SASYR Symposium of Applied Science for Young Researchers

2nd Symposium of Applied Science for Young Researchers

BOOK OF ABSTRACTS 2022

June 22 , 2022

Applications of the Analog Ensembles Method to Meteorological Data Reconstruction in the Northeast of Portugal

Murilo Montanini Breve[1], José Rufino[1], Carlos Balsa[2], and Luís de Sousa Costa[2]

[1] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
{murilo.breve,rufino,balsa}@ipb.pt
[2] Mountain Research Center (CIMO), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
lcosta@ipb.pt

Abstract

The observation of weather states has always been a human need. Our most distant ancestors already tried to understand and predict the weather, but did not have reliable methods. In the 19th century, modern meteorology took its first steps: the French government, motivated by the sinking of ships near the coast of Crimea, because of a heavy rainstorm, created a network of 24 stations spread across Europe, which began to observe the weather. In recent years, due to computational advances, different methods of predicting weather states have begun to emerge, increasing the forecast extent and its accuracy.

The Analog Ensembles method (AnEn), introduced by Luca Delle Monache in 2011 [1], is a post-processing tool that has shown good results to improve whether predictions or perform hindcasting (reconstruction of missing meteorological data). The goal of this study is to use the AnEn method to perform hindcasting, in order to reconstruct past weather conditions in a specific area of the northeeast of Portugal and verify its similarity with the actual forecast.

The AnEn method uses a two different time series: one with historical data (from a predictor station) and another with observed data (concerning a predicted station). The historical data is complete, while the observed data is missing or sparse in the prediction period. In Figure 1, which illustrates the methodology, a number of analogs are selected from the historical data set, according to their similarity to a predictor value (see step 1). At the same time instant, but at the predicted station, the corresponding observed data is selected and is used to produce a predicted value (step 3). This process is performed successively until the end of the prediction period data, and thereby it is possible to reconstruct the full predicted data from station 2. The AnEn method also allows using more than one predictor station (or more than on variable from the same station); in this scenario, the data from the predictor stations (or variable) can be used either dependently or independently (i.e., with the analogs selected in different predictor series having to overlap in time, or not).

The data for this research comes from weather stations managed by IPB and located in the northeast region of Portugal, near the villages of Edroso (*latitude* : 41.912778;

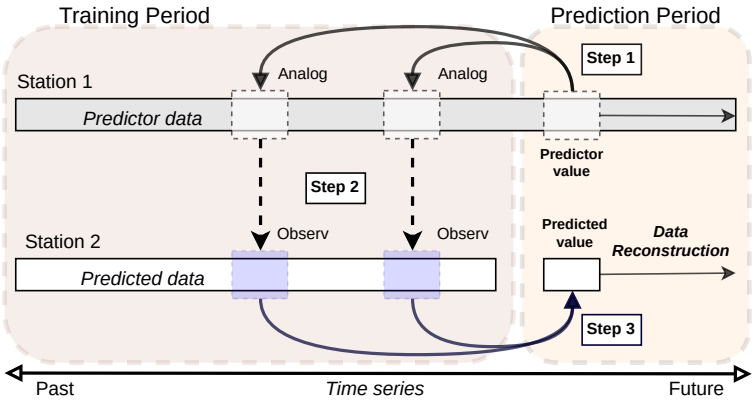

Fig. 1. Hindcasting with the Analogues Ensemble method.

longitude : −7.15283), Soutelo (*latitude* : 41.9211; *longitude* : −6.80852) and Valongo (*latitude* : 41.92305; *longitude* : −6.95083). Although reliable, the data series are often incomplete. To correct this problem, an interpolation was performed on the data, in order to standardise the sampling intervals to every 30 minutes. This interpolation process is limited to a maximum of 4 missing time intervals, since longer periods could distort the data too far beyond reality.

The AnEn simulations conducted were based on R [2] code adapted from previous work [3]. A new tool for the interpolation was also created in R. An algorithm for the conversion from csv to NetCDF format was developed (since NetCDF is the world standard format for meteorological data), and this conversion was performed for each station mentioned.

The year intervals were chosen based on the intersection between the weather station data. Then, the tests were run using the years of 2000 to 2006 as the predictor period, and the years 2006 to 2007 for the prediction period. As far as the climatic variables are concerned, the variables $PRES$ (Pressure), GST (Gust Speed) and $WSPD$ (Wind Speed) and $ATMP$ (Temperature) were used.

Table 1 presents the results of the experiments, which were divided into the error measures $Bias$ ($\frac{1}{n}\sum_{n}^{i=1}(x_i - y_i)$) and $RMSE$ ($\sqrt{\frac{1}{n}\sum_{n}^{i=1}(y_i - x_i)^2}$). Only Valongo was predicted, while Soutelo and Edroso stations were the predictors, either combined or not. These results were compared with previous work [3], in which the Ykt was predicted by the Dom and Ykr weather stations.

Soutelo and Edroso showed lower error rates in $WSPD$ and GST variables, while Dom and Ykr [3] showed superior performance in $PRES$ and $ATMP$ variables. In overall, using two stations simultaneously showed better results. In this scenario, the dependent-station method was also superior.

The data time intervals of the Dom and Ykr stations are shorter in comparison to this study (6 minutes versus 30 minutes), which promotes higher data resolution. Despite this, the prediction results obtained in the selected portuguese villages were similar to those observed at the stations of Dom and Ykr, indicating the effectiveness of the Monache method.

Variable	Dependency	**Soutelo**		**Edroso**		**Soutelo/Edroso**		**Dom/Ykr**	
		BIAS	*RMSE*	*BIAS*	*RMSE*	*BIAS*	*RMSE*	*BIAS*	*RMSE*
WSPD	Yes	-0.064	0.707	**-0.032**	0.851	-0.049	**0.670**	-0.206	2.075
	No					-0.061	0.725	-0.166	1.572
GST	Yes	-0.219	1.646	**-0.037**	1.870	-0.125	**1.467**	-0.530	2.197
	No					-0.141	1.644	-0.421	1.733
PRES	Yes	0.657	1.323	-0.832	2.008	-0.137	1.520	0.278	**0.497**
	No					**-0.061**	1.216	0.579	0.853
ATMP	Yes	-0.077	2.667	-0.142	2.690	-0.037	2.308	**0.001**	**1.071**
	No					-0.133	2.498	0.208	1.437

Table 1. Valongo variables predicted by Soutelo and Edroso. The results were compared to [3] (which predicted Ykt with the Dome and Ykr stations).

Therefore, the goal of this study was accomplished: the AnEn experiments with data from the stations in the villages of Edroso, Soutelo and Valongo were performed, and the results were compared with previous related work. Although the data series from the stations used in this study are not complete, they still managed to perform as well as higher data quality data series. To infer more reliable results, further tests with other variables and stations are needed.

Keywords: Analog Ensembles · Meteorological Data · Hindcasting.

References

1. Luca Delle Monache, Thomas Nipen, Yubao Liu, Gregory Roux, and Roland Stull. Kalman filter and analog schemes to postprocess numerical weather predictions. *Monthly Weather Review*, 139(11):3554 – 3570, 2011.
2. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
3. Leonardo Araújo, Carlos Balsa, C. Veiga Rodrigues, and José Rufino. Parametric study of the analog ensembles algorithm with clustering methods for hindcasting with multistations. In *Trends and Applications in Information Systems and Technologies*, pages 544–559, Cham, 2021. Springer International Publishing.