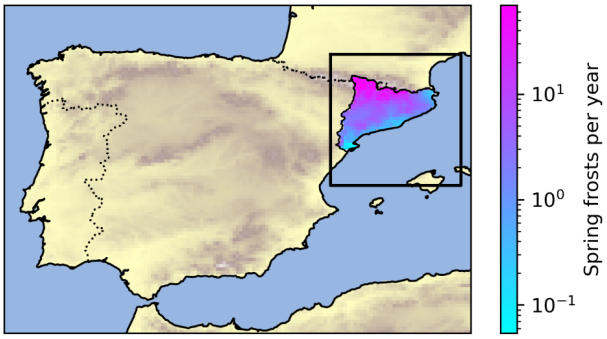
Funded by the European Union

# Statistical downscaling of seasonal spring-frost events for the agricultural case study

This archive contains statistically downscaled retrospective seasonal predictions of spring frost events in Catalonia. Essential details on the archived data are provided below. Please refer to Roncoroni and Athanasiadis (2025) for an extended discussion (preprint attached to this document).




## Scope and rationale

The data was produced in the framework of the European project ASPECT (Adaptation-oriented Seamless Predictions of European ClimaTe) and contributes towards ASPECT's agricultural case study: more information on the project and its case studies can be found at the ASPECT website https://www.aspect-project.eu/. The agricultural case study is led and coordinated by the Barcelona Supercomputing Center. This work is part of a coordinated effort to demonstrate the value of statistically-downscaled climate information for a number of societal sectors. Please refer to ASPECT's dedicated infosheet for information on ASPECT's statistical downscaling approach.

## Context and definitions

*Case-study:* the end user for the agricultural case study is a winemaking company based in the Spanish region of Catalonia.
*Domain*: the geographical domain is Catalonia. Grid points at high surface elevation (approximately above 700 m elevation) are not considered in the analysis.
*Spatial resolution:* the target horizontal spatial resolution for the downscaled climate information is of approximately 10 km.
*Predictand:* the predictand is March Frost Occurrence (MFO), a categorical variable which takes the values 1 (frost) and 0 (no frost).
*Time horizon:* Retrospective predictions are issued for the standard 1993-2016 verification period with a lead time of one month (February initialisation).
*Type of predictions:* The predictions are probabilistic and are continuously valued between 0 and 1. For example, a value of 0.99 denotes high predicted probability of MFO.

**Input data**

The archived data is computed based on observational, reanalysis and seasonal prediction data.
*Observations:* obtained from the E-OBS dataset, version 28.0. The spatial resolution is approximately 10 km. We use minimum temperature data over the period 1950–2021.
*Reanalysis data:* obtained from the ERA-5 dataset. The spatial resolution is approximately 100 km. We use 500 hPa geopotential height data over the period 1950–2021.
*Seasonal predictions data:* obtained from the C3S seasonal predictions portal. We use data produced by five different institutions: the Euro-Mediterranean Centre on Climate Change SPS3.5 (CMCC), the German Climate Forecast System 2.1 (DWD) the European Centre for Medium-Range Weather Forecasts SEAS5.1 (ECMWF), the Météo-France System8 (MF), and the UK Met Office GloSea6-GC3.2 version 602 (UKMO), for a total of 141 ensemble members. We analyse both the individual systems separately and the multi-system ensemble mean. The spatial resolution of the data is approximately 100 km. We use minimum temperature and 500 hPa geopotential height data over the period 1993–2016. The predictions are initialised around February 1st each year.
For all datasets, the original temporal resolution is daily, but the data is aggregated to monthly resolution during the analysis. The input data is not included in the archive.

**Methodology**

Two distinct sets of probabilistic March Frost Occurrence predictions are considered in Roncoroni and Athanasiadis (2025) and are made available in this archive:
*Dynamical predictions:* obtained by interpolating the daily minimum temperature data predicted by the C3S seasonal prediction systems from the original coarse grid (100 km) to the finer target grid (10 km), and subsequently computing MFO. The predictions from the best performing system (UKMO) only are available in this archive.
*Hybrid statistical–dynamical predictions:* obtained by leveraging: *(i)* a statistical relationship between MFO in Catalonia and large-scale patterns of monthly atmospheric blocking frequency. This relationship is identified using observational and reanalysis data. *(ii)* the predictability of such large-scale anomalous blocking patterns one month in advance. The predictability is assessed using renalysis and seasonal predictions data. Mathematically, the statistical relationship between blocking and frost is modelled with a logistic regression. The logistic regression model is trained with observational and reanalysis data. A leave-one-out cross-validation protocol is adopted to protect against overfitting. The final predictions are computed using multi-system seasonal prediction data. The detailed steps followed to produce the predictions are documented in Roncoroni and Athanasiadis (2025).

**Forecast evaluation**

The quality of the predictions is measured with two metrics, the ROC score and the Brier score, and contrasted against the climatology. The climatology is a common benchmark prediction which in most cases corresponds to the form of climate information already available to users.

Roncoroni and Athanasiadis (2025) found that both the dynamical and the hybrid statistical-dynamical predictions can improve with statistical downscaling in terms of forecast quality assessed against the climatology, although no individual prediction strategy is optimal across the domain and metrics considered. The improvements achieved with the hybrid statistical-dynamical method lack strong statistical significance, which hinders a conclusive statement on the robustness of the method in an operational setting. The same methodology may be applied also to downscale predictions in different areas and for different extremes in different calendar seasons.

**Output data format**

The data is archived as NetCDF files. The archive contains three files, named according to the convention outlined in the main deliverable document:
*obs_EOBS_MFO_m3.nc*: observed MFO in the target region over the verification period 1993–2016.
*dwnhyb_C3SMM_MFP_sm2_ft2.nc:* probability of MFO predicted with the hybrid statistical–dynamical method in the target region over the verification period 1993–2016.
*dwndyn_UKMO_MFP_sm2_ft2.nc*: probability of MFO predicted with the dynamical method (UKMO system only) in the target region over the verification period 1993–2016.
The data structure is identical for the three files. The data arrays are endowed with three logical dimensions corresponding to time (time), latitude (lat), and longitude (lon). Each dimension is associated with corresponding physical coordinates. Invalid values (e.g., for grid points at high surface elevation) are represented with NaNs. Units and additional information are stored in the data attributes. For illustration, Figure 1 shows the data structure of *obs_EOBS_MFO_m3.nc* as represented by the Python library xarray (left), and the time series of the observations, dynamical, and hybrid predictions at a specific grid point (right), together with the associated climatology.

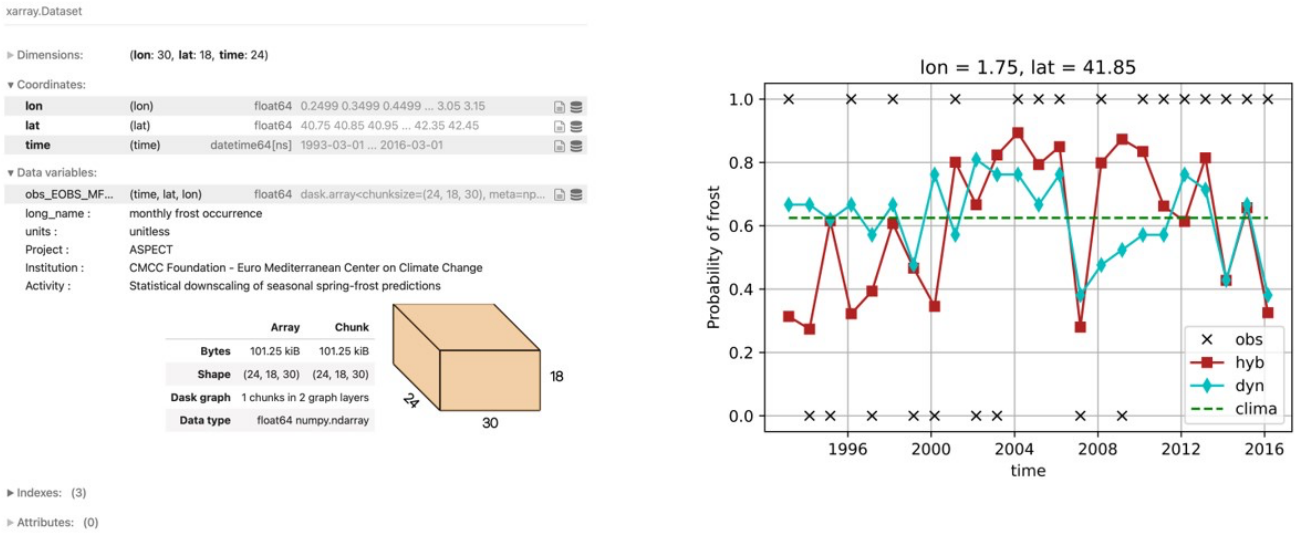



**Figure 1:** *Left: data structure of obs_EOBS_MFO_m3.nc as represented by the Python library xarray. Right: time series of the observations, dynamical, and hybrid predictions at a specific grid point, together with the associated climatology.*

## References

S. Roncoroni, P. Athanasiadis, *Statistical downscaling of seasonal spring-frost predictions*, Weather and Climate Extremes, 2025 (in review).

## Acknowledgements

This research was supported by the ASPECT project, which has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081460. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

*E-OBS data:* We acknowledge the E-OBS dataset and the data providers in the ECA&D project (https://www.ecad.eu).
Cornes, R., G. van der Schrier, E.J.M. van den Besselaar, and P.D. Jones. 2018: An Ensemble Version of the E-OBS Temperature and Precipitation Datasets, J. Geophys. Res. Atmos., 123. doi:10.1029/2017JD028200".

*ERA-5 data:* Contains modified Copernicus Climate Change Service information, 2024.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.bd0915c6.

*Seasonal predictions data:* Contains modified Copernicus Climate Change Service information, 2024.

Copernicus Climate Change Service, Climate Data Store, (2018): Seasonal forecast daily and subdaily data on single levels. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.181d637e.

Copernicus Climate Change Service, Climate Data Store, (2018): Seasonal forecast subdaily data on pressure levels. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). DOI: 10.24381/cds.50ed0a73.