Analysis of time series characteristics using machine learning model and correlation matrix in the tasks of forecasting the state of forest ecosystems

Pavel Gusev[1], Alexander Tavolzhanskij[1]*, Vladimir Zolnikov[2] and Antonina Deniskina[3]

[1]Voronezh State Technical University, 84, 20-letiya Oktyabrya St., Voronezh, 394006, Russia
[2] Voronezh State University of Forestry and Technologies named after G.F. Morozov, Voronezh, Russia
[3] Moscow Aviation Institute (National Research University)., Moscow, Russia

Abstract. This paper considers the problem of choosing the required set of characteristics for time series forecasting. The method of solving this problem on the basis of correlation matrix is proposed. A correlation matrix is constructed based on the prepared data, after which a list is formed for each variable, ordered by decreasing modulus of the correlation degree. Then linear regression models are trained and the quality of predictions for different sets of variables from the sorted list is compared. Next, a comparison is made for different values of sampling time to determine the optimal value for each variable. To apply the considered algorithm, information from various measuring sensors taking readings of climate variables in the forest lands was used.

1 Introduction

In recent years, time series forecasting has become one of the most urgent issues in the field of data analysis [1-4]. However, despite significant progress in this area, many tasks remain without proper attention from the scientific community. One such problem is identifying the minimum set of characteristics required for time series forecasting.

Pre-defined targets and relevant data are often used for time series forecasting [5]. The choice of indicators and data is based on the process requirements and the actual number of measuring devices. However, this approach has some limitations. One of them is the empirical determination of the required data set, which may degrade the quality of the results when the number of variables is large. In addition, there are cases where the magnitude and behavior of the target indicator cannot be accurately predicted from the available data. Often, in such cases, the use of neural network apparatus is resorted to [6-10].

Forest ecosystems absorb carbon, regulate water balance and maintain biodiversity, which is important for climate stabilization. Nature-based solutions, such as forest restoration, increase their resilience to climate change. Modern digital technologies,

* Corresponding author: tavolzhanskij.a@yandex.ru

including satellite surveillance, GIS and artificial intelligence, improve the monitoring, management and protection of forest resources.

The use of predictive analytics in the control of such systems can help optimize the production process and improve product quality.

This paper proposes a method that allows the identification of a set of characteristics based on a correlation matrix. The application of this method allows to identify the most significant characteristics for predicting the state of forest land climate [11-12] exclude those that do not affect the accuracy of the prediction. In addition to the theoretical justification, the paper investigates the forecast accuracy depending on the number of variables involved in the forecast, as well as the sampling time, taking into account the different frequency of obtaining measurements.

2 Materials and methods

2.1 Defining the problem and designing the experiment

Climate predictive analysis is an important planning step that can predict potential problems and improve the condition of forest lands [11-14]. Predictive analyses can be used to determine the potential risks of fire occurrence as well as to identify the causes of fires.
Let us consider the aspects of the problem formulation under study.

Firstly, it is required to obtain a prediction for a target indicator based on the measurements of the remaining indicators, excluding its own measurements.

Secondly, it is necessary to be able to divide the variables into two sets: those for which forecasting is possible with acceptable accuracy and those for which it is impossible.

Third, for the first set it is necessary to define the attributes required for prediction as well as the optimal conditions for training the model[15-17].

Climate gauge readings are used as input data. The aim of the study is to analyze the magnitude of the mean square error (MSE) under the following conditions [18]:

Under different depth of indirect correlation; Under different number of original variables; Under different number of variables used for prediction; Under different sampling time.

Table 1. Parameters of experiments

Experiment	1	2	3	4
Number of initial variables, n	13	23		
Sampling step, c, δ	5			3, 5, 10
Number of values for a time series row, q	10			
Forecasting timeframe, w	500			
Number of used variables, $\hat{n}$	12, 6	22, 12, 6	22, 12, 8, 6, 5, 3	8, 6, 5
Number of variables for indirect correlation, γ	0, 1, 2		1	
Number of original variables involved in the indirect correlation, ζ	1, 2, 12	1, 2, 22	1	

2.2 Research methodology

Let the initial data be presented in the following form (Table 2):

Table 2. Raw data

	F_1	F_2	…	F_n
DT_1	$V_{1,1}$	$V_{1,2}$	…	$V_{1,n}$
DT_2	$V_{2,1}$	$V_{2,2}$	…	$V_{2,n}$
…	…	…	…	…
DT_m	$V_{m,1}$	$V_{m,2}$	…	$V_{m,n}$

where DT_m is the moment in time, F_n is the indicator, $V_{m,n}$ is the value of the indicator F_n at the time T_m.

We then present the correlation matrix R from Table 2 in the form (1):

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{pmatrix} \quad (1)$$

where r_{ij} are the correlation coefficients.

We then form a new matrix whose elements represent the elements of the matrix R taken modulo (2):

$$\hat{R}_{ij} = \left| R_{ij} \right|, \quad i, j \in [1, n], \quad i, j, n \in \square \quad (2)$$

Subsequently, a list, K, is formed for each variable. This is a sorted column of the $\hat{R}$ matrix in descending order, with the correlation value replaced by the name of the characteristic.

It is recommended that two approaches to variable selection be considered.

1. The list of variables to be predicted is generated by selecting the first k elements from the list K. However, to improve the accuracy of the predictions, it is proposed that the method of determining the set of parameters be used, taking into account the indirect correlation [19-20].

2. The key concept of the method is to supplement the existing list of k parameters with a specific number of γ variables that are indirectly correlated with it. This can be applied to all selected variables, or only to one or two of the most correlated with the target feature. We will refer to the number of such variables as ζ.

Let us consider a target indicator F_x, which we will denote as T. The structure of the time series for training will contain data collected over z time intervals. In each interval, we will consider the number of discrete values q for $\hat{n}$ different indicators, excluding the parameter T. In this case, the target indicator is shifted relative to the current time by w time intervals. The discretization step is defined as follows:

$$DT_{a+1} = DT_a + \delta, \; a \in [k, zk] \quad (3)$$

The transformed structure of the initial form from Table 2 will then take the form shown in Table 3.

Table 3. Obtained data structure

	$1F_1$	$2F_1$	…	KF_1	…	$1F_j$	$2F_j$	…	KF_j	T
DT_k	$V_{1,1}$	$V_{2,1}$	…	$V_{k,1}$	…	$V_{1,j}$	$V_{2,j}$	…	$V_{k,j}$	$V_{kw,1}$
DT_{2k}	$V_{k+1,1}$	$V_{k+2,1}$	…	$V_{2k,1}$	…	$V_{k+1,j}$	$V_{k+2,j}$	…	$V_{2k,j}$	$V_{(k+1)w,1}$
…	…	…	…	…	…	…	…	…	…	…
DT_{zk}	$V_{(z-1)k+1,1}$	$V_{(z-1)k+2,1}$	…	$V_{zk,1}$	…	$V_{(z-1)k+1,j}$	$V_{(z-1)k+2,j}$	…	$V_{zk,j}$	$V_{zkw,1}$

Subsequently, the generated data structure will be employed to train a linear regression model. The linear regression model is a fundamental component of machine learning, providing a solid foundation for more sophisticated models. This approach allows us to estimate the influence of individual factors on the dependent variable, thereby ensuring the interpretability of the results.

The mean square error (MSE) is used as the target function for model training.

To evaluate the precision of forecasts generated under varying conditions, we will utilize MSE, as this approach is more discerning in identifying significant discrepancies.

3 Results and analysis

Forest land climate condition gauges provide the basis for the input data. The data set comprises records of measurements taken over a nine-day period.

The first day's data are used to train the model, after which the model is tested on all other readings. The resulting MSE error is then averaged across all nine days.

The initial experiment will examine the MSE at varying levels of indirect correlation depth. This will enable us to assess the impact of this parameter on the modelling accuracy (Table 4).

Table 4. Results of experiment 1 (MSE)

Exp.	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
	12UID	6UID	For one var. $\zeta = 1$		For two var. $\zeta = 2$		For every var. $\zeta = 12$	
UID / γ	0	0	1	2	1	2	1	2
1	17.7125	11.8997	11.9143	11.9711	11.9711	11.3495	11.9711	11.3495
2	2.043 E-05	3.479 E-05	3.431 E-05	3.488 E-05	3.488E-05	3.494 E-05	3.488 E-05	3.494 E-05
3	0.1962	0.2132	0.2265	0.2267	0.2267	0.2255	0.2153	0.2255
4	4.465 E-04	4.461 E-04	4.709 E-04	4.564 E-04	4.629 E-04	5.135 E-04	4.567 E-04	5.135 E-04
5	2.727 E-04	3.637 E-04	3.639 E-04	4.394 E-04	4.394 E-04	3.714 E-04	4.394 E-04	3.714 E-04
6	164.523	164.433	164.634	159.876	165.290	158.925	158.456	158.925
7	36.5125	36.4552	36.4552	36.5194	35.7301	44.9435	35.9917	44.9435
8	1.193 E-02	9.662 E-03	9.662 E-03	1.129 E-02	1.108 E-02	1.185 E-02	1.323 E-02	1.185 E-02
9	1.034 E-03	1.031 E-03	1.133 E-03	1.063 E-03	1.131 E-03	9.568 E-04	1.131 E-03	9.568 E-04
10	2.052 E-01	2.131 E-01	2.065 E-01	2.026 E-01	2.026 E-01	2.529 E-01	2.005 E-01	2.529 E-01
11	12.3346	12.3235	12.3654	12.4390	12.4390	7.2828	12.2481	7.2828
12	1.8232	1.5104	1.4188	1.6518	1.4365	1.5763	1.6925	1.5763
13	464.901 5	456.235 7	475.188 7	478.188 8	478.347 2	498.584 5	478.347 2	498.584 5

In Tables 4-7, the cell highlighted in green represents the minimum error for the given trait, while the cell highlighted in red represents the maximum error.

The results of the initial experiment demonstrate that for the majority of UIDs, the error rate is lower than that observed in the benchmark experiment 1.1. Furthermore, it can be observed that as the number of variables and the number of ζ values employed are augmented, the error magnitude is concomitantly increased. The identical outcomes for $\zeta = [2, 12]$ can be attributed to the limited number of variables. It is notable that the lists of selected features are consistent across some UIDs.

The following experiment was conducted with the number of initial variables increased from 13 to 23 (Table 5). The results were compared using all 22 UID variables, half of the 12 UID variables, and six UIDs. This allowed for comparison with the first experiment and with the remaining cases within the second experiment.

Table 5. Results of experiment 2 (MSE)

Exp.	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
	22UID	12UID	6UID	For one var. $\zeta = 1$		For two var. $\zeta = 2$		For every var. $\zeta = 22$	
UID / γ	0	0	0	1	2	1	2	1	2
1	21.925	19.480	29.343	29.343	29.343	29.343	21.82	30.1484	44.279
2	3.27 E-05	3.61 E-05	2.74 E-05	2.73 E-05	2.73 E-05	2.74 E-05	2.86 E-05	2.78 E-05	1.93 E-04
3	0.1729	0.1722	0.1689	0.1689	0.1702	0.1689	0.171	0.1695	0.1711
4	4E-4	3E-04	4E-04	4E-04	4.1E-4	4E-04	4E-04	4E-04	4E-04
5	2.09 E-04	3.81 E-04	2.22 E-04	2.22 E-04	2.22 E-04	2.22 E-04	1.58 E-04	1.65 E-04	1.58 E-04
6	140.35	146.43	137.72	142.71	144.56	142.65	148.8	145.861	148.8
7	33.702	35.897	36.693	36.144	38.217	40.183	37.56	39.41	40.9
8	0.0087	0.0089	0.0084	0.0084	0.0085	0.0084	0.009	0.008	0.0076
9	1.12 E-03	1.37 E-03	1.33 E-03	1.33 E-03	9.47 E-04	9.50 E-04	8.33 E-04	9.97 E-04	7.26 E-04
10	0.5232	0.3554	0.4905	0.4905	0.4416	0.4416	0.451	0.4143	0.4510
11	7.4523	10.193	6.0139	6.0139	5.0011	6.4733	2.040	6.4941	3.2699
12	13.427	9.2192	0.8823	10.104	8.4208	8.4208	7.252	12.3142	7.2524
13	434.08	446.43	409.82	409.82	435.35	439.53	432.1	445.601	432.17
14	4.4989	5.7393	5.4793	5.4793	5.1002	5.4793	5.294	5.4793	4.1589
15	2.0182	2.2962	1.2002	1.2002	1.2002	1.2002	1.041 0	1.0410	0.8727
16	0.0031	0.0034	0.0025	0.0025	0.0025	0.0025	0.002	0.0022	0.0014
17	7.27 E-04	7.66 E-04	7.15 E-04	7.72 E-04	7.67 E-04	7.67 E-04	7.67 E-04	8.32 E-04	7.86 E-04
18	0.0263	0.0253	0.0310	0.0310	0.0309	0.0309	0.053	0.0534	0.0286
19	6.94E-06	6.74E-06	3.46E-06	3.46E-06	3.53E-06	3.53E-06	3.97E-06	3.53E-06	3.97E-06
20	69.975	58.111	34.565	34.565	34.565	34.565	49.52	34.5650	12.449
21	8.0063	5.5163	13.646	13.646	13.597	13.597	13.59	14.6681	13.597
22	0.2487	0.2447	0.1919	0.2775	0.2831	0.2831	0.282	0.2981	0.2465
23	329.10 13	310.80 12	860.37 34	860.37 34	845.37 71	836.60 39	446.1 616	836.603 9	435.13 62

The second experiment demonstrates that utilizing all variables in conjunction with the simple exclusion of half of the features yields inferior outcomes in comparison to the proposed method (Experiments 2.1 and 2.2). It can be concluded that the favorable outcomes of experiments 1.1 and 1.2 are attributable to the limited number of initial variables.

To further examine this phenomenon, the parameters of Experiment 2.4, specifically $\zeta = 1$, will be utilized to investigate the impact of varying the number of utilized features, $\hat{n}$, on the error (Table 6).

Table 6. Results of experiment 3 (MSE)

Exp.	3.1	3.2	3.3	3.4	3.5	3.6
UID / $\hat{n}$	22UID	12UID	8UID	6UID	5UID	3UID
1	21.9251	19.4801	18.6285	29.3438	29.3438	38.5931
2	3.27E-05	3.61E-05	3.64E-05	2.74E-05	2.74E-05	3.07E-05
3	0.1729	0.1722	0.1806	0.1689	0.1702	0.1707
4	4.05E-04	3.65E-04	3.89E-04	4.00E-04	4.16E-04	4.89E-04
5	2.09E-04	3.81E-04	3.69E-04	2.22E-04	2.22E-04	1.58E-04
6	140.3576	146.4357	142.5102	142.7147	142.6531	145.2973
7	33.7028	35.8976	36.0262	36.1448	39.6931	39.1850
8	0.0087	0.0090	0.0088	0.0084	0.0085	0.0086
9	1.12E-03	1.37E-03	1.35E-03	1.33E-03	9.95E-04	9.71E-04
10	0.5232	0.3554	0.4962	0.4905	0.4416	0.4279
11	7.4523	10.1934	6.6301	6.0139	6.0139	6.7104
12	13.4273	9.2192	8.9920	10.1041	8.4208	5.5121
13	434.0808	446.4379	408.8020	409.8250	439.5382	440.5795
14	4.4989	5.7393	5.6559	5.4793	5.4793	6.0817
15	2.0182	2.2962	1.2118	1.2002	1.2002	2.1915
16	0.0031	0.0034	0.0023	0.0025	0.0025	0.0059
17	7.27E-04	7.66E-04	7.59E-04	7.72E-04	7.67E-04	8.37E-04
18	0.0263	0.0253	0.0265	0.0310	0.0309	0.0562
19	6.94E-06	6.74E-06	3.38E-06	3.46E-06	3.53E-06	2.70E-06
20	69.9759	58.1114	32.4097	34.5650	34.5650	47.7579
21	8.0063	5.5163	12.1952	13.6467	13.5971	15.9503
22	0.2487	0.2447	0.2870	0.2775	0.2831	0.5510
23	329.1013	310.8012	325.7551	860.3734	874.1462	128.6250

The results of Experiment 3.2 demonstrate an improvement in comparison to those of Experiment 2. A reduction in the quantity of input data to an excessive degree has a detrimental impact on the accuracy of the resulting prediction.

The results for 3.3–3.6 are of a superior quality to those of other studies. The parameters within this range will be subjected to further investigation at varying sampling intervals from the current time point.

Given the disparate sampling rates of the features, which range from approximately 1.5 to 20 minutes, an experiment with the following sampling intervals is planned: The sampling intervals were 3, 5, and 10 seconds (Table 7).

Table 7. Results of experiment 4 (MSE)

Exp.	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
δ, c	3			5			10		
$\hat{n}$ / UID	8UID	6UID	5UID	8UID	6UID	5UID	8UID	6UID	5UID
1	21.748	32.001	32.001	18.628	29.343	29.343	13.858	30.5874	30.5874
2	3.21 E-05	2.67 E-05	2.67 E-05	3.64 E-05	2.74 E-05	2.74 E-05	8.91 E-05	4.19 E-05	4.19 E-05
3	0.2366	0.2395	0.2279	0.1806	0.1689	0.1702	0.4516	0.4394	0.4408
4	3.86 E-04	3.84 E-04	3.84 E-04	3.89 E-04	4.00 E-04	4.16 E-04	3.96 E-04	3.85 E-04	3.88 E-04

5	2.51 E-04	2.52 E-04	1.72 E-04	3.69 E-04	2.22 E-04	2.22 E-04	6.7 2E-04	4.30 E-04	2.81 E-04
6	144.80	144.26	144.22	142.51	142.71	142.65	150.69	149.53	149.61
7	35.857	35.430	35.381	36.026	36.144	39.693	38.960	39.5227	38.7386
8	0.0091	0.0099	0.0082	0.0088	0.0084	0.0085	0.0133	0.0133	0.0109
9	1.12 E-03	1.05 E-03	1.06 E-03	1.35 E-03	1.33 E-03	9.95 E-04	9.92 E-04	1.03 E-03	1.89 E-03
10	0.4890	0.4822	0.4063	0.4962	0.4905	0.4416	0.5831	0.5492	0.5401
11	5.3425	5.3884	5.3884	6.6301	6.0139	6.0139	9.6795	9.2327	9.2327
12	0.7874	0.8137	1.5397	8.9920	10.104	8.4208	17.985	25.5719	20.9965
13	381.64	385.47	419.47	408.80	409.82	439.53	568.61	573.606	522.766
14	5.6690	5.8360	5.8360	5.6559	5.4793	5.4793	7.0997	5.6031	5.6031
15	1.2475	1.2171	1.2171	1.2118	1.2002	1.2002	2.2430	2.2641	2.2641
16	0.0015	0.0016	0.0016	0.0023	0.0025	0.0025	0.0052	0.0054	0.0054
17	9.51 E-04	8.75 E-04	8.61 E-04	7.59 E-04	7.72 E-04	7.67 E-04	1.40 E-03	1.50 E-03	8.83 E-04
18	0.0288	0.0352	0.0351	0.0265	0.0310	0.0309	0.0256	0.0230	0.0231
19	3.80 E-06	3.79 E-06	3.60 E-06	3.38 E-06	3.46 E-06	3.53 E-06	3.19 E-06	3.07 E-06	3.02 E-06
20	19.625	21.385	21.385	32.409	34.565	34.565	3.9206	2.6399	2.6399
21	11.588	12.638	14.557	12.195	13.646	13.597	16.712	17.7916	11.8957
22	0.5905	0.5663	0.5740	0.2870	0.2775	0.2831	1.2787	1.2741	1.2793
23	289.97 19	554.87 95	555.57 68	325.75 51	860.37 34	874.14 62	703.88 43	1023.97 10	1030.60 59

The fourth experiment demonstrated that for a considerable number of variables, reducing the time interval between measurements resulted in enhanced data quality. Furthermore, in certain instances, an increase in the measurement interval to 10 seconds was observed to correlate with enhanced prediction accuracy. This suggests that modifying the sampling interval can have a considerable impact on the ultimate outcome.

In order to conduct a comparative analysis of the effectiveness of the method, a summary table has been constructed, which presents the best results without the use of the proposed method (experiments 2.1-2.3) and with the use of the proposed method (experiments 2.4-4.9) (Table 8).

Table 8. Summary of results (MSE)

UID	Without method	With method	Difference
1	19.4801	13.858	-28.86%
2	2.74E-05	2.67E-05	-2.55%
3	0.1689	0.1689	0%
4	3.65E-04	3.65E-04	0%
5	2.09E-04	1.58E-04	-24.40%

6	137.7209	142.5102	3.48%
7	33.7028	35.3815	4.98%
8	0.0084	0.0076	-9.52%
9	0.00112	0.000726	-35.18%
10	0.3554	0.3554	0%
11	6.0139	2.0407	-66.07%
12	0.8823	0.7874	-10.76%
13	409.825	381.6416	-6.88%
14	4.4989	4.1589	-7.56%
15	1.2002	0.8727	-27.29%
16	0.0025	0.0014	-44.00%
17	0.000715	0.000759	6.15%
18	0.0253	0.023	-9.09%
19	3.46E-06	0.0000027	-21.97%
20	34.565	2.6399	-92.36%
21	5.5163	5.5163	0%
22	0.1919	0.2447	27.51%
23	310.8012	128.625	-58.62%

4 Conclusions

The analysis of the data presented in the summary table illustrates that the proposed methodology improves the quality of forecasts for most variables. However, for some attributes there is both preservation of the previous results and decrease in forecasting accuracy. This may be due to the peculiarities of the data structure, the specifics of the processes under study, or other factors that require additional study. It is important to note that despite these exceptions, in general, the proposed methodology shows its effectiveness and can be recommended for use in similar researches.

References

1. V.Z. Kovalev, S.Yu. Shvetsov, O.V. Arkhipova, Inzenernyj vestnik Dona **100(4)**, 127-141 (2023)
2. O.G. Neizvestny, Bulletin of Voronezh State Technical University **19(6)**, 32-40 (2023)
3. A. Shehadeh, O. Alshboul, R. E. Al Mamlook, O. Hamedat, Automation in Construction **129(2),** 103827 (2021)
4. S.L. Semakov, I.S. Semakov, Trudy MAI **(100)**, 25 (2018)
5. I.A. Sedykh, I.V. Strugov, Bulletin of Voronezh State Technical University **(19)**, 72-78 (2023)
6. A.A. Chernikov, Trudy MAI **(129)**, 26 (2023)
7. E. Chigrinets, A. Verchenko, Trudy MAI **(104)**, 22 (2019)
8. A. Manoharan, K. M. Begam, V. R. Aparow, D. Sooriamoorthy, Journal of Energy Storage **(55)**, 105384 (2022)
9. G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, An introduction to statistical learning: With applications in python. (Springer Nature. 2023)
10. N.V. Rumyantsev, S.V. Solovyov, D.V. Pavlov, Trudy MAI **(136)** (2024)
11. M. I. Ashraf, F. R. Meng, C. P. A. Bourque, D. A. MacLean, PloS one **10(7)**, e0132066 (2015)

12. L. Chen, B. Han, X. Wang, J. Zhao, W. Yang, Z. Yang, Applied Sciences **13(21)**, 12019 (2023)
13. L. J. Slater, L. Arnal, M. A. Boucher, A. Y. Y. Chang, S. Moulds, C. Murphy,... M. Zappa, Hydrology and earth system sciences **27(9)**, 1865-1889 (2023)
14. M. Avand, H. Moradi, Journal of Hydrology **595**, 125663 (2021)
15. P.Yu. Gusev, A.V. Tavolzhanskij, *Instrumenty predikativnoy analitiki [Predicate analytics tools].* Collection of scientific articles of the 12th International Youth Scientific Conference, ZAO "Universitetskaya kniga " pp. 96-99 (2023)
16. P.Yu. Gusev, A.V. Tavolzhanskij, Bulletin of Voronezh State Technical University **19(2)**, 14-19 (2024)
17. P.Yu. Gusev, A.V. Tavolzhanskij, Anthropocentric Sciences in education: challenges, transformations, resources 396-400 (2024)
18. J. Qi, J. Du, S. M. Siniscalchi, X. Ma, & C. H. Lee, IEEE Signal Processing Letters **(27)**, 1485-1489 (2020)
19. E.R. Rizvanova, Izvestiya St. Petersburg State University of Economics **108(6)**, 159-163 (2017)
20. V.L. Egoshin, S.V.Ivanov, N.V. Savvina, A.R. Ermolaev, S.A. Mamyrbekova, L.M. Zhamaliyeva, A.M. Grjibovski, Human ecology **12**, 55-64 (2018)