Seventeenth International Conference Zaragoza-Pau on Mathematics and its Applications
Jaca, September 4–6th 2024

A selection algorithm of spatio-temporal models for the occurrence of temperature records

Elsa Barrio-Torres[1], **Zeus Gracia-Tabuenca**[1], **Jorge Castillo-Mateo**[1], **Jess Asn**[1], **Ana C Cebrin**[1], **Jess Abaurrea**[1]

SUMMARY

Evidence of global warming appears in extreme events of daily maximum temperature (Tx), particularly in record-breaking events. In the Iberian Peninsula, [1] has shown that record occurrence evolves with a trend above the expected stationary behavior and varies over space. A new methodology is proposed to model and understand the spatial variability and spatio-temporal dependence of this phenomenon.

Daily Tx series for 1960-2023 from 36 Spanish stations were downloaded from the European Climate Assessment & Dataset. Daily geopotential variables at 12 p.m. at pressure levels of 300, 500 and 700 hPa, on a 1°x1° grid, [45^o N, 10^o W, 35^o S, 5^o E], from ERA5 reanalysis data, served as the predictor database. Summer (JJA) days were used.

An algorithm was developed to obtain an optimal model for each station using logistic regression models. The target variable was the binary daily record indicator, defined as $I_{t,l} = 1$ if $Tx_{t,l} > max\{Tx_{1,l}, ..., Tx_{t-1,l}\}$, where $Tx_{t,l}$ is the maximum temperature on the year $t \in \{1, ..., T\}$ and day $l \in \{1, ..., L\}$. Geopotential levels' series at grid points of the four furthest corners and the closest grid point to the specific station were considered as predictors, along with a lag and second-order polynomial for these variables. The selection algorithm involved several steps: 1) A stepwise regression was used for feature selection at each station, resulting in 36 optimal models; 2) Based on these models, the most frequent and meaningful predictor variables were chosen to create a global model; 3) An interaction model incorporating second-order polynomial terms for latitude and longitude, along with the selected predictors, was fitted. Data from the first 51 years were used as training data, and the final 13 years were used as testing data. Given the class imbalance, the AUC was used to assess the model performance both globally and for each station.

Results showed that the local model achieved high prediction scores (average AUC=0.917), but used k=781 parameters. The simplest global model had a good overall performance (AUC=0.86; k=13), but showed low scores in stations near the coast, and the interaction model (AUC=0.89; k=43) showed substantial improvement in those stations.

In conclusion, the simplest global model was enough to successfully capture the behavior in the inner region. Stations near the coast displayed complex behavior, indicating the necessity of a model with spatial interactions.

Keywords: Statistical downscaling, weather extremes, records, modeling selection.

AMS Classification: 62312, 62P12, 62M30

References

[1] Castillo-Mateo, Jorge and Cebrián, Ana C and Asín, Jesús. Statistical analysis of extreme and record-breaking daily maximum temperatures in peninsular Spain during 1960–2021. *Atmospheric Research* **106934**(2023).

[1]Department of Statistical Methods, University of Zaragoza
email: elsa.barrio@unizar.es