SUNSET
the SUbseasoNal to decadal climate forecast post-processiNg and asSEssmenT suite

Victòria Agudetse[1], Lluis Palma[1], An-Chi Ho[1], Carlos Delgado-Torres[1], Nadia Milders[1], Eren Duzenli[1], Jaume Ramon[1], Alba Llabrés-Brustenga[1], Eva Rifà[1], Bruno de Paula Kinoshita[1], Pierre-Antoine Bretonnière[1], Núria Pérez-Zanón[1]

[1]Barcelona Supercomputing Center (BSC)

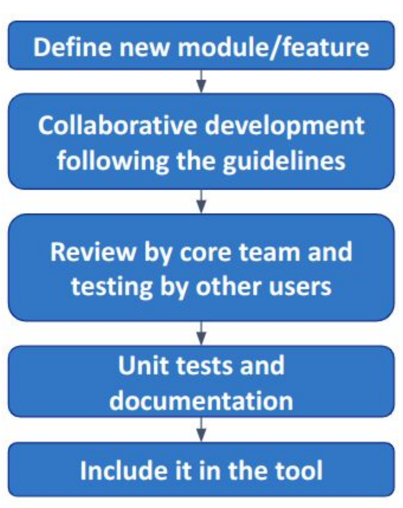
Introduction

SUNSET is an R-based tool that aims to provide climate services for sub-seasonal, seasonal and decadal time scales. The tool reads and post-processes climate forecast outputs by applying state-of-the-art methodologies to tailor climate products for each application and sector (e.g.: agriculture, energy, water management, or health).

Its modular design allows the technicians and researchers flexibility in defining the required post-processing steps, as well as the products definition by deciding on the forecast system and reference datasets, variables, and forecast horizon among others. The tool also allows for the creation and visualization of climate forecast products, such as maps for the most likely terciles, and performs the verification of the products, which can be visualized on maps and scorecards. The inclusion of a launcher script allows users to easily parallelize the computation in HPC machines.

```
>_ bash launch_SUNSET.sh <path_to_recipe>
   <path_to_script> --wallclock="01:00:00"
   --custom_directives="--constraint=medmem
   --exclusive"
```

Development Strategy



The development strategy for the next features to include is defined in regular user meetings and the priority of each feature is discussed by the core-team, which provides technical support for all user developments.

The GitLab pipeline integrates unit tests for different segments of the workflow, to ensure that changes to the code do not affect any functionality in unexpected ways.

Fig 1: Overview of the workflow to include a new feature or module in the tool.

The necessary information about the usage of the tool is stored in the Wiki section of the GitLab repository. Tutorials are provided in the form of use cases: specific examples of reproducible code with step-by-step instructions and detailed descriptions.

Modular structure

The SUNSET code features a modular design that provides flexibility with a fixed structure for module input and output, so that parts of the workflow can be easily arranged in the desired order, added, or skipped.

Loading

The data requested by the user is retrieved from netCDF or GRIB files and loaded into R as an array with named dimensions, with the associated metadata attached. Regridding and unit conversion are performed if requested.

Anomalies | **Calibration** | **Downscaling** | **Indices** | **Indicators**

Several modules are available to perform required post-processing steps such as anomaly computation,nbias adjustment or downscaling, with various methodologies. The tool can also perform the computation of climate indices (e.g. NAO, El Niño) and tailored sectoral indicators for climate services (e.g. SPI and SPEI).

Probabilities | **Skill**

The output of any of the above modules can be used to compute probabilities based on specified thresholds and to evaluate the model skill against the reference dataset.

Visualization | **Scorecards**

The outputs of the modules can be visualized in maps, scorecards and time series plots.

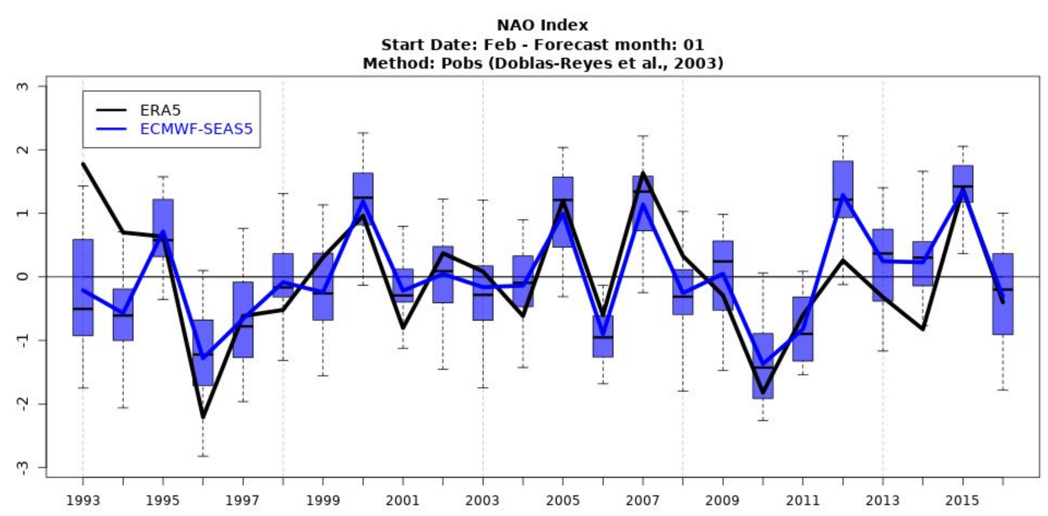


Fig 2: Example of Indices module output. Time series of standardized NAO index, comparing the reference dataset (ERA5) with the ECMWF SEAS5 mean and ensemble spread.

How to use it

SUNSET is based on user-defined configuration files, named **recipes**. The recipe contains the information of the variables and systems to evaluate, the region and time period of the evaluation, which results to save and to which directory, as well as the details of each step of the workflow.

Users build a simple script calling the modules needed for their workflow: each module is an R function that takes the recipe as an input parameter. The execution of the recipe and script is handled by the SUNSET Launcher. For a smaller test or an analysis of a single dataset, the code can also be run directly on the terminal or interactively in an R session.

Parallelization

Often, the same analysis needs to be performed for different models, variables, and initialization dates. SUNSET allows the user to create one recipe defining a common workflow for multiple datasets. When running the launcher script, the recipe is split into individual verifications (atomic recipes), which are run in parallel on the computer cluster with the option to use SLURM directly or through Autosubmit as the workflow manager.

To take full advantage of HPC capabilities and deal with memory limitations, an additional layer of parallelization can be added to each atomic workflow, through lazy loading and domain decomposition. The data can be chunked along temporal or spatial dimensions, depending on the specific analysis, and each chunk is processed separately until the data size is reduced. Afterwards, the numeric arrays and corresponding metadata are pieced back together and the workflow can continue.

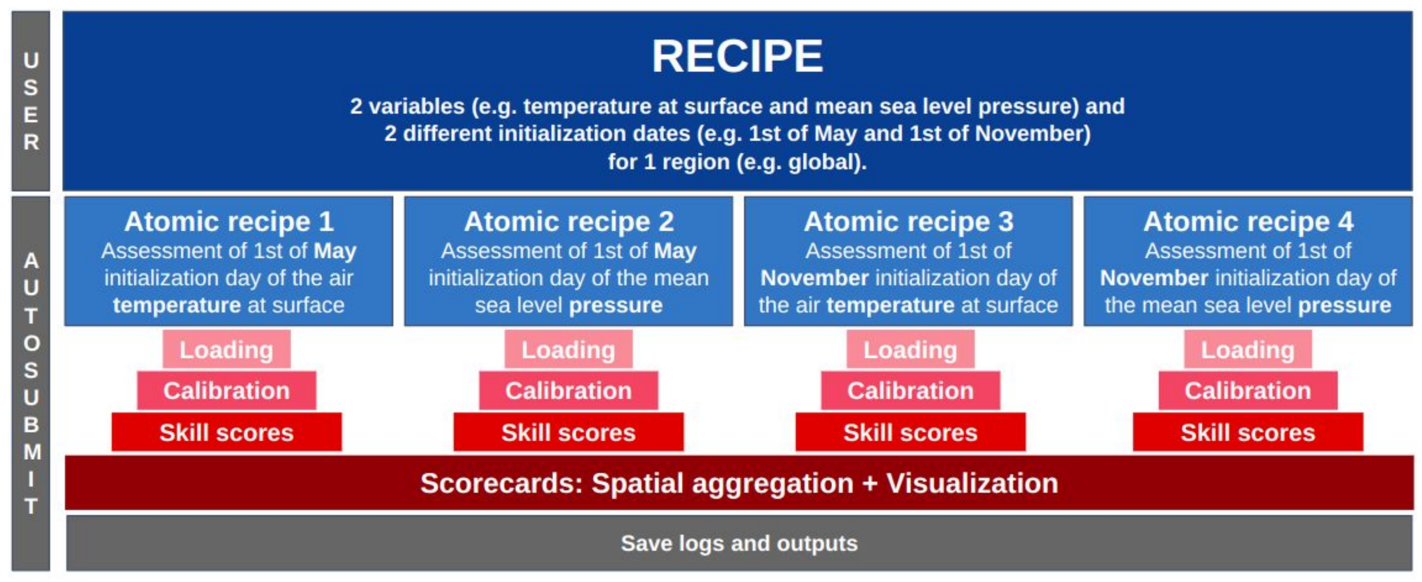


Fig 3: Diagram showing an example of recipe splitting and parallelization of the workflow on an HPC cluster.

Future plans

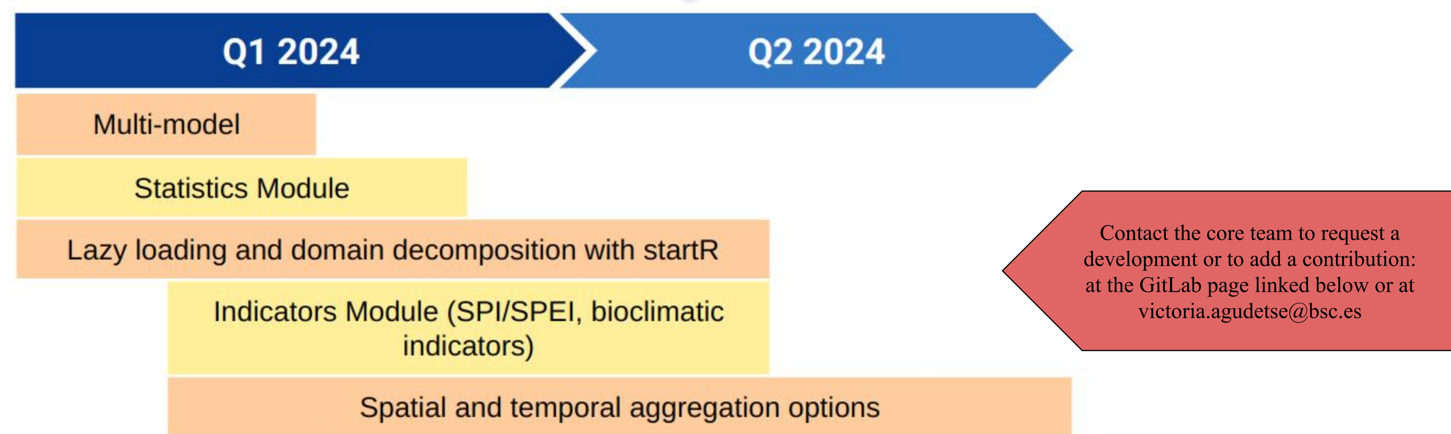


Fig 4: Current developments conducted by users and the core-team developers.

Portability

SUNSET uses a configuration file, the archive, which contains information about the datasets available in the filesystem: variable names, paths, grid descriptions, and other relevant details. Multiple entries can be added to the archive to allow usage in environments with different file structures or data availability.

The software requirements include several R packages aimed towards climate forecast post processing such as startR (Manubens et al., 2023), s2dv (Manubens et al., 2018), CSTools (Pérez-Zanón et al., 2022) and CSIndicators (Pérez-Zanón et al., 2023). Climate Data Operator (CDO; Schulzweida, 2019) is also a requirement, and the Autosubmit workflow manager (Manubens-Gil et al., 2016) is optional for more user-friendly parallelization.

A conda environment including R and all of the software dependencies has been created to facilitate portability. The team is also working on a Docker container for easy installation.

Visualization

With the Visualization module, the forecast products and skill assessment metrics computed during the workflow can be displayed in the form of maps and scorecards, with options for different color palettes, map projections and single-panel or multi-panel layouts. It is also possible to generate Scorecards (fig. 7) to display an overview of model performance across regions, initialization dates and forecast times.

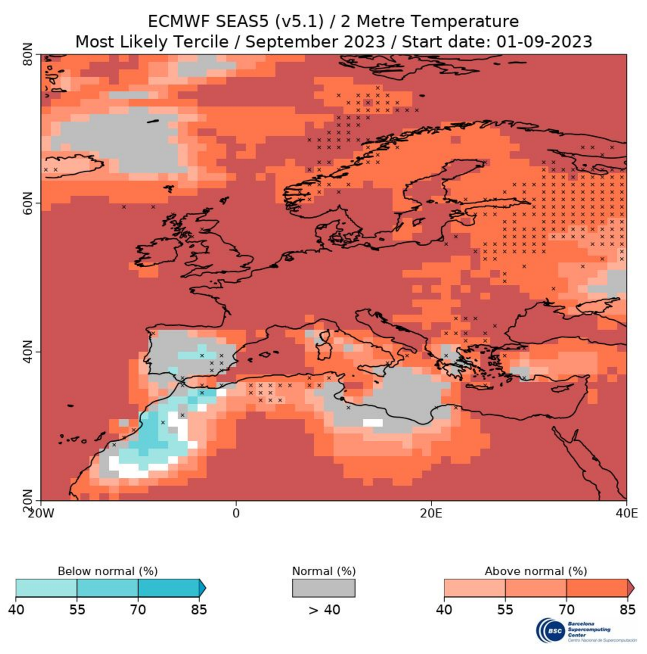


Fig 5: Most-likely terciles for September 2023 2 metre temperature, for the ECMWF SEAS5 (v5.1) September-initialized forecast. Crosses indicates grid points with negative Rank Probability Skill Score (RPSS).

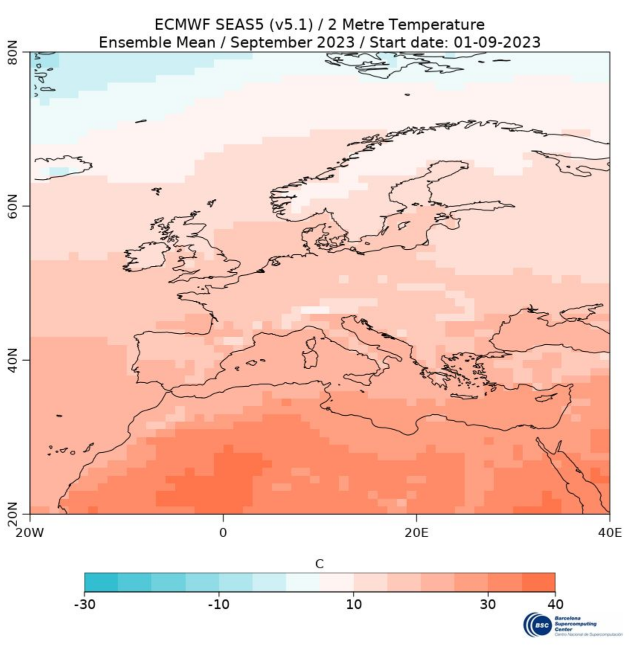


Fig 6: Forecast ensemble mean for September 2023 2 metre temperature, for the ECMWF SEAS5 (v5.1) September-initialized forecast.

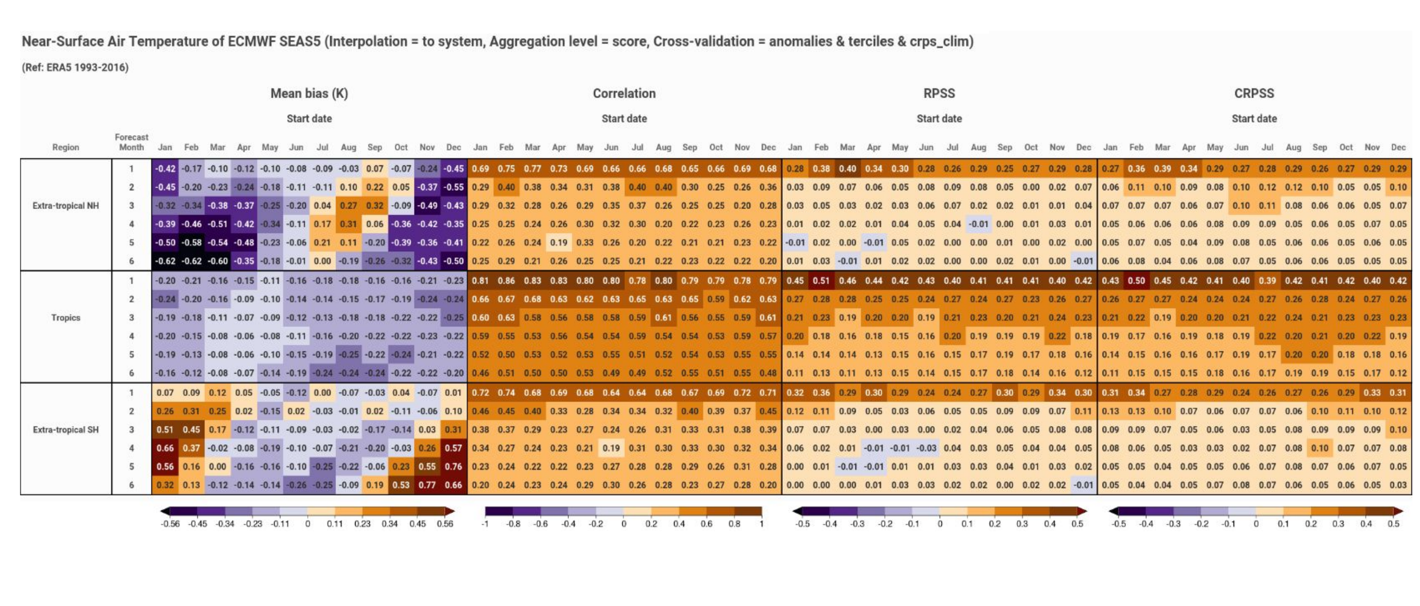


Fig 7: Scorecard displaying the near-surface air temperature mean bias, correlation, RPSS and continuous rank probability skill score (CRPSS) values for the 1993-2016 hindcast period in three regions (tropics and extra-tropics) for each initialization date of ECMWF SEAS5, computed using ERA5 as the reference dataset. The user can decide which metrics and regions to display in the scorecard.

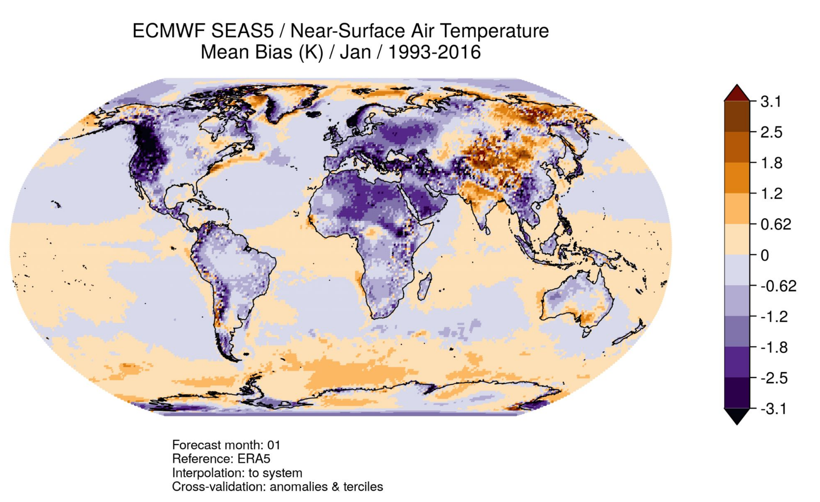


Fig 8: Robinson projection map, showing the mean bias of January near-surface air temperature for ECMWF SEAS5 evaluated against ERA5 for the 1993-2016 hindcast period, initialized in January.

Acknowledgments

The CERISE project (grant agreement No101082139) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union

COORDINATED BY ECMWF

Corresponding authors:
victoria.agudetse@bsc.es, nuria.perez@bsc.es

References

Doblas-Reyes, F. J., Pavan, V., & Stephenson, D. B. (2003). The skill of multi-model seasonal forecasts of the wintertime North Atlantic Oscillation. Climate Dynamics, 21(5–6). https://doi.org/10.1007/s00382-003-0350-4

Manubens-Gil, D., Vegas-Regidor, J., Prodhomme, C., ... (2016). Seamless management of ensemble climate prediction experiments on HPC platforms. 2016 International Conference on High Performance Computing and Simulation, HPCS 2016. https://doi.org/10.1109/HPCSim.2016.7568429

Manubens, N., Caron, L.-P., Hunter, A., … (2018). An R package for climate forecast verification. Environmental Modelling and Software, 103, 29–42. https://doi.org/10.1016/j.envsoft.2018.01.018

Manubens, N., Ho, A., Pérez-Zanón, N., … (2023). startR: Automatically Retrieve Multidimensional Distributed Data Sets. https://CRAN.R-project.org/package=startR

Pérez-Zanón, N., Caron, L.-P., Terzago, S., ... (2022). Climate Services Toolbox (CSTools) v4.0: from climate forecasts to climate forecast information. Geoscientific Model Development, 15(15), 6115–6142. https://doi.org/10.5194/GMD-15-6115-2022

Pérez-Zanón, N., A. Ho, C. Chou, L. Lledó, R. Marcos-Matamoros, E. Rifà and N. González-Reviriego (2023). CSIndicators: Get tailored climate indicators for applications in your sector. Climate Services. https://doi.org/10.1016/j.cliser.2023.100393

Schulzweida, U. (2019). CDO User Guide (Version 1.9.8). https://doi.org/http://doi.org/10.5281/zenodo.3539275

The **SUNSET** repository:
https://earth.bsc.es/gitlab/es/sunset