Guidelines for the blending methods: inconsistencies to address, reference quality metrics and target outputs

Work Package:	5
Deliverable:	5.1
Due Date:	29.02.2024
Submission Date:	29.02.2024
Dissemination Level:	Public
Type:	Report
Responsible:	DMI
Author(s):	Rashed Mahmood (DMI), Roberto Bilbao (BSC), Julien Boé (CERFACS), Rémy Bonnet (CERFACS), Swen Brands (CSIC), Christophe Cassou (CERFACS), Bo Christiansen (DMI), Jesus Fernandez (CSIC), Emilia Sanchez (CERFACS), Pablo Ortega (BSC), Shuting Yang (DMI)]

Funded by the European Union

Contents

1 Summary for Publication

This deliverable documents some of the key inconsistencies that need to be addressed by the blending methodologies to be developed in Task 5.2 to produce truly seamless predictive information. It describes results from previous literature as well as from new analyses performed in Task 5.1, all relevant for the blending of decadal climate predictions and multidecadal climate projections.

2 Contribution to the top-level objectives of Impetus4Change

This deliverable contributes to the following I4C specific objectives:

SO5: Develop seamless homogeneous predictive information from very short (subseasonal) to climate change (several decades) timescales, both at the global and regional scale and additionally i) advance novel approaches to blending and aligning forecast information across timescales in a way that improves the forecast skill and therefore underpins strategic decision-making and ii) provide this information to I4C demonstrator projects and stakeholder with clear guidance and best practices.
How: by revealing the main inconsistencies between climate predictions (providing our best information from months up to a decade) and climate projections (our main source of future information beyond a decade) that need to be addressed by the blending methodologies to produce truly seamless predictive information at all timescales.
SO2: Identify factors degrading dynamical predictions and propose solutions to i) improve dynamical prediction systems on seasonal-to-decadal timescales using novel methods to mitigate deficiencies and close the gap between current skill and potential predictability, and thus to ii) enable delivery of predictions of improved quality in particular for Europe and for variables of high societal relevance.
How: by identifying the regions and variables for which dynamical initialised predictions and transient climate projects show marked differences, which can help to identify the mechanisms and processes that develop initialization shocks or adjustments.

3 Detailed Report

3.1 Introduction and literature review

Before exploring the various methodologies for blending climate information from different sources/experiments, such as initialised climate predictions and transient historical simulations, it is important to investigate the potential uncertainties and errors that can arise when concatenating data that are not fully consistent with each other. The rationale for blending information from these sources is related to the fact that the initialised climate predictions (which deliver the most trustworthy information about the imminent future as they leverage the predictive capacity from both internal variability sources and external forcings) only cover the next 10 years after initialization (Boer et al., 2016). Beyond 10 years, our best source of information derives from ensembles of transient climate simulations (i.e. historical + future projections) that only capture the externally forced predictable signals. Both sources can be potentially combined to produce seamless predictions at multidecadal timescales to assist in the development of successful strategic plans for climate change adaptation. However, some necessary adjustments need to be addressed by the blending approaches to deal with potential inconsistencies between the datasets.

The concept of blending climate information from initialised predictions and the uninitialised projections is fairly new with only one study (Befort et al., 2022) attempting to explore potential benefits and challenges that may arise during the blending process. Both physical and probabilistic inconsistencies between the decadal predictions and the historical+projection experiments can indeed hinder the blending process. Befort et al. (2022) suggested tackling some of the probabilistic inconsistencies (See Figure 1 for an example based on Northern European surface air temperatures) by applying a calibration technique that is based on a well-known variance inflation method (Doblas-Reyes et al., 2005). However, variance inflation alone cannot resolve issues arising from blending data with non-stationary ensemble spreads and/or ensembles and neglects the problem of physical consistency. A separate study (Verfaillie et al., 2021), also revealed additional inconsistencies between the decadal predictions and historical+projections of regional air surface temperatures for different scores that measure the reliability of the datasets, including rank histograms and Jolliffe and Primo (2008) test statistics.
This deliverable investigates other potential inconsistencies between initialised predictions and future projections, presenting results from new analyses that have been conducted in Task 5.1 of the project, which will be helpful in refining the existing blending methods and also developing new approaches following recent constraining methods (e.g. Mahmood et al. 2021) in the coming months. These results are provided in the following sections, as individual contributions from the different institutions involved in WP5.

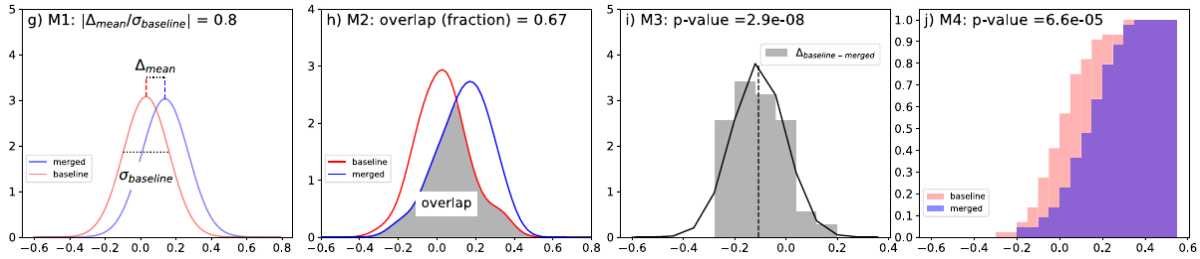

Figure 1. Bottom panels from Figure 2 of Befort et al., (2022). (g) Gausian fit for baseline and merged incremental distributions, with dashes lines representing mean and dotted lines standard deviations. (h) PDFs (probability distribution functions) of the baseline and merged distributions. The grey area indicates overlap between the two distributions. (i) Distribution for the time series of baseline and merged increments. (j) Cumulative distribution of the merged and baseline increments.

3.2 Data and Methods

For the sake of homogeneity, all the new analyses presented in this deliverable that explore the consistency between retrospective initialised decadal climate predictions and historical simulation ensembles are based on three models, CNRM-ESM2.1, EC-Earth3 and NorCPM1, which have been performed by I4C partners involved in WP2 and WP5. Some of the analyses presented in this deliverable also consider the large ensembles of historical simulations (LENS) and decadal predictions (DPLE) performed with the CESM model (Kay et al., 2015; Yeager et a., 2018), as these provide 40 members per type of experiment, which is substantially higher than for the rest.

The common analysis period for all the new analyses covers the years 1970 to 2014 (with some exceptions properly indicated in the relevant sections). Likewise, in the decadal predictions we have only considered the last forecast year as (1) this is when the model is expected to be closer to the mean state of the historical run which should minimise the inconsistencies between the two ensembles and (2) this is the forecast year that will be later used by the blending approaches to leverage as much as possible the added predictive value of initialization in the final blended products. The last forecast year corresponds to year 10 in models EC-Earth3, NorCPM1, CESM and year 5 in CNRM-ESM2.1. Some institutions have also analysed forecasts 1 and 5 in addition to forecast year 10.

From the three models, both EC-Earth and NorCPM1 have more than one retrospective prediction set, which mostly differ in the initialization strategy. We use the first and the second initializations of EC-Earth, as they are very different in nature, and only the first of NorCPM1 as the second is very similar to the first. From the four decadal prediction systems considered, CNRM-ESM2.1 decadal predictions are full-field initialised from a reconstruction based on a 3D nudging of the ocean component of the model EC-Earth predictions performed at BSC (i1) are full-field initialised, while the one with EC-Earth by DMI/SMHI (i2), and the set considered from NorCPM are both anomaly initialised (an strategy that could potentially improve the consistency with the historical ensemble). The details of these models are summarised in Table 1.

Model	Resolution (Atmosphere/Ocean)	DCPP-A (s1960-s2014)			Historical Simulations	
		Initialisation	Members	Ref.	Members	Ref.
CNRM-ESM2-1	T255L91/ ORCA1L75	Full-Field	15	Séférian et al. (2019)-	10	Séférian et al. (2019)-
EC-Earth3-i2	T255L91/ ORCA1L75	Anomaly	4	Tian et al. (2021)	22	Döscher et al. (2020)
EC-Earth3-i1	T255L91/ ORCA1L75	Full-Field	10	Bilbao et al. (2021)	22	Döscher et al. (2020)
NorCPM1	1.9°x2.5°L26 / 1°L53	Anomaly (EnKF assimilation)	10	Bethke et al. (2021)	30	-
CESM1	1.°x1°	Full-Field	40	Yeager et al. (2018)	40	Kay et al., (2015)

Table 1. Details of the climate models used in this deliverable. Note that both EC-Earth3 decadal predictions have the same reference historical simulations.

3.3 Work Carried Out by the different partners

DMI

1. A literature review on the main inconsistencies between initialised decadal predictions (DEC) and historical (HIST) simulations as revealed in previous works, with several examples to illustrate them (integrated in the introduction).
2. A new analysis that focuses on the DEC vs HIST inconsistencies in the representation of the North Atlantic atmospheric circulation modes and their relationship with the AMV (described in the Results section) .

BSC

1. A new analysis that documents the inconsistencies between the probabilistic distributions of the anomaly sea surface temperature (SST) patterns in the DEC and HIST ensembles (described in the Results section).
2. A new analysis that documents the inconsistencies between the main global modes of SST variability (as described by the first five Empirical Orthogonal Functions of the anomaly SST patterns) in the DEC and HIST ensembles (described in the Results section).

CSIC

1. A new analysis revealing the inconsistencies in the representation of city-relevant weather types in reanalysis, DEC and HIST ensembles. Weather type frequencies from decadal predictions at different lead times were compared to historical simulations and reanalysis data, considering the signal-to-noise ratio to identify significant predicted anomalies using 10-member ensembles for each of the sources (described in the Results section).

CNRS-CERFACS

1. A new analysis that illustrates how (in)consistently DEC and HIST ensembles represent surface air temperature trends in the main IPCC European regions (described in the Results section).

3.5 Main Results Achieved

3.5.1 Inconsistencies in the representation of the main modes of variability of the North Atlantic circulation (DMI)

We analysed the historical and prediction runs based on LENS and DPLE ensembles from CESM which both include 40 ensemble members. The analyses have been repeated with 15-members ensembles from EC-Earth. We use the period 1960-2017.

Christiansen et al. (2023) found that for surface air temperature (TAS) in CESM model the increased skill, measured as the correlation between ensemble mean and observations, of decadal forecasts compared to the historical ensemble decays fast with lead-time. For lead-times larger than 4 years the initialisation adds little skill in most geographical areas, which could imply that the predictions have converged towards the historical trajectory.

Here we have extended the analysis by looking at the ensemble spreads. Even though the skill remains low, the initialised ensemble could potentially provide a more precise prediction. Figure 2 shows the geographical distribution of the ensemble spread for the historical ensemble and for the initialised ensemble at lead-times 1 and 10 years. We find that for small lead-times the initialisation does decrease the spread. For larger lead-times, however, the effect is negligible and the spread for lead-time10 appears to be consistent with the spread in the historical ensemble. However, some minor inconsistencies might be hard to detect in this type of plot.

For the CESM ensembles, the fraction of the globe where initialisation reduces the historical spread changes more or less linearly from 1 in the first forecast year to around 0.4 at forecast year ten (Figure 3), which is still far from zero (the value that would indicate that both spreads are fully consistent for all grid points). For the EC-Earth ensembles we find a much larger effect of initialization on the spread, as evidenced by the fact that the same metric stabilises around 0.8. Interestingly, the spread in the initialised EC-Earth ensemble is comparable to the spread in the DPLE ensemble (not shown for the sake of conciseness). The main differences between EC-Earth and CESM happen for the spread in the historical ensemble, which is comparatively larger in EC-Earth at high latitudes.

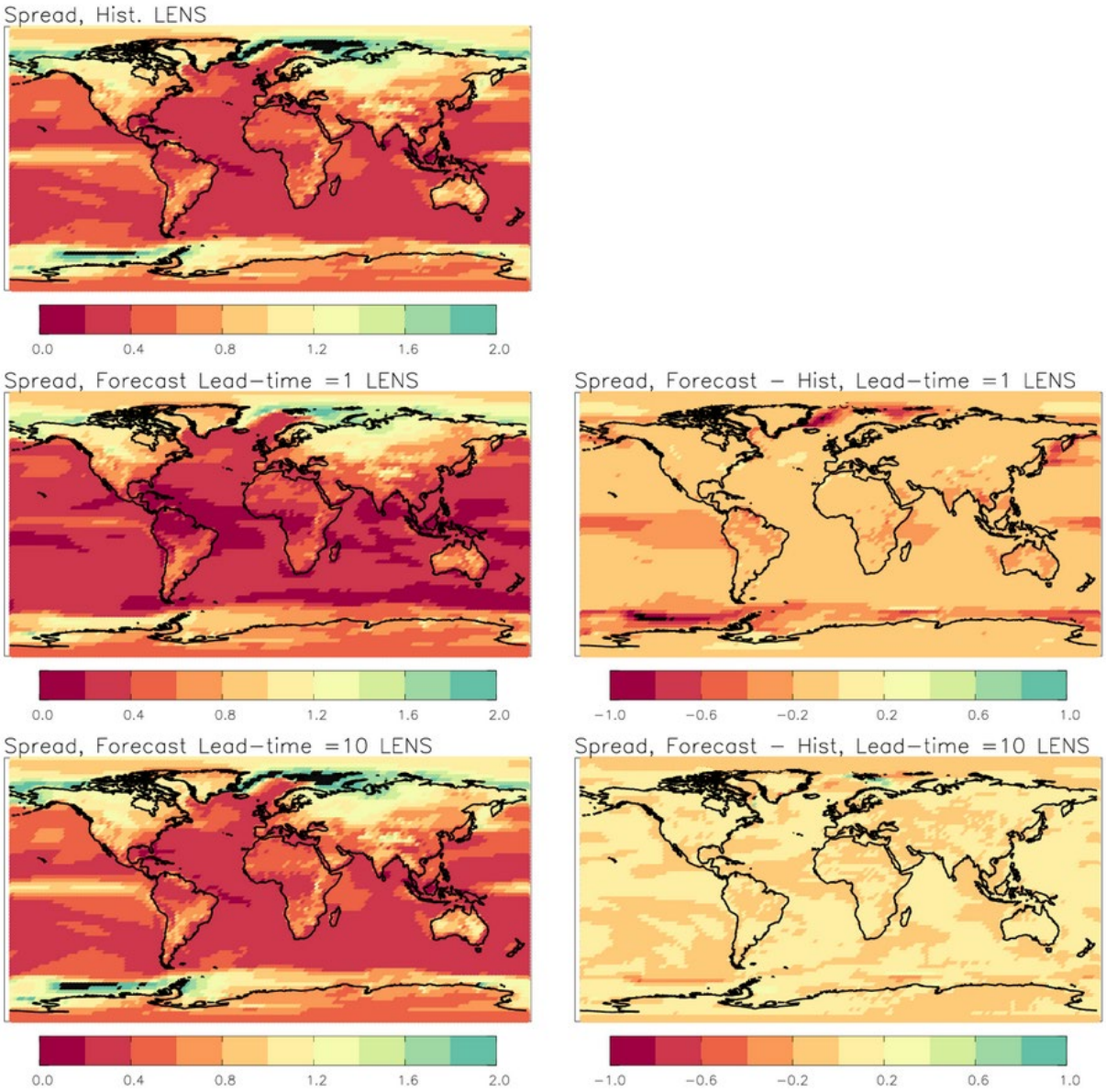

Figure 2. Left panels show the geographical distribution of the ensemble spread (standard deviation) in surface air temperature for (top) the historical ensemble LENS, (middle) the initialised ensemble DPLE at forecast year 1, and (bottom) the initialised DPLE ensemble at forecast year 10. Right panels show the differences between initialised and historical ensembles. The spread is averaged over all years.

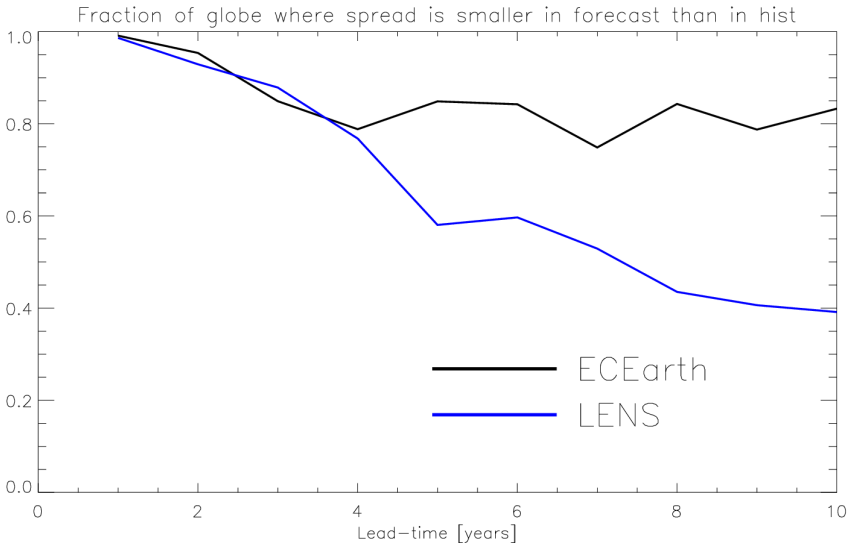

Figure 3. The fraction of the globe where the spread is smaller in the initialised ensemble than in the historical one. Values are shown as a function of forecast-time (in years).

We now explore whether the connection between the Atlantic Multidecadal Variability (AMV) and the North Atlantic Oscillations (NAO) is consistent between the decadal predictions and the historical ensembles. To this end we calculate their lagged correlations using both unsmoothed and decadally smoothed time series. The results are shown in Figure 4.

For the unsmoothed time series, which mostly describes their interannual connection (left) the LENS ensemble shows a negative correlation between both variables at lag 0. In observations this correlation is weaker and takes maximum values when the NAO leads by 2 years. In DPLE, no clear negative correlations are seen for lag 0 in the first forecast year, although the negative lag 0 correlation is somewhat recovered at forecast year 10, indicating that by the end of the predictions their consistency with the historical ensemble has qualitatively improved, even if none of the ensembles is consistent with the observations.

For the decadally smoothed time series high positive correlations are found when NAO leads the AMV by one decade. This result is not found in the historical ensembles, nor in the last year of the predictions which are therefore consistent with each other. By contrast, a positive correlation is indeed found when the NAO leads the AMV by 10 years in the first year of the predictions, which indicates the relationship is partly corrected via initialization.

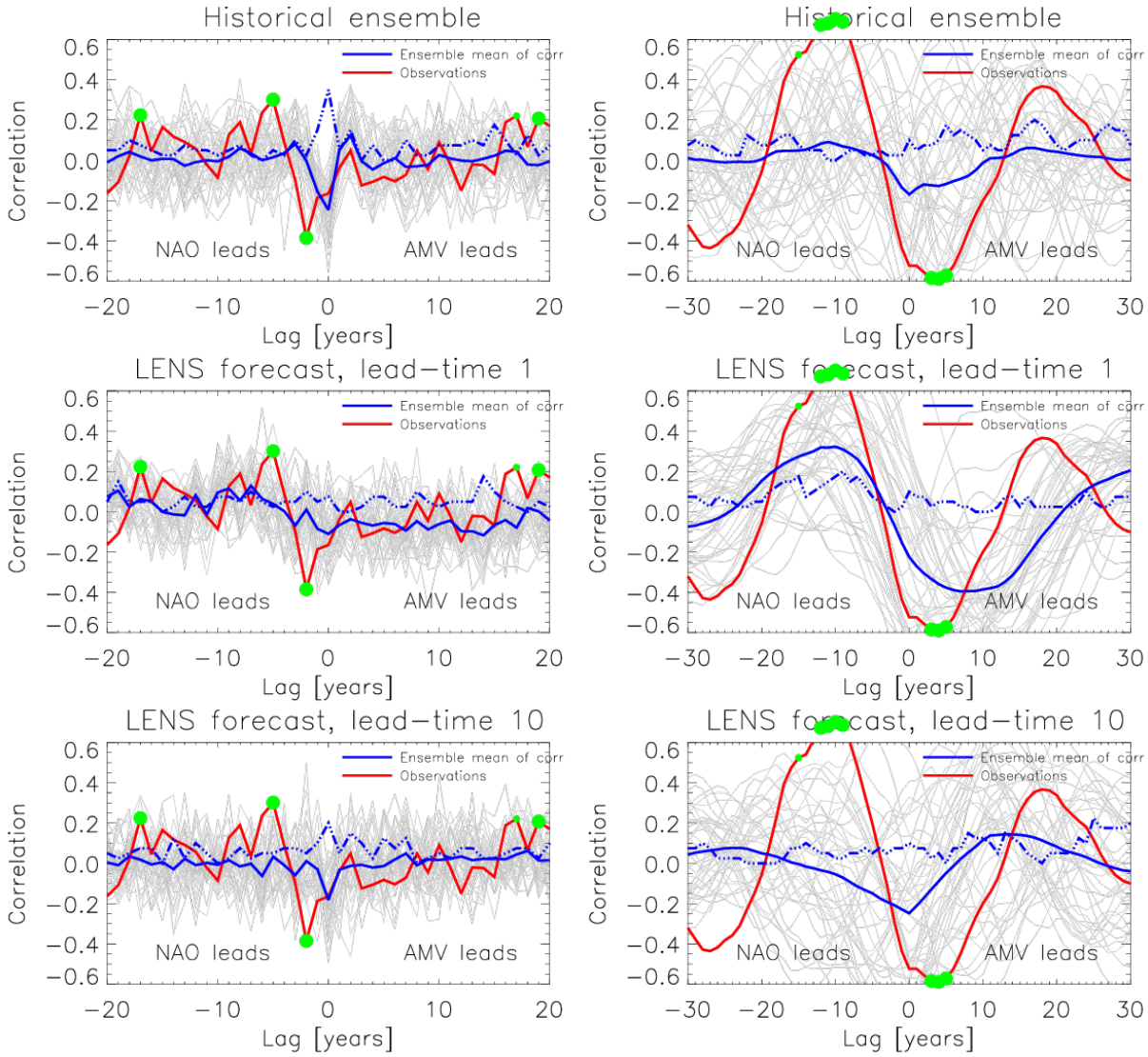

Figure 4. Correlations between NAO and AMV as a function of time lag. NAO leads AMV for negative lags. Red curve corresponds to results from observations (20CRv2c). The thin grey lines represent individual ensemble members and the thick solid blue line the ensemble mean of the correlations. For observations big/small, green filled circles indicate that correlations are significant at the 5/10% level ($p < 0.05$ / $p < 0.1$). For the model ensemble the dashed blue curve gives the fraction of ensemble members that are significant to the 5% level ($p < 0.05$). Left: annual values. Right: Decadally smoothed values (10-years moving average filter). Top row: LENS historical. Bottom rows: DPLE for forecast-time 1 and 10 years.

3.5.2 Probabilistic and physical inconsistencies in the representation of SST anomaly fields (BSC)

To inform the blending methodologies for the decadal predictions and climate projections, which will be implemented later in this WP, we compare the sea surface temperature (SST) anomaly field in decadal predictions and historical simulations for the four prediction systems described in the methods section. All anomalies correspond to the period 1970-2014, except for CNRM-ESM2-1 for which the period is 1965-2014 (since the hindcasts are 5 years long instead of 10), and in the case of the

prediction systems are computed for the selected forecast time. For this analysis we have used the updated i4 initialisation version of the EC-Earth3 full-field initialised system, instead of i1.

We start by comparing the temporal mean of the intra-ensemble spread (as characterised by the standard deviation across members) of the SST anomaly fields both in the decadal predictions and the historical simulations. Figure 5 shows that the main features of the resulting patterns are very similar across all the experiments and models. The tropical and North Pacific Ocean and the North Atlantic Ocean standout as the regions with the largest intra-ensemble spread. There are differences however in the magnitude of the standard deviation in these regions, in particular, in the North Atlantic. In the CNRM-ESM2-1 and EC-Earth3 models, the intra-ensemble variability is greater in the historical simulations with respect to the decadal predictions, while for NorCPM, the differences are much smaller and probably not statistically significant.

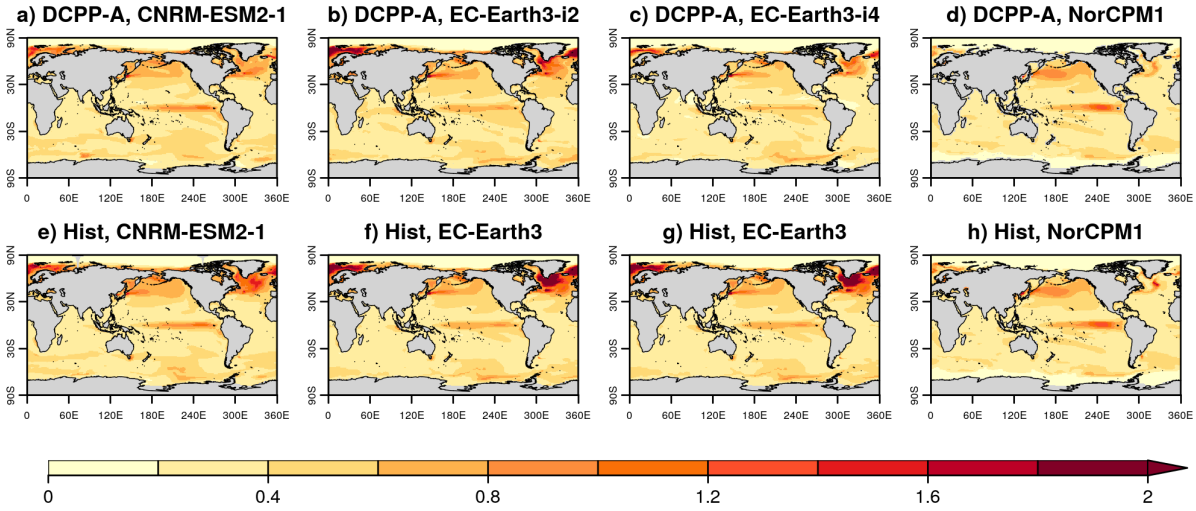

Figure 5. Temporal mean of the intra-ensemble standard deviation for the SST anomaly fields (°C) in (a-d) the decadal prediction systems at forecast year 10 (except for CNRM-ESM2-1, for which forecast year 5 is shown) and (e-h) for their corresponding historical simulations. The reference period used for the computation of the anomalies, and the period covered by the experiments are both 1970-2014. Note that panels f and g are the same since there is only one set of historical simulations.

Next we compare the SST probability distributions in a different way. For each prediction system and model we generate an ensemble that combines the SST anomaly fields for all ensemble members and start years. The analogous ensemble for the historical simulations is in turn composed by all the ensemble members and simulated years. For the resulting distributions of SST anomaly fields, we computed the 5th, 20th, 50th, 80th and 95th percentile to make a comprehensive comparison of the most extreme anomaly values in the historical and decadal prediction ensembles, as well as their medians. Figure 6 shows the results for the full-field initialised predictions and the historical ensemble of EC-Earth (i4). In most regions of the world, the different percentiles show similar values in both ensembles of simulations. However, the North Atlantic stands out as a region with important differences, which become larger for the most extreme percentiles, a result that was also found for different variables and metrics in Befort et al. (2022). Similar results are also found for the CNRM-ESM2-1 and the anomaly initialised predictions of EC-Earth3 (not shown). In contrast, the

differences in the percentile patterns for NorCPM1 are considerably smaller, and probably, as shown in Figure 6, statistically not significant (Figure 7).

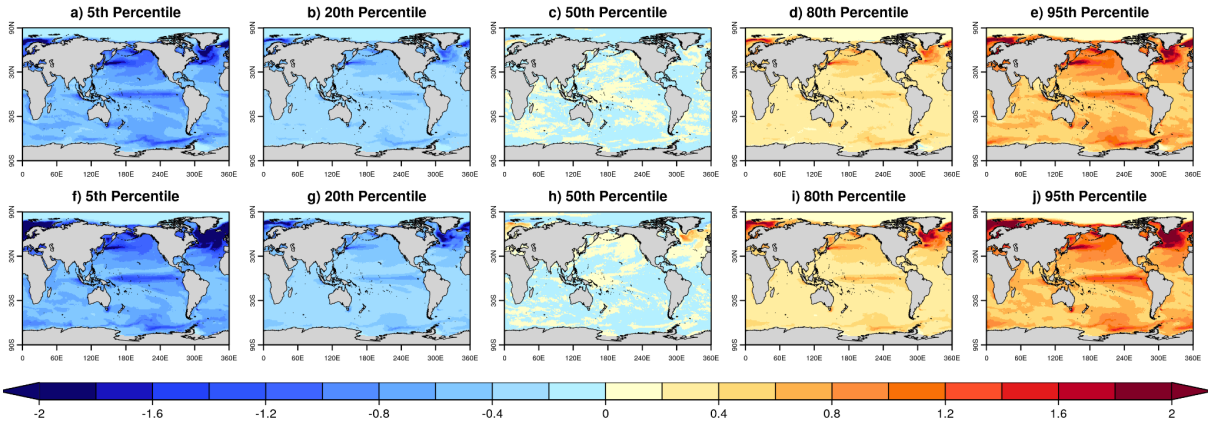

Figure 6. Selected percentiles in the distributions of SST anomaly fields (°C) in the full-field initialised predictions (a-e) and historical simulations (f-j) of EC-Earth3 for the period 1970-2014.

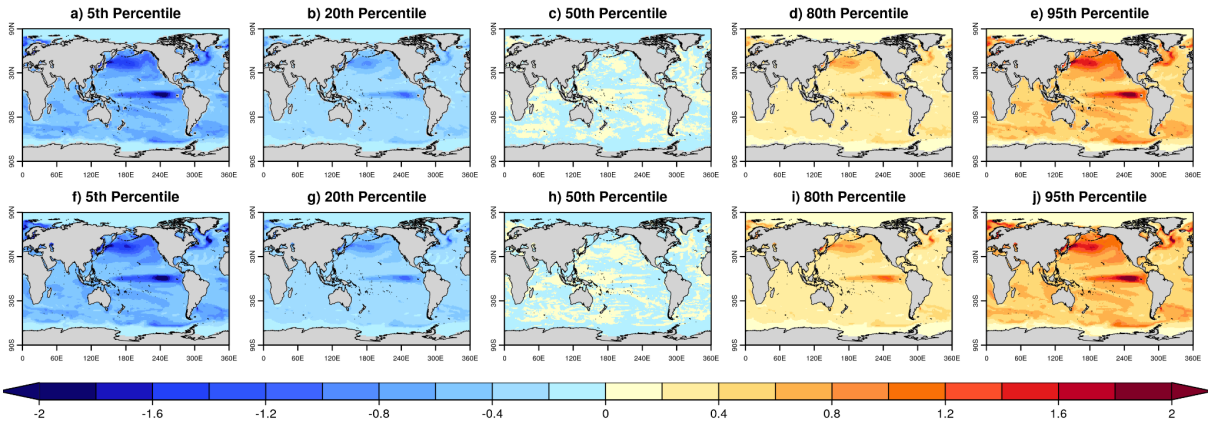

Figure 7. As Figure 6, but for NorCPM1.

The large differences found in the North Atlantic for CNRM-ESM2-1 and EC-Earth3, which use NEMO3.6 as their ocean component, could be related to the overly large multi-decadal variability in the North Atlantic that their piControl and historical simulations exhibit, a feature variability that is somewhat damped by the initialization process (Bilbao et al 2021). The strong agreement between the historical and climate predictions for the NorCPM1 model could be partly related to the initialization technique used in the decadal predictions, as with anomaly initialisation the model is constrained by observed anomalies deviating from the historical simulations climatology. The anomaly-initialised predictions of EC-Earth3 (i2) also use anomaly initialisation, and while differences with the historical ensemble are still evident in the North Atlantic, it is also true that these are smaller with respect to the full-field initialised version (i4). This suggests that anomaly-initialised decadal predictions systems could enhance the consistency and thus be more suitable for the blending decadal predictions and climate projections.

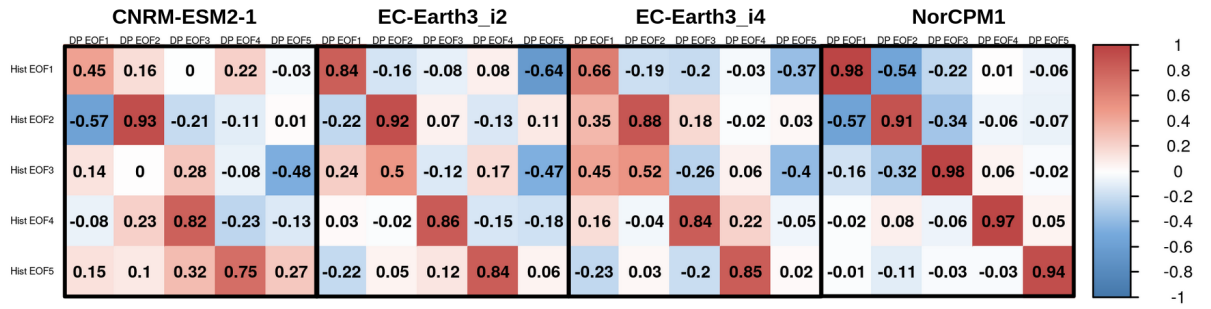

Figure 8. Area weighted spatial correlation between the first five EOFs of the SST anomaly fields in the decadal predictions and the historical simulations for the four systems considered.

While, so far, all blending methodologies targeting the decadal predictions and climate projections have addressed their statistical consistency (e.g. Befort et al., 2022), their physical consistency has been neglected, and is indeed not guaranteed. To illustrate to what extent physical consistency can be a problem, we compare the dominant modes of variability for the previous SST anomaly fields variability as derived from Empirical Orthogonal Function (EOF) analysis. Figure 8 shows the area-weighted spatial correlations between the EOFs of the decadal predictions and the historical simulations in the four selected prediction systems. Interestingly, the spatial correlations across EOFs for the EC-Earth3 and the CNRM-ESM2-1 models show important differences between the two ensembles, as the historical EOFs are not always in good correspondence with their analogous ones in the predictions. More importantly, NorCPM1 stands out again as a model of high consistency between the historical and the predictions, as all the EOFs with the same number are very strongly correlated with each other. For NorCPM1, the first three EOFS explain most of the variance, while for the other models it is mostly the first two.

For the first EOF, which explains most of the variance, the systems that are full-field initialised (CNRM-ESM2-1 and EC-Earth3 i4) show lower correlation values with respect to the historical counterpart than the two other decadal prediction systems that are anomaly initialised (EC-Earth3 i2 and NorCPM1). This suggests a strong impact on the physical consistency of the initialisation method used. To further explore this, figures 9 and 10 show the first five EOFs for the EC-Earth3 and NorCPM decadal predictions systems and the historical simulations. For both EC-Earth3 prediction systems (Figure 9a, g, l), the first EOF resembles the global warming pattern associated with increasing greenhouse gas (GHG) emissions, however there are strong differences in the North Atlantic Ocean which are likely associated with the presence of strong centennial variability in this region in the historical simulations. Similar patterns and differences are found for CNRM-ESM2-1. In contrast, NorCPM1 predictions and historical simulations show very strong agreement and a pattern that is highly dominated by El Niño Southern Oscillation variability (Figure 10a, g).

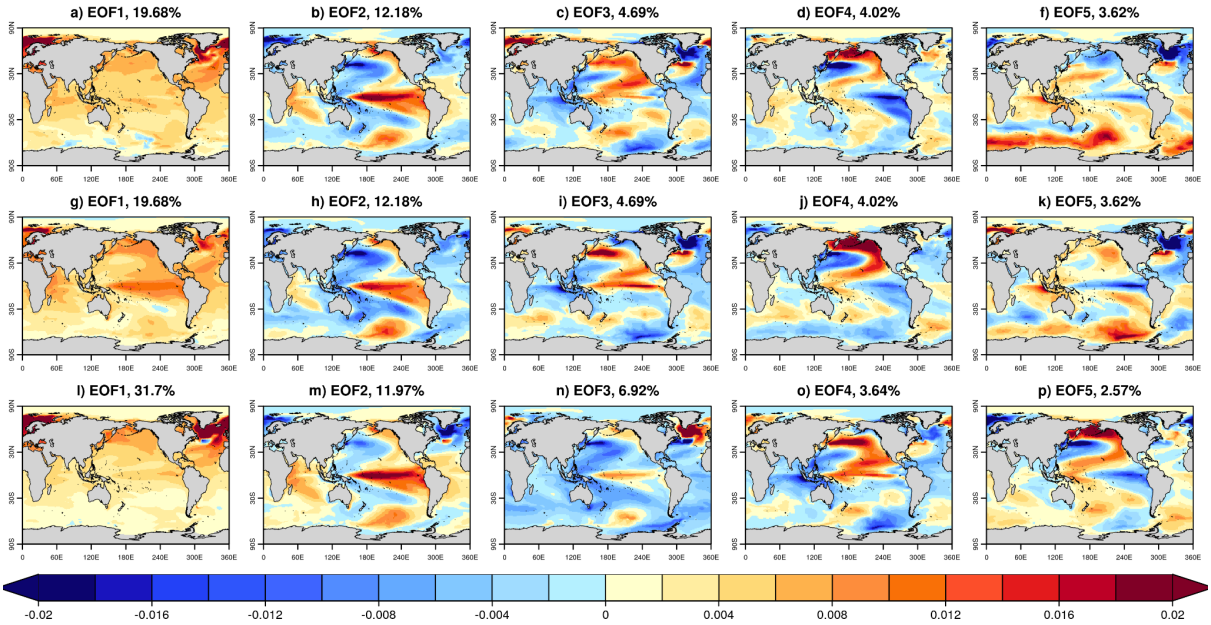

Figure 9. EOFs of the SST anomaly fields for (a-f) EC-Earth3 anomaly initialised (i2) decadal predictions, (g-k) EC-Earth3 full field initialised (i4) decadal predictions and (l-p) EC-Earth3 historical simulations.

In contrast with the first EOF, the second EOF is very consistently simulated between the historical simulations and the predictions in the four models (Figure 8). In EC-Earth3 and CNRM-ESM2-1 this is the EOF that is dominated by El Niño Southern Oscillation variability (Figure 9b, h, m), while in NorCPM1 the second EOF captures most of the global warming signal.

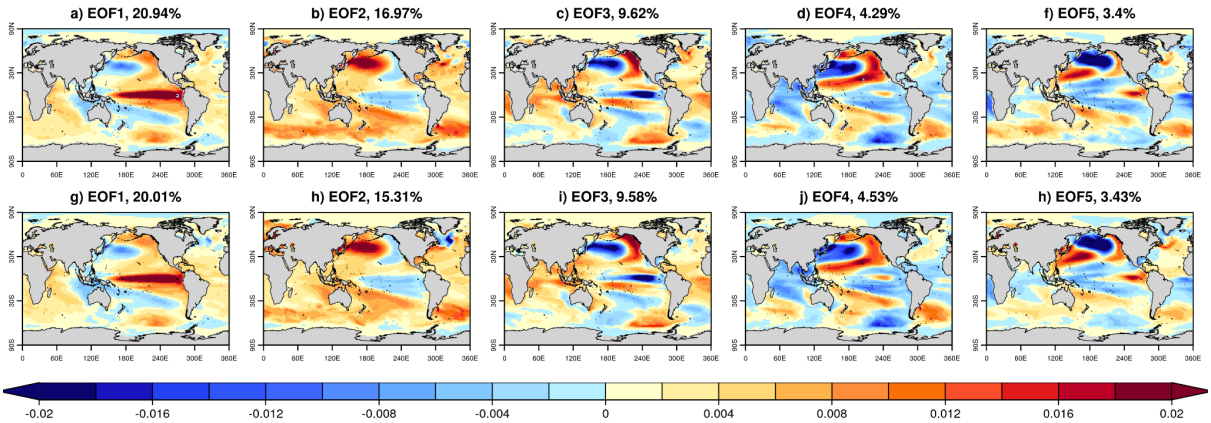

Figure 10. EOFs of the SST anomaly fields for NorCPM1 (a-f) decadal predictions and (g-k) historical simulations.

3.5.3 Inconsistencies in the representation of city-relevant weather types (CSIC)

The Jenkinson-Collision (JC) approach is a well-known weather type (WT) classification method used in synoptic climatology that assigns an SLP field centred on a given location at a given time-step to one out of 27 possible discrete classes, which can be physically interpreted and have a discernible impact on the local weather conditions (Brands 2022, Brands et al. 2023, Fernández-Granja et al. 2023). There are *8 pure*

directional types characterised by a pronounced pressure gradient over the region of interest which, in combination with the relative positions of the low and high pressure systems, determine the direction of atmospheric flow (NE, E, SE, S, SW, W, NW and N) and, indirectly, the properties of the advected air masses reaching the target location. Furthermore, there is a *pure cyclonic* and a *pure anticyclonic* class and 8 hybrid classes for each of them which combine (anti)cyclonic conditions with a predominant atmospheric flow that, however, is weaker than during pure directional situations. Finally, the *Unclassified* type refers to synoptic conditions that do not meet any of the aforementioned conditions, characterised by a weak pressure gradient and negligible vorticity over the target location. As an example for the impact of the JC classes on local weather, Figure 11 shows the signature of cold events over Prague on the weather types centred on that location as reproduced in our observation-based reference (ERA5). In general, the pure anticyclonic (25%), westerly (9%) and cyclonic (9%) types are the most common in Prague. Cold events are largely associated with high anticyclonic type frequencies (52%), while, e.g., westerlies occur rarely (2%) during these events. This link has a physical background through atmospheric blocking situations and decreased mild air advection. Therefore, the JC types encompass spatial scales which are resolved by GCMs (they are computed on a 30° x 20° lon-lat domain) and have a direct, local-scale impact, as shown here for an I4C demonstrator city.

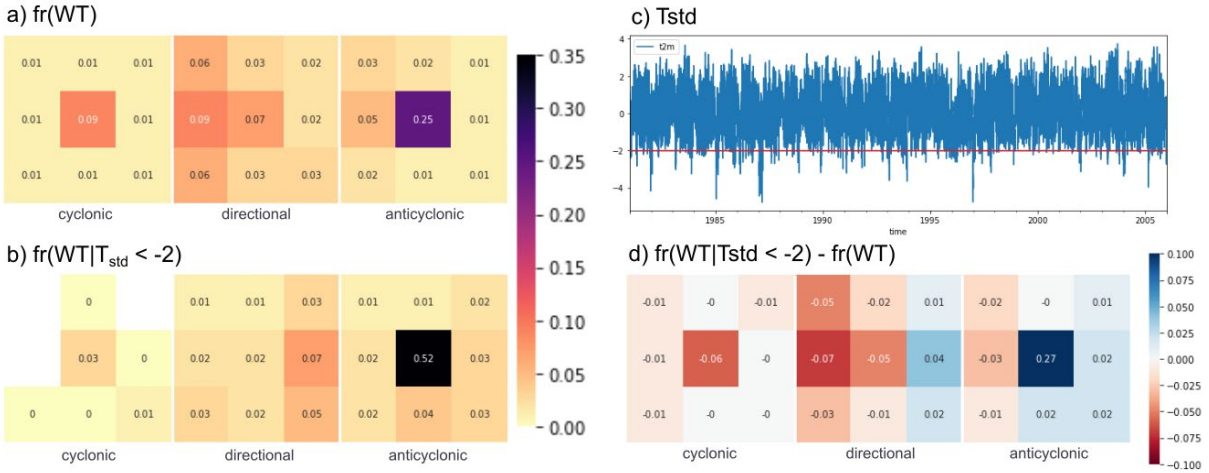

Figure 11: (a) Marginal frequency of the 27 JC weather types centered over Prague, according to the ERA5 reanalysis for the period 1979-2005. (b) Conditional frequency of the corresponding types during cold events with (c) standardised deseasonalized temperature anomalies below 2 units. (d) Conditional minus marginal weather type frequency.

The computation of the JC WTs requires 6-hourly instantaneous mean sea-level pressure data. These sub-daily data are not commonly saved in decadal forecasts, so we focused on the full-field initialised version of Ec-Earth3, the only one of the preselected prediction systems providing this data.

These datasets were used to classify the circulation, at each time step, into JC WTs centred over each of the I4C demonstrator cities. For the dcppA experiment, forecast years 1, 5 and 10 were considered, covering the periods 1961-2019, 1965-2023 and 1970-2028, respectively. The historical simulations were extended with their respective SSP245 runs to cover the period 1961-2028. In order to have a comparable ensemble size, we selected also 10 members from the historical simulation: r1i1p1f1, and the

corresponding r4, r10, r12, r14, r16, r17, r18, r21 and r24. All model data were retrieved from the ESGF data portals. JC types were also calculated upon 6-hourly instantaneous SLP data from the 10 members of the CERA-20C reanalysis dataset (Laloyaux et al. 2018) and from ERA5 (Hersbach et al. 2020), covering the period 1961-2010 and 1961-2022, respectively. Both reanalysis datasets were retrieved from the Copernicus Climate Data Store. They serve as observational references and also permit estimating observational uncertainties along the indicated time period. City-scale 6-hourly JC type time series were aggregated into annual relative frequencies for each type. Individual forecast years of consecutive decadal hindcast initializations were assembled to obtain quasi-continuous time series for forecast years 1, 5 and 10, covering the aforementioned periods so that they can be directly compared with the respective time series of historical simulations and reanalyses. No *a-posteriori* model correction was applied. To leverage the anticipated large effects of internal variability on the results (Deser et al. 2012, Maher et al. 2019), the annual JC type frequency time series were further averaged using a 10-year running window.

To evaluate to which degree a forced signal arises from internal variability (or climate noise) on decadal time-scales, signal-to-noise ratios were computed as follows: Firstly, the yearly JC type relative frequencies were transformed into anomalies by removing the 1970-2019 mean value from each yearly value. Then, 10-year running mean values were calculated on the resulting yearly anomalies. This was done separately for each historical experiment member and for the EC-Earth3 DCPP hindcasts (i1) re-formatted to form a quasi-continuous time series with fixed forecast year 1, 5 and 10, respectively. The signal-to-noise ratio (SNR) for an individual 10-year period of any of these series is then defined as:

$$SNR = \mu / \sigma,$$

with μ being the mean-value and σ the standard deviation of the 10 individual anomalies for the decade and experiment in question. The significance of SNR, i.e. emergence of a forced signal on the background of climate noise for this decade, is then given by Deser et al. (2012):

$$SNR \geq \pm 2 / \sqrt{(n-1)} ,$$

depending on the sample size (n), which in this case is the number of ensemble members (n = 10). In practice, this means that SNR values exceeding ⅔ are significant, i.e. the weather type frequency anomaly is forced either by external agents or by internal agents from climate system components outside the atmosphere (e.g. SSTs, sea-ice extent, snow cover or soil moisture).

Figure 12 shows 10-yr running mean frequencies of the pure anticyclonic type centred over Barcelona. Thick solid lines show ensemble means for the dcppA experiment at forecast years 1, 5 and 10, as well as for the historical experiment. For DCPP-A forecast year 10 and the historical experiments, the running decadal-mean values of the individual members are also depicted as dashed lines. The observation-based reference and their respective uncertainty, as provided by CERA-20C, are shown in black. Decadal forecasts show very limited skill for weather types. However, individual EC-Earth3 simulations show very low biases and comparable variability to the observations (see e.g. the thicker dashed lines highlighting a single member from DCPP-A FY10 and from the historical experiment).

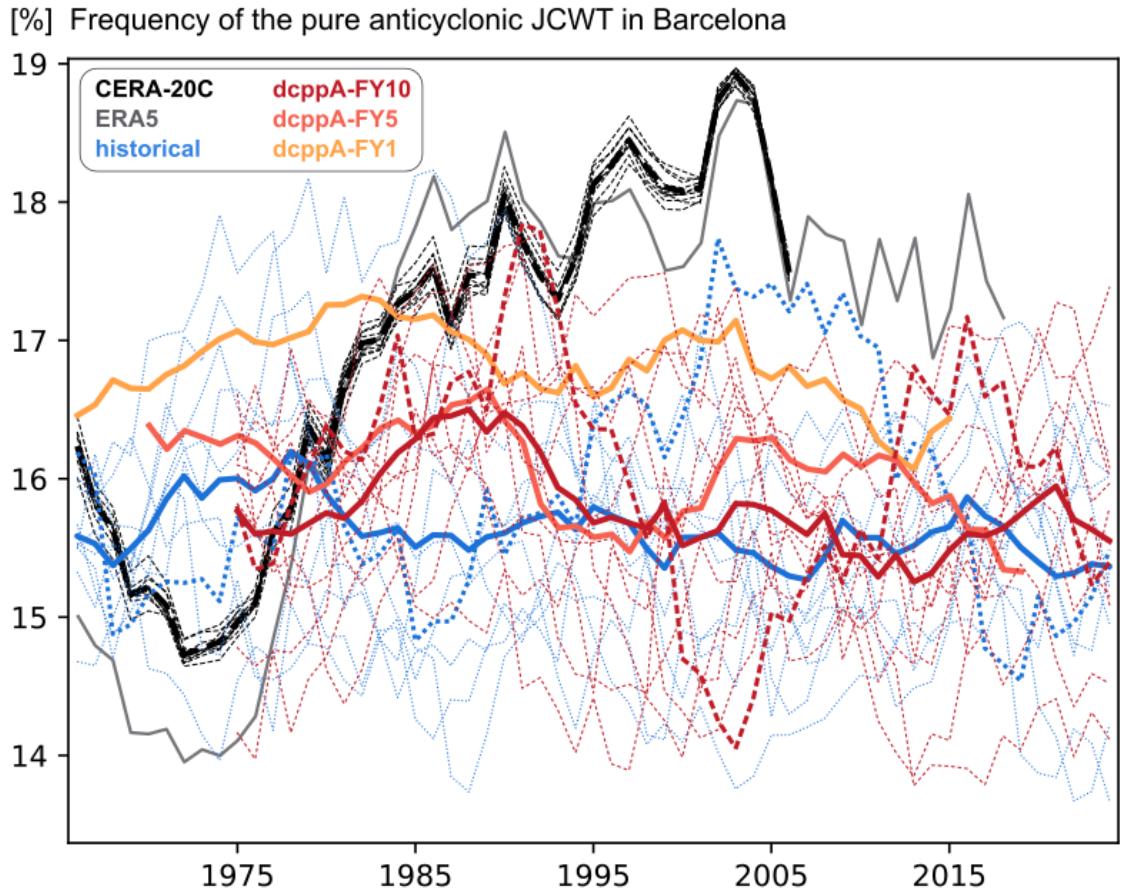

Figure 12: Running decadal mean frequencies of the Pure Anticylonic type in Barcelona according to 10-members ensembles from the CERA-20C reanalysis, historical EC-Earth3 simulations and DCPP hindcast (dcppA) simulations for forecast years 1, 5 and 10. Dashed and dotted lines show individual members and solid lines represent their ensemble mean. ERA5 is also shown as reference.

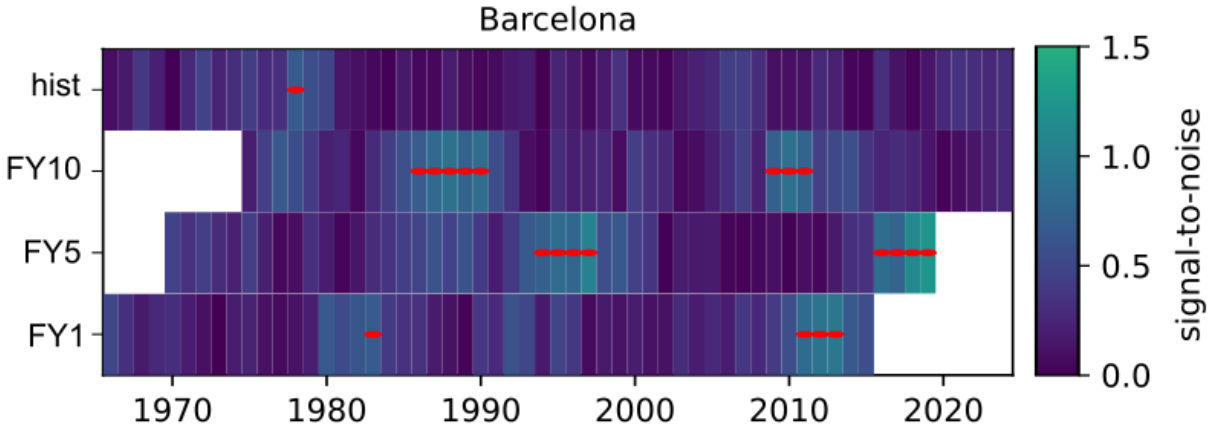

Figure 13: Running signal-to-noise ratios (SNRs) for the decadal-mean relative frequencies of the Pure Anticyclonic type in Barcelona for the dccpA experiment and forecast years 1, 5 and 10, as well as for the historical experiment (y-axis) along the period 1961-2028 (x-axis). Significant SNRs ($\alpha = 0.05$) are indicated by red dots.

Regarding the signal in decadal predictions, as expected from previous studies assessing SLP variability on multi-annual to decadal time-scale (Deser et al. 2012, Maher et al. 2019), the noise component dominates the predictable signal even for 10-year mean anomalies and, consequently, SNR is rarely significant (Figure 13). The running predicted response of DCPP-A, as represented by the ensemble-mean in Figure 12, differs from that of the historical experiment for short lead-times (forecast year 1 in this case), indicating that initialization has an impact on the local-scale atmospheric circulation even if the background noise is substantial. As expected, these differences are much reduced for forecast years 5 and 10, where the running predicted response moves towards the corresponding values of the non-initialised historical simulations. For the case of Barcelona, the most notable indication of a forced response in the dccpA ensemble is found for the decades centred around 1985-90 and forecast years 5 and 10, as revealed by significant signal-to-noise ratios with a confidence interval of 68 and 95% (Figure 13), respectively. If such relatively weak confidence intervals are considered, significant signal-to-noise ratios appear consistently in all considered forecast years and demonstrator cities during a number of heterogeneous time periods. Across these cities, the clearest surplus of significant responses along the study period derived from the dccpA ensemble, if compared with the historical ensemble, is obtained in Barcelona. This indicates an influence of the observationally constrained internal climate system components in the dcppA ensemble on top of the external forcing agents' influence, prescribed in both the dcppA and historical ensemble which, however, are generally not consistently present in all forecast years. In this context, Mediterranean SSTs in the initialised ensemble might be colder or warmer than in the historical ensemble, leading to suppressed or enhanced land-sea temperature contrasts that favour or disfavour the build-up of thermal pressure cells and associated cold high-pressure and warm low-pressure systems.

As a final summary, Figure 14 compares the spread of the 10 temporal-mean relative LWT frequencies from the historical experiment with the respective ensemble-mean values for the initialised forecast years 1, 5 and 10, and the mean value of ERA5 is included as observational reference. This is shown for all I4C demonstrator cities and for the two dominant JC WTs (PA and PDW). For comparison with the other sections, the 1970-2014 period is considered here. For the *Pure Anticyclonic* weather type, the ensemble-mean of the initialised predictions tend to thrive away from the observed values as the forecast year (or lead-time) increases, being generally shortest for FY1. For the *Pure Directional Westerly* type, however, this is not the case. The reasons for such a distinct behaviour of the weather types are unclear at this point and should be further explored in the future. The FY10 climatology tends to be consistent with the historical non-initialized simulations, falling within the range of the historical ensemble members for all I4C demonstrator cities and both dominant weather types. Figure 14 also shows that the spread of the *Pure Anticyclonic* type climatologies in Barcelona and Bergen is considerably smaller for FY1 than the spread of the historical climatologies, pointing to a possible constraint to observations in the initialised model runs. The existence of such a constraint is further strengthened by an exact match between the dcppA mean climatology at FY1 and the respective value taken from ERA5, found for Barcelona.

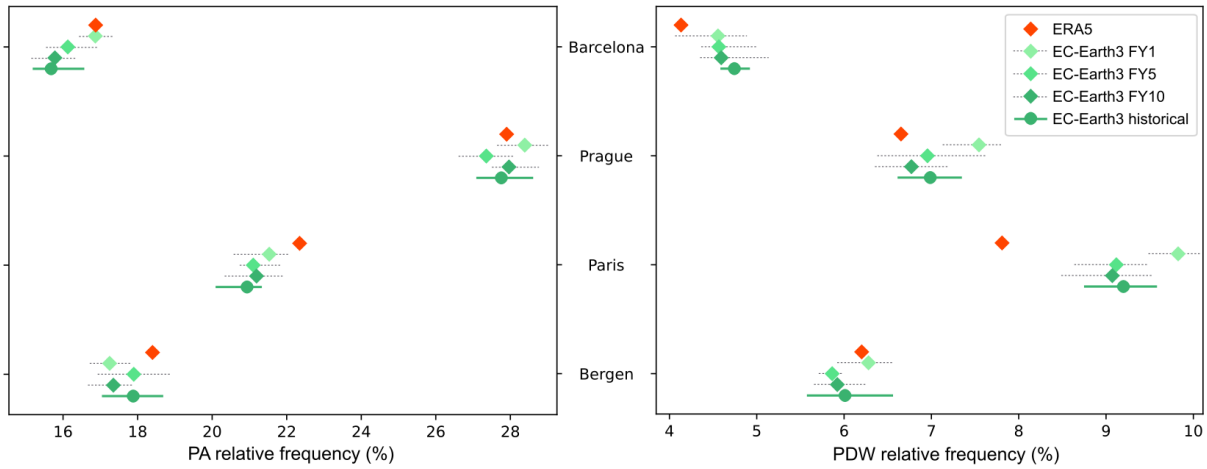

Figure 14: Climatological-mean relative frequency of the Pure Anticyclonic (left) and Pure Directional West (right) weather types for the four I4C demonstrator cities during the common period 1970-2014. For each city and weather type, the ensemble-mean (dark green dot) and ensemble-spread (dark green line) of the 10 historical EC-Earth3 experiments are compared with the respective values of the dcppA experiment run with the same model at forecast years 1, 5, and 10 (see legend). For reference, the climatological-mean frequency from ERA5 is also shown.

3.5.4 Inconsistencies in the representation of European surface air temperature trends (CNRS-CERFACS)

To assess the capacity of decadal predictions and climate projections to be used for the blending methodologies implemented later in this WP, we compare their performance in simulating the observed trends in surface temperature over Europe over 1970-2014 (Figure 15). Blending methods often consist of finding the uninitialised members among the available pool of projections that are the closest to the decadal predictions (e.g. Mahmood et al. 2021), which requires the trends of the two data sets to be compatible (i.e. the hindcasts to be within the spread of the uninitialised simulations). We consider three prediction systems: CNRM-ESM2-1, NorCPM1 and the full-field initialised predictions from EC-Earth3 (see Data and Methods section). We find that for each region and each forecast year, the hindcast ensemble mean trend is within the spread of the historical simulations for each model.

A strong influence of the internal climate variability is visible at decadal timescales, with a large spread of surface temperature trends over the three regions, especially in winter. At forecast year 5, regardless of the season, the surface temperature trends for the hindcast simulations are generally close to the observed ones over the West Central Europe (WCE) and Northern Europe (NEU) region for the three models used, with slight improvement in comparison to the historical ensemble mean (Figure 15a, b). This highlights the added-value of initialization in constraining modelled climate trends either by accounting for the phases of internal variability and/or correcting errors in model climate sensitivity. Results are more contrasted over the Mediterranean (MED) region, with some hindcasts having larger differences in the surface temperature trend than the historical ensemble mean in comparison to the observations, for example in summer for the CNRM-ESM2-1 model (Figure 15c).

A deterioration in the representation of surface temperature trends in decadal predictions is visible at forecast year 10 (Figure 15d, e, f). For example, the winter surface temperature trend over the NEU region in the historical ensemble mean of EC-Earth3 is closer to the observed one than for the one predicted at year 10, which has a larger trend. Such a difference was not observed at forecast year 5, which might indicate that any added value of the initialization of internal variability is lost by the end of the forecast. This can also be the fingerprint of interaction between forecasted signal and drifts as hindcast simulations are not at equilibrium state. More importantly, that difference in surface temperature trends between the historical simulations and the forecasts highlights a potential inconsistency that would need to be addressed by the blending methods. Interestingly, other regions and seasons (e.g. the yearly trends in the MED area) appear to have more consistent values between the mean of the historical and prediction ensembles.

Beyond the blending, we also see that the added value of decadal predictions in the representation of surface temperature trend in comparison to the historical simulations is however sometimes limited, in particular at the longest forecast times. This should be taken into account when implementing a blending method, for example by implementing the blending at shorter forecast times, to thus leverage as much as possible the predictability that arises from initialization.

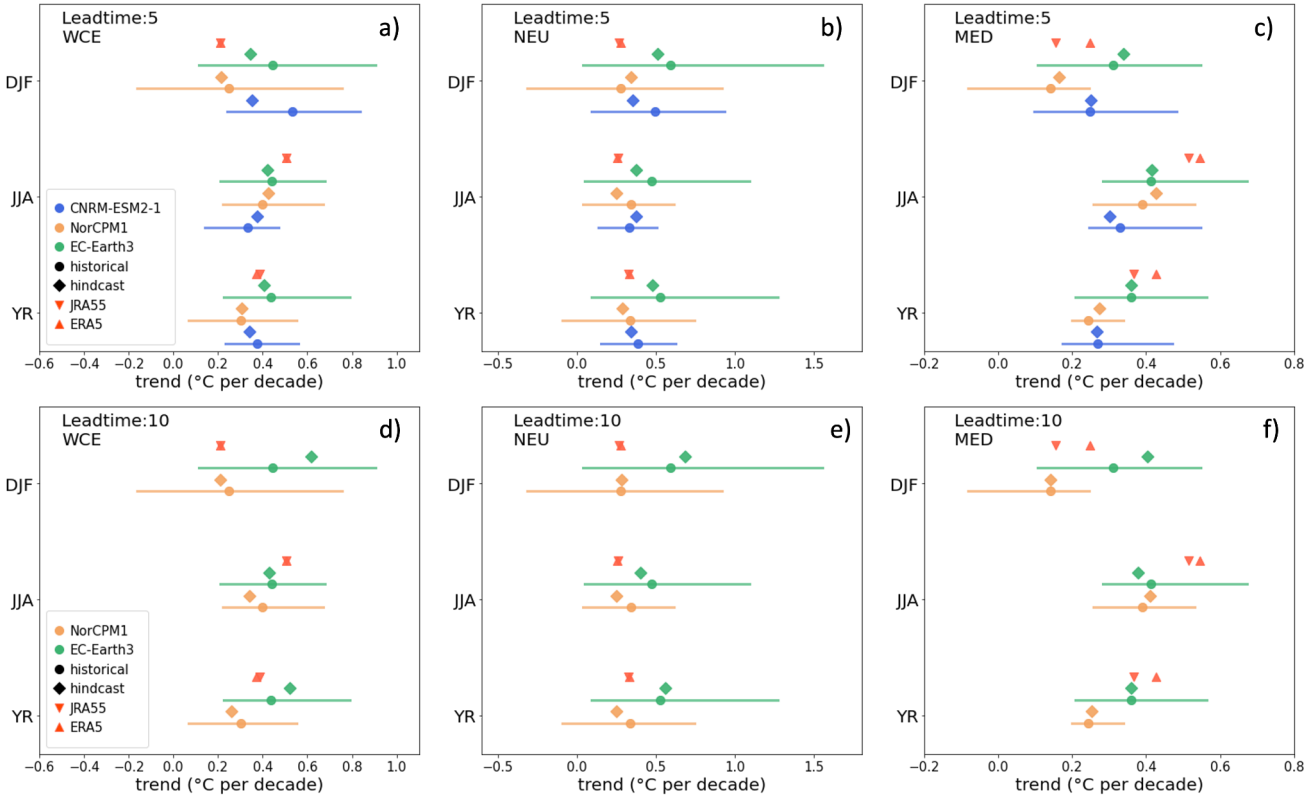

Figure 15: Winter, summer and annual surface temperature trends calculated over the 1970-2014 period for the (blue) CNRM-ESM2-1, (beige) NorCPM1 and (green) EC-Earth3 models. The dot shows the historical ensemble mean and the line shows the minimum and maximum among the ensemble members. The diamond shows the hindcast ensemble mean for a leadtime (i.e. forecast year) of (a,b,c) 5 years and (d,e,f) 10 years. Trends in the ERA5 and JRA55 reanalyses are shown with red up

triangle and red down triangle, respectively. The trends are calculated over three European IPCC regions (a, d) WCE, (b, e) NEU and (c, f) MED.

3.6 Progress Beyond State of the Art

This deliverable has presented novel work that has revealed new important inconsistencies that could prevent an effective implementation of seamless predictions that blend information from initialised decadal predictions and multi-decadal climate projections, which have been so far neglected. These inconsistencies will be tackled in the blending methodologies to be developed in Task 5.2 to underpin the reliability and applicability of the final blended products.

3.7 Discussion and Next Steps

The inconsistencies revealed in this deliverable cover a broad range of variables and metrics, which reflect the scientific interests and backgrounds of the different partners involved. The analyses have also been tailored to assess the consistency for variables that describe processes that are relevant for the different blending approaches that will be developed by each group, from the dominant modes of atmospheric and oceanic variability that drive the regional climate responses (e.g. over Europe), to the regional local trends and the main weather-types affecting the demonstrator cities.

The main lessons derived from the different analyses performed are now summarised:

- Important inconsistencies between decadal predictions and multi-decadal projections have been identified both for their statistical properties (e.g. ensemble spread, extreme and median percentiles of their probabilistic distribution, trends) and physical processes (e.g. weather types frequencies, NAO-AMV interactions, global modes of SST variability).
- By the end of the predictions (typically finishing on forecast year 10) their inconsistencies with the historical ensemble tend to be smaller, as the model has had a longer time to move from an initial state constrained by observations to its own equilibrium. Implementing the blending for the last forecast year could be therefore an optimal strategy to improve the seamlessness of the final blended product.
- The North Atlantic has emerged as a region that is more prone to develop strong inconsistencies between the two datasets of interest. This is probably explained by the fact that it is a key source region of decadal variability, where the predictive value of initialization is expected to last longer.
- Anomaly initialised prediction seems to reduce the inconsistencies between the decadal predictions and the multi-decadal projections, and could be prioritised in the production of the blended products. This result, however, still needs to be confirmed with a larger ensemble of models, as current analyses have only considered two anomaly initialised and two full-field initialised prediction systems.

4 Impact

We will share these main lessons with other colleagues and collaborators in the sister project ASPECT, so they can leverage them to develop more effective blending and constraining methodologies. Some of the lessons are also relevant for decadal prediction centres, and will therefore be shared with the I4C partners involved in WP2.

In this way the outcomes of this deliverable will contribute to I4C Expected Outcomes 3 ("Improved assessment of risks for people and systems exposed to extreme weather a climate events") and 4 ("Enhanced scientific collaboration and exploitation of synergies across the EU and Associated Countries").

5 Links Built

Lessons learned in this deliverable will inform the blending methodologies that will be developed in Task 5.2, fully assessed in Task 5.3, and reported in Deliverables 5.2 and 5.3. It will also inform WP3 regarding large scale predictors for use in statistical downscaling of decadal and historical/scenario simulations that will be developed in T3.1.

6 Communication, Dissemination and Exploitation

6.1.1.1 Participation to conferences

- The work by CSIC on Weather types will be presented in the next EGU General Assembly 2024 in Vienna, Austria.
- The contributions from DMI will also be presented in the EGU General Assembly 2024 in Vienna, Austria.

6.1.1.2 Participation to dissemination project activities:

- Results from this deliverable will be included in a general presentation to the rest of the project partners on the blending of predictions (date and final format still to be confirmed).

6.2 Peer Reviewed Articles

There are articles by all groups in preparation related to the corresponding work shown in this deliverable.

1. Christiansen, B., Yang, S., & Drew, A. (2024), The Atlantic Multi-decadal Variability in observations and in a large historical multi-model ensemble: Forced and internal variability, J. Clim., **in review**. Submitted version available at: www.researchgate.net/publication/374632170

7 References

Befort, D. J., Brunner, L., Borchert, L. F., O'Reilly, C. H., Mignot, J., Ballinger, A. P., Hegerl, G. C., Murphy, J. M., & Weisheimer, A. (2022). Combination of Decadal Predictions and Climate Projections in Time: Challenges and Potential Solutions. *Geophysical Research Letters*, *49*(15), e2022GL098568. https://doi.org/10.1029/2022GL098568

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., Passos, L., Bentsen, M., Guo, C., Gupta, A., Tjiputra, J., Kirkevåg, A., Olivié, D., Seland, Ø., Solsvik Vågane, J., … Eldevik, T. (2021). NorCPM1 and its contribution to CMIP6 DCPP. *Geoscientific Model Development*, *14*(11), 7073–7116. https://doi.org/10.5194/gmd-14-7073-2021

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Cruz-García, R., Cvijanovic, I., Doblas-Reyes, F. J., Donat, M., Dutra, E., Echevarría, P., Ho, A.-C., Loosveldt-Tomas, S., Moreno-Chamarro, E., Pérez-Zanon, N., Ramos, A., … Vegas-Regidor, J. (2021). Assessment of a full-field initialised decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth System Dynamics*, *12*(1), 173–196. https://doi.org/10.5194/esd-12-173-2021

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., & Eade, R. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. Geoscientific Model Development, 9(10), 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Brands, S., Fernández-Granja, J. A., Bedia, J., Casanueva, A., & Fernández, J. (2023). A Global Climate Model Performance Atlas for the Southern Hemisphere Extratropics Based on Regional Atmospheric Circulation Patterns. *Geophysical Research Letters*, *50*(10), e2023GL103531. https://doi.org/10.1029/2023GL103531

Brands, S. (2022). A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the Northern Hemisphere mid-to-high latitudes. *Geoscientific Model Development*, *15*(4), 1375–1411. https://doi.org/10.5194/gmd-15-1375-2022

Christiansen, B., Yang, S., & Matte, D. (2023). Estimating the Significance of the Added Skill from Initializations: The Case of Decadal Predictions. *Journal of Climate*, *36*(9), 2781–2793. https://doi.org/10.1175/JCLI-D-22-0605.1

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3–4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Doblas-Reyes, F. J., Hagedorn, R., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. Tellus A: Dynamic Meteorology and Oceanography, 57(3), 234. https://doi.org/10.3402/tellusa.v57i3.14658

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., … Zhang, Q. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geoscientific Model Development*, *15*(7), 2973–3020. https://doi.org/10.5194/gmd-15-2973-2022

Fernández-Granja, J. A., Brands, S., Bedia, J., Casanueva, A., & Fernández, J. (2023). Exploring the limits of the Jenkinson–Collison weather types classification scheme: A global assessment based on various reanalyses. *Climate Dynamics*, *61*(3–4), 1829–1845. https://doi.org/10.1007/s00382-022-06658-7

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., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Jolliffe, I. T., & Primo, C. (2008). Evaluating Rank Histograms Using Decompositions of the Chi-Square Test Statistic. *Monthly Weather Review*, *136*(6), 2133–2139. https://doi.org/10.1175/2007MWR2219.1

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., … Vertenstein, M. (2015). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Laloyaux, P., De Boisseson, E., Balmaseda, M., Bidlot, J., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach, H., Kosaka, Y., Martin, M., Poli, P., Rayner, N., Rustemeier, E., & Schepers, D. (2018). CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, *10*(5), 1172–1195. https://doi.org/10.1029/2018MS001273

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., … Marotzke, J. (2019). The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2050–2069. https://doi.org/10.1029/2019MS001639

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., & Ruprich-Robert, Y. (2021). Constraining Decadal Variability Yields Skillful Projections of Near-Term Climate Change. *Geophysical Research Letters*, *48*(24), e2021GL094915. https://doi.org/10.1029/2021GL094915

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J., Moine, M., Msadek, R., … Madec, G. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182–4227. https://doi.org/10.1029/2019MS001791

Tian, T., Yang, S., Karami, M. P., Massonnet, F., Kruschke, T., & Koenigk, T. (2021). Benefits of sea ice initialization for the interannual-to-decadal climate prediction skill in the Arctic in EC-Earth3. *Geoscientific Model Development*, *14*(7), 4283–4305. https://doi.org/10.5194/gmd-14-4283-2021

Verfaillie, D., Doblas-Reyes, F. J., Donat, M. G., Pérez-Zanón, N., Solaraju-Murali, B., Torralba, V., & Wild, S. (2021). How Reliable Are Decadal Climate Predictions of Near-Surface Air Temperature? Journal of Climate, 34(2), 697–713. https://doi.org/10.1175/JCLI-D-20-0138.1

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., & Lovenduski, N. S. (2018). Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model. *Bulletin of the American Meteorological Society*, *99*(9), 1867–1886. https://doi.org/10.1175/BAMS-D-17-0098.1

IMPETUS4CHANGE (I4C)

IMPROVING NEAR-TERM CLIMATE PREDICTIONS FOR SOCIETAL TRANSFORMATION

Grant agreement ID: 101081555

Call: HORIZON-CL5-2022-D1-02

Type of Action: HORIZON-RIA

Start date: 1 November 2022

Duration: 48 months

Website	impetus4change.eu
Twitter	@I4C_eu
LinkedIn	Impetus4Change
Zenodo repository for I4C open access documents	Impetus4Change Community