Tourist and recreational complex elements dynamics research data analysis qualitative methods

A M Kumratova¹, E V Popova¹, L O Velikanova¹, I I Vasilenko¹ and M I Popova¹

¹ Kuban State Agrarian University, 30, Kalinin street, Krasnodar, 350089, Russia
E-mail: kumratova.a@edu.kubsau.ru

Abstract. This article presents the results of the adapted complex methodology operation for the pre-estimating analysis of the dynamics of tourist flow time series decomposition in the Dombai ski village, the features of which are in the combined use of both classical and new "nonlinear" statistics. The methods proposed and tested by the authors are presented in the form of a pre-estimating model for assessing the tourist flow time series trend stability. The following methods of nonlinear dynamics were tested: the method of the normalized Hurst range, phase-plane analysis. The methods of fractal analysis used and adapted by the authors ensure the identification and assessment of a number of socio-economic time series fundamental qualitative and quantitative pre-estimating characteristics, namely, the presence of memory, including long-term memory, its depth, which in turn can determine the process as persistent (antipersistent, trend-resistant or reverse) to reveal the noise color.

1. Introduction
The relevance of this research is due to the state need to create favorable conditions for the tourism market dynamic development. In accordance with government decree No. 434 of 06/04/2020, tourism has become one of the sectors of the Russian economy most affected by the coronavirus. The website of the Federal Agency for Tourism of the Russian Federation presents current and adopted support measures for small and medium-sized businesses in the tourism industry. In the context of the epidemiological situation in the world, much attention is paid to the development of domestic tourism, in connection with which the tourist flow size quantitative forecast will make it possible to qualitatively solve the issues of tourist and recreational activities management, for example, in planning the occupancy of beds.

2. Materials and methods
Using the analysis, generalized conclusions and results description, the researcher forms "structured" information about the time series. Namely this structured knowledge allows the analyst to develop sound predictive analysis methods [1, 2]. The article uses the values of the tourist flow decomposed daily time series in the Dombay ski village by days of the week for the period from 2015-2017 as the calculation base. Let us note the fact that the tourist flow volume quantitative data is the basic indicator of the country, region and individual organization’s tourist and recreational activity. The authors investigated: the initial time series (TS) of the daily tourist flow, and time series decomposition (by days of the week, separately the weekends time series (Saturday, Sunday), weekdays). The practical importance of studying separately the weekends time series is due to the fact...
that the majority of tourists are the neighboring regions for the Karachay-Cherkess Republic: Stavropol Territory, Rostov Region, Krasnodar Territory.

Separately, we note that for conclusions about the quantitative measurement of the any nature process dynamics, the research of averaged (typical) values is not effective [1, 3].

In classical statistics, three main indicators calculated for the temporal sequence of data: kurtosis, asymmetry, and variation, should be distinguished. It is believed that these three coefficients act as a triad for assessing the trend stability of the process [1, 4]. The tourist flow time series different decompositions summary research results are presented in the Table 1. Based on the time series decomposition by days of the week initial data graphical presentation visualization, the authors proposed a mechanism for carrying out the data normalization procedure for the further procedure for obtaining pre-forecasting information: isolated cases of outliers in the “Monday”, “Thursday”, “Friday” time series were removed. It can be noted that outliers correspond to single event components and create a "blurry" image in dynamics.

Nobel laureate G. Markowitz defined two main risk indicators: mathematical expectation and variance or standard deviation. In later works, asymmetry coefficients were also defined as indicators of the risk measure

$$A = \sigma^{-3} \sum_{s=1}^{n} (W_s - M)^3 P_s$$

and excess kurtosis

$$E = \sigma^{-4} \sum_{s=1}^{n} (W_s - M)^4 P_s$$

where $P_s$ is the probability (relative frequency) of the random variable value $W_s$, $1 \leq s \leq n$.

3. Investigation of decomposition time series by methods of classical statistics and nonlinear dynamics

The analysis of the calculated risk indicators (Table 1) allows to draw a number of conclusions:
- for time series decomposition on weekdays, on average, tourists are expected to arrive in the amount of 2-3 excursion buses. A sharp increase in the tourist flow is observed on weekends, which is confirmed by practice;
- with relatively equal values of the mathematical expectation of "Thursday" and "Friday" time series there is a significant difference in the variance index. This is a consequence of the difference in the form of their distribution;
- the coefficient of variation for all investigated time series $\nu > 0.33$. The latter means that the degree of data dispersion is considered significant;
- for all considered time series, the asymmetry index is higher than 0.5, i.e. there is a significant asymmetry in data distribution;
- the values of the kurtosis index, which are highlighted in the table, exceeding the value 3, which characterizes the "normality" of the distribution, for the normalized "Monday" and "Friday" time series, as well as for the "Tuesday" and "Wednesday" time series, confirm the fact that their probability density functions do not obey Gauss's law;
- the advantage of the median is that it is not affected by outliers. A significant range in the values of the mathematical expectation and the median confirms the fact that the investigated time series do not obey the normal distribution law.

Considering the presence of "heavy tails" in the initial time series and time series decomposition (Monday, Tuesday, Wednesday, Friday), it is proposed to use the apparatus of nonlinear dynamics methods to identify pre-forecasting characteristics and calculate predicted values not only with long (large) sampling, but also with short time series [1, 4].

A detailed description of the algorithm for the operation of nonlinear dynamics methods: R/S-analysis, phase-plane analysis, linear cellular automaton is presented in the sources [1, 2, 4].
Table 1. Statistical indicators for decomposed tourist flows time series.

| Days of week | MX     | DX       | σ   | V   | A    | E    | Median |
|--------------|--------|----------|-----|-----|------|------|--------|
| Monday       | 63.16  | 2082.4   | 45.63 | 0.72 | 2.2  | 6.4  | 49.5   |
| Tuesday      | 61.4   | 2154.2   | 46.4 | 0.75 | 2.03 | 5.6  | 48     |
| Wednesday    | 57.4   | 1184.5   | 34.4 | 0.6  | 1.8  | 5.21 | 48.5   |
| Thursday     | 91.8   | 5098.5   | 71.4 | 0.77 | 1.6  | 2.28 | 73     |
| Friday       | 92.57  | 6962.9   | 83.4 | 0.9  | 2.24 | 5.6  | 69     |
| Saturday     | 136    | 10116.2  | 100.6 | 0.74 | 1.45 | 2.43 | 108    |
| Sunday       | 207.3  | 30179.8  | 173.7 | 0.84 | 1.33 | 1.39 | 151    |
| Weekdays     | 370.8  | 29158.8  | 170.76 | 0.46 | 1.18 | 1.63 | 333    |
| Weekends     | 344.7  | 43774    | 209.2 | 0.61 | 0.91 | 0.25 | 291.5  |
| Initial time serial «Tourists» | 1021.1 | 11195.5 | 105.81 | 1.03 | 2.74 | 9.65 | 67     |

Table 2 presents the results of the calculated indicators obtained on the basis of nonlinear dynamics methods for the initial and decomposed tourist flows time series.

Table 2. The results of the calculated indicators obtained on the basis of nonlinear dynamics methods for the decomposed tourist flow time series.

| Pre-forecasting analysis | R/S-analysis | Phase-plane analysis |
|--------------------------|--------------|----------------------|
| Indices                  | Breakdown point | Regression equation | Hurst exponent | Length of quasi-cycles with the highest frequency | Frequency of quasi-cycles length |
| Day of week              |               |                      |                |                                           |                                  |
| Monday                   | 5             | \( y = 0.64 \cdot x - 0.42 \) | 0.64           | 5                                           | 10                               |
| Tuesday                  | 5             | \( y = 0.67 \cdot x - 0.46 \) | 0.67           | 5                                           | 7                                |
| Wednesday                | 7             | \( y = 0.62 \cdot x - 0.4 \)  | 0.62           | 5                                           | 7                                |
| Thursday                 | 6             | \( y = 0.65 \cdot x - 0.43 \) | 0.65           | 4.5                                         | 8                                |
| Friday                   | 6             | \( y = 0.65 \cdot x - 0.43 \) | 0.65           | 5                                           | 9                                |
| Saturday                 | 5             | \( y = 0.65 \cdot x - 0.43 \) | 0.65           | 4                                           | 9                                |
| Sunday                   | 5             | \( y = 0.65 \cdot x - 0.42 \) | 0.65           | 5                                           | 10                               |
| Weekdays                 |               | \( y = 0.64 \cdot x - 0.42 \) |                | 4                                           | 7                                |
| Weekends                 | 5             | \( y = 0.61 \cdot x - 0.39 \) | 0.61           | 5                                           | 10                               |
| Initial time serial «Tourists» | 6            | \( y = 0.63 \cdot x - 0.41 \) | 0.63           | 5                                           | 46                               |

The analysis of the calculated data in Table 2 allows to draw up the following conclusions:
- for each of the studied time series, the fifth or sixth point of breakdown from the R/S-trajectory is mainly fixed. On average, the duration of 5–6 weeks characterizes a month for time series decomposition, which in turn provides pre-forecasting information for determining the memory depth of the time series;
- for all time series the value of the Hurst exponent belongs to the zone of "gray" noise. The latter means that the data series do not follow random walks and have sufficient trend stability [1];
- the regression equations are of the same type for all studied time series, this conclusion allows to analyze the coefficient \( a \), which value ranges from 0.61 to 0.67, i.e. the entire investigated time series model has the property of "fractality";
- the lengths of the quasi-cycles with the highest frequency are 4 and 5, respectively, which is equal to the calendar month for time series decomposition by days of the week and the number of weekdays for the initial time series “Tourists”, for example, the quasi-cycles lengths frequencies histogram for "Friday" time series is shown in Figure 1 c);
- one of the pre-predictive characteristics is the quasi-cycles length highest frequency, which characterizes the "const" property of the process under study. "Tourists" time series has the highest value of this indicator, which means the presence of frequently encountered 5-day quasi-cycles. In this context, we can talk about the presence of the process trend stability property under study. For decompositional time series, the highest frequency is 10 (weeks), which on average will make it possible to make a medium-term forecast for 2 months.

Figure 1 shows the author's "Methods of nonlinear dynamics" work out, which contains the above algorithms for methods of nonlinear dynamics.

4. Conclusion
When carrying out a comparative analysis of classical forecasting methods and methods of nonlinear dynamics, it can be concluded that for volatile time series, additional research is required in the context of forming formations (aggregation, increment). Using the mechanism of listed algorithms operation, the synergistic effect obtained from the study of complex socio-economic processes in the context of the triad results is important: the initial time series, a number of its increments and aggregated data. The results of the pre-forecasting analysis are needed to select or develop a predictive model. The accuracy of the quantitative forecast determines the sequence of data, which allows to identify the possibility of the next value appearance in time.

Approbation of the author's software package was carried out on the basis of real data on the tourist flow volume of the. Quantitative values and qualitative characteristics of the pre-forecasting information were obtained for the tourist and recreational activities modeling lower level values, which in turn are input data for the predictive models of the tourist and recreational activities management upper level (planning the occupancy of beds), which becomes a particularly important fact in domestic tourism development conditions in the existing framework of the epidemiological situation.

5. Acknowledgment
The reported study was funded by RFBR according to the research project № 19-010-00134 A.
References
[1] Peters E 2000 Chaos and order in the capital markets Mir 333 p
[2] Kumratova A, Popova E, Costa L and Zamotajlova D 2019 Methods of nonlinear dynamics as a hybrid tool for predictive analysis and research of risk-extreme levels International Journal of Hybrid Intelligent Systems 15(4) 221-241
[3] Krichevsky M L 2005 Intellectual methods in management St. Petersburg: Peter 305 p
[4] Kumratova A Forecasting development of economic processes using adapted nonlinear dynamics methods / Kumratova A., Popova E., Temirova L., Shaposhnikova O. // International Journal of Engineering and Advanced Technology, 2019. № 9(1), c. 3082-3085.