring sites. The calculation results showed that the hypothesis about the homogeneity of the variances of the analyzed row statistics can only be accepted when comparing the number of birds in the second and third sites, where the Fisher criterion $F_{23} = 1.4$ is less than the critical value $F_{kr} = 5.0$. In other cases, the homogeneity hypothesis is not fulfilled.

Let us test the hypothesis about the equality of the average number of birds in the second and third sites. The Student’s $T$-test calculated by the formula (6) is 0.0088, the critical value of $T_{kr}$ for the 5% significance level at 11 degrees of freedom is 2.2. Since the value of $T$ is several times less than the critical value $T_{kr}$, it can be argued that the mathematical expectations for the number of birds in the second and third sites are the same.

3.2. Correlation and regression analysis

Methods of correlation and regression analysis are a powerful tool in identifying relationships between different parameters or determining the influence of one factor on another. Let’s consider two samples $X$, $Y$ with the same number of elements $n$. The presence of a linear correlation between them is determined by the correlation coefficient

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - nx_a y_a}{ns_x s_y}, \quad (8)$$

where $s_x = \sqrt{s^2_x}$, $s_y = \sqrt{s^2_y}$ - standard deviations.

The significance of the coefficient (8) is checked using the Student’s $T$-test for $q=n-2$ degrees of freedom:

$$T = \left| r_{xy} \sqrt{n-2} / \sqrt{1 - r^2_{xy}} \right|, \quad (9)$$

If the value of $T$ is greater than the critical value $T_{kr}$ at the chosen significance level $q$, then there is a relationship between the parameters $X$ and $Y$. Otherwise, the dependency is considered not installed. The existence of a correlation dependence allows us to represent it in the form of a linear regression equation

$$y = a + bx, \quad (10)$$

where $a = y_a - r_{xy} x_a s_y / s_x, b = r_{xy} s_y / s_x$. 

6
The coefficients $a$ and $b$ can also be found by the least squares method (LSM). The LSM method allows us to find the regression equation in the case of linear and nonlinear multifactorial models in a more general form

$$y(x_1, x_2, ..., x_n) = a_0 + b_1x_1 + b_2x_2 + ... + b_{12}x_{12} + b_{13}x_{13} + \ldots b_{14}x_{14} + \ldots,$$  (11)

The adequacy of the regression equations (10), (11) is checked using the Fisher criterion, based on a comparison of two variances. One of them $s_y^2 = \sum(y_i - y_a)^2/(n - 1)$ determines the dispersion of the observation results $y_i$ from the average value $y_a$. The second dispersion of adequacy $s_{ad}^2 = \sum(y_i - y_a)^2/(n - 2)$ characterizes the degree of deviation of the recorded data $y_i$ from the values calculated using the regression equation at $x = x_i$:

$$F = s_y^2/s_{ad}^2.$$  (12)

If the value of $F$ is greater than the critical value $F_{kr}$ with a given significance level $q$ with two degrees of freedom equal to $n-1$ for the variance $s_y^2$ and $n-2$ for the variance $s_{ad}^2$, then mathematical models (10), (11) adequately describe the situation under study. Otherwise, the equations are not adequate.

As an example, we will study the possibility of the existence of a correlation dependence between the activity coefficient $K^{(3)}$ on the total number of birds $n^{(3)}$ per 1 km$^2$ based on the results of monitoring in the spring migration season in the third site of the Prymorsk-1 wind farm. The initial data are given in table 4.

**Table 4.** The number of birds $n^{(k)}$ and their activity coefficients $K^{(k)}$ in the third site of the wind farm on different days of observation.

| Days       | 13.03.17 | 14.03.17 | 25.03.17 | 02.04.17 | 03.04.17 | 04.04.17 | 28.04.17 |
|------------|----------|----------|----------|----------|----------|----------|----------|
| $n^{(3)}$  | 32.5     | 39.0     | 108.9    | 85.4     | 23.6     | 52.0     | 63.4     |
| $K^{(3)}$  | 1407.3   | 1634.0   | 3669.0   | 1974.9   | 894.4    | 1879.6   | 2083.8   |

The correlation coefficient $r_{xy}$ and the Student’s $T$-test calculated by formulas (8), (9) are 0.92 and 5.2, respectively. The critical value of $T_{kr}$ at 5 degrees of freedom for the 5% significance level is 2.57. Since the Student’s criterion is greater than the critical value of $T_{kr}$, then the relationship between the statistical samples in table. 4 exists and we can proceed to the construction of the regression equation (10). In the case under consideration, it has the form

$$K^{(3)} = 432.9 + 26.00n^{(3)},$$  (13)

The Fisher criterion calculated by the formula (12) is 5.3, which is greater than the critical value $F_{kr} = 4.95$. The use of the Student’s and Fisher’s criteria in this case allows us to draw the same conclusion: the mathematical model (13) is adequate.

### 3.3. Trend analysis

Let us pay attention to the situation, often encountered in the ornithology, when the measured Student’s criterion is comparable to or somewhat less than its critical value. In this case, it is recommended to use the following methods of processing observations:

- evaluate the significance of the regression coefficient $b$ in equation (10),
- apply the Fisher criterion,
- use the method of the trend analysis of the study of time series analysis.

Time series differ from row statistics in that the values of its elements depend on the previous elements [15]. Nevertheless, the trend analysis algorithm, in our opinion, can be extended to the results of ornithological research with some reservations. Of practical interest, for example, is the assessment of the possibility of changing the number of birds in a given region under the influence of anthropogenic factors over time.

The main task of the trend analysis is to identify the trend of changing one factor as another one changes. The mathematical side of the trend model is also expressed by formulas like (10), (11). When identifying a trend in ornithological research, we single out two components: the main one that determines the development trend, called the trend, and the random one \(e_i\), responsible for the deviation from the trend. The deviation value \(e_i\) is determined by the difference between the measurement result \(y_i\) and the value of functions (10), (11):

\[e_i = y_i - y(x_i)\].

A necessary condition for successful data processing using the trend is the fulfillment of the following requirements [15]:

- \(e_i\) deviations are random;
- different \(e_i\) deviations do not depend on each other, i.e. there is no autocorrelation;
- deviations obey the normal distribution law.

Let us consider the procedure for carrying out the trend analysis. At the first stage, the randomness of deviations is checked. The randomness is tested using turning points. A turning point is the result of a measurement, for which one of the following conditions is met:

\[y_{i-1} < y_i > y_{i+1} \text{ or } y_{i-1} > y_i < y_{i+1},\]  \[\text{(14)}\]

If the number of turning points \(m\) is greater than the critical value \(m_{kr}\), then the deviations are considered random. The critical value for the 95% confidence level is given by the formula

\[m_{kr} = 2 \times (n - 2)/3 - 2\sqrt{(16n - 29)/90},\]  \[\text{(15)}\]

where \(n\) - the number of elements of the row statistics. Checking deviations for autocorrelation is carried out using the Durbin-Watson test \(d\) according to the formula

\[d = \sum_{i=1}^{n-1} (e_i - e_{i+1})^2 / \sum_{i=1}^{n} e_i^2,\]  \[\text{(16)}\]

The calculated value of \(d\) is compared with the tabular values \(d_1\) and \(d_2\). There may be two cases here:

a) The value of \(d\) does not fall within the interval from 2 to 4. For \(d > d_2\), there is no autocorrelation. If \(d < d_1\), then there is a relationship between the deviations. When the condition \(d_1 \leq d \leq d_2\) is fulfilled, it is impossible to make an unambiguous conclusion.

b) The value of \(d\) falls within the range from 2 to 4. In this case, the parameter \(d=4-d\) is determined. Further, the analysis is carried out in accordance with the algorithm given in paragraph (a), where the parameter \(d\) is used instead of \(d\).

Checking the compliance of deviations with the normal distribution law is performed using the Pearson’s \(\chi^2\)-criterion (4). In addition to it, the Westergaard method, the DR criterion and other criteria are used [15].

When the above conditions are met, the trend analysis is started. The initial data from \(n\) measurements are divided into two approximately identical groups. The first group includes the first \(n_1\) measurements, the second - the remaining \(n_2\) elements: \(n = n_1 + n_2\). Let us introduce the notation for the average values and variances of elements in each group:

\[y^{(1)}_i = \frac{y^{(1)}_1}{n_1}, y^{(2)}_i = \frac{y^{(2)}_1}{n_2}, s_1^2 = \frac{(y^{(1)}_1 - y^{(1)})^2}{(n_1 - 1)}, s_2^2 = \frac{(y^{(2)}_1 - y^{(2)})^2}{(n_2 - 1)}\].

First, the hypothesis about the equality of variances in groups is tested using the Fisher criterion
When the value of $F$ is less than the critical value $F_{kr}$ with a given level of significance, the general variances are equal and it is possible to proceed to assessing the presence of the trend using the Student’s $T$-test (6). If the value of $T$ is greater than the critical value of $T_{kr}$, then the trend exists. Otherwise, there is no trend.

As an example, we will study the possibility of a trend in the number of birds $N$ based on the results of $n = 11$ observations in 2010-2011. The initial data are presented in table 5. The equation for the dependence of the number of birds in thousands on the year of observation $x$, obtained by the LSM method, has the form

$$N = 454.5 - 0.223x,$$  \hspace{1cm} (17)

The last line in the table 5 contains the deviations $e_i$ of the observed values of $N_i$ from the values of $N(x_i)$ obtained using the regression equation (3.4). The standard deviation was $s_x = 0.36$.

**Table 5.** The initial data for determining the possibility of existence of a trend in the change in the number of birds $N_i$ in thousands based on the results of $n=11$ observations from 2000 to 2010.

| $x_i$, Year | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|-------------|------|------|------|------|------|------|------|------|------|------|------|
| $i$         | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   |
| $N_i$       | 9.3  | 8.7  | 8.6  | 8.1  | 8.5  | 8.0  | 7.4  | 7.7  | 6.3  | 7.2  | 7.3  |
| $e_i = N_i - N(x_i)$ | 0.30 | -0.08 | 0.041 | -0.24 | 0.09 | 0.11 | -0.27 | 0.25 | -0.92 | 0.20 | 0.52 |

Let us carry out a rejection of possible anomalous values by the Irwin method. The critical value of the Irwin parameter $\lambda_a$ in accordance with the data in table 1 equals about 1.5. The maximum value of the parameter $\lambda_i$ refers to the deviation for number 9 related to measurements in 2008: $\lambda_9 = |y_9 - y_8|/s_x = |-0.92 - 0.25|/0.36 = 3.25$. Since $\lambda_9$ is greater than $\lambda_a$, the result of the measurement in 2008 should be discarded. In all other measurements, the parameter $\lambda_i$ is less than the critical value, so they are saved for further analysis. The corrected data after rejection are given in table 6.

**Table 6.** Adjusted data to determine the possibility of the existence of a trend in the number of birds $N_i$ in thousands based on the results of $n=10$ observations from 2000 to 2010.

| $x_i$, Year | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2009 | 2010 |
|-------------|------|------|------|------|------|------|------|------|------|------|
| $i$         | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| $N_i$       | 9.3  | 8.7  | 8.6  | 8.1  | 8.5  | 8.0  | 7.4  | 7.7  | 7.2  | 7.3  |
| $e_i = N_i - N(x_i)$ | 0.31 | -0.10 | 0.005 | -0.31 | 0.28 | 0.02 | -0.43 | 0.07 | -0.04 | 0.25 |

The regression equation obtained by the LSM method for the dependence of the corrected data has the form
Using the regression equation (18), it can be found that the average annual number of birds in the first 5 years was estimated at the level of 8640 individuals, then in the last five years number of birds in the period from 2000 to 2010 is declining. If the average annual number of birds in the first 5 years was about 8610, and in the last 5 years it was about

\[ N = 397.80 - 0.1944x, \]  

where \( x \) is the year, \( N \) is the number of birds in thousands.

The last line in table 6 contains the \( e_i \) deviations of the observed values of \( N_i \) from the values of \( N(x_i) \) obtained using the regression equation (18). As expected, the standard deviation of the new model has decreased compared to equation (17) from 0.36 to 0.23.

At the first stage, we check the requirement for randomness of deviations by the method of turning points. In accordance with condition (14), the turning points are the results of seven deviations \((m=7)\) with numbers 2-5, 7-9 in table 6. The critical value \( m_{kr} \) for ten measurements \( n=10 \) is 2.92 according to the formula (15). Since \( m \) is less than \( m_{kr} \), the deviations are random.

Next, we check the deviations for autocorrelation. The lower and upper critical limits of the Durbin-Watson test for 1% significance are \( d_1 = 0.32, d_2 = 1.30 \). The Durbin-Watson criterion calculated by the formula (16) is \( d=1.95 \). The value of \( d \) is greater than the upper bound of \( d_2 \), so there is no autocorrelation.

Let us check the correspondence of deviations to the normal distribution law using the Pearson's \( \chi^2 \)-criterion. Let us break down the sample of \( n=10 \) deviations of table 6 on \( m=4 \) intervals of size \( h = (e_{\text{max}} - e_{\text{min}})/m \). From table 6 it follows that the maximum and minimum values of deviations are equal to \( e_{\text{max}} = 0.31, e_{\text{min}} = -0.43 \). Therefore, \( h=0.185 \). The Pearson \( \chi^2 \)-criterion calculated by the formula (4) is 2.36. The critical value \( \chi^2_{kr} \), for the 5% significance level is 3.8. Since the inequality \( \chi^2 < \chi^2_{kr} \) is satisfied, the random variables \( e_k \) obey the normal distribution law.

All the conditions for the trend analysis are met, so we can proceed to the study of the adequacy of the regression equation (18). The correlation coefficient \( r_{xy} \) between the row statistics \( N \) and \( x \) and the Student’s \( T \)-test calculated by formulas (8), (9) for the data of table 3.2 are -0.93 and 7.54 respectively. The critical value of \( T_{xy} \) at 8 degrees of freedom at a 5% significance level is 2.31. Since the Student’s criterion is greater than the critical value of \( T_{kr} \), then the relationship between the statistical samples in table 6 exists.

The Fisher criterion calculated by the formula (12) is equal to 7.20, more than the critical value \( F_{kr} = 4.95 \), which indicates the adequacy of the mathematical model (18). In the case under consideration, the conclusions obtained using the Student’s and Fisher’s criteria coincide.

Fisher’s method is more stringent compared to the Student’s method with small correlation coefficients at the level of 0.5-0.7, which can lead to different results. In this case, it is advisable to conduct additional trend studies. Let us split the data in table 6 into two identical groups. In the first group we will include the first five measurements, in the second - the remaining measurements. Means and variances in each group are \( N^{(1)} = 8.64, N^{(2)} = 7.52, s_{(1)}^2 = 0.15, s_{(2)}^2 = 0.086 \). We will check the homogeneity of the variances of the groups under consideration using the Fisher criterion (7) \( F = 0.15/0.086 = 1.74 \). The critical value of \( F_{kr} \) for the 5% significance level at 4 degrees of freedom of the numerator and 4 degrees of freedom of the denominator is 6.39. Since \( F_{kr} > F \), the variances are homogeneous.

Let us evaluate the presence of the trend using the Student’s \( T \)-test (6). After substituting the average values of deviations and variances of each group into the formula (6), we have: \( T=12.9 \). This value is greater than the critical value \( T_{kr} = 2.31 \) for 5% significance level at \( n_1 + n_2 - 2 = 8 \) degrees of freedom, so the hypothesis of a trend is accepted. Consequently, the number of birds in the period from 2000 to 2010 is declining. If the average annual number of birds in the first 5 years was estimated at the level of 8640 individuals, then in the last 5 years it decreased to 7520 individuals.

Let us compare the results obtained with the help of regression and the trend analyses. Using the regression equation 18, it can be found that the average annual number of birds in the first 5 years of observations was about \( N_{Trend}^{(1)} = 8610 \), and in the last 5 years it was about
$N_{Trend}^{(2)} = 6540$ individuals. The difference between the two methods for assessing the trend is at the level of 0.3

3.4. Development of an algorithm for statistical analysis
Numerous studies of the avifauna on the territory of the wind farms indicate great difficulties in organizing monitoring, collecting information, and in the process of statistical processing of the data obtained [1,5,6,8,9,11,13]. The impossibility of duplicating observations under absolutely identical conditions and the influence of uncontrollable factors associated with meteorological and other conditions are partially compensated by long-term observations up to 1-10 years. However, even in this case there are situations when the measured parameters of the dynamics of ornithocomplexes do not fit into the classical schemes of statistical regularities. In this case, it is necessary to pay special attention to rejecting erroneous measurements, identifying individual anomalous data, and careful checking the adequacy of the obtained mathematical models.

The reliability of conclusions based on the results of statistical analysis depends, first of all, on the correctness of the application of the criteria used. The use of each of them requires the fulfillment of certain conditions. For example, the use of the Student’s, Irwin’s, Durbin -Watson’s, Pearson’s, Fisher’s and other criteria implies the obligatory compliance of the studied sample with the chosen random variable distribution laws. Based on the material presented in this paper and the accumulated experience of statistical studies in a number of publications [1,6,8,11,12,15–18], we can propose the following algorithm for statistical processing of the results of monitoring the wind farm territory.

Step 1. Selection or development of an information system to ensure the storage of observation results in the form of tables or databases, convenient for systematizing the information received and their subsequent analysis. Examples of such information and computer systems are presented in papers [5,7,8].

Step 2. Primary statistical processing of monitoring results, which consists in determining the average values of row statistics, variances, rejecting the results of erroneous measurements and identifying the random variable distribution law.

Step 3. Testing hypotheses about the coincidence of mathematical expectations in a comparative analysis of the number of birds and other parameters that determine their behavior in different monitoring sites and in different seasons.

Step 4. Calculation of correlation coefficients between various parameters that characterize the dynamics of ornithocomplexes depending on the census method, weather, seasonal or other conditions.

Step 5. Correlation-regression analysis. Construction of mathematical models that determine the dependence of one parameter on another parameter.

Step 6. Trend analysis to identify the possibility of changes in the number of birds over time or the direction of change in one parameter as another parameter changes.

The proposed algorithm is focused on processing the results of monitoring the dynamics of ornithocomplexes on the territory of wind farms, obtained by the SNH and RAM methods, although they can also be used in the case of other observation methods.

4. Conclusions
An algorithm for analyzing the results of monitoring the dynamics of ornithological complexes on the territory of wind farms has been developed using statistical methods. The proposed algorithm allows solving the following problems: 1. Carrying out primary processing of information on the number of birds of various species, flight altitude and time spent on observation sites using the Student’s T-test, Fisher and Pearson criteria. 2. Construction of regression equations that determine the dependence of the number of birds on various factors using the Student’s T-test, Fisher and Irwin criteria. 3. Trend detection when studying the
time dependence of the number of birds or the direction of change of one parameter as another parameter changes using turning points and the Student’s $T$-test, Fisher, Irwin and Durbin-Watson tests. The proposed statistical methods have been tested in the analysis of the results of monitoring ornithocomplexes on the territory of the Prymorsk-1 wind farm.

**Acknowledgement**
The authors of this paper are grateful to Anastasia Gorlova (Horlova) for translation into English.

**ORCID iDs**
V V Osadchyi https://orcid.org/0000-0001-5659-4774
V S Yeremieiev https://orcid.org/0000-0002-0131-0049

**References**

[1] Powlesland R 2009 Impacts of wind farms on birds: a review (Publishing Team, Department of Conservation)
[2] Liechti F, Guélat J and Komenda-Zehnder S 2013 Biological Conservation 162 24–32
[3] Bicknell A W, Sheehan E V, Godley B J, Doherty P D and Witt M J 2019 Marine environmental research 147 126–137
[4] Wang S, Wang S and Smith P 2015 Renewable and Sustainable Energy Reviews 44 599–607
[5] Osadchyi V, Siokhin V, Gorlov P, Yeremieiev V and Osadcha K 2019 Eastern-European Journal of Enterprise Technologies 4 29–40
[6] Siokhin V, Osadchyi V, Gorlov P and Yeremieiev V 2021 Identification of “Prymorsk-1” wind power plant impact on the ecological situation connected with the behavior of ornithoфаuna on the Azov Sea coast Journal of Physics: Conference Series vol 1946 (IOP Publishing) p 012007
[7] Annenkov O, Gorlov P, Siokhin V, Salnikova-Budenko I and Siokhin E 2014 Method of using the Web application WebBirds for monitoring seasonal ornithological complexes and computer modeling of wind impact assessment Scientific and methodological bases of protection and assessment of the impact on the environment during the design, construction, operation of wind and solar power plants, power lines: a guide (Bohdan Khmelnytsky Melitopol State Pedagogical University) pp 93–107
[8] Siokhin V, Gorlov P I, Polishchuk I K, Podorozhny S M, Dolinna O M and Salnikova-Budenko I B 2018 Provedennya suchasnych monitorynhovykh doslidzhenn’ ta rozробка ekspertnoho vynozhnu ta naukovoiy dopovidi shchodo vplyvu buduvannya VES na prirodno-ekologichni kompleksy, ornitolohichni kompleksy ta perelitnykh ptakhiv, kakhazhin u mezhakh Prymors’koho rayonu Zaporiz’koi oblasti (Conducting modern monitoring research and developing an expert opinion and scientific report on the impact of wind farm construction on natural environmental complexes, ornithological complexes and migratory birds, bats within the Primorsky district of Zaporozhye region) (Melitopol: Research and Production Enterprise “Ekoresurs i monitorynh”) pp 93–107
[9] Siokhin V D, Chernichko Y I, Sidorenko A I, Gorlov P I and Aleinikova K G 2020 Monitorynh ta okhorona bioriznomantity v Ukrayini : Prykladni aspekti monitorynhu ta okhorony bioriznomantity 16 168–195 URL https://umcg.org.ua/wp-content/uploads/2020/05/T3_WEB_MonOchBioriz_Konference_2.pdf
[10] Arikan K and Turan S L 2017 Fresenius Environmental Bulletin 26 6543–6550
[11] Siokhin V D, Gorlov P I, Polishchuk I K, Podorozhny S M, Dolinna O M and Salnikova-Budenko I B 2017 Substantiation of the impact of the project area of the Primorye wind farm on natural environmental complexes, ornithological complexes and migratory birds, bats, vegetation within the Primorye district of Zaporozhye region (Melitopol: Public Ecological Organization “Lagoon”) pp 93–107
[12] 2017 Recommended bird survey methods to inform impact assessment of onshore windfarms URL https://tinyurl.com/mr46u453
[13] Yeremieiev V, Osadchyi V, Siokhin V and Gorlov P 2020 E3S Web Conf. 166 01003 URL https://doi.org/10.1051/e3scconf/202016601003
[14] Gorlov P I and Siokhin V D 2012 Biologichnyi Visnyk Melitopolskoho Derzhavnoho Pedahohichnoho Universytetu Imeni Bohdana Khmel’ntskoho 2 37–47
[15] Fedosseev V V, Garmash A N, Davit’hegov D M, Orlova I V and Polonovykh V A 1999 Economikomatematicheskije metody i prikladnyje modeli (Economic and Mathematical Methods and Applied Models) (Yunitsi-Dana)
[16] Zhaldak M I, Kuzmina N M and Mikhailin G O 2009 Teorija veroyatnostej i matematicheskaya statistika (Probability theory and mathematical statistics) (Poltava: Environment)
[17] Kremer N S 2004 Teoriya veroyatnostey i matematicheskaya statistika (Theory of Probability and Mathematical Statistics) (Yunit-Dana)

[18] Bevanger, Kjetil and Flagstad, Øystein and Follestad, Arne and Gjershaug, Jan and Halley, Duncan and Hanssen, Frank and May, Roel and Nygård, Torgeir and Pedersen, Hans and Reitan, Ole and Vang, Roald and Johnsen, Lars 2009 Pre- and post-Construction Studies of Conflicts between Birds and Wind Turbines in Coastal Norway (BirdWind) Report NINA Report URL https://www.researchgate.net/publication/239567330_Pre_and_post-Construction_Studies_of_Conflicts_between_Birds_and_Wind_Turbines_in_Coastal_Norway