УДК 9

Arushanov M. L., Doctor of Geographical Sciences, Professor
Head of the Laboratory
Scientific Research Hydrometeorological Institute,
Tashkent, Uzbekistan.
Eshmuratova G. Sh.
Engineer
JSC "Hydroproject"
Tashkent, Uzbekistan.

STOCHASTIC MODEL FOR FORECASTING ATMOSPHERIC DROUGHT IN UZBEKISTAN USING REGRESSION ON CHARACTERISTIC ROOTS

Abstract. *This paper presents a one-month lead-time predictive model of atmospheric drought developed at the Scientific Research Hydrometeorological Institute. The model is based on a dynamic-stochastic approach to constructing a regression predictive equation. From the set of existing methods for constructing regression equations, the method based on the characteristic roots (eigenvalues) of the correlation matrix, including the predictand column and predictor columns (extended matrix) was applied.*

The standardized drought index SPI serves as the predictand, while the predictors are the average monthly precipitation for the 3 months preceding the forecast month, the average monthly value of variations in solar activity (Wolf numbers) and the average monthly value of the Southern Oscillation index for the month preceding the forecast.

The predictors were selected based on mutual correlation and applied time series analyses between the aridity index SPI and the indicated heliogeophysical values. The performed estimates of the investigated dependence of the aridity index SPI on the state of solar activity, the influence of El Niño (La Niña) and precipitation preceding the forecast date showed their high correlation.

Estimates of the accuracy of the SPI forecast with a monthly advance lead time for the territory of Uzbekistan, performed on an independent sample, were quite high, which was the basis for the introduction of this model into the operational work of the hydrometeorological service of Uzbekistan (Uzhydromet).

Keywords: *precipitation, Wolf numbers, Southern oscillation index, standardized aridity index SPI, correlation function, causality function, proper numbers, eigenvectors.*

Introduction

Atmospheric drought is a natural phenomenon characterized by abnormally low amount of precipitation and elevated temperatures [Khromov, Mamontova, 1974], manifested in various geographical regions across the globe, with an extremely uneven spatial distribution. The demand for the forecast of atmospheric drought is driven by the significant economic damage it inflicts, the considerable social consequences, and the frequent escalation into humanitarian disasters.

When developing a predictive model based on the construction of a regression equation, the key factors for its accuracy are, firstly, the stability of the resulting equation when transitioning to data on an independent sample and the optimal selection of predictors. The predictors included in the regression equation are determined primarily from physical considerations responsible for the dynamics of the predictand, supported by a statistical analysis of the dependance of the latter on the selected set of predictors.

It is important to choose a method for constructing a regression equation, in particular, when constructing multiple regression, which often yields low results on independent samples.

In this paper, the construction of a regression model is based on the use of the method of characteristic roots (eigenvalues) of the inverse correlation matrix of the predictand and predictors [Vuchkov, Boyadzhieva, Solakov, 1987; Draper, Smith, 1986].

1. The data used and the rationale for the selected predictors

The data used in this study comprises observed precipitation data from the Uzhydromet meteorological network for the period 1966-2023, covering 12 regional centers and the city of Tashkent. The average monthly precipitation values for each month of the year were derived from these data. The standardized index of aridity SPI [Lloyd-Hughes, Saunders, 2002], recommended by the World Meteorological Organization (WMO), as one of the most informative indices of the state of aridity (humidity) of the atmosphere [WMO, 2016] was calculated based on these data. The specified index in the developed model serves as the predictand. During the same period, the monthly average values of the Wolf numbers were entered into the generated database, which are freely available on the website (http://www.kosmofizika.ru/spravka/spots.htm) characterizing the dynamics of variations in solar activity. Additionally, average monthly values of the Southern Oscillation index (*SOI*), which are also freely available on the website https://www.ncdc. noaa.gov/teleconnections/enso/indicators/soi/data.csv, were incorporated. The latter, along with precipitation data for the 3 months preceding the forecast period, are used as predictors in the model, but with a monthly delay, i.e. the month preceding the forecast month is selected from the database.

The Wolf Number (W) characterize the state of solar activity and is calculated as:

$$W = k\,(f + 10g), \qquad (1)$$

where f - is the number of observed spots, g- is the number of observed groups of spots, and k - is a normalization coefficient.

Studies [Pokrovskaya, 1969, 1973; Pokrovskaya, Mandel, 1974] have established a significant influence of the year's position within the 11-year solar activity cycle on drought occurrence. However, the impact of solar activity variations on atmospheric conditions, particularly drought (wetness), varies across different geographic regions of the planet, even to the extent of a change in the sign of influence.

For instance, spring-summer droughts in the European territory, according to the catalog of N.G. Kamenkova [Pokrovskaya, 1969], occur on the ascending branch of solar activity, while droughts in Central Asia, taken from the catalog of A.S. Uteshev [Uteshev, 1965, 1972], are grouped on the descending branch of solar activity. For Central Asia, the ratio of droughts on the ascending branch to those on the descending branch is 1:11 in percentage terms. This relationship is clearly illustrated in Figure 1 by the cross-correlation function and the causality function [Arushanov, Korotaev, 1994; Arushanov, 2023] between the Wolf numbers and the *SPI* index as functions of time lag. In selecting predictors for constructing the regression equation, a crucial fact is that the causality function remains within the domain of normal causality for the entire range of time lags, meaning solar activity variations unequivocally influence drought variability in the specified region. Furthermore, the negative values of the cross-correlation function in the time lag range of ±3 years confirm the inverse relationship between solar activity variations and the processes of drought formation.

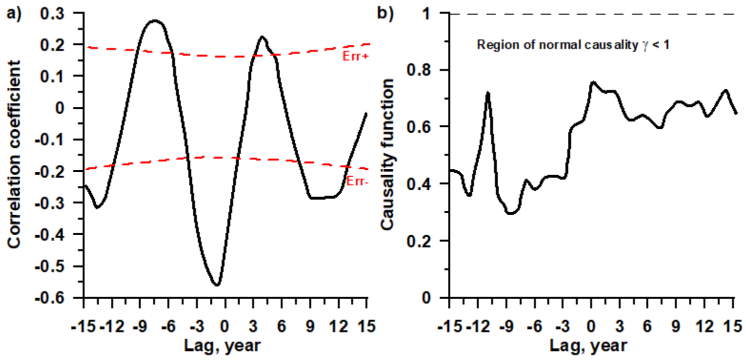

Fig. 1.Correlation function (a) and causality function (b) between Wolf numbers *W* and the aridity index *SPI*, as a function of time lag.

The Southern Oscillation index (SOI). The climate system is characterized by large-scale self-oscillatory processes, such as the Southern oscillation (redistribution of air masses in the low latitudes of the Southern Hemisphere between the Indian and Pacific Oceans - ENSO) and ocean fluctuations – El Niño (warm phase) and La Niña (cold phase).

A quantitative characteristic of the ENSO is the Southern Oscillation Index (*SOI*), introduced by G. Walker in 1924 [Walker, 1924]. The southern oscillation is an atmospheric component of air currents and represents fluctuations in air pressure near the surface layer of the atmosphere between the waters of the eastern and western parts of the Pacific Ocean. The SOI is calculated based on the difference in surface air pressure between the area of Tahiti Island (French Polynesia) and Darwin (Australia), and is determined by the following relationships [Walker, 1924]:

$$SOI = \frac{P_T - P_D}{\sigma_{TD}},$$

$$P_T = \frac{P_T^{sur} - \overline{P}_T}{\sigma_T}, \quad P_D = \frac{P_D^{sur} - \overline{P}_D}{\sigma_D},$$

$$\sigma_T = \sqrt{\frac{\sum_{i=1}^{N}\left(P_{Ti}^{sur} - \overline{P}_T\right)^2}{N}}; \quad \sigma_D = \sqrt{\frac{\sum_{i=1}^{N}\left(P_{Di}^{sir} - \overline{P}_D\right)^2}{N}}; \quad \sigma_{TD} = \sqrt{\frac{\sum_{i=1}^{N}\left(P_{T_i}^{sur} - P_{D_i}^{sur}\right)^2}{N}}. \qquad (2)$$

In (2), P_T^{sur} - is surface pressure at Tahiti point, P_D^{sur}- is surface pressure at Darwin point, $\overline{P}_T . \overline{P}_D$ - is average surface pressures over the base period in the corresponding points.

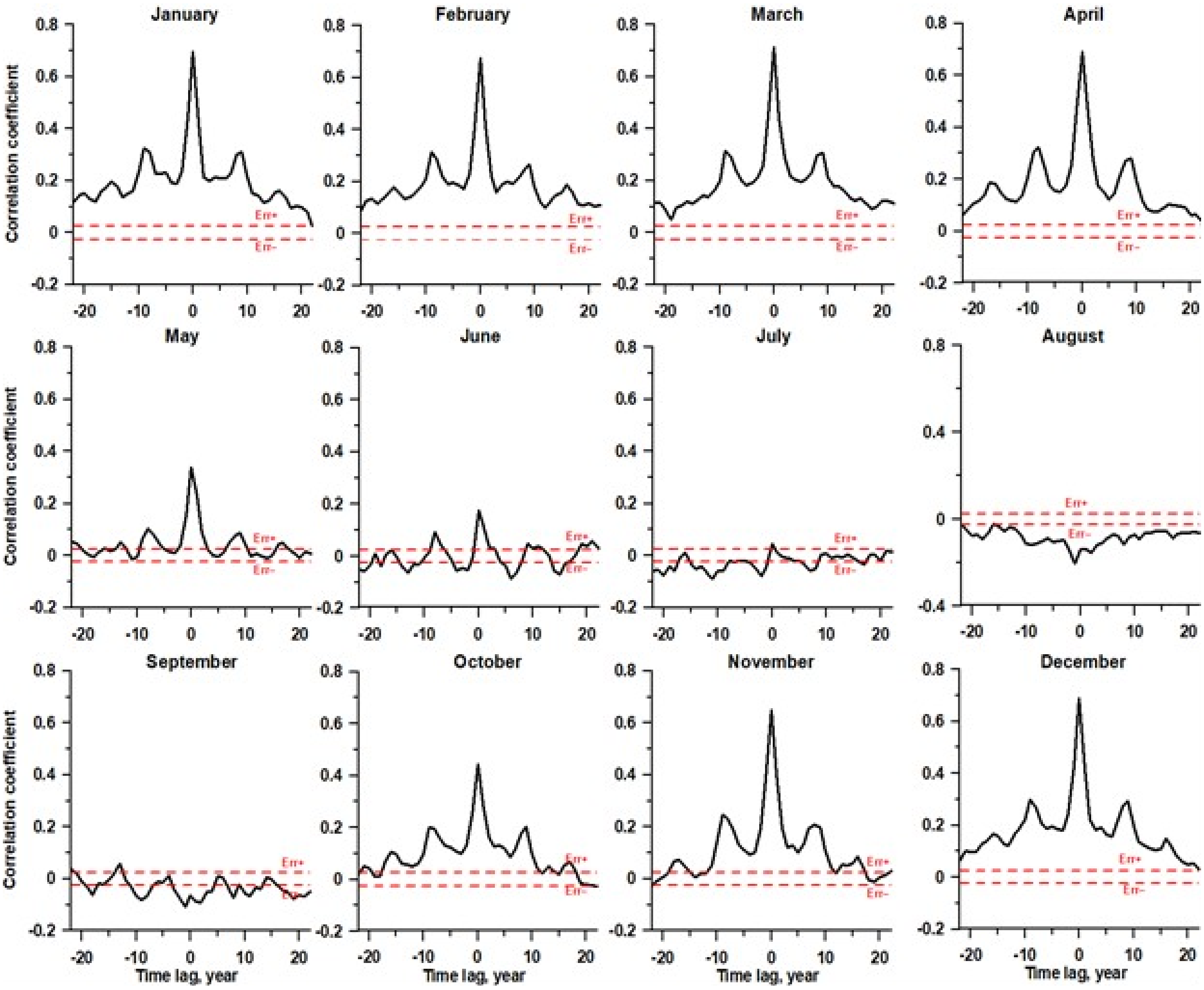

Fig. 2. Correlation functions between *SPI* and *SOI* as a function of time lag.

Figure 2 shows the cross-correlation functions between SPI and SOI for each month of the year in Uzbekistan. The influence of ENSO on atmospheric processes that stimulate aridity in the territory of Uzbekistan varies: it is less pronounced from June to September, and most significant during the cold period of the year.

2. Methods of analysis and construction of the regression equation

Construction of a predictive regression equation based on characteristic roots. The regression procedure on characteristic roots was developed by R. Webster, G. Hans and R. Mason [Webster, Guns, Mason] and independently of them by D. Hawkins [Hawkins, 1973]. This method is described in detail in [Arushanov, 2009; Vuchkov, Boyadzhieva, Solakov, 1987; Draper, Smith, 1986]. Following, for example, [Arushanov, 2009], we present an algorithm that implements the regression construction procedure on characteristic roots (eigenvalues).

The time series of predictors $F_i(t_j)$, i=1, 2, …, 5, j=1, 2, …, N (precipitation for the three months preceding the forecast, Southern Oscillation index, and Wolf numbers) and predictand $Z_i(t_j)$ (SPI index) are standardized:

$$X_i(t_j) = \frac{F_i(t_j) - \overline{F_i}}{\sigma_{iF}}, \quad Y_i(t_j) = \frac{Z_i(t_j) - \overline{Z_i}}{\sigma_{iZ}}, \tag{3}$$

where $\sigma_{iF} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(F_i(t_j) - \overline{F_i}\right)^2}$, $\sigma_{iZ} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(Z_i(t_j) - \overline{Z_i}\right)^2}$ – is the mean square deviations; the top line is averaging. An extended matrix $\boldsymbol{A}$ of size N×(i+1) is constructed:

$$\boldsymbol{A} = (\boldsymbol{Y}, \boldsymbol{X}) = \begin{bmatrix} Y_{11} & X_{11} & X_{21} & \dots & X_{i1} \\ Y_{12} & X_{12} & X_{22} & \dots & X_{i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{1N} & X_{1N} & X_{2N} & \dots & X_{iN} \end{bmatrix} \tag{4}$$

Having the matrix (4), the correlation matrix R = (ATA)-1 is calculated by the measure (i+1) × (i+1). The matrix R is a symmetric matrix with 1 diagonal:

$$\mathbf{R} = \begin{bmatrix} 1 & R_{11} & R_{21} & \dots & R_{i1} \\ & 1 & R_{22} & \dots & R_{i2} \\ & & \vdots & \vdots & \vdots \\ & & & \dots & 1 \end{bmatrix} \tag{5}$$

The next step is to calculate the eigenvalues and eigenvectors of the correlation matrix (5) and arrange the eigenvectors in accordance with the decreasing magnitude of the eigenvalues. The estimates of the regression coefficients a_j are based on the formula:

$$a_j = -\frac{c\sum_{m=1}^{k} v_{om} v_{jm}}{\lambda_m}, \quad j = 1, 2, \dots, k$$

(6)

where v_{om} - is the 1st eigenvector, v_m - are the subsequent eigenvectors arranged in descending order of eigenvalues λ_m, and c - is determined from the following expression:

$$c = \left(\sum_{m=1}^{k} (v_{\sigma m}^2) / \lambda_m \right)^{-1} \times \sigma_z.$$

3. Numerical experiments

To conduct numerical experiments for one-month-ahead atmospheric drought forecasting, a database was used, consisting predictors: average monthly precipitation, the SOI, Wolf numbers and the predictand - calculated values of the SPI aridity index. The training sample period for all predictors and predictand was 50 years, and the independent (test) sample was 10 years. Forecasts using the model were calculated for 12 regional centers and Tashkent city.

As quantitative estimates of the accuracy of the forecast, the following were used:

- standard error: $\sigma_{for} = \frac{1}{N} \sum_{i=1}^{N} (SPI_{akt} - SPI_{for})^2$ (7)
- maximum absolute error: $\Delta_{max} = max|SPI_{akt} - SPI_{for}|$ (8)
- sign agreement: $\rho = \frac{n_+ - n_-}{n}$. (9)

In (7), (8) SPI_{akt} – is the actual value of the aridity index; SPI_{for} - is the prognostic value of the aridity index. In (9) n+ – is the number of *SPI* values with matched signs, n_- – is the number of *SPI* values with non–matched signs, n - is the total number of cases.

Table 1 summarizes the quantitative assessment of the accuracy of one-month lead-time forecasts for the average monthly SPI, averaged across 13 locations according to formulas (7) - (9).

As follows from the estimates presented in the table, on average, the level of justification of the forecast for the parameter of atmospheric aridity with a month's advance is more than 80%, which is approaching the verification level of short-term forecasts, which currently holds the highest accuracy.

Table

Average scores on all points (summary table)

Forecast month	Estimated parameter			
	$\lvert\Delta_{max}\rvert$	ρ	$R_{фп}$	$\sigma_{пр}$
January	1,152	0,886	0,771	0,383
February	1,085	0,954	0,818	0,246
March	0,741	0,980	0,881	0,154
April	1,588	0,913	0,787	0,376
May	2,091	0,812	0,703	0,556
June	2,038	0,692	0,664	0,579
July	1,768	0,793	0,661	0,471

August	1,399	0,791	0,736	0,367
September	1,414	0,771	0,736	0,413
October	2,184	0,665	0,629	0,644
November	1,803	0,852	0,742	0,543
December	1,859	0,874	0,654	0,585
Average	**1,593**	**0,832**	**0,732**	**0,443**

Conclusions

Overall, it can be stated that the presented assessments of the model, taking into account the monthly forecast lead time, can be considered sufficiently high, which made it possible to introduce this predictive model of atmospheric aridity into the operational work of the hydrometeorological service of Uzbekistan.

References

1. Arushanov M. L., Korotaev S. M. Causal analysis and its application to the study of physical processes in the atmosphere //Meteorology and Hydrology N6. - 1994. - pp. 15-22.
2. Arushanov M. L. Natural causes of modern climate change on the planet. Hamburg: Lambert Academic Publishing House (LAP), 2012. – 178 p.
3. Arushanov M. L. Climate dynamics. Space factors. Hamburg: Lambert Academic Publishing House. - 2023. – 145 p.
4. Arushanov M. L. Causes of Earth climate change, as a result of space impact, dispelling the myth about antropopogenic global warming //German International Journal of Modern Science №53, – pp. 4-15. DOI: 10.5281/zenodo.7795979
5. Arushanov M. L. Construction of regression equations for forecasts of hydrometeorological elements in a wide range of time using regression on characteristic roots //Construction of regression equations for forecasts of hydrometeorological elements in a wide range of time using regression on characteristic roots // Proceedings of NIGMI, issue 14(259), 2009. – pp. 3-19.
6. Vuchkov I., Boyadzhieva L., Solakov E. Applied linear regression analysis. – M.: Finance and Statistics, 1987. – 238 p.
7. Draper N., Smith S. Applied regression analysis. – M.: Finance and Statistics, 1986. – 365 p.
8. Climate change: A summary report by the Intergovernmental Panel on Climate Change / edited by R.T. Watson et al. Geneva. - 2007. – 104 p.
9. Pokrovskaya T. V. Synoptic-climatological and heliogeophysical long-term weather forecasts. – L.: Hydrometeoizdat, 1969. – 254 p.
10. Pokrovskaya T. V. On the connection between the onset of droughts in European territory USSR with computer radiation of the Sun // Proceedings of the GGO, vol. 299. -1973, pp. 126-135.
11. Pokrovskaya T. V., Mandel D. G. On the relationship of precipitation in the European territory of the USSR and in Northern Kazakhstan with geomagnetic activity and the possibility of probabilistic precipitation forecast // Proceedings of

the 1st All-Union Meeting "Solar-atmospheric relations in climate theory and weather forecasts". - L., Hydrometeoizdat, 1974. – pp. 347-352.
12. Handbook of indicators and indices of aridity. – Geneva: WMO, Weather, Climate, Water, No. 1173. – 2016. – 53 p.
13. Uteshev A. S. Drought and grain harvest. – Alma-Ata: publishing house "Kainar", 1965. – 60 p.
14. Uteshev A. S. Atmospheric droughts and their impact on natural phenomena. – Alma-Ata: Nauka, 1972. – 176 p.,
15. Khromov S. P., Mamontova L. I. Meteorological dictionary. – L. Hydro-meteorological publishing house, 1974. – 568 p.
16. Hawkins D. On the study of alternative regression using the analysis of basic components // Appal. Statistics. vol. 22. – 1973. – pp. 275-286.
17. McKee, T. B., N. J. Doken, and J. Kleist. Dependence of the frequency and duration of drought on the time scale // Proceedings of the Eighth Conference on Applied Climatology: – Anaheim, California, Boston. – American Meteorological Society. - 1993. – pp. 179-184. DOI: 10.4236/jwarp.2016.87058
18. Lloyd-Hughes B., Saunders M. A. Climatology of drought in Europe // International Journal of Climatology, vol. 22, No. 13. – 2002. – pp. 1571-1592. DOI: 10.1002/joc.846
19. Lloyd-Hughes B., Prudhomme K., Hannaford J., Parry S., Keefe K. and Rees H. G.: Drought Catalogs for Great Britain and Europe, Scientific Report of the Environment Agency SC070079/SR, Environment Agency, Bristol, 2010. – pp. 59-76.
20. Gunst R., Mason R. Analysis of hidden root regressions // Technometrica, vol. 16. – 1974. – pp. 513-522.
21. Walker G. T. Correlation of seasonal weather fluctuations // Meme. Meteorology of India. Department. 24. – 1924. – pp. 275-332. DOI: 10.4236/ajps.2022.135037