

CopERNicus climate change Service Evolution



D5.6 Conclusions of the common assessment of the impact of improvements in land initialization on forecast performance

Due date of deliverable	31-Oct-2026
Submission date	22-Oct-2026
File Name	CERISE-D5-6-V3-0
Work Package /Task	WP5 / Task 5.3
Organisation Responsible of Deliverable	Barcelona Supercomputing Center (BSC)
Author name(s)	Núria Pérez-Zanón
Revision number	V3.0
Status	Issued
Dissemination Level	Public



The CERISE project (grant agreement No 101082139) 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.

1 Executive Summary

This deliverable presents the work carried out under Task 5.3, focusing on the common assessment of climate prediction demonstrators. The assessment involved the deployment of the SUNSET software suite on ECMWF infrastructure to perform the assessment of all seasonal forecast demonstrators developed under Task 5.2, the selection of reference datasets, variables and metrics, and the processing of climate predictions to evaluate forecast performance across different temporal and spatial scales. Five institutions provided 19 experiments across three different phases. For each of them, more than 1152 maps were obtained (i.e. considering 6 variable, 4 initialisation dates, 6 forecast times - 2 of them being 3-monthly aggregations - and 8 metrics).

The main outcomes of the work are:

- The capabilities of SUNSET for the post-processing and verification of climate predictions have enabled a common assessment of all CERISE seasonal forecast demonstrators.
- The scorecards provide a useful tool for exploring the overall performance of the demonstrators, while spatial visualisations remain necessary for a more detailed understanding of the results.
- The web application provides an accessible way for model developers and other interested members of the community to explore and analyse the assessment results.

This deliverable summarises the description of the methodology, the different visualisation products and the access to the web application. These results contribute to the overall objectives of the project by providing a common framework for the assessment of seasonal forecast demonstrators and facilitating the exploration and interpretation of their performance.

Table of Contents

1	Executive Summary	2
2	Introduction	4
2.1	Background.....	4
2.2	Scope of this deliverable	5
2.2.1	Objectives of this deliverable	5
2.2.2	Work performed in this deliverable.....	5
2.2.3	Deviations and counter measures.....	6
2.2.4	Reference Documents	6
2.2.5	CERISE Project Partners:.....	6
3	Data and Methods.....	7
3.1	Data	7
3.2	Methods.....	9
4	Results.....	10
5	Conclusion	12
6	References	13

2 Introduction

Ensemble seasonal forecasts capture many important sources of predictability, but forecast models are affected by systematic errors and biases that can accumulate over time and degrade forecast quality. As model errors can differ between forecasting systems, combining outputs from multiple systems can partially compensate for these errors and provide more robust forecast products. This has led to the development of the C3S multi-system seasonal forecast system. The different systems included in the multi-system vary not only in the models they use, but also in their representation and analysis of different components of the initial conditions, in particular the sub-surface ocean. Land initial conditions and their associated uncertainties are also important for seasonal prediction, but they have received comparatively less attention to date.

The seasonal prediction demonstrators developed within CERISE have spanned three phases, each testing different approaches to the initialisation of the land surface in seasonal prediction systems. In each phase, participating centres tested sets of initial conditions appropriate to the formulation of their respective forecasting systems. The demonstrators therefore provide an opportunity to assess the impact of different land initialisation approaches on the quality of seasonal forecasts of a set of atmospheric Essential Climate Variables (ECVs).

This deliverable (D5.6) presents the results of the common assessment of the seasonal forecast demonstrators carried out under Task 5.3. The assessment provides a consistent framework for evaluating the different demonstrators across the three development phases, using a common set of reference datasets, variables and verification metrics. The analysis covers different forecast initialisation dates, forecast times and spatial scales, enabling the performance of the demonstrators to be assessed from a common perspective.

Three complementary activities were carried out to meet the objectives of Task 5.3. First, a common scientific protocol was defined, specifying the datasets, variables, metrics and evaluation procedures used in the assessment. Second, the SUNSET software suite was deployed on ECMWF infrastructure to enable the systematic processing and verification of the seasonal forecasts from all demonstrators. Finally, the resulting assessment products were made available through a web application, providing an accessible interface for exploring and analysing the results. Together, these activities establish a common framework for the assessment of the CERISE seasonal forecast demonstrators and facilitate the comparison, interpretation and exploration of their forecast performance.

2.1 Background

The scope of CERISE is to enhance the quality of the C3S reanalysis and seasonal forecast portfolio, with a focus on land-atmosphere coupling.

It will support the evolution of C3S, over the project's 4-year timescale and beyond, by improving the C3S climate reanalysis and the seasonal prediction systems and products towards enhanced integrity and coherence of the C3S Earth system Essential Climate Variables.

CERISE will develop new and innovative ensemble-based coupled land-atmosphere data assimilation approaches and land surface initialisation techniques to pave the way for the next generations of the C3S reanalysis and seasonal prediction systems.

These developments will be combined with innovative work on observation operator developments integrating Artificial Intelligence (AI) to ensure optimal data fusion fully integrated in coupled assimilation systems. They will drastically enhance the exploitation of past, current, and future Earth system observations over land surfaces, including from the Copernicus Sentinels and from the European Space Agency (ESA) Earth Explorer missions,

moving towards an all-sky and all-surface approach. For example, land observations can simultaneously improve the representation and prediction of land and atmosphere and provide additional benefits through the coupling feedback mechanisms. Using an ensemble-based approach will improve uncertainty estimates over land and lowest atmospheric levels.

By improving coupled land-atmosphere assimilation methods, land surface evolution, and satellite data exploitation, R&I inputs from CERISE will improve the representation of long-term trends and regional extremes in the C3S reanalysis and seasonal prediction systems.

In addition, CERISE will provide the proof of concept to demonstrate the feasibility of the integration of the developed approaches in the core C3S (operational Service), with the delivery of reanalysis prototype datasets (demonstrated in pre-operational environment), and seasonal prediction demonstrator datasets (demonstrated in relevant environment).

CERISE will improve the quality and consistency of the C3S reanalysis systems and of the components of the seasonal prediction multi-system, directly addressing the evolving user needs for improved and more consistent C3S Earth system products.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

The main objective of this deliverable is to provide a common and consistent assessment of the seasonal forecast demonstrators developed under Task 5.2. To achieve this objective, the deliverable aims to:

- define a common scientific protocol for the assessment of the seasonal forecast demonstrators, including the selection of reference datasets, variables and verification metrics;
- systematically process and verify the forecasts produced by the different demonstrators across the three development phases;
- assess forecast performance across different forecast times and spatial scales;
- provide a set of standardised visualisation products, including scorecards and spatial maps, to facilitate the interpretation of the assessment results; and
- provide an accessible web application through which model developers and other interested members of the community can explore and analyse the results.

2.2.2 Work performed in this deliverable

In this deliverable the work outlined in The Description of Action (WP5 T5.3) is described. In particular, Task 5.3 foresees the design, implementation and execution of a general evaluation of the seasonal forecast demonstrators using common metrics, software and reference data, as a first step towards assessing the impact of improved land initial conditions on forecast quality.

To fulfil this objective, a common scientific protocol was defined and applied to the seasonal forecast demonstrators developed across the three phases of WP5. The assessment was implemented using the SUNSET software suite (Pérez-Zanón et al., 2022, 2024) and deployed on ECMWF infrastructure, enabling the systematic processing and verification of the forecast data provided by the participating institutions. The resulting assessment products, including scorecards and spatial visualisations, were subsequently made accessible through a web application. This also supports the dissemination and exploration of the results by the project partners and the wider community and provides an output that can be used by WP6, which will integrate this assessment with a range of other tests of seasonal prediction system fidelity to create an overview of the benefits of initialisation with land surface assimilation.

The work therefore addresses the planned activities of Task 5.3 by providing a common evaluation framework and associated tools for the systematic assessment of the seasonal forecast demonstrators.

2.2.3 Deviations and counter measures

No deviations have been encountered.

2.2.4 Reference Documents

[1] Project 101082139- CERISE-HORIZON-CL4-2021-SPACE-01 Grant Agreement

2.2.5 CERISE Project Partners:

ECMWF	European Centre for Medium-Range Weather Forecasts
Met Norway	Norwegian Meteorological Institute
SMHI	Swedish Meteorological and Hydrological Institute
MF	Météo-France
DWD	Deutscher Wetterdienst
CMCC	Euro-Mediterranean Center on Climate Change
BSC	Barcelona Supercomputing Centre
DMI	Danish Meteorological Institute
Estellus	Estellus
IPMA	Portuguese Institute for Sea and Atmosphere
NILU	Norwegian Institute for Air Research
MetO	Met Office

3 Data and Methods

3.1 Data

During the project, each of the five institutions (i.e. CMCC, DWD, ECMWF MetOffice, MeteoFrance) providing the seasonal forecast demonstrators has described their different set of experiments in previous deliverables (i.e. D5.1 to D5.5). Key required for the common assessment, such as the available initialisation dates, are described in Table 1. For phase 0, CMCC, MeteoFrance and MetOffice (UKMO) used their operational forecast system at the time of starting the project in 2022.

Table 1. Summary of the 19 seasonal climate prediction experiments conducted by each institution and project phase, including the study period, forecast length, ensemble size, initialisation dates, and experiment identifier.

Phase	Institution	Period		Forecast length	Ensemble size	Ini. dates	Exp ID
ZERO	DWD	1993	2022	4	25	Feb/May/Aug/Nov	DWD-Phase0
	CMCC	1993	2016	6	30	All	CMCC-SPS3.5
	MeteoFrance	1993	2016	6	25	All	MeteoFrance-System 8
	UKMO	1993	2018	6	28	All	UKMO-S602
	ECMWF	1981	2022	4	51	May/Nov	ECMWF-Systemi2o2
ONE		1993	2022	4	25	Feb/Aug	
	DWD	1993	2022	4	25	Feb/May/Aug/Nov	DWD-Phase1
	CMCC	1993	2022	4	30	Feb/May/Aug/Nov	CMCC-Phase1
	MeteoFrance	1993	2022	4	30	Feb/May/Aug	MeteoFrance-Phase1
		1993	2021			Nov	
	UKMO	1993	2022	4	20	Feb/May/Aug/Nov	UKMO-Phase1
	ECMWF	1993	2023	4	51	May/Nov	ECMWF-System201
						Feb/May/Aug/Nov	ECMWF-System202
						May/Nov	ECMWF-System203
						May/Nov	ECMWF-System204
TWO	DWD	1993	2022	4	25	Feb/May/Aug/Nov	DWD-Phase2
	CMCC	2002	2021	4	25	Feb/May/Aug/Nov	CMCC-Phase2
	MeteoFrance	1993	2022	4	30	Feb/May/Aug	MeteoFrance-Phase2
		1993	2021			Nov	
	UKMO	2000	2019	4	20	Feb/May/Aug/Nov	UKMO-Phase2
	ECMWF	2000	2019	4	25	May/Nov	ECMWF-Systm301
							ECMWF-System302

CMCC Phase 0 corresponds to the C3S-operational CMCC-CM2-v20191201 system (CESM1.2-based), which is used as a benchmark for evaluating subsequent system upgrades. Phase 1 introduces the CMCC-SPS4 prototype based on the CMCC-CM3 Earth System model (CESM2-based) and constitutes the new operational system. In this configuration, the land component includes interactive biogeochemistry and dynamic vegetation, replacing the prescribed satellite phenology used in previous versions. Phase 2 builds upon SPS4 by updating land initial conditions using the CMCC Land Reanalysis, which assimilates Leaf Area Index (LAI), Snow Water Equivalent (SWE), and Soil Moisture (SM), providing physically consistent land states and improving land–atmosphere coupling.

DWD phase 0 corresponds to a set of baseline experiments conducted within the CERISE project, representing the reference configuration of the seasonal prediction system. All phases are based on the ICON general circulation model. Initial conditions are derived from ocean temperature and salinity from EN4, while the atmosphere is constrained through nudging against ERA5. Phase 1 builds on this baseline by additionally applying a two-dimensional

variational (2DVAR) analysis for snow depth, improving the representation of land surface initial conditions. Phase 2 further extends the system by incorporating a variational grid point analysis of LAI, enhancing the consistency of vegetation-related initial conditions within the coupled system.

Météo-France phase 0 corresponds to the operational System 8 configuration, in which the atmosphere is initialized through a coupled nudging strategy and land initial conditions are only indirectly constrained. In Phase 1, soil moisture and soil temperature initial conditions are updated using LDAS-Monde simulations with LAI assimilation. Phase 2 further improves the land initialization by incorporating snow fields from ERA5 while retaining the Phase 1 soil initialization.

Met Office phase 0 used the GloSea6 GC3.2 seasonal forecasting system, which was the operational forecasting system provided through C3S at the beginning of the project. Land initial conditions were generated using the JULES-GL8 land model forced with atmospheric data from JRA-55, resulting in a model-consistent but only weakly constrained land surface state. Phase 1 improved the land initialisation by replacing soil moisture, soil temperature and snow fields with values derived from the ERA6-Land-Pv1.3 reanalysis. Phase 2 further refined the land initial conditions using ERA6-Land-Pv2.0, incorporating seasonally and interannually varying LAI and land use/land cover information, together with improved soil moisture initialisation.

ECMWF phase 0 used ERA5-based land initial conditions together with Integrated Forecast System (IFS) cycle 48r1, representing the state-of-the-art configuration available at the start of the project. Phase 1 upgraded the atmospheric model to cycle 49r1 and investigated the impact of land initialisation through four alternative configurations based on ERA5, snow and soil moisture assimilation, snow-only assimilation, and an open-loop simulation forced by ERA5. Phase 2 further extended the system by incorporating time-varying vegetation and land use/land cover information within cycle 49r2b, allowing the impact of evolving vegetation conditions and land surface characteristics on seasonal forecast skill to be assessed.

To assess the simulations from each institution across phases, the common period for each of them has been selected to perform the verification (Table 2). Although several ECVs have been considered for the common assessment, this report focuses on the monthly mean of 2-metre air temperature and monthly total accumulated precipitation.

Table 2. Common period across experiments and phases for each institution.

Institution	Period
DWD	1993-2022
CMCC	2002-2016
MeteoFrance	1993-2016
UKMO	2000-2018
ECMWF	2000-2019

Reference datasets used for the verification of the seasonal climate predictions include ERA5, JRA-3Q and GPCP. ERA5 (Hersbach et al., 2020), the latest global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) within the Copernicus Climate Change Service (C3S), provides a physically consistent reconstruction of the climate system through the assimilation of a wide range of observations into a numerical weather prediction model. ERA5 data were used at a horizontal resolution of $0.25^\circ \times 0.25^\circ$. As an independent reanalysis reference, JRA-3Q (Kobayashi et al., 2024) was also employed. Produced by the Japan Meteorological Agency, JRA-3Q provides a homogeneous representation of atmospheric conditions and was used at its native horizontal resolution of approximately $0.625^\circ \times 0.625^\circ$.

Precipitation was verified against the Global Precipitation Climatology Project Version 3.2 (GPCP V3.2) dataset (Adler et al., 2018; Huffman et al., 2023). The monthly GPCP V3.2 product provides a long-term, globally consistent precipitation record at $0.5^\circ \times 0.5^\circ$ spatial resolution by combining multiple satellite retrievals with rain-gauge analyses through a suite of bias-adjustment and merging procedures. Over land, satellite estimates are merged with gauge observations from the Global Precipitation Climatology Centre (GPCC), while over the oceans the product relies primarily on microwave and infrared satellite observations.

3.2 Methods

A set of verification metrics was selected to account for association, accuracy, resolution, and reliability of the forecast systems. These metrics have been calculated grid point wise and the reference datasets have been regridded using a conservative method to the system grid, i.e. $1^\circ \times 1^\circ$ using CDO (Schulzweida, 2019). Except for the mean bias, the rest of the metrics have been calculated over anomalies. The climatological information, i.e. climatological means, probability thresholds, and reference climatological forecast, has been obtained by applying leave-one-out cross-validation.

The mean bias is the mean difference between the forecast ensemble mean and the reference dataset. The statistical significance of the mean bias was assessed using Welch's test (Welch, 1947), with the null hypothesis that the mean bias is equal to zero. All the significant tests, including the Welch's test, are considered statistically significant at the 95% confidence level ($p < 0.05$).

The rest of the metrics have been calculated over the monthly anomalies. The ensemble mean correlation, using Pearson's method (Wilks, 2011), has been considered as a measure of the association between forecast and reference datasets. The statistical significance of the correlation coefficients was assessed using a one-sided Student's t-test.

To assess the accuracy, resolution and reliability of the forecast systems, two probabilistic metrics have been calculated: the Ranked Probability Skill Score (RPSS) on tercile categories and the Continuous Ranked Probability Skill Score (CRPSS). Both metrics are calculated using their fair versions (Ferro, 2014), and statistical significance is assessed using the two-sided approximation of the Random Walk test (DelSole and Tippet, 2016), under the null hypothesis that the forecast system and the climatological forecast have equal skill. The corresponding RPS and CRPS are also reported to aid the interpretation of the skill scores, as they provide a measure of the absolute forecast performance rather than the improvement relative to climatology.

Finally, the spread-to-error ratio quantifies ensemble reliability through the consistency between ensemble spread and the error of the ensemble mean. Statistical significance is assessed using a two-sided Fisher's test under the null hypothesis of a unit spread-to-error ratio. The root mean square error is also provided to better facilitate its interpretation.

Given the large number of maps generated by the assessment, scorecards were used to provide a compact overview of the general performance of the experiments. For example, considering one provider and one variable across the three phases, the four initialisation dates and four forecast times resulted in 384 maps when the eight metrics were considered ($4 \text{ initialisation dates} \times 4 \text{ forecast times} \times 3 \text{ phases} \times 8 \text{ metrics}$). The scorecards summarise this information by applying a weighted mean to the metric values over the North and South Extra-Tropics and the Tropics. This provides a concise overview of the overall performance while retaining the information from the different metrics, initialisation dates, forecast times and phases.

4 Results

The results are available at a web application developed by BSC that allows friendly access and visualisation of maps and scorecards (Figure 1).

Earth System Services

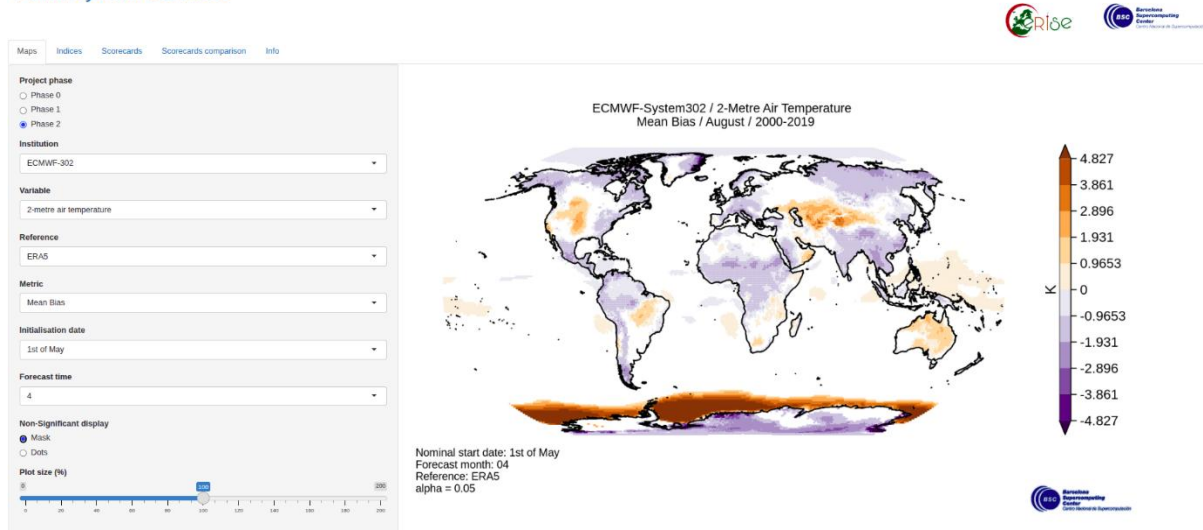


Figure 1. Screenshot of the web application <https://earth.bsc.es/shiny/cerise/>. The top bar allows the selection of different visualisation formats and documentation, while the left panel displays a menu to select experiments and assessment options. To get access contact nuria.perez@bsc.es.

The Maps tab enables users to explore the spatial characteristics of the seasonal forecast verification results. The user can select the project phase, forecasting institution, variable, reference dataset, verification metric, initialisation date, and forecast lead time. The application dynamically displays the corresponding verification map generated by the post-processing and evaluation workflow. In addition, users can control the representation of statistically non-significant results, either by masking them or by indicating them with dots and adjust the plot size. This functionality provides an interactive way to examine the spatial distribution of forecast performance across different forecast systems, variables, initialisation dates and lead times.

The Scorecards tab presents a summary of 5 metrics for each experiment and variable. As an example, Figure 2 shows a scorecard of 2-metre air temperature for the CMCC SPS3.5 system, which was the phase 0 demonstrator of the project.

2-metre air temperature of CMCC-SPS3.5

(Ref: ERA5 2002-2016)

		Mean bias (K)				Correlation				RPSS				CRPSS				Spread-to-error ratio			
		Start date				Start date				Start date				Start date				Start date			
Region	Forecast Month	Feb	May	Aug	Nov	Feb	May	Aug	Nov	Feb	May	Aug	Nov	Feb	May	Aug	Nov	Feb	May	Aug	Nov
Extra-tropical NH	1	-1.15	-0.67	0.11	-0.53	0.69	0.58	0.55	0.58	0.27	0.17	0.13	0.15	0.25	0.13	0.08	0.15	0.99	1.03	0.94	0.93
	2	-0.92	-0.84	-0.37	-0.74	0.42	0.21	0.26	0.32	0.07	-0.04	-0.02	0.00	0.07	-0.07	-0.02	0.02	1.06	1.17	1.05	1.00
	3	-1.13	-0.47	-0.39	-1.02	0.30	0.19	0.12	0.17	-0.02	-0.05	-0.07	-0.04	0.00	-0.13	-0.04	-0.02	1.06	1.14	0.93	1.03
	4	-1.21	-0.52	-0.18	-0.78	0.19	0.18	0.15	0.23	-0.04	-0.06	-0.06	-0.03	-0.03	-0.07	-0.02	-0.01	1.11	1.10	0.93	0.98
Tropics	1	0.22	0.26	0.17	0.24	0.76	0.70	0.71	0.68	0.33	0.23	0.26	0.24	0.30	0.21	0.23	0.20	0.83	0.80	0.78	0.84
	2	0.29	0.16	0.20	0.17	0.60	0.53	0.57	0.56	0.18	0.11	0.15	0.17	0.17	0.09	0.12	0.15	0.94	0.87	0.89	0.96
	3	0.30	0.09	0.10	0.13	0.55	0.47	0.51	0.52	0.12	0.06	0.11	0.12	0.12	0.02	0.09	0.12	0.99	0.89	0.95	0.98
	4	0.23	-0.01	0.00	0.14	0.44	0.41	0.45	0.50	0.03	0.00	0.08	0.11	0.04	-0.03	0.07	0.10	0.94	0.89	1.02	0.94
Extra-tropical SH	1	0.56	-0.57	-0.41	0.39	0.67	0.60	0.61	0.65	0.26	0.16	0.18	0.19	0.24	0.17	0.17	0.20	0.91	0.91	0.84	0.93
	2	-0.05	-0.65	-0.36	1.07	0.38	0.26	0.32	0.37	0.04	-0.05	0.01	0.02	0.05	-0.02	0.03	0.01	0.88	0.91	0.88	0.87
	3	-0.24	-0.56	-0.19	1.22	0.27	0.21	0.32	0.33	-0.04	-0.05	-0.02	0.00	-0.01	-0.02	0.02	-0.01	0.85	0.93	0.95	0.95
	4	-0.40	-0.74	0.24	0.78	0.20	0.16	0.29	0.32	-0.07	-0.10	-0.02	-0.01	-0.04	-0.06	0.00	0.03	0.82	0.85	0.91	0.91

Figure 2. Scorecard showing the verification results for 2-metre air temperature from the CMCC-SPS3.5 forecast system against ERA5 for 2002–2016. Results are presented for forecast months 1–4 and the extra-tropical Northern Hemisphere (NH), Tropics and extra-tropical Southern Hemisphere (SH), and for initialisation dates in February, May, August and November. Columns show mean bias, correlation, RPSS, CRPSS and the spread-to-error ratio. The colour shading indicates the magnitude and sign of each metric.

The **Scorecards comparison** tab provides a compact summary of the forecast verification results for each of the regions (i.e. Tropics, extra-tropical Southern Hemisphere or extra-tropical Northern Hemisphere), allowing users to compare the performance of different forecast phases for a given institution, variable or index and reference dataset (Figure 3). This representation facilitates the comparison of forecast performance across forecast lead times, initialisation dates and system phases.

2-metre air temperature. Region: Extra-tropical NH

(Ref: ERA5 1993-2016)

Forecast month	System	Mean bias (K)				Correlation				RPSS				CRPSS				Spread-to-error ratio			
		Start date				Start date				Start date				Start date				Start date			
		Feb	May	Aug	Nov	Feb	May	Aug	Nov	Feb	May	Aug	Nov	Feb	May	Aug	Nov	Feb	May	Aug	Nov
1	Meteo-France System 8	-0.93	1.21	1.54	0.83	0.50	0.40	0.45	0.36	0.15	0.07	0.09	0.06	0.11	0.05	0.06	0.04	1.02	1.05	1.10	0.90
	MeteoFrance-Phase1	-1.10	0.50	0.66	0.54	0.63	0.50	0.55	0.54	0.20	0.07	0.11	0.11	0.11	0.04	0.05	0.10	0.68	0.73	0.78	0.72
	MeteoFrance-Phase2	-0.37	0.76	0.54	0.66	0.63	0.51	0.55	0.55	0.21	0.09	0.12	0.12	0.15	0.06	0.06	0.10	0.69	0.76	0.79	0.73
2	Meteo-France System 8	-0.79	2.01	1.13	0.07	0.32	0.19	0.22	0.16	0.05	-0.01	-0.02	-0.03	0.05	-0.02	-0.01	-0.03	1.06	1.08	1.09	1.02
	MeteoFrance-Phase1	-0.95	1.33	0.62	-0.03	0.30	0.21	0.26	0.27	0.03	-0.01	0.00	0.02	0.02	-0.01	0.01	0.01	1.03	1.03	1.06	1.05
	MeteoFrance-Phase2	-0.56	1.53	0.58	0.02	0.31	0.20	0.23	0.25	0.03	-0.01	-0.02	0.02	0.03	-0.02	-0.01	0.01	1.02	1.02	1.05	1.04
3	Meteo-France System 8	-0.11	2.22	0.86	-0.75	0.26	0.23	0.15	0.18	0.02	0.00	-0.02	-0.01	0.03	0.00	-0.01	-0.02	1.04	1.12	0.99	1.11
	MeteoFrance-Phase1	-0.36	1.54	0.57	-0.80	0.23	0.16	0.14	0.17	0.00	-0.02	-0.02	-0.03	0.01	-0.02	-0.01	-0.03	1.03	1.05	0.98	1.11
	MeteoFrance-Phase2	0.04	1.60	0.59	-0.77	0.23	0.19	0.13	0.10	0.01	-0.01	-0.04	-0.04	0.02	-0.01	-0.03	-0.04	1.04	1.06	0.97	1.10
4	Meteo-France System 8	0.99	2.19	0.85	-1.06	0.18	0.25	0.18	0.21	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.00	1.06	1.11	0.90	1.05
	MeteoFrance-Phase1	0.66	1.60	0.77	-1.20	0.23	0.21	0.17	0.19	0.01	0.00	-0.02	-0.02	0.01	-0.01	-0.01	-0.01	1.06	1.08	0.92	1.05
	MeteoFrance-Phase2	1.10	1.62	0.76	-1.21	0.22	0.24	0.11	0.21	0.00	0.01	-0.04	-0.01	0.00	0.01	-0.03	0.00	1.06	1.10	0.89	1.07

Figure 3: Scorecard comparing the verification results of the Meteo-France forecast systems for 2-metre air temperature against ERA5 over the extra-tropical Northern Hemisphere for 1993–2016. Results are shown for forecast months 1–4 and initialisation dates in February, May, August and November. Columns show mean bias, correlation, RPSS, CRPSS and the spread-to-error ratio. Colour shading indicates the magnitude and sign of each metric, allowing the relative performance of the three forecast systems to be assessed across forecast lead times and initialisation dates.

Users of the platform can further restrict the analysis by applying a sea mask.

To complement the assessment of forecast performance with climate-relevant information, the Index tab includes the calculation of the NAO and Niño 3.4 climate indices. These indices provide an additional way of assessing the representation of large-scale climate variability in the forecast systems and, for some experimental configurations, are expected to remain consistent across different phases of the demonstrators. Figure 4 shows an example for the NAO index, comparing the Meteo-France System 8 prediction with ERA5 for boreal winter (December–February, DJF) from the 1 November initialisation. The figure displays the ensemble distribution together with the ERA5 reference and provides the CRPSS and ensemble-mean correlation, allowing the forecast performance of the index to be assessed alongside the spatial metrics presented in the other sections of the application.

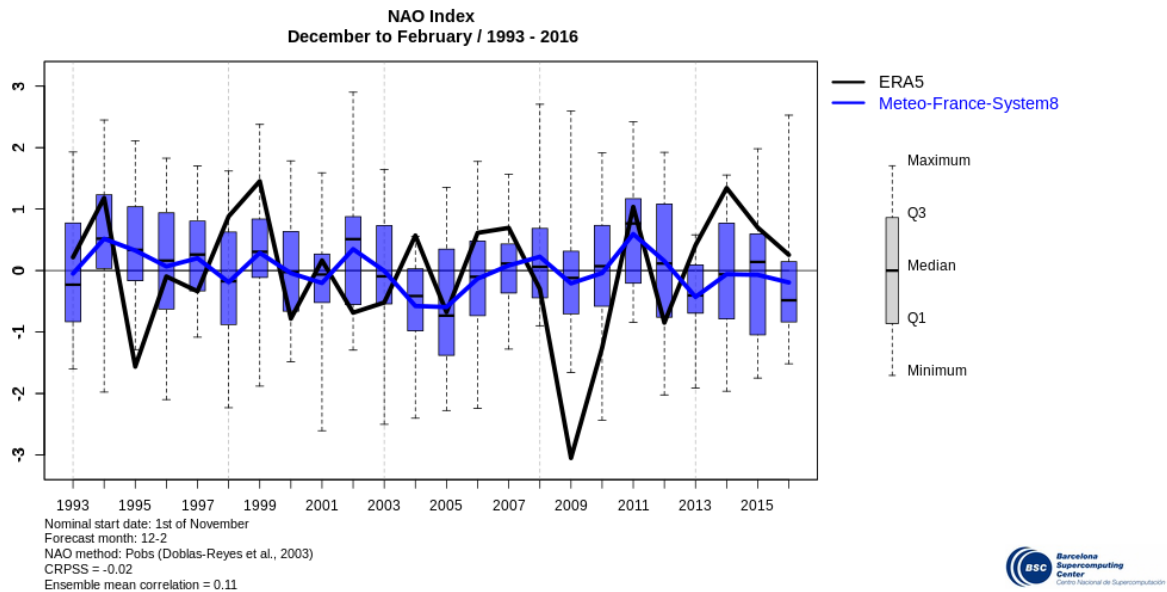


Figure 4. NAO index for December–February from the 1 November initialisation for 1993–2016. The box-and-whisker plots show the Meteo-France System 8 ensemble distribution, with the ensemble median, interquartile range and minimum–maximum range, while the black line represents ERA5 observations. The CRPSS and ensemble-mean correlation with respect to ERA5 are shown below the plot.

5 Conclusion

This deliverable has presented the common assessment of the seasonal forecast demonstrators developed under Task 5.2. The assessment was conducted using a common set of reference datasets, variables and verification metrics, covering the different experiments and phases provided by the participating institutions. The deployment of the SUNSET software suite on ECMWF infrastructure enabled the processing and verification of the climate predictions using a consistent methodology across the demonstrators.

All analyses defined within the assessment framework have been completed, and the resulting products are available through the Shiny web application developed as part of the task. The application provides access to the spatial visualisations, scorecards and additional climate-index analyses, allowing users to explore forecast performance across variables, forecast times, initialisation dates and experimental phases. While the scorecards provide a compact overview of the general performance of the demonstrators, the spatial visualisations allow users to examine the geographical distribution of the results in greater detail.

Overall, the work provides a common framework for assessing and exploring the performance of the CERISE seasonal forecast demonstrators. The combination of the SUNSET-based assessment workflow and the web application facilitates the consistent evaluation, visualisation and interpretation of the results and provides a common resource for model developers and other members of the community.

6 References

- Adler, R. F., and Coauthors, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, 9, 138, doi: 10.3390/atmos9040138
- DelSole, T., and M. K. Tippett, 2016: Forecast comparison based on random walks. *Mon. Wea. Rev.*, 144, 615–626, doi: 10.1175/MWR-D-15-0218.1
- Ferro, C. A. T., 2014: Fair scores for ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, 140, 1917–1923, doi: 10.1002/qj.2270
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 146, 1999–2049. doi: 10.1002/qj.3803
- Huffman, G. J., Adler, R. F., Behrangi, A., Bolvin, D. T., Nelkin, E. J., Guojun, G. U., & Ehsani, M. R. (2023). The New Version 3.2 Global Precipitation Climatology Project (GPCP) Monthly and Daily Precipitation Products. *Journal of Climate*, 36, 7635–7655. doi: 10.1175/JCLI-D-23-0123.1
- Kosaka, Y., S. Kobayashi, Y. Harada, C. Kobayashi, H. Naoe, K. Yoshimoto, M. Harada, N. Goto, J. Chiba, K. Miyaoka, R. Sekiguchi, M. Deushi, H. Kamahori, T. Nakaegawa, T. Y. Tanaka, T. Tokuhiro, Y. Sato, Y. Matsushita, and K. Onogi, 2024: The JRA-3Q reanalysis. *J. Meteor. Soc. Japan*, 102, doi:10.2151/jmsj.2024-004
- Pérez-Zanón, N., Caron, L.-P., Terzago, S., Van Schaeybroeck, B., Lledó, L., Manubens, N., Roulin, E., Alvarez-Castro, M. C., Batté, L., Bretonnière, P.-A., Corti, S., Delgado-Torres, C., Domínguez, M., Fabiano, F., Giuntoli, I., von Hardenberg, J., Sánchez-García, E., Torralba, V., and Verfaillie, D.: Climate Services Toolbox (CSTools) v4.0: from climate forecasts to climate forecast information, *Geosci. Model Dev.*, 15, 6115–6142, doi: 0.5194/gmd-15-6115-2022
- Pérez-Zanón, N., Agudetse, V., Palma, L., Ho, A.-C., Delgado-Torres, C., Milders, N., Duzenli, E., Llabrés-Brustenga, A., de Paula Kinoshita, B., Bretonnière, P.-A., and G. Muñoz, Á, 2024.: SUNSET: SUBseasonal to decadal climate forecast post-processing and asSEssmenT suite, EMS Annual Meeting 2024, Barcelona, Spain, 1–6 Sep 2024, EMS2024-361, doi: 10.5194/ems2024-361
- Schulzweida, Uwe. (2023). CDO User Guide (2.3.0). Zenodo. doi: 10.5281/zenodo.10020800
- Welch, B. L., 1947: "The generalization of "Student's" problem when several different population variances are involved". *Biometrika*. 34 (1–2): 28–35. doi:10.1093/biomet/34.1-2.28
- Wilks, D.S., 2011. *Statistical methods in atmospheric science*. Academic Press, Ithaca, New York (USA).

Document History

Version	Author(s)	Date	Changes
1.0	Núria Pérez-Zanón	30-Sep-2026	Initial version
2.0	Núria Pérez-Zanón	16-Oct-2026	Address comments and add references
3.0	Núria Pérez-Zanón	22-Oct-2026	Address comments

Internal Review History

Internal Reviewers	Date	Comments
Luis Gustavo G de Gonçalves (CMCC), Jeff Knight (UKMO)	October 2026	Initial version
Patricia de Rosnay (ECMWF)	October 2026	V2.0

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.