

STATISTICAL MODEL OF TEMPERATURE ANOMALIES USING A NEW INSTABILITY INDEX

Hristo Chervenkov, Valery Spiridonov

Received on June 10, 2022

Presented by K. Ganev, Corresponding Member of BAS, on September 27, 2022

Abstract

A statistical model (relationship) between a new circulation instability index and the registered temperature anomalies in two regions of Europe has been proposed and tested. The average values of this index, calculated at levels of 850 hPa and 500 hPa, were used as predictors in the modelling of temperature anomalies by months for the period 2011–2020. The average monthly values of temperature and geopotential during this period were utilized to calculate the index and anomalies in these regions. For each of them, the following statistical characteristics were calculated: correlation, root mean square error, and the coefficient of determination (R-squared). They evaluate the performance of the statistical model. The obtained results demonstrate the principal applicability of the instability index in a similar way as the well established circulation estimator North Atlantic Oscillation.

Key words: instability index, temperature anomalies

Introduction. In [1] an index characterizing the instability of atmospheric processes in a horizontal surface is derived, (hereafter referred to as the index).

This work has been carried out in the framework of the National Science Programme “Environmental Protection and Reduction of Risks of Adverse Events and Natural Disasters”, approved by the Resolution of the Council of Ministers No. 577/17.08.2018 and supported by the Ministry of Education and Science of Bulgaria (Agreement No. D01-279/03.12.2021) and with the financial support by Grant No. BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Programme (2014–2020) and co-financed by the European Union through the European structural and Investment funds.

DOI:10.7546/CRABS.2022.11.09

It was obtained by transforming the hydrostatic system in spherical coordinates. The index I can be written in the form:

$$(1) \quad I \doteq \Gamma \frac{dH}{dT} - 1,$$

where H is the geopotential and T the air temperature. $\Gamma \doteq \frac{g}{C_p} = 0.0097 \text{ K.m}^{-1}$, $g = 9.81 \text{ m.s}^{-1}$ is the acceleration of the gravity, and $C_p = 1012 \text{ m}^2.\text{s}^{-2}.\text{K}^{-1}$ is the specific heat capacity of dry air at constant pressure. The derivative $\frac{dH}{dT}$ is of the geopotential H with respect to temperature where T and H are along the meridian. The negative values of the index characterize the unstable state of the atmosphere. Positive values correspond to a steady state in the considered region.

It should be noted that there is no direct functional relationship between instabilities and anomalies, in particular anomalies of the tropospheric temperature distribution. For example, the instability of EADY [2] or the waves of Rossby [6] cannot be directly related to the anomalies. Deviations from norms in either direction can occur under both stable and unstable weather conditions. In our previous work [11] a significant correlation was found between I and the North Atlantic Oscillation (NAO). This suggests that the index can be used as an indicator of the prevailing atmospheric circulation (similar to the NAO) as well as a measure of stability. This is a reason to look for a relationship between the index and temperature anomalies. It is worth emphasizing, that deviations from norms in either direction can occur under both stable and unstable weather conditions.

In this paper the 500 hPa level is considered additionally, alongside 850 hPa. This makes it possible to create a statistical model using the index to predict temperature anomalies. Another feature here is the use of monthly averages of geopotential and temperature, as well as deviations from the average for the levels for which the index is calculated.

1. Method and data. There are various possibilities to find a relationship between the index and temperature anomalies. One of them will be presented here, without claiming to be the most appropriate. The aim is to illustrate the possibility of finding a relationship between the index and anomalies that is suitable and easy to implement. For the anomalies (defined as the difference of the monthly means with respect to the reference period 1981–2010), data from the ERA5 reanalysis [5] in regular $0.25^\circ \times 0.25^\circ$ grid downloaded from the Copernicus Climate Data Store was used. The estimations are performed on monthly basis, for the 10 years long period 2011–2020. Again, it should be noted that the index in [11] is calculated for the daily values of geopotential and temperature. The possibility of using their monthly average values will be checked here. If this proves successful, the method will be very easy to implement.

The use of the index on two levels creates a premise for appearance of a signal indicating a barotropic or baroclinic state of the atmosphere, dominating

in a given month. The conditions for baroclinic and barotropic instability are discussed, for example, in [10], where various linear and nonlinear theories are analyzed. Through the values of the index at these levels, predictors can be formed to create a statistical model that predicts temperature anomalies. If the levels used do not match the level of the anomalies, the task of finding a relationship between the index and the anomalies becomes very complex. Furthermore, the index is calculated based on monthly averages to simplify the method and make it more practical. To find the relationship with NAO in [11], the average index values for the area under consideration were used. The obtained significant correlations of the index with the NAO stimulate the idea of using the index to find a relationship with temperature anomalies.

The hypothesis accepted here is that in a certain area, for a certain period, the presence of instability above certain values may form a prevailing circulation, which will form a certain anomaly. It is essential that the anomalies are calculated as deviations from the mean over the period considered, and that the index is calculated from the corresponding temperatures used. The level at which the anomalies are calculated is also important. The level of 850 hPa is accepted here. The levels of 850 hPa and 500 hPa are widely used in practice in the analysis of the development of atmospheric processes and are available for the reanalysis of most models. The predictors of the model are the average monthly values of the indices at the two isobaric levels. The predictants are the corresponding temperature anomalies. These values are determined for each region individually. In other words, for each region there is a different statistical model with different statistical estimates as shown in Table 3. A modification of multiple linear regression (MLR) method is used as the statistical model here by applying the so-called SWEEP operator [3,9]. This operator allows the creation of regression models. The sweep method is particularly advantageous for degenerate or singular matrix (the determinant is equal to zero), which can happen with correlated (i.e. improperly selected) predictors.

The location of the area as well as its size is essential. The area must be of a size that allows the formation of a predominant circulation corresponding to the average values of the index. In other words, the anomalies in the whole area must have predominant values in a given month (whether positive or negative). Such cases of hot and cold waves affecting large areas of Europe have been described in many papers, for example [4]. There are also detailed studies in sub-areas of the considered areas [7].

The correlation and estimates for RMSE and R-squared are used to assess how good the regression model is (i.e. how good the predicted anomalies are). As is known, they have the following properties.

The Pearson's correlation coefficient, called further for the sake of brevity correlation, is a statistical measure that shows the extent to which two or more variables differ from each other. Although the correlation does not imply neces-

sarily causal dependence between the predictants and the predictor it is frequently used in many geophysical and engineering disciplines for evaluation of the agreement between examined and reference datasets [1,8,12]. There are three possible results of a correlational study: a positive correlation, a negative correlation, and no correlation.

Root Mean Square Error (RMSE) is a standard way to measure the error of a model in predicting quantitative data [1,12]. RMSE is the square root of Mean Squared Error. It measures the standard deviation of residuals.

The coefficient of determination R^2 or R-squared represents the proportion of the variance in the dependent variable which is explained by the linear regression model. It is a scale-free score, i.e. irrespective of the values being small or large, the value of R-squared will be less than one. R-squared shows how well the data fit the regression model.

The three measures complement each other. For example, correlation values may be low (bad) but RMSE may be low (good). High correlation values (which is good) may be accompanied by high RMSE values (which is bad). The closer R^2 is to one, the better the model works. If R^2 is negative, the model result is not good.

2. Results. In line with the above conditions, two regions covering parts of Europe were studied. In the Mercator projection, the first region covers the area defined by the coordinates 15.0°E, 30.0°E; 35.0°N, 45.0°N. The second region in the same projection is defined by the coordinates 10.0°W, 15.0°E; 35.0°N, 55.0°N. The first region covers the bigger part of Western Europe and the western Mediterranean. The second region, which is significantly smaller, is situated over Southeastern Europe.

For each of the regions, the results obtained are described below. They illustrate the extent to which the temperature anomalies are consistent with those recovered by the statistical model. All values below are rounded to the decimal place. The degree of independence of the selected regions in the formation of the anomalies can be verified by the correlations between the corresponding regional anomalies. High correlations signal shows dependence between regions. The correlation by month is shown in Table 1. The correlations are low, indicating that the two regions have different conditions for anomaly formation.

The second important question is whether the selected predictors (in this case, the averaged index values for the two levels 850 hPa and 500 hPa) are

T a b l e 1

Correlation by months between the anomalies of the two regions

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
0.0	0.1	-0.2	0.0	0.1	0.4	0.0	-0.1	0.1	0.1	0.2	0.2

independent, i.e., not correlated with each other. This is a prerequisite for good model estimates. Table 2 shows the index correlations for these levels for both regions. The correlation between the two levels provides additional information

T a b l e 2

Correlation of indices for the two levels for region 1 (upper section) and for region 2 (lower section)

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
0.2	0.6	0.6	0.6	0.3	0.2	0.1	−0.2	0.1	−0.4	0.5	0.1
0.6	0.2	0.4	0.4	0.6	0.1	0.4	0.5	−0.2	0.0	0.3	0.5

on the extent to which the instability is barotropic or baroclinic. It can be argued that the high correlation between the two levels is a signal of barotropic processes in a given month, overall for the period under consideration. Correspondingly, low or negative correlations signal baroclinicity. The proof of this proposition is quite extensive and goes beyond the objectives set here. For the remaining criteria, the following results were obtained. For the first region, the correlation, RMSE and R-squared by month are shown in Table 3. In addition to monthly

T a b l e 3

Correlation, RMSE and R^2 for region 1 (upper section) and for region 2 (lower section)

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
<i>corr</i>	0.4	0.7	0.2	0.7	0.3	0.4	0.2	0.4	0.8	0.2	0.2	0.6
RMSE	1.7	0.8	0.9	1.4	1.1	1.1	0.9	0.5	0.6	1.4	1.2	0.9
R^2	−3.7	0.4	0.4	0.3	0.6	0.6	0.7	0.9	0.9	0.5	0.7	0.8
<i>corr</i>	0.6	0.8	0.8	0.8	0.9	0.7	0.8	0.5	0.4	0.2	0.3	0.7
RMSE	1.0	0.7	0.8	0.8	0.5	0.7	0.5	0.6	0.8	0.9	1.2	1.0
R^2	−0.8	0.6	0.7	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.7	0.8

correlations, correlations by year can also be compared. In Table 4 they are given for the two regions. Generally speaking, the better estimates are for the second

T a b l e 4

Annual correlation for both regions

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Region 1	0.5	0.8	0.6	0.8	0.8	0.9	0.7	0.3	−0.3	0.6
Region 2	0.7	0.6	0.7	0.9	0.6	0.7	0.6	0.6	0.6	0.9

region. Neither region had good R-squared scores in January. The second region has significantly better scores for the remaining months. The RMSE values for both regions are similar.

3. Conclusion. According to the obtained and presented in this article results, the following conclusions can be drawn:

- The two regions have different scores on the three indicators, both by month and by year. This shows that the regions differ in terms of anomaly formation in them.
- Although the index is calculated using monthly averages for geopotential and temperature, the correlations in both regions are significant and for some months very good. This indicates that monthly averages can be successfully used instead of daily values of temperature and geopotential. The explanation for this requires a separate study. In this work it was only demonstrated that the index can be successfully used in statistical models with predictors formed using it.
- The NAO has only one value from which to determine the area, extent, and sign of anomalies. The index has values at each point on the computational grid. This allows to determine the different areas where persistent circulation is observed. This was demonstrated by studying two regions. One covers the Balkan Peninsula and the other is positioned in Central and Western Europe. This advantage of the index allowed the creation of the presented model, with which two different areas were analyzed, and the results proved to be encouraging both for use and for the creation of models with a similar ideology. It is interesting to note that if terminology such as “about the average (mean, norm)”, “above the average”, and “below the average” is used, as in the practice of long-range forecasts, the model proposed here will prove very successful.

Acknowledgement. Since this study is entirely based on free available data and software, the authors would like to express their deep gratitude to all organizations and institutions (the European Centre for Medium-Range Weather Forecasts, the Max Planck Institute for Meteorology, Copernicus Climate Data Store), which provide free of charge software and data.

REFERENCES

- [¹] CHERVENKOV H., K. SLAVOV (2022) NEX-GDDP Multimodel Ensemble vs. E-OBS — Evaluation of the Extreme Temperatures and Precipitation over Southeast Europe: Historical Comparison, *Atmosphere*, **13**(4), 581, <https://doi.org/10.3390/atmos13040581>.
- [²] EADY E. (1949) Long waves and cyclone waves, *Tellus*, **1**, 33–52.
- [³] GENTLE J. E. (1998) *Numerical Linear Algebra for Applications in Statistics*, New York, Springer, <https://doi.org/10.1007/978-1-4612-0623-1>.

- [4] HAFEZ Y. (2017) On the Relationship between Heat Waves over the Western and Central Europe and NAO, SOI, El-Niño 3.4 in Summer 2015, *Journal of Geoscience and Environment Protection*, **05**(04), 31–45, <https://doi.org/10.4236/gep.2017.54004>.
- [5] HERBACH H., B. BELL, P. BERRISFORD, S. HIRAHARA, A. HORÁNYI et al. (2020) The ERA5 global reanalysis, *Q. J. R. Meteorol Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- [6] LORENZ E. N. (1972) Barotropic Instability of Rossby Wave Motion, *Journal of Atmospheric Sciences*, **29**(2), 258–265, [https://doi.org/10.1175/1520-0469\(1972\)029<0258:BIORWM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<0258:BIORWM>2.0.CO;2).
- [7] MALCHEVA K., L. BOCHEVA, H. CHERVENKOV (2021) Climatology of the Extremely Hot Spells in Bulgaria (1961–1990), *Proceedings of the 21th International Multidisciplinary Scientific Geoconference SGEM 2021*, 16–22 August 2021 Albena, Bulgaria, vol. 4.1, 311–318, <https://doi.org/10.5593/sgem2021/4.1/s19.40>.
- [8] MOBERG A., P. D. JONES, D. LISTER, A. WALTHER, M. BRUNET et al. (2006) Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000, *Journal of Geophysical Research*, **111**(D22), D22106, <https://doi.org/10.1029/2006JD007103>.
- [9] NEYTCHIEV P. N. (1995) SWEEP operator for least-squares subject to linear constraints, *Computational Statistics & Data Analysis*, **20**(6), 599–609, [https://doi.org/10.1016/0167-9473\(94\)00067-8](https://doi.org/10.1016/0167-9473(94)00067-8).
- [10] READ P., N. LEWIS, D. KENNEDY, H. SCOLAN, F. TABATABA-VAKILI (2019) Baroclinic and barotropic instabilities in planetary atmospheres – Energetics, equilibration and adjustment [Preprint], *Nonlinear Waves, Pattern Formation, Turbulence/Climate, Atmosphere, Ocean, Hydrology, Cryosphere, Biosphere/Simulation*, <https://doi.org/10.5194/npg-2019-53>.
- [11] SPIRIDONOV V., H. CHERVENKOV (2021) A New Criterion for the Stability of the Atmospheric Processes and Its Relationships with NAO, *C. R. Acad. Bulg. Sci.*, **74**(9), 1363–1369, DOI:10.7546/CRABS.2021.09.12.
- [12] TORMA C. Z. (2019) Detailed validation of EURO-CORDEX and Med-CORDEX regional climate model ensembles over the Carpathian Region, *IDŐJÁRÁS*, **123**(2), 217–240, <https://doi.org/10.28974/idojaras.2019.2.6>.

National Institute of Meteorology and Hydrology
 66 Tsarigradsko Shosse Blvd
 1784 Sofia, Bulgaria
 e-mail: hristo.tchervenkov@meteo.bg
 valery.spiridonov@meteo.bg