UDC 004.032.26

Investigation of Parameters of Meteorological Models Based on Patterns

Andrey K. Gorshenin

Federal Research Center "Computer Science and Control" of RAS
44-2 Vavilov str., Moscow, 119333, Russian Federation

Email: agorshenin@frccsc.ru

The probabilistic characteristics and the forecasts for precipitation on the basis of a special transformation of the initial data, which makes it possible to reveal patterns in observations, are briefly discussed. Patterns in data analysis can be used to improve the accuracy and speed of forecasting. Moreover, pattern's methodology is a convenient approach to the solution of various climatological problems. The issues of testing the Markov property of data, probabilistic and neural network forecasting for statistical observations without involvement of any additional information about meteorological conditions are investigated. The initial data is volumes of daily precipitation observed during 60 years. The best accuracy for neural networks trained on patterns based on sequences of «D» (dry days, i.e. without precipitations) and «W» (wet ones, i.e. with any nonzero volume) is 97% for one-day and 89% for two-day forecasts. Few directions for further investigations are suggested. The paper continues the author's research in the fields of creation of mathematical models and data mining algorithms for meteorological observations.

Key words and phrases: precipitation, patterns, forecast, neural networks, deep learning, probabilistic forecasting.

In: K. E. Samouylov, L. A. Sevastianov, D. S. Kulyabov (eds.): Selected Papers of the VIII Conference "Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems", Moscow, Russia, 20-Apr-2018, published at http://ceur-ws.org

1. Introduction

Precipitation is an important parameter for meteorological models (see, for example, papers [1–3]), so the development of an adequate mathematical models (including probabilistic and statistical) and the creation of software tools for processing a significant amount of accumulated observations using modern methods are in demand. In this case, probabilistic approaches can be used to solve forecasting problems (see, for example, paper [4]) as well as neural networks that are very effective in a wide range of application areas (see, for example, papers [5–7]). Moreover, at present, the research of various precipitation processes in the context of global warming and climate change problems is quite popular (see, for example, [8–11]).

In this paper, the probabilistic characteristics and the forecasts for precipitation on the basis of a special transformation of the initial data, which makes it possible to reveal patterns in observations, are briefly discussed. Patterns in data analysis can be used to improve the accuracy and speed of forecasting. Moreover, pattern's methodology is a fairly common tool in the solution of various climatological problems. This paper continues the previous author's research in the fields of creation mathematical models and data mining algorithms for meteorological observations [12–15].

2. Investigation of the probabilistic characteristics of precipitation based on patterns

The volumes of daily precipitation observed during 60 years in Potsdam are the initial data. Let's consider transformation of non-negative data V_{daily} according to the following rule: if any positive value was observed in the i-th day, it is replaced by one ($\widetilde{V}^{(i)}_{\text{daily}} = 1$), otherwise the value of $\widetilde{V}^{(i)}_{\text{daily}}$ equals zero. Thus, the initial series consisting of continuous values becomes discrete, taking two possible values $\{0, 1\}$. This simplification makes it possible to analyze the presence or absence of precipitation irrespective of their volume. Thus, any sequence of dry (without precipitations) and wet (with any nonzero volume) days can be represented as a «0–1» (or «D–W») chain (the pattern).

For each pattern within the historical data it is possible to determine the frequencies of appearance as the ratio of the number of such sets of fixed length N to the total number of possible chains (obviously 2^N). In fact, these are the probabilities according to the classical definition. Within the framework of the research, observations for 60 years for Potsdam were analyzed for the values of the parameter N from 1 to 14. For each set, the frequencies (probabilities) were obtained, the pattern with a maximum value was determined [12]. Fig. 1 demonstrates an example of frequencies for patterns of length $N = 5$. The numerical values of the corresponding probabilities are indicated in Table 1.

In the most of the papers devoted to the statistical analysis of meteorological data, it is assumed that the duration of the period of precipitation, measured in days (that is, the number of successive wet days), has the geometric distribution. Perhaps, these assumptions are based on the classical interpretation of the geometric distribution in terms of Bernoulli's tests as the distribution of the number of successive wet days («success») to the first day without precipitation («failure»). With the use of patterns, it was demonstrated [12] that the sequence of wet and dry days is not even Markovian, so using Bernoulli's scheme based on independence of data is incorrect. Alternative probabilistic models are proposed in the papers [14, 15].

3. Forecasts for precipitation based on patterns

Using the data on frequencies (probabilities) for patterns, it is possible to calculate the values of the conditional probability of occurrence in the future of certain combinations,

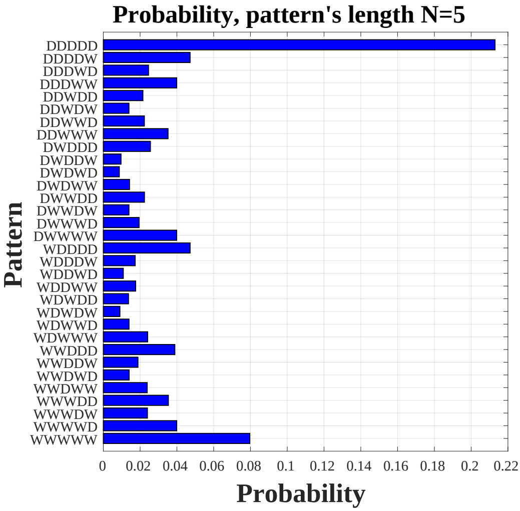

Figure 1. Probabilities/frequencies for patterns with length that equals 5

Table 1

Probabilities/frequencies for patterns with length that equals 5

Pattern	**Probability**	**Pattern**	**Probability**
DDDDD	0,21	WDDDD	0,05
DDDDW	0,05	WDDDW	0,02
DDDWD	0,02	WDDWD	0,01
DDDWW	0,04	WDDWW	0,02
DDWDD	0,02	WDWDD	0,01
DDWDW	0,01	WDWDW	0,01
DDWWD	0,02	WDWWD	0,01
DDWWW	0,04	WDWWW	0,02
DWDDD	0,03	WWDDD	0,04
DWDDW	0,01	WWDDW	0,02
DWDWD	0,01	WWDWD	0,01
DWDWW	0,01	WWDWW	0,02
DWWDD	0,02	WWWDD	0,04
DWWDW	0,01	WWWDW	0,02
DWWWD	0,02	WWWWD	0,04
DWWWW	0,04	WWWWW	0,08

that is, to obtain probabilistic forecasts for certain events. For example, if current observations is «`Wet-Wet-Dry-Dry`» (that is, there were precipitations for two days in a row, on the next two days volumes were equal zero), the following statements can be formulated as: «The probability of precipitation through 2 days in Potsdam at current observations is 0,3961, and the probability of precipitation absence through 2 days is 0,6039». Unlike the standard practice for data analysis, when the predicted window should not exceed the size of input observations, this rule can be violated for historical values.

As an alternative forecasting tool, feed-forward neural networks with several hidden layers and various activation functions were used [12]. The patterns are used as the training sets. However, the frequency of each of the sets is not used explicitly, and the corresponding procedures are implemented in the hidden layers of the neural network. As a result of the work, a forecast is obtained for the following $1-2$ days. The best obtained prediction accuracy for a neural network with a sigmoid activation function and two hidden layers was 82% for a one-day and 74% for a two-day forecast (with `"PHP"`implementation). The adding of hidden layer, the changing of the activation function to the `rectifier`, the increasing of a size of the input sample and the use of the deep learning library `Keras` (with `Python` implementation) lead to enhance the forecast accuracy for the same data to 97% for one-day and to 89% for two-day forecasts.

For the chosen architecture of the neural network, there is no overfitting: the error value is the same for both the training part and for the test part, which does not participate directly in the process of building the neural network. Thus, we can expect that the model will work correctly not only for the training part, but also for real data.

Fig. 2 demonstrates an example of the accuracy of precipitation prediction for the next day taking into account the month of data. The numerical values of the corresponding forecast errors are indicated in Table 2.

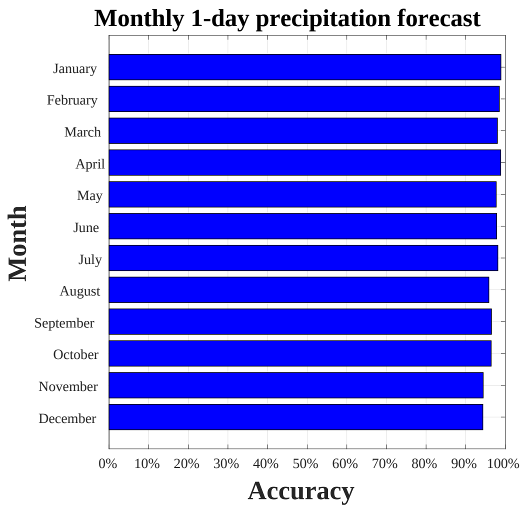

Figure 2. Accuracy of the monthly 1-day precipitation forecast

Table 2

Monthly 1-day precipitation forecast errors

Month	Error (1-day)
January	1,1%
February	1,5%
March	2%
April	1,2%
May	2%
June	2,2%
July	1,9%
August	4,2%
September	3,5%
October	3,6%
November	5,6%
December	5,7%

4. Conclusions

For the analysis of the probabilistic behavior of the precipitation process and the forecasting, it is suggested to use the chains of events (patterns) extracted from the data. High accuracy for neural networks forecasting is demonstrated, wherein the analysis is based solely on basic statistical data without any additional information about meteorological conditions.

Working with patterns is of interest in terms of verification ensemble of forecasts. Also, this methodology can be used to predict the behavior of the moment characteristics of finite normal mixtures of probability distributions [16] to determine the trend direction within the framework of modeling of physical experiments [17]. Such data are different from observations considered in this work (for example, there is no a seasonal factor), but in general the task seems to be quite similar, although the architecture of neural network should be modified.

As one of the directions for further research, it is possible to propose a transition from the binary model of the discretization of events to base-k numeral system. It could allow to solve more complex forecasting tasks, for example, to predict the amount of precipitation in terms of falling into pre-selected ranges of values. That is, if any positive value was observed in i-th day, it is replaced by j from the range $1, \ldots, k-1$ ($\widetilde{V}^{(i)}_{\text{daily}} = j$) corresponding to the precipitation volume partition by $k-2$ intervals; otherwise, the value $\widetilde{V}^{(i)}_{\text{daily}}$ is assigned a zero value.

To maximize the automation of the research process, the developed forecasting methods will be integrated into the service of stochastic data analysis [18–20] as a special tool for data mining.

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project No 17-07-00851) and the Ministry of Education and Science of the Russian Federation (No 538.2018.5). The author is grateful to Corresponding Member of the Russian Academy of Sciences Prof. S. K. Gulev for provided data, to Prof. V. Yu. Korolev for useful discussions within the framework of joint studies of meteorological phenomena and V. Kuzmin for training of neural networks.

References

1. L. V. Alexander, X. Zhang, T. C. Peterson et al., Global observed changes in daily climate extremes of temperature and precipitation, Journal of Geophysical Research-Atmospheres, **111** (D5) (2006) D05109. doi:10.1029/2005JD006290.
2. J. H. Christensen, F. Boberg, O. B. Christensen, P. Lucas-Picher, On the need for bias correction of regional climate change projections of temperature and precipitation, Geophysical Research Letters, **35** (20) (2008) L20709. doi:10.1029/2008GL035694.
3. R. W. Portmann, S. Solomon, G. C. Hegerl, Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States, Proceedings of the National Academy of Sciences of the United States of America, **106** (18) (2009) 7324–7329. doi:10.1073/pnas.0808533106.
4. R. Krzysztofowicz, The case for probabilistic forecasting in hydrology, Journal of Hydrology, **249** (1–4) (2001) 2–9. doi:10.1016/S0022-1694(01)00420-6.
5. N. Q. Hung, M. S. Babel, S. Weesakul, N. K. Tripathi, An artificial neural network model for rainfall forecasting in Bangkok, Thailand, Hydrology and Earth System Sciences, **13** (8) (2009) 1413–1425. doi:10.5194/hess-13-1413-2009.
6. T. Partal, H. K. Cigizoglu, Prediction of daily precipitation using wavelet-neural networks, Hydrological Sciences Journal, **54** (2) (2009) 234–246. doi:10.1623/hysj.54.2.234.
7. K. P. Moustris, I. K. Larissi, P. T. Nastos, A. G. Paliatsos, Precipitation Forecast Using Artificial Neural Networks in Specific Regions of Greece, Water Resources Management, **25** (8) (2011) 1979–1993. doi:10.1007/s11269-011-9790-5.
8. P. A. O'Gorman, T. Schneider, The physical basis for increases in precipitation extremes in simulations of 21st-century climate change, Proceedings of the National Academy of Sciences of the United States of America, **106** (35) (2009) 14773–14777. doi:10.1073/pnas.0907610106.
9. K. E. Trenberth, Changes in precipitation with climate change, Climate Research, **47** (1–2) (2011) 123–138. doi:10.3354/cr00953.
10. K. E. Kunkel, T. R. Karl, D. R. Easterling et al., Probable maximum precipitation and climate change, Geophysical Research Letters, **40** (7) (2013) 1402–1408. doi:10.1002/grl.50334.
11. N. Ban, J. Schmidli, C. Schar, Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster?, Geophysical Research Letters, **42** (4) (2015) 1165–1172. doi:10.1002/2014GL062588.
12. A. K. Gorshenin, Pattern-based analysis of probabilistic and statistical characteristics of precipitations, Informatics and Applications, **11** (4) (2017) 38–46. doi:10.14357/19922264170405.
13. A. K. Gorshenin, On some mathematical and programming methods for construction of structural models of information flows, Informatics and Applications, **11** (1) (2017) 58–68. doi:10.14357/19922264170105.
14. V. Yu. Korolev, A. K. Gorshenin, S. K. Gulev, K. P. Belyaev, A. A. Grusho, Statistical Analysis of Precipitation Events, AIP Conference Proceedings **1863** (2017) 090011. doi:10.1063/1.4992276.
15. V. Yu. Korolev, A. K. Gorshenin, The probability distribution of extreme precipitation, Doklady Earth Sciences, **477** (2) (2017) 1461–1466. doi:10.1134/S1028334X17120145.
16. A. K. Gorshenin, Concept of online service for stochastic modeling of real processes, Informatics and Applications, **10** (1) (2016) 72–81. doi:10.14357/19922264160107.
17. G. M. Batanov, A. K. Gorshenin, V. Yu. Korolev, D. V. Malakhov, N. N. Skvortsova, The Evolution of Probability Characteristics of Low-Frequency Plasma Turbulence, Mathematical Models and Computer Simulations, **4** (1) (2011) 10–25. doi:10.1134/S2070048212010048.
18. A. Gorshenin, V. Kuzmin, Online system for the construction of structural models of information flows, in: Proceedings of the 7^{th} International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, 2015, 216–219.

19. A. Gorshenin, V. Kuzmin, On an interface of the online system for a stochastic analysis of the varied information flows, AIP Conference Proceedings **1738** (2016) 220009. doi:10.1063/1.4952008.
20. A. K. Gorshenin, V. Yu. Kuzmin, Research support system for stochastic data processing, Pattern Recognition and Image Analysis, **27** (3) (2017) 518–524. doi:10.1134/S1054661817030117.