A Range of Outcomes: The Combined Effects of Internal Variability and Anthropogenic Forcing on Regional Climate Trends over Europe

Clara Deser* and Adam S. Phillips

National Center for Atmospheric Research, Boulder CO USA

EGU Nonlinear Processes in Geophysics Special Issue

"Interdisciplinary perspectives on climate sciences – highlighting past and current scientific achievements"

Submitted October 14, 2022

Revised December 27, 2022

* Corresponding author: Clara Deser cdeser@ucar.edu

Abstract

Disentangling the effects of internal variability and anthropogenic forcing on regional climate trends remains a key challenge with far-reaching implications. Due to its largely unpredictable nature on long timescales, internal climate variability limits the accuracy of climate model projections, introduces challenges in attributing past climate trends, and complicates climate model evaluation. Here, we highlight recent advances in climate modeling and physical understanding that have led to novel insights on these key issues. In particular, we synthesize new findings from Earth System Model and Observationally-based Large Ensembles (LEs), along with empirical "dynamical adjustment" methodologies. Using the new 100-member Community Earth System Model version 2 (CESM2) LE, we show that internal climate variability imparts considerable uncertainty to past and future 50-year trends in wintertime temperature and precipitation over Europe. Quantitatively similar levels of uncertainty in internally-generated 50-year trends are found for the Observational-LE. The observed thermodynamic-residual trends based on "dynamical adjustment" compare well with the CESM2-LE forced response, which is dominated by thermodynamic processes. Combining the internal variability of trends from the Observational-LE with the observed thermodynamic-residual trend yields a purely observationally-based range of trend outcomes, and provides a powerful test of the range of simulated trends in the CESM2-LE.

1. Introduction

a. Internal variability and forced climate change

The climate system is highly variable in both space and time. This variability originates from processes within the coupled ocean-atmosphere-cryosphere-land-biosphere system, as well as from external influences such as solar and orbital cycles, volcanic eruptions, and anthropogenic emissions of greenhouse gases and sulfate aerosols. A primary source of internally-generated variability is the atmospheric general circulation, which produces familiar day-to-day and week-to-week weather fluctuations. The non-linear nature of atmospheric dynamics limits predictability to less than a few weeks; beyond this time scale, atmospheric motions may be considered as random stochastic processes, often termed "weather noise" (e.g., Lorenz, 1963; Leith, 1973; James and James, 1992). It is important to note that such "weather noise" imparts variability on a continuum of time scales, from sub-monthly to decadal and longer (e.g., Madden, 1975; Deser et al. 2012; Thompson et al. 2015).

Another important source of internally-generated variability is the coupling between the ocean and atmosphere. Large-scale air-sea interactions give rise to distinctive patterns (or "modes") of variability on interannual and longer time scales, including phenomena such as "El Niño – Southern Oscillation" (ENSO; Wang et al. 2017), "Pacific Decadal Variability" (PDV; Newman et al. 2016) and "Atlantic Multi-decadal Variability" (AMV; Zhang et al. 2019). Like the atmospheric general circulation, these coupled modes are governed by non-linear dynamical processes which limit their predictability. For example, forecast skill is generally limited to 1-2 years for ENSO (Jin et al., 2008; DiNezio et al. 2017; Wu et al. 2021), 5 years for PDV (Teng and Branstator, 2010; Meehl et al., 2016; Gordon and Barnes, 2022) and 10 years for AMV (Griffies and Bryan, 1997; Trenary and DelSole, 2016; Yeager et al., 2018). Beyond these predictability time horizons, internally-generated variability can be thought of as a "roll of the dice", introducing unavoidable uncertainty to climate model projections especially at local and regional scales (e.g., Deser et al. 2012, 2014 and 2020a).

Not only does unpredictable internal variability cause irreducible uncertainty in future climate projections, it also confounds interpretation of the historical climate record. For example, internal variability may partially obscure the regional climate response to external forcings including industrial greenhouse gas emissions, stratospheric ozone depletion and volcanic eruptions (Wallace et al., 2013; Schneider et al. 2015; Lehner et al. 2016; McGraw et al. 2016). In some areas, climate trends driven by internal processes may even outweigh those due to anthropogenic influences over the past 30-60 years (Deser et al., 2012, 2016 and 2017; Wallace et al., 2013; Swart et al. 2015; Lehner et al. 2017). It is important to note that such internally-generated multi-decadal trends need not originate from slow processes within the ocean or coupled ocean-atmosphere system: indeed, random fluctuations of the atmospheric circulation independent of oceanic influences have been shown to drive a large fraction of long-term precipitation and temperature trends over North America and Eurasia (Deser et al. 2012; McKinnon and Deser, 2018). The co-existence of internal and anthropogenic factors necessitates a probabilistic approach to detection and attribution of the human contribution to extreme weather events.

The prevalence of internal climate variability also complicates model evaluation efforts, since the simulated temporal sequence of (unpredictable) internal variability need not match observations even if the model's physics are realistic. Further, the brevity of the instrumental record provides only a limited sampling of internal variability, hindering robust model evaluation. Thus, climate models may show an apparent bias with respect to observations, but this could be entirely attributable to sampling issues rather than indicative of a true bias due to incorrect model physics. Apparent model bias due to sampling uncertainty must be kept in mind when assessing fidelity of simulated modes of internal variability (e.g., Wittenberg et al. 2009; Deser et al. 2017; Capotondi et al. 2020; Fasullo et al. 2021; McKenna and Maycock, 2021), transient climate sensitivity (Dong et al. 2021; Andrews et al. 2022), and "signal-to-noise" properties of initial-value predictions and forced responses (e.g., Scaife and Smith, 2018; Smith et al., 2020; Klavans et al. 2021). In particular, even with 100 years of data, sampling uncertainty is a limiting factor for evaluating ENSO properties in climate models, including its global atmospheric teleconnections and associated climate impacts (Deser et al. 2017 and 2018; Capotondi et al. 2020) and forced changes thereof (Stevenson et al. 2012; Maher et al. 2018; Maher et al. 2022; O'Brien and Deser, 2022). This issue is particularly acute for model assessment of modes of decadal variability such as PDV and AMV due to the paucity of samples in the short instrumental record (Deser and Phillips 2021; Fasullo et al. 2021).

b. Initial-condition Large Ensemble Simulations with Earth System Models

To overcome the issue of sampling uncertainty, a recent thrust in climate modeling is to run a large number of simulations (30-100) with the same coupled model and the same radiative forcing protocol (historical and/or future scenario) but vary the initial conditions. The initial-condition variation can be accomplished by introducing a random perturbation to the atmosphere on the order of the model's numerical round-off error (e.g., 10^{-14} K in the case of atmospheric temperatures; Kay et al. 2015) or it can be done by selecting a different ocean state from a long control run of the coupled model, or a combination of the two (Deser et al. 2020a and Rodgers et al. 2021). Regardless of the method used, the initial-condition perturbation serves to create ensemble spread once the memory of the initial state is lost, typically within a month for the atmosphere and a few years to a couple of decades for the ocean (Yeager et al., 2018). The ensuing ensemble spread is thus solely attributable to random internal variability (e.g., the "butterfly effect" in chaos theory); see Lorenz (1963) and Tel et al. (2019). Because the temporal sequences of internal variability unfold differently in the various ensemble members once the memory of the initial conditions is lost, one can estimate the forced component at each time step (at each location) by averaging the members together, provided the ensemble size is sufficiently large. The internal component in each ensemble member is then obtained as a residual from the ensemble-mean. Note that a larger ensemble may be needed for some aspects of the forced response than others, depending on the relative magnitudes of the forced response and internal variability (Milinski et al., 2020). For example, forced changes in ocean heat content may be readily detected with just a few members (Fasullo and Nerem, 2018), while forced changes in atmospheric circulation (Deser et al., 2012) or precipitation and temperature extremes (Tebaldi et al. 2021) may require 20-30 members. Detecting forced changes in the characteristics of internal variability itself, such as its amplitude, spatial pattern and remote teleconnections, may necessitate even larger ensembles (Milinski et al., 2020; Bódai et al., 2020; Bódai et al., 2022; O'Brien and Deser, 2023).

Initial-condition Large Ensembles (LEs for short) have proven enormously useful for separating internal variability and forced climate change on regional scales in models, and for providing robust sampling of models' internal variability by pooling together all of the ensemble members (e.g., Deser et al., 2012; Kay et al., 2015; Maher et al., 2019; Deser et al. 2020a; Lehner et al., 2020). They have also been used to assess externally-forced changes in the characteristics of simulated internal variability, including extreme events for which large sample sizes are crucial (e.g., Tebaldi et al., 2021; O'Brien and Deser, 2023). Additionally, they have served as methodological testbeds for evaluating approaches to detection and attribution of anthropogenic climate change in the (single) observational record (e.g., Deser et al., 2016; Barnes et al., 2019; Sippel et al., 2019 and 2021; Santer et al. 2019; Bonfils et al., 2019; Wills et al., 2020). Until the advent of LEs, it was problematic to identify the sources of model differences in the Coupled Model Intercomparison Project (CMIP) archives due to the limited number of simulations (generally < 3) for each model (i.e., structural uncertainty was confounded with uncertainty due to internal variability). This concern has been largely alleviated thanks to the recent availability of LEs with multiple earth system models (e.g., Deser et al. 2020a; Lehner et al., 2020).

c. Observationally-based Large Ensemble

Just as in a model LE, the sequence of internal variability in the real world could have unfolded differently. That is, the observational record traces only one of many possible climate histories that could have happened under the same external radiative forcing. For example, El Niño and La Niña events could have occurred in a different set of years, and positive or negative regimes of PDV and AMV could have taken place in different decades. This concept of alternate chronologies, sometimes referred to as the “Theory of Parallel Climate Realizations” (Tel et al., 2019) or the notion of “Contingency” (Gould, 1989), has major implications that call for a reframing of perspective. For example, it means that a single model simulation of the historical period need not match the observed record, even if the model is “perfect” in its physical representation of the real world’s climate. However, the statistical characteristics of the model’s internal variability must agree with those of the real world, taking into account sampling uncertainty (uncertainty due to limited sampling in the short observational record). Thus, while a single ensemble member need not match observations, the ensemble as a whole should encompass the instrumental data, provided there are enough members to adequately span the range of possible sequences of internal variability (Suarez-Guttierez et al. 2021).

Another implication of the concept of “parallel climate realizations” is that the climate trends we have experienced are not the only ones that could have occurred under the same radiative forcing conditions. In analogy with a model LE, the observational record is just one “member” of a larger set of possible “members”, each with a different (and largely unpredictable) chronology of internal variability. Although one cannot replay the “tape of history”, one can construct an “Observational LE” by generating alternate synthetic sequences of internal variability from the instrumental data. Conceptually, this involves removing an estimate of the forced component from the data and then randomizing the residual (internal) variability in time. Importantly, the randomization procedure must be done in a way that preserves the statistical properties of the observed variability including its variance, temporal autocorrelation, and spatial patterns. The resulting synthetic sequences of internal variability derived from the observational record can then be added back to the time-evolving forced response obtained from a climate model LE.

The development of statistically-based Observational LEs is just beginning, with recent efforts targeting surface climate fields (McKinnon et al., 2017; McKinnon and Deser, 2018 and 2021) and carbon dioxide fluxes across the air-sea interface (Olivarez et al. 2022). Here, we focus on the work of McKinnon and Deser (2018 and 2021) who constructed an Observational LE for global sea level pressure (SLP) and terrestrial precipitation and temperature based on ~100 years of monthly gridded instrumental data. To test the skill of their method, they applied it independently to each member of a climate model LE and then compared the results to the “true” statistical properties of the model's internal variability based on the full set of ensemble members. According to this test, their approach was found to be accurate to within 10-20% at most locations. They then constructed a large (1000 member) ensemble of plausible "parallel worlds" of what the observational record might have looked like had a different sequence of internal variability unfolded by chance. Their Observational LE has been used for many applications, including evaluation of internal variability in climate model LEs, assessment of uncertainty in observed 50-year climate trends, and quantification of extreme precipitation risk over the Upper Colorado River basin, a critical water resource for the western US (McKinnon and Deser 2018 and 2021).

d. Dynamical Adjustment

Determining the forced contribution to observed changes in climate remains an ongoing challenge. Most "Detection and Attribution" methods rely on climate models to provide a set of spatial and temporal "fingerprints" of forced climate change that are distinct from patterns of internal variability (Hegerl et al. 2007; Santer et al. 2019; Sippel et al. 2019). These model-based "fingerprints" are then used to assess the proportion of observed climate change that is due to external forcing. However, model shortcomings may limit the accuracy of such methods. Thus, it is also desirable to develop complementary approaches to attribution that do not rely on climate model information. Two such methods, Linear Inverse Modeling (Newman, 2007) and Low-Frequency Pattern Analysis (Wills et al. 2020), leverage the assumption that forced climate change evolves slowly compared to the time scales of internal variability. However, decadal shifts in regional anthropogenic aerosol emissions (Deser et al. 2020b; Persad et al. 2018), in addition to decadal changes in solar and volcanic activity and the rate of greenhouse gas rise, present challenges to this assumption and may complicate interpretation of the results.

A complementary, physically-based approach to isolating the externally-forced response in observations without reliance on climate model information is the technique of "Dynamical Adjustment". This method aims to remove the influence of atmospheric circulation variability from observed temperature and precipitation data, thereby revealing the thermodynamically-induced component of observed climate change (Wallace et al. 2013; Smoliak et al. 2015; Deser et al. 2016; Guo et al. 2019). According to the current generation of coupled climate models, the forced component of extra-tropical atmospheric circulation changes is small relative to internal variability (Deser et al. 2012; Shepherd, 2014). If models are correct in this regard, then dynamical adjustment can be used to parse the relative contributions of internal dynamics and forced thermodynamics to observed climate changes at middle and high latitudes (Wallace et al. 2013; Deser et al. 2016). A variety of dynamical adjustment algorithms have been developed and tested within the framework of a model LE (Deser et al., 2016; Lehner et al., 2017 and 2018; Smoliak et al., 2015; Guo et al. 2019; Merrifield et al., 2017; Terray 2021; Sippel et al. 2019). These protocols are all based on statistical associations between patterns of SLP and temperature or precipitation deduced from long observational records. Generally, the data are high-pass filtered or detrended so as to avoid aliasing any potential forced component onto the statistical relationships. These procedures generally work best for large-amplitude SLP anomaly patterns, and are more effective for temperature than precipitation due to higher levels of noise in the latter (Guo et al. 2019).

2. Data and Methods

We make use of a state-of-the-art 100-member LE conducted with the National Center for Atmospheric Research (NCAR) Community Earth System Model version 2 (CESM2), described in Rodgers et al. (2021). This publicly-available LE resource is unprecedented for its combination of large ensemble size, high spatial resolution (approximately 1° in both latitude and longitude), and length of simulation (1850-2100). Each ensemble member is driven by the same radiative forcing scenario (historical from 1850-2014, and SSP3-7.0 from 2015-2100), but begins from a different state on 1 January 1850, taken from a long pre-industrial control simulation. We analyze linear trends in air temperature, precipitation and sea level pressure over the past 50 years (1972-2021) and projected for the next 50 years (2022-2071). It should be noted that memory of the initial state is negligible by the middle of the 20th century for the quantities we analyze; thus, diversity in trends amongst the individual ensemble members is solely due to different random samples of internal variability, which are superimposed upon a common forced response.

For consistency with the 100-member CESM2 LE, we make use of the first 100 members of the Observational LE (OBS LE) constructed by McKinnon and Deser (2018) to illustrate the diversity of past 50-year trends consistent with the statistical spatio-temporal properties of internal variability in the observational record. For the purpose of comparing directly to the CESM2 LE, we have added the model's forced trend to the internal trend of each OBS LE member. The OBS LE is based on the Berkeley Earth Surface Temperature (BEST) dataset (Rohde et al. 2013), the Global Precipitation Climatology Centre (GPCC) dataset (Schneider et al. 2008), and the Twentieth Century Reanalysis version 2c (20CR) sea level pressure (SLP) dataset (Compo et al. 2011).

We apply the dynamical adjustment methodology of Deser et al. (2016) based on SLP "constructed circulation analogues" to monthly temperature and precipitation during 1900-2021, using the same observational data sets as in the OBS LE. The reader is referred to Deser et al. (2016) for details of the methodology, and to Lehner et al. (2017 and 2018), Guo et al. (2019) and Terray (2021) for additional applications.

For each ensemble member of the CESM2 and OBS LEs, we form monthly anomalies by subtracting the long-term means for each month individually, and then form seasonal averages (December-February) of the monthly anomalies. We compute 50-year trends of the wintertime anomalies using linear least-squares regression analysis. All results shown in this study are original findings.

3. European climate trends

We begin by illustrating the diversity of winter temperature and precipitation trends over Europe during the past 50 years (1972-2021) in the CESM2 and OBS LEs (Sections 3a and b), and projected for the next 50 years (2022-2071) in the CESM2 LE (Section 3c). We then provide a more quantitative view of the relative contributions of forced climate change and internal variability to past and future climate trends using a variety of signal-to-noise metrics, with comparison between the CESM2 and OBS LEs (Section 3d). We summarize the CESM2 LE results by showing the "expected range" of trend outcomes in Section 3e. Finally, we apply the technique of "dynamical adjustment" to estimate the forced component of observed temperature trends (Section 3f), and then use this estimate in conjunction with the OBS LE to produce a purely observational estimate of the plausible range of temperature trend outcomes over the past 60 years (Section 3g).

a. Past trends (1972-2021) in the CESM2 LE

The CESM2 model simulates a wide range of wintertime temperature trend patterns for the past 50 years due to the combined effects of internal variability and forced response, as illustrated by the first 28 members of the LE (Fig. 1). Recall that the only reason that these trend maps are not identical is because of random differences in internal variability between the members. While moderate warming is seen over most of the European continent in the majority of cases, as expected, some members show regions of considerably greater temperature increase (in excess of 1°C per decade for example members 1, 10 and 18), while others exhibit weak cooling in some locations (for example, members 17, 23 and 26; Fig. 1). The relative contributions of internal variability and forced response can be readily discerned by comparing the individual member trends with the ensemble-mean trend (see "EM" panel in Fig. 1). The observed trend ("OBS" panel in Fig. 1) bears a close resemblance to the model's forced trend in both amplitude and spatial pattern. This correspondence may be coincidental, as individual members of the CESM2 LE also resemble the forced response (for example, members 6 and 21), or it may suggest that the model overestimates the amplitude of internally-generated 50-year trends relative to forced trends. The OBS LE results shown below will shed some light on these two possibilities.

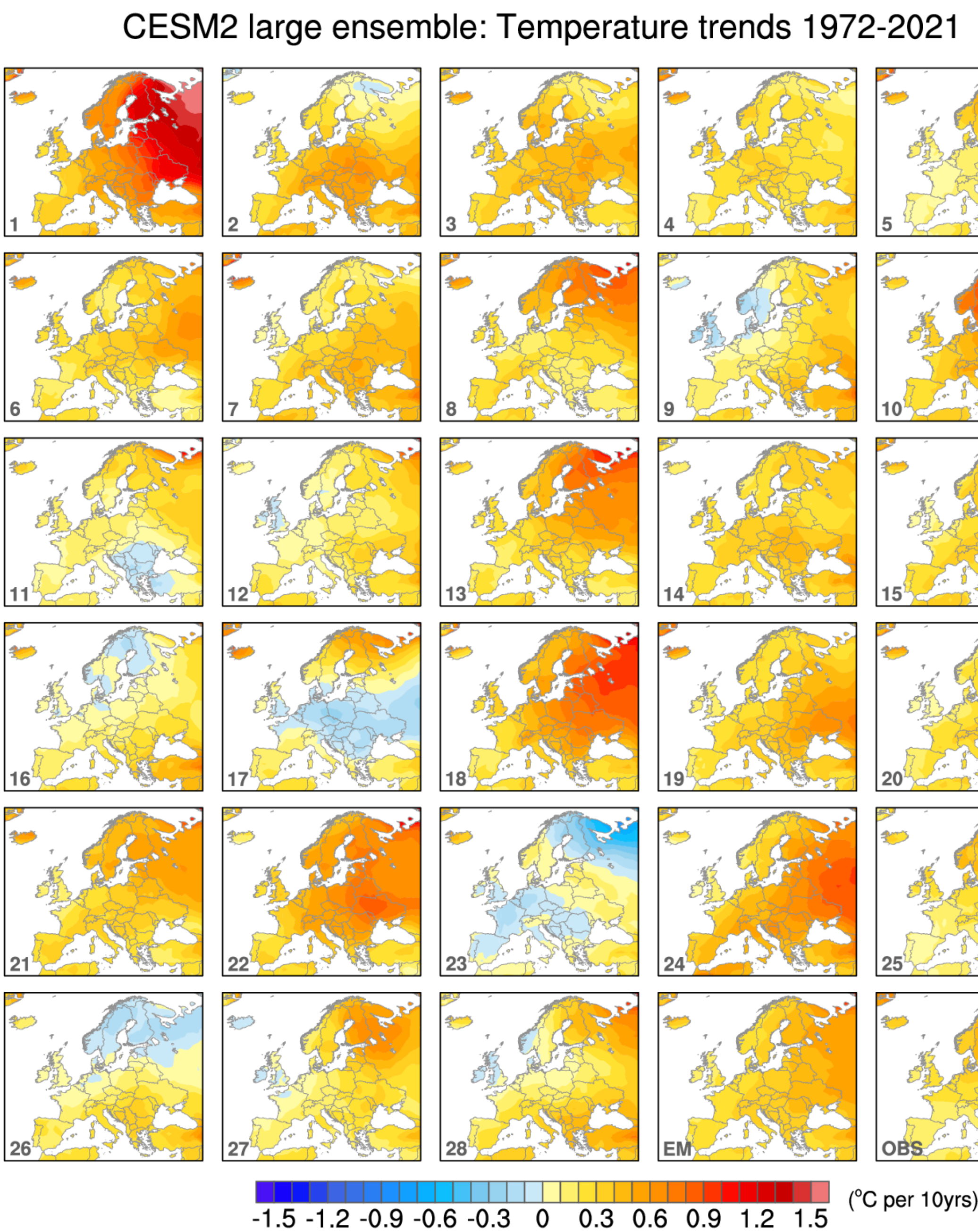

Figure 1. Winter air temperature trends (°C per decade) for the period 1972-2021 as simulated by the first 28 members of the CESM2 Large Ensemble (number in the lower left of each panel denotes the ensemble member) and the 100-member ensemble-mean (panel labeled "EM"). Observed trends are shown in the lower right (panel labeled "OBS").

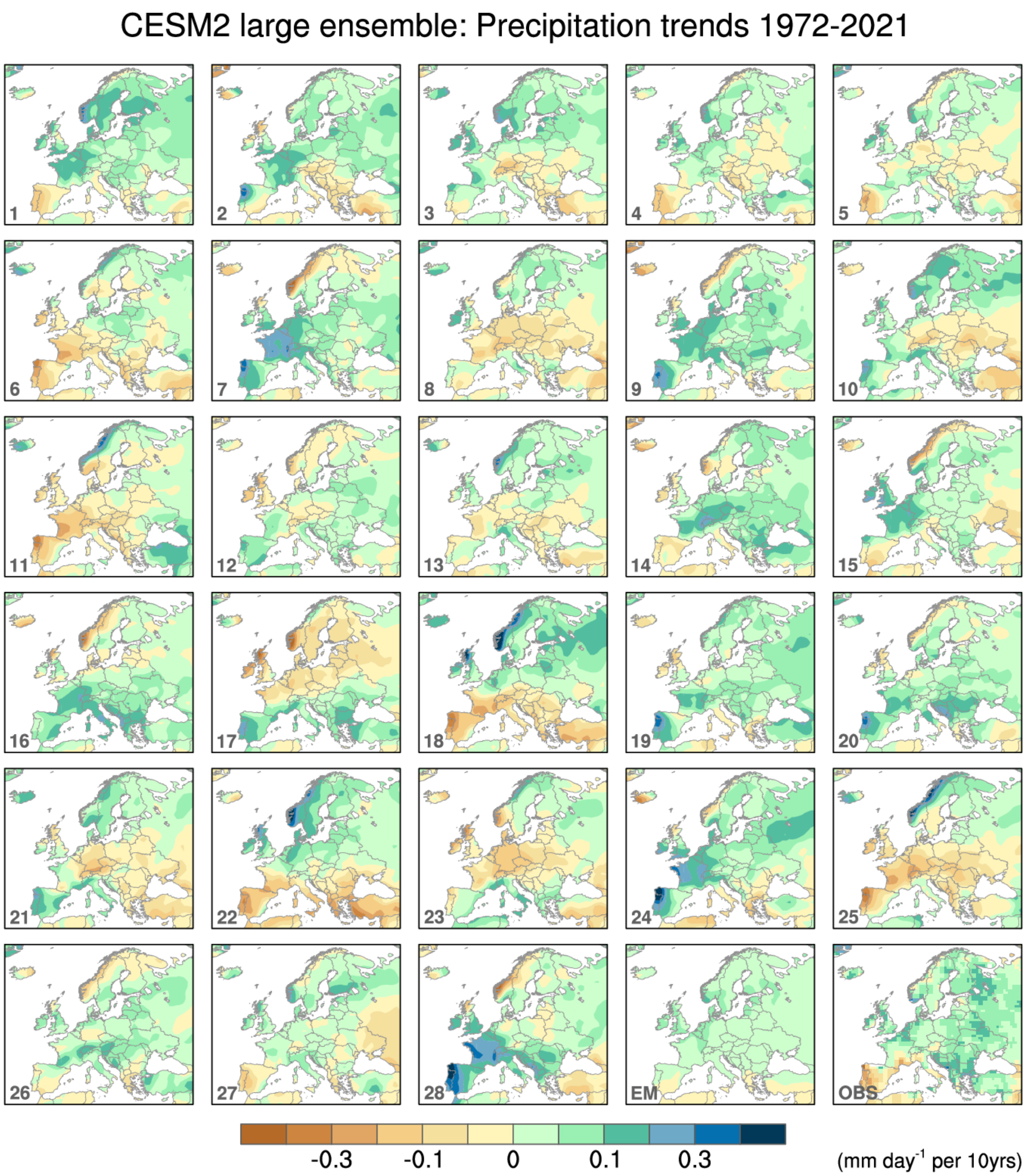

Figure 2. As in Fig. 1 but for precipitation (mm d-1 per decade).

Like temperature, precipitation trends also vary considerably across ensemble members (Fig. 2). While the ensemble-mean trend shows modest increases in precipitation throughout Europe (except for the southernmost fringes), internal variability can evidently overwhelm the forced response in individual simulations. For example, some members show drying over large parts of the continent, while others depict enhanced wetting in the same regions (compare, for example, members 22 and 28, which show nearly opposite patterns). Observed precipitation trends are generally positive, except over Spain, Portugal, southern France and other parts of the western Mediterranean (Fig. 2). The observed precipitation increases, while of the same sign as the model's forced response, are approximately twice as large in many areas. Again, the interpretation of the observed trends is ambiguous, since there are individual members that resemble observations (for example, member 1).

b. Past trends (1972-2021) in the OBS LE

The individual members of the OBS LE show a qualitatively similar diversity of 50-year temperature trends as the CESM2 LE (Fig. 3). Like CESM2, some members show weak cooling in some areas while others show widespread moderate or strong warming. This suggests that the resemblance between the observed trend and the model's forced response may be purely coincidental. Precipitation trends in the OBS LE also display large contrasts between members, similar to CESM2 (Fig. 4). For example, nearly opposite patterns are found between members 6 and 11 (or 8 and 9). Trend amplitudes also vary considerably across the OBS LE, with larger magnitudes in some members (for example, members 3 and 20) compared to others (e.g., members 21 and 13). While no single member of the 28 OBS LE samples shown matches the model's forced trend, member 21 with its relatively muted trends comes close.

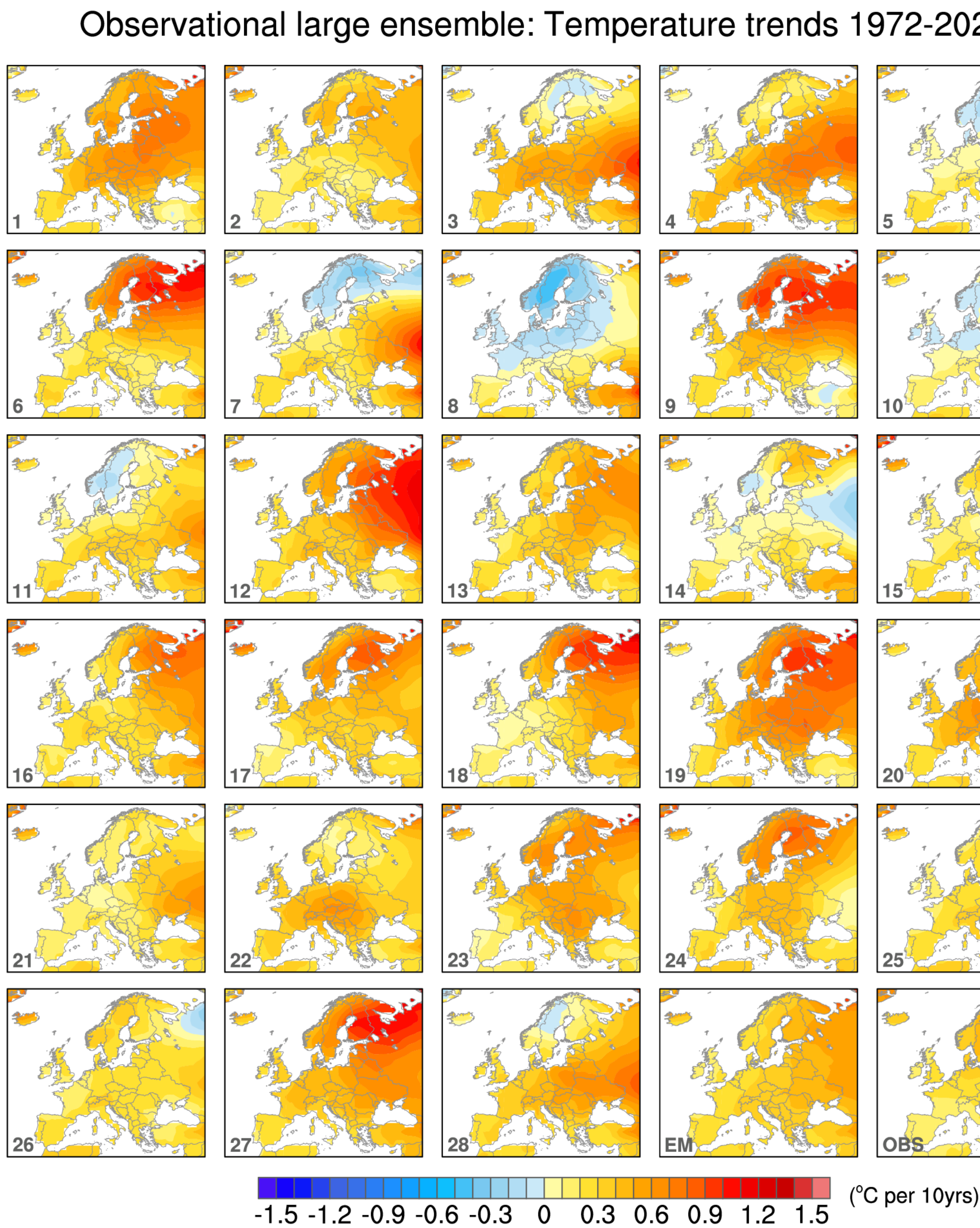

Figure 3. As in Fig. 1, but for the Observational Large Ensemble of McKinnon and Deser (2018) with the ensemble-mean from the 100-member CESM2 Large Ensemble. See text for details.

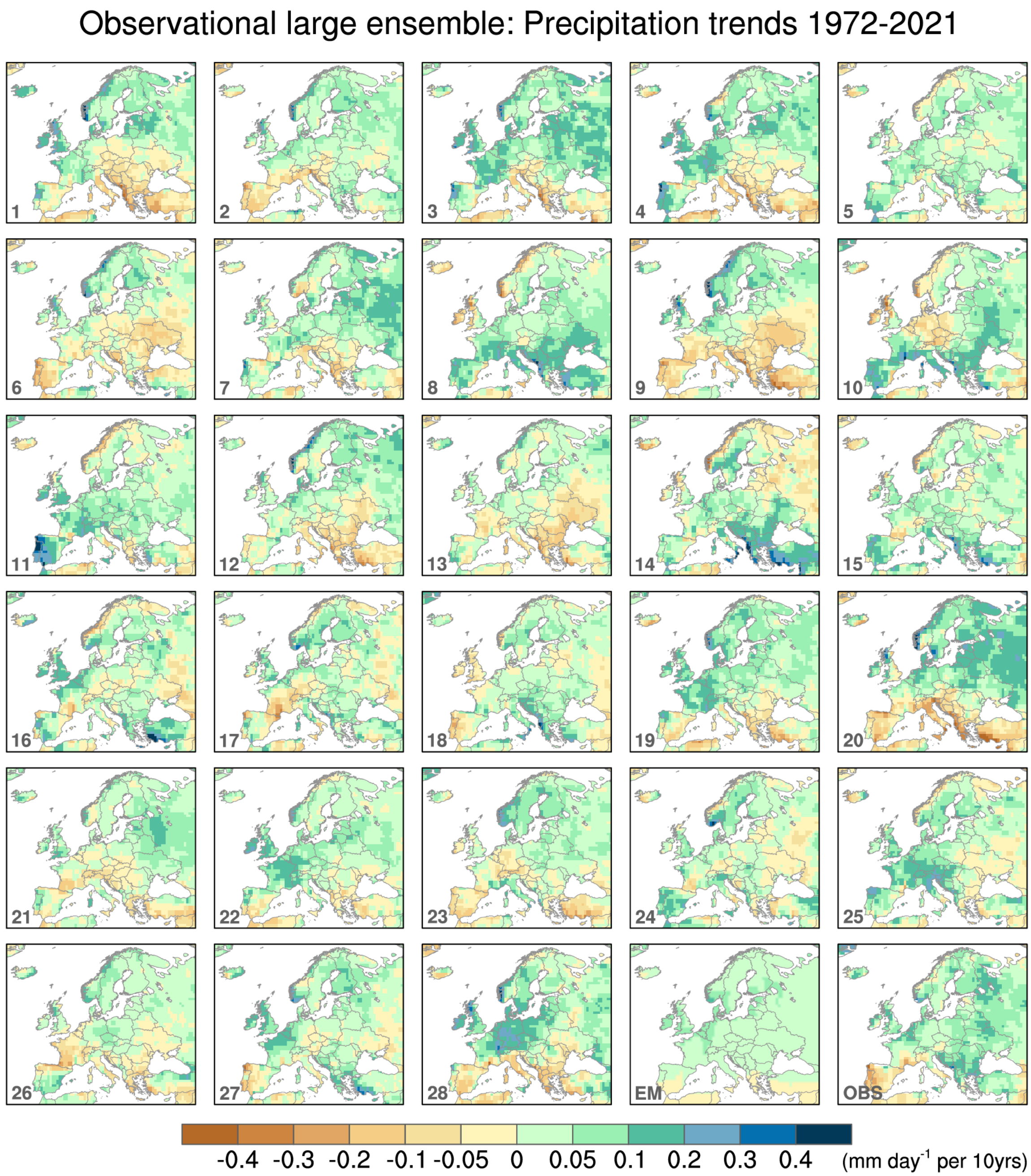

Figure 4. As in Fig. 2, but for the Observational Large Ensemble of McKinnon and Deser (2018) with the ensemble-mean from the 100-member CESM2 Large Ensemble. See text for details.

c. Future Trends (2022-2071) in the CESM2 LE

As expected, temperature trends projected for the next 50 years show larger amplitudes than those for the past 50 years in the CESM2 LE (Fig. 5). This is due to the fact that the forced (ensemble-mean) component of warming increases as greenhouse gas emissions accelerate. In most regions, the forced warming trend increases by approximately 0.2°C per decade in the future compared to the past. Notable exceptions are Iceland and the British Isles, which show less warming in the future due to a circulation-induced forced cooling trend (see Section 3e). Despite a larger forced component, temperature trends projected for the next 50 years still show a wide range of amplitudes across individual members of the CESM2 LE. For example, member 13 is striking for its muted warming (generally < 0.5°C per decade) across Europe (and absolute cooling over the UK and Iceland), while member 28 shows highly amplified warming, with values exceeding 1.3 °C per decade over western Russia.

Forced trends in precipitation are projected to amplify over the next 50 years, with greater wetting over northern Europe and drying over southern Europe and the Mediterranean (Fig. 6). In addition, the region with a forced drying trend is projected to expand northward into Spain, Italy and the Balkan Republics. While the forced pattern of future drying in the south and wetting in the north is generally evident in most of the simulations shown, there are notable differences in amplitude across the members. For example, member 28 shows precipitation trends in excess of 0.1 mm d^{-1} per decade over most of northern Europe, while member 11 shows positive precipitation trends of less than half this amount. Members 27 and 28 illustrate that the mid-section of the European continent may get wetter or drier depending on the unpredictable sequence of internal variability that unfolds. Thus, internal variability can still make a sizeable contribution to the projected patterns and amplitudes of winter precipitation trends over the next 50 years.

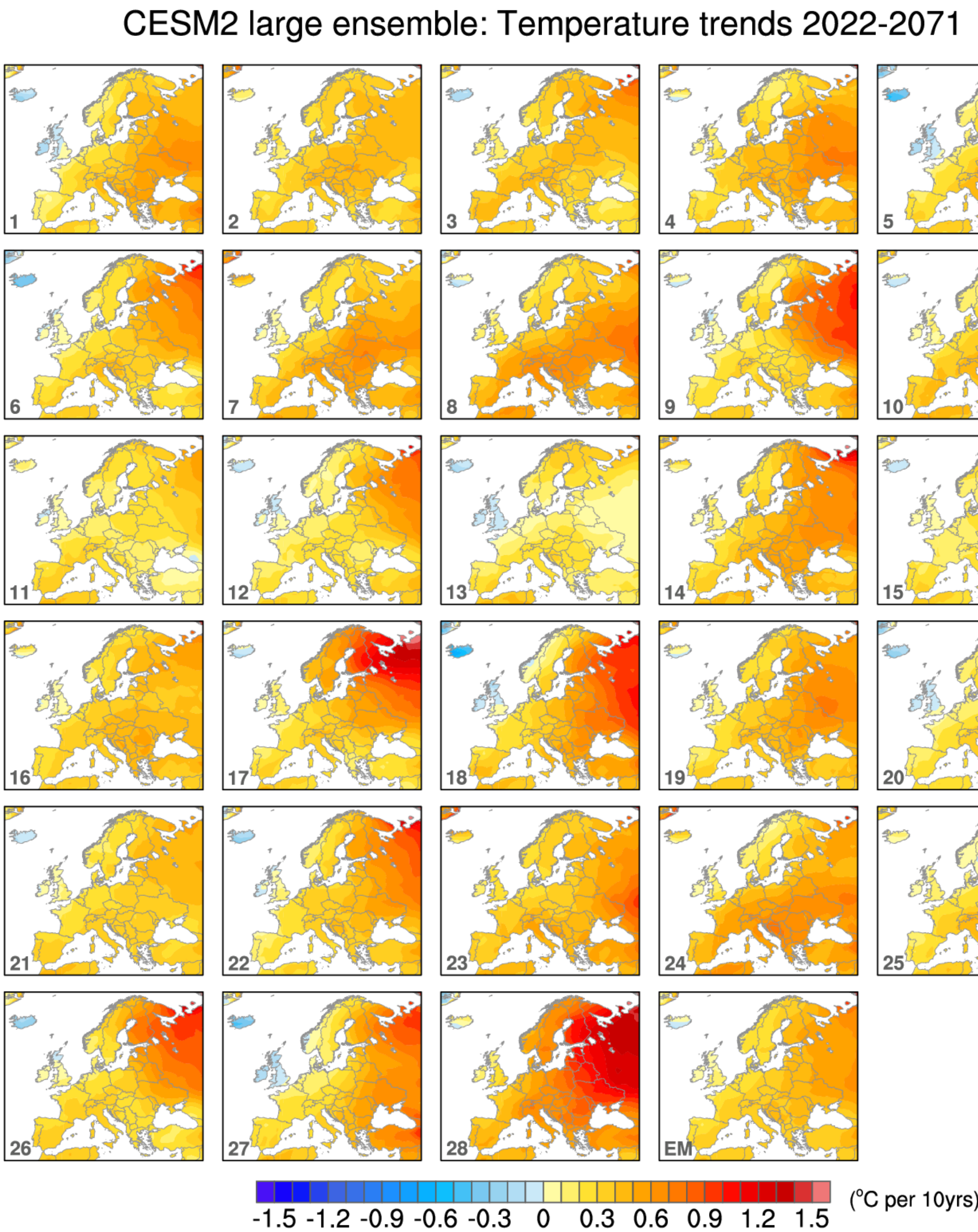

Figure 5. As in Fig. 1, but for the period 2022-2071.

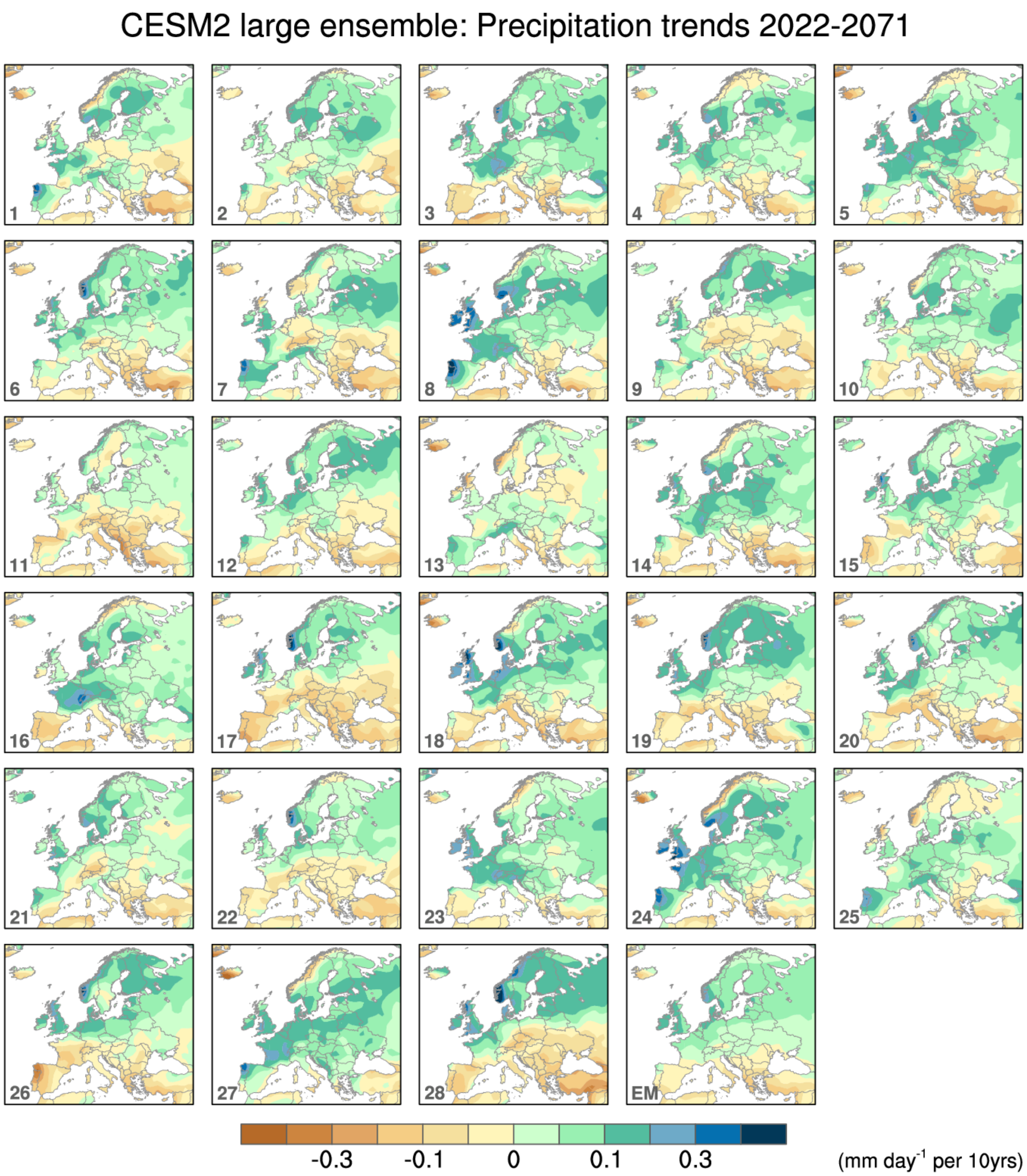

Figure 6. As in Fig. 2, but for the period 2022-2071.

d. Signal-to-noise metrics and model evaluation.

In the previous section, we conveyed a qualitative impression of the possible range of 50-year trends due to the superposition of internal variability and forced climate change in the CESM2 and OBS LEs. Here, we provide a more quantitative view, beginning with a comparison of the standard deviation (σ) of trends over the period 1972-2021 computed across the ensemble members of each LE. In the CESM2 LE, the ensemble σ of temperature trends increases from southwest to northeast, with minimum values (0.05-0.10 K per decade) over Spain and northern Africa, and maximum values (0.30-0.35 0.5°C per decade) over northwestern Russia (Fig. 7a). A similar pattern is found in OBS LE, with some regional differences in amplitude (Fig. 7b). In particular, the ensemble σ values are significantly smaller (20-40%) over Scandinavia, Germany and Poland, and significantly larger (20-40%) in areas near the Mediterranean and Black Seas, in the OBS LE compared to the CESM2 LE (Fig. 7c). For precipitation trends, the two LEs show similar patterns of ensemble σ, with largest amplitudes generally along the west coasts (0.10 - 0.25 mm d^{-1} per decade) and over southwestern Europe (values 0.05 – 0.10 mm d^{-1} per decade: Figs. 7d and e). However, CESM2 LE significantly underestimates the OBS LE by more than 40% along the Mediterranean and Black Seas and parts of Russia, and significantly overestimates the OBS LE by 20-40% in many areas of western Europe (Fig. 7f).

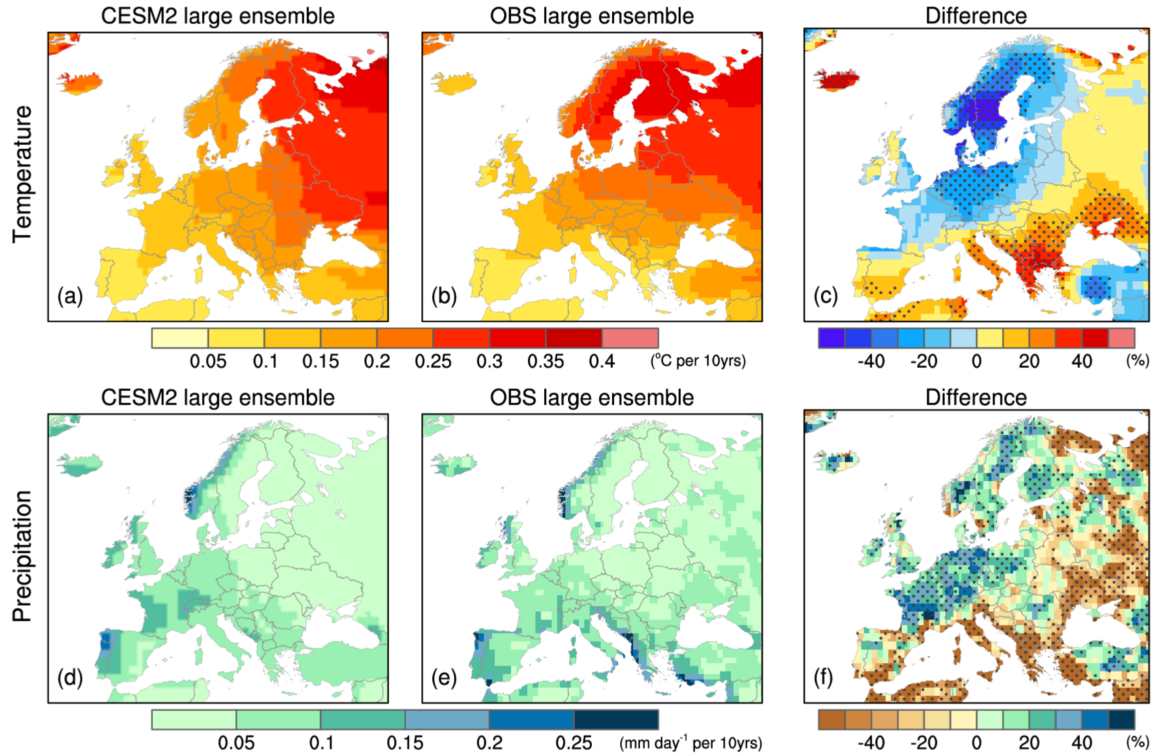

Figure 7. Standard deviation of 50-year trends (1972-2021) across 100 members of the CESM2 Large Ensemble (a,d) and 100 members of the Observational Large Ensemble (b,e), and their difference (c,f) for winter air temperature (top; °C per decade) and precipitation (bottom; mm d^{-1} per decade). Stippling in panels c and f indicates that the differences are statistically significant at the 95% confidence level according to an f-test.

Next, we assess the relative magnitude of the forced and internal components of trends by computing a "signal-to-noise" ratio defined as the CESM2 ensemble-mean trend divided by the σ of trends across the 100 members of each LE. This "signal-to-noise" ratio provides a metric of the likelihood that the ensemble-mean (e.g., forced) trend might be overwhelmed by the internally-generated trend in any given ensemble member (and by extension, the real world). Assuming that the 100-member set of 50-year trends follows a normal distribution (not shown, but see related results in Deser et al. 2012; Thompson et al. 2015; Deser et al. 2020a), a signal-to-noise ratio greater than one (two) indicates that the magnitude of the ensemble-mean (forced) trend is larger than (more than twice as large as) that of a typical (e.g., one standard deviation) internal trend, and a signal-to-noise ratio less than one indicates that the amplitude of a typical internal trend exceeds the magnitude of the forced trend. In the CESM2 LE, the signal-to-noise of forced temperature trends over the past 50 years generally ranges from 1.5 - 2 over central and northern Europe, and from 2-3 over southern Europe (Fig. 8a). Forced precipitation trends over the past 50 years exhibit much lower signal-to-noise ratios than temperature, with values generally < 1 and nearly always < 1.5 (Fig. 8d).

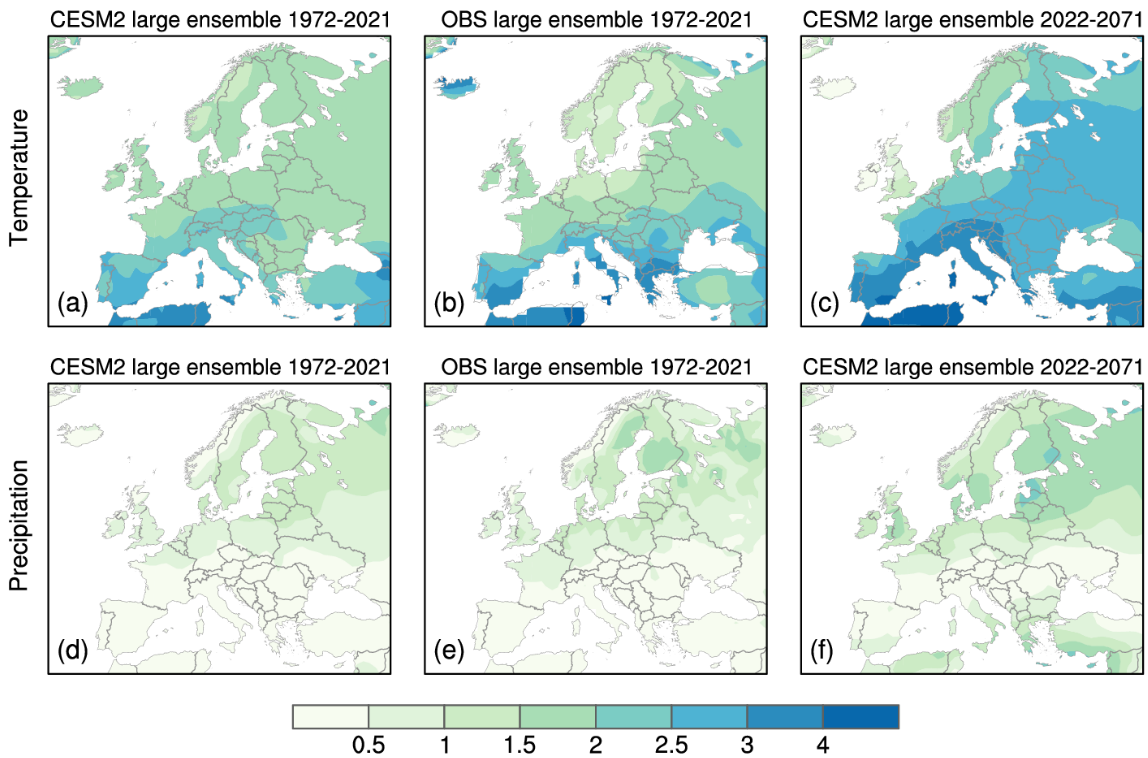

Figure 8. Signal-to-noise of forced trends in winter (top) air temperature and (bottom) precipitation based on the 100-member CESM2 Large Ensemble during 1972-2021 (a,d), the Observational Large Ensemble during 1972-2021 (b,e), and the CESM2 Large Ensemble during 2022-2071 (c,f). See text for details.

How much do model biases in ensemble σ shown previously affect the signal-to-noise of the model's forced trends? We address this question by using the OBS LE σ values in place of the model's σ values in the signal-to-noise calculation (note that the "signal" in the two LEs is identical by construction). This substitution results in an enhancement of signal-to-noise of past forced temperature trends over southern Europe and a reduction in signal-to-noise over Scandinavia, Germany and Poland, with a net increase from 38% to 60% in the area with values > 2 (Fig. 8b). The impact of model biases in ensemble trend σ is much less pronounced for precipitation than temperature, with signal-to-noise values in all locations remaining below 2 (Fig. 8e).

As expected, signal-to-noise values are higher for forced trends in the future than in the past. Ninety-seven percent of the area of the continent (excluding Iceland and Greenland) shows a signal-to-noise value > 2 for forced temperature trends during 2022-2071 (Fig. 8c), compared with 38% for trends during 1972-2021. Forced precipitation trends in the future remain uncertain, with only 2% of the land area showing a signal-to-noise value > 2 (Fig. 8f).

Another way to view the relative impacts of internal variability and external forcing on trends is by computing the fraction of ensemble members at each location that show a positive trend (e.g., warming or wetting). This metric conveys the likelihood of having a positive (or negative) trend in any single ensemble member, which is analogous to the single "realization" of the real world. At nearly all locations, more than 95% of ensemble members in the CESM2 LE show warming in both the past and future periods, with slightly lower percentages (85-95%) over western Scandinavia and parts of Great Britain (and < 75% over Ireland, Scotland and Iceland in the future); (Figs. 9a and c). Similar percentages are obtained when the internal component of past temperature trends in the OBS LE is used in place of the model's internal trends, with some reduction (75-95%) over Scandinavia, northern Russia, Germany and Poland (Fig. 9b).

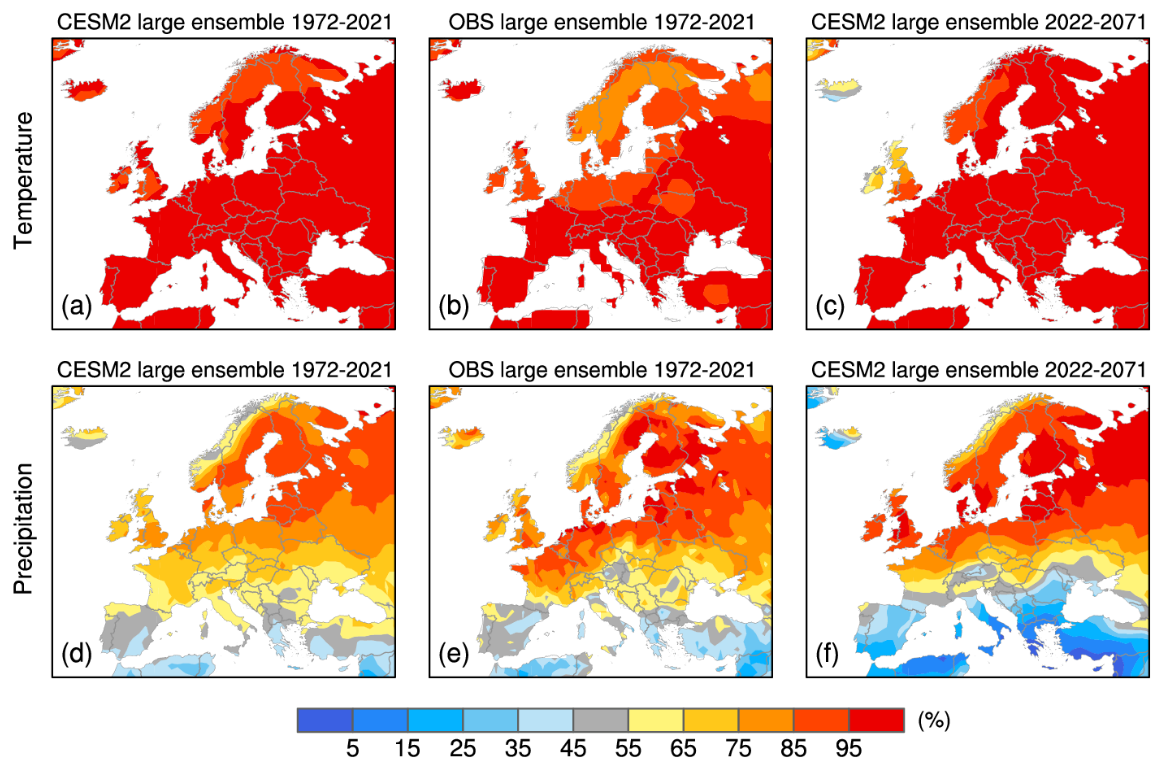

Figure 9. The percentage of ensemble members with a positive trend in winter (top) air temperature and (bottom) precipitation trends based on (a,d) the 100-member CESM2 Large Ensemble during 1972-2021, (b,e) the 100-member Observational Large Ensemble during 1972-2021, and (c,f) the 100-member CESM2 Large Ensemble during 2022-2071.

The sign of the trend in any given ensemble member is more uncertain for precipitation than for temperature. The highest chances (> 85%) of a positive precipitation trend are found over the northernmost third of the continent excluding Norway, both in the past and future (Figs. 9d and f). Similarly high chances of a negative precipitation trend (equivalent to < 15% of a positive trend) occur in areas near the Mediterranean Sea, but only in the future. The central portion of the continent shows roughly equal chances of having a positive or negative trend, both in the past and future. The area with a > 85% chance of a positive precipitation trend in the past 50 years expands southward into northern France, Germany and areas bordering the Baltic Sea when internal variability is derived from the OBS LE compared to the CESM2 LE (Fig. 9e).

Taken together, the results shown in Fig. 9 indicate that warming is virtually guaranteed at nearly all locations, both in the past 50 years and the next 50 years, according to the CESM2 LE. However, the sign of the precipitation trend (past and future) is robust only over the northern tier of the continent, and only in the future over the Mediterranean region. The model results for past trends are found to be generally credible as measured against the OBS LE, with some overestimation in north-central Europe.

e. Range of outcomes and the role of the atmospheric circulation

As the saying goes, "climate is what we expect, weather is what we get". This adage is also applicable to climate change, where "human-induced climate change is what we expect, internal variability plus human-induced climate change is what we get" (Deser 2020). Here, we illustrate "what we expect" and the range of "what we get" for past and future 50-year trends in the CESM2 LE, using the ensemble-mean for "what we expect" and two contrasting ensemble members for the range of "what we get". We select the contrasting members from the bottom and top 5th percentiles of the distribution of 100-member trends averaged over the European continent for each period separately. This selection criterion is somewhat arbitrary and does not necessarily capture the wide range of trend amplitudes that may occur at a single location or sub-region, nor does it portray the full range of spatial patterns that occur within the ensemble.

There is a large range in temperature trend outcomes ("what we get") for both the past 50 years and the next 50 years as depicted by the "warm" and "cool" end-members (Fig. 10). For past trends, the "warm" end-member shows temperature increases of 0.9-1.1 °C per decade over the eastern portion of the continent (Fig. 10b), while the "cool" end-member displays muted warming

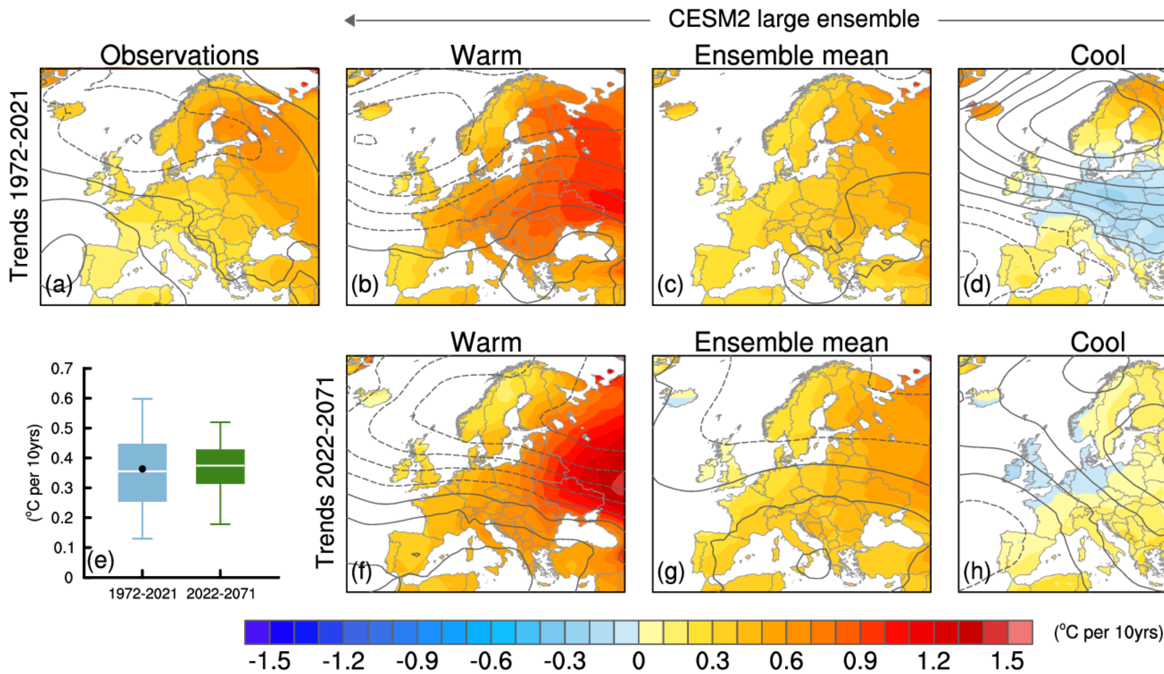

Figure 10. **A Range of Outcomes.** Trends in winter air temperature (color shading; °C per decade) and sea-level pressure (SLP) (contours; contour interval of 0.25 hPa per decade, negative values dashed) for the period (top) 1972-2021 and (bottom) 2022-2071. Panel (a) shows observed trends (1972-2021) and remaining panels show simulated trends from the 100-member CESM2 Large Ensemble: (c,g) ensemble-mean; (b,f) "warm" end-member; (d,h): "cool" end-member. See text for details. Note that panels (a) and (c) are identical to the "OBS" and "EM" panels in Fig. 1, respectively. Panel (e): Distribution of European-average trends for 1972-2021 (blue) and 2022-2071 (green) from the CESM2 Large Ensemble (box outlines 25th-to-75th percentile range, whiskers mark the 5th-to-95th percentile range, the horizontal white line denotes the median value, and the black circle marks the observed value).

(< 0.3 °C per decade) and even slight cooling through the midsection of the continent (Fig. 10d). Clearly, the forced trend ("what we expect"), which depicts moderate warming (0.2-0.6°C per decade) across the continent does not tell the whole story (Fig. 10c). Analogous results are found for trends projected over the next 50 years: the "warm" member shows temperature increases of 1.0-1.5 °C per decade over west-central Russia (Fig. 10f) while the "cool" member depicts < 0.2°C per decade warming over most of the continent (Fig. 10h), in marked contrast to the forced trend which ranges from 0.3-0.6°C per decade (Fig. 10g). As discussed previously, the observed temperature trend map resembles the model's ensemble-mean, but this could be by chance (Fig.

10a). In terms of European averages, the observed trend (0.36 °C per decade) is nearly coincident with the median value of the model's trend distribution, which has a 5th-to-95th percentile range of 0.13-0.60 °C per decade for past 50-year trends (Fig. 10e). Curiously, the model's median trend value for Europe as a whole increases only slightly in the future compared to the past, while the 5th-to-95th and 25th-to-75th percentile ranges narrow (Fig. 10e). Further work is needed to understand why this is the case.

As mentioned in Section 1d, previous work has shown that internal variability of the large-scale atmospheric circulation causes much of the member-to-member differences in temperature trends in model LEs. Here, we provide a qualitative indication of the circulation influence by superimposing SLP trends upon the maps in Fig. 10. In the case of past trends, the "warm" member shows a positive North Atlantic Oscillation (NAO)-like pattern (Hurrell et al. 2003), with negative SLP trends centered near Iceland and positive SLP trends centered over the Mediterranean (Fig. 10b). This SLP pattern is indicative of stronger westerly/southwesterly flow, which brings relatively warm maritime air over the continent. The "cool" member shows a largely opposite flow configuration (albeit with longitudinal shifts in the SLP centers-of-action), which advects relatively cold air from the east over the continent (Fig. 10d). In comparison, the forced response shows negligible atmospheric circulation change (Fig. 10c). Striking contrasts in circulation are also found for the future period, with a large positive NAO-like trend pattern in the "warm" member and a blocking continental "High" in the "cool" member (Figs. 10f and h). Future trends in SLP also contain a modest forced component indicative of enhanced westerlies over the continent (Fig. 10g).

The "wet" and "dry" end-members also show striking regional contrasts in both precipitation and circulation (Fig. 11). For example, for past trends, the "wet" member shows precipitation increases of 0.2-0.3 mm d^{-1} per decade over France, southern Germany, Portugal and the UK, and precipitation declines over northern Norway and along the Mediterranean Sea (Fig. 11b). A nearly opposite pattern is found for the "dry" member (Fig. 11d). These contrasting precipitation trends can be understood in the context of the overlying atmospheric circulation changes, with wetter areas coinciding with anomalous westerly/southwesterly flow and drier areas located under blocking anticyclones. Analogous patterns are found for future trends, with pronounced increases in precipitation over western Europe associated with the low pressure trend centered over the British Isles in the "wet" member (Fig. 11f), and generally reduced precipitation in the "dry" member associated with the blocking High centered over southern Europe (Fig. 11h).

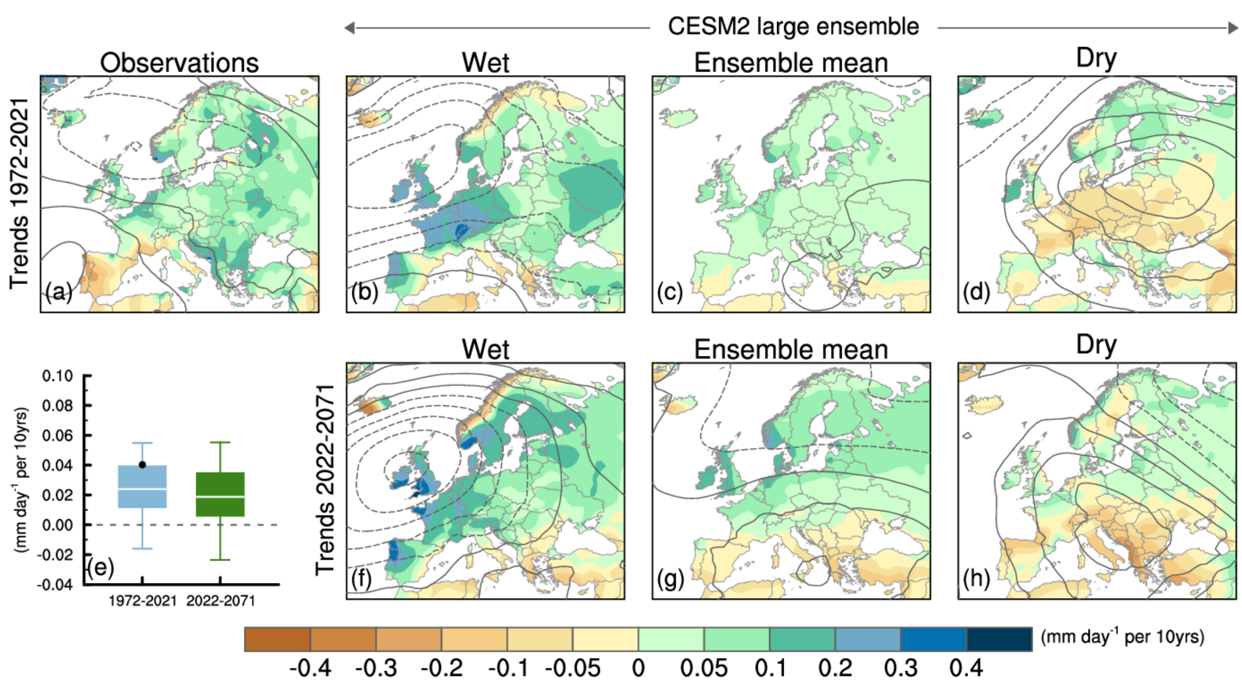

Figure 11. As in Fig. 10 but for precipitation (mm d^{-1} per decade). Note that panels (a) and (c) are identical to the "OBS" and "EM" panels in Fig. 2, respectively.

f. Unmasking forced climate change in observations via "Dynamical Adjustment"

The empirical method of "dynamical adjustment" introduced in Section 1d can be used to estimate the circulation-induced component of observed temperature anomalies; this dynamically-induced contribution can then be subtracted from the original anomaly to obtain the thermodynamically-induced component as a residual. Since this method uses no information from climate models, it provides an independent estimate of the thermodynamic component of observed temperature trends, which can be compared with the forced response simulated by climate model LEs.

Figure 12 shows the decomposition of observed DJF temperature trends into their dynamical and residual thermodynamic contributions. For this example, we have used the 60-year period 1962-2021 when observed SLP trends are more than twice as large as those during 1972-2021 on a per decade basis (compare SLP contours in Figs. 10a and 12a). Observed SLP trends during the past 60 years show a pronounced positive NAO-like pattern, with maximum negative values of -1.25 hPa per decade near Iceland and maximum positive values of +0.75 hPa per decade west of Spain (Fig. 12a). Enhanced westerly/southwesterly flow associated with this pattern advects warm air, raising surface temperatures by 0.1- 0.3°C per decade (with maximum warming over northern Europe) according to the dynamical adjustment algorithm (Fig. 12b). Removing this dynamically-induced component from the total trend reveals the residual thermodynamic contribution to the observed warming trend (Fig. 12c). This observed thermodynamic trend is much closer in amplitude (and arguably pattern) to the model's forced response, given by the CESM2 LE ensemble-mean trend (Fig. 12d), than is the total observed trend. Further, the lack of an appreciable forced SLP trend in CESM2 indicates that the model's forced temperature trend is nearly all thermodynamically-driven. The level of agreement between the observed thermodynamic temperature trend and the model's forced thermodynamic trend leads to two powerful conclusions: 1) the model's forced temperature trend is realistic; and 2) removing the circulation-induced component from the observed trends can effectively reveal the influence of anthropogenic forcing. Analogous results have been found for North America (Deser et al. 2016). It may seem surprising that the model's forced temperature trend agrees so well in amplitude with the observed thermodynamic-residual trend, given that CESM2 has been characterized as a "high climate sensitivity" model (Gettelman et al., 2019). However, this characterization refers specifically to the model's equilibrium climate sensitivity (diagnosed as the model's response to an instantaneous doubling of CO_2 based on a slab-ocean configuration), and does not translate to a high transient climate sensitivity over the 1962-2021 period of record analyzed here, as evidenced by the fact that the observed global-mean temperature increase lies within the ensemble-spread of global-mean temperature trends simulated by the CESM2-LE for this time period (not shown).

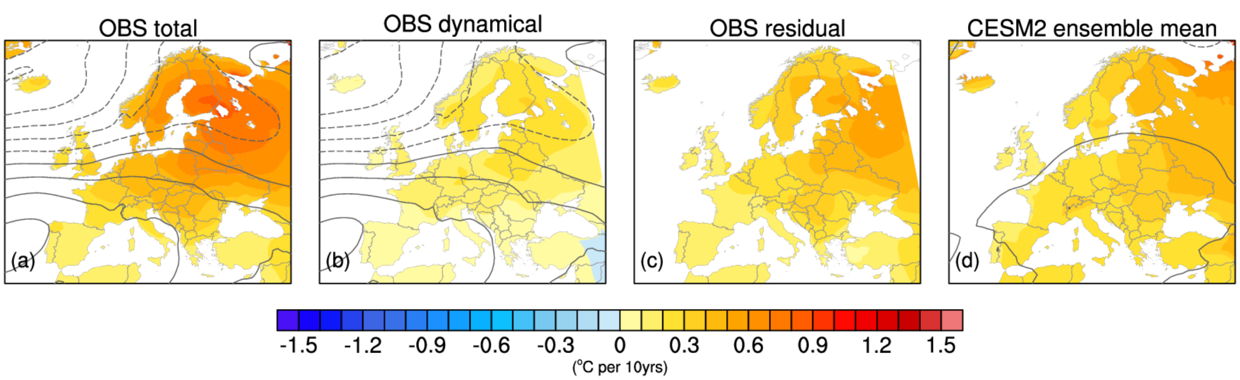

Figure 12. Decomposition of (a) observed winter air temperature trends (1962-2021; °C per decade) into (b) dynamical and (c) residual thermodynamic contributions using the "dynamical adjustment" procedure of Deser et al. (2018) based on constructed circulation analogues (see text for details). Contours in (a) show observed sea-level pressure (SLP) trends (contour interval of 0.25 hPa per decade, negative values dashed); contours in (b) show the observed SLP trends

estimated from the constructed circulation analogues; contours in (c) based on the difference between (a) and (b) are near-zero and not shown. Panel (d) shows the ensemble-mean temperature and SLP trends from the 100-member CESM2 Large Ensemble (note that only the zero contour shows up in panel d).

Precipitation is an inherently noisier field than temperature in both time and space, making it challenging to extract the forced signal via "dynamical adjustment"; indeed, only one previous study has attempted dynamical adjustment of observed precipitation trends (Guo et al. 2019). Keeping in mind that the estimate of the circulation-induced component of precipitation trends may be less robust than for temperature, we present the results as a proof-of-concept. Observed precipitation trends during 1962-2021 are mainly driven by changes in atmospheric circulation, with a small thermodynamic residual component (Fig. 13). This residual component bears some resemblance to the forced response in CESM2, particularly in terms of amplitude (~ 0.05 mm d-1 per decade; Fig. 13d). Notable areas of agreement in the sign of trends include drying over most of Spain, Portugal, Algeria, Turkey and Syria, and wetting over parts of northern and north-central Europe; disagreement in sign is found over many central European countries (France, Germany, Switzerland, Austria, Ukraine, Romania and southern Russia) where the signal-to-noise is low (Figs. 8d,e) due to a combination of low signal in the transition region between southern drying and northern wetting (Fig. 11c) and high noise (Figs. 7d,e). The low signal and high noise in these areas limits the accuracy of the dynamical adjustment results, where the error of the method is of the same amplitude as the thermodynamic-residual trend (see Guo et al., 2019 for details).

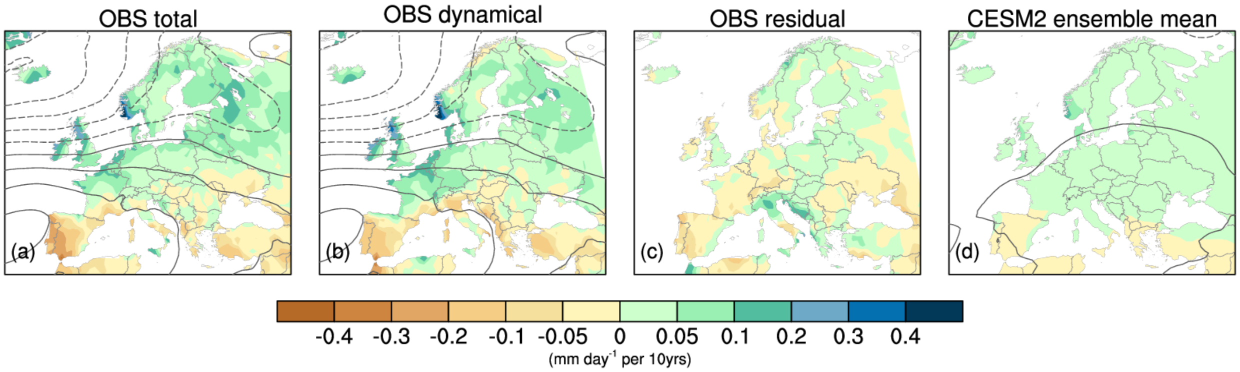

Figure 13. As in Fig. 12 but for precipitation (mm d^{-1} per decade).

g. Toward an observationally-based "range of outcomes"

We conclude by bringing together the results of the Observational LE and "dynamical adjustment" to produce a fully observationally-based estimate of the range of past 60-year trends in temperature and precipitation. To the best of our knowledge, this is first time that these two approaches have been combined. Specifically, we add the internal component of trends from each member of the OBS LE to the thermodynamic-residual trend (the estimated observed forced response) obtained from dynamical adjustment. As before, we select two contrasting ensemble members from the tails of the distribution based on European-wide averages to illustrate the range of trend outcomes. The "warm" end-member shows pronounced temperature increases over the northern two-thirds of the continent, with maximum values in excess of 0.9 °C per decade, while the "cool" end-member warms less than 0.2 °C per decade in most areas and even cools slightly over Ukraine and neighboring countries (Figs. 14 b and d, respectively). These divergent temperature trends are associated with contrasting SLP trends, with a positive NAO-like pattern in the "warm" member a negative (and eastward-shifted) NAO pattern in the "cool" member (Figs. 14 b and d). Qualitatively, this range of trend outcomes for both temperature and SLP is remarkably similar to that obtained directly from the CESM2 LE, with some regional differences in the location of cooling in the "cool" end-member (Figs. 14 e and g). There is no guarantee that the patterns and amplitudes of trends sampled in our selected end-members will agree between the model and observationally-based results, since there are many configurations that produce extremes in European-wide averages (not shown). That there is a strong qualitative resemblance between them is a testament to both the realism of the model's forced response and internal variability, and the efficacy of the OBS LE and dynamical adjustment approaches.

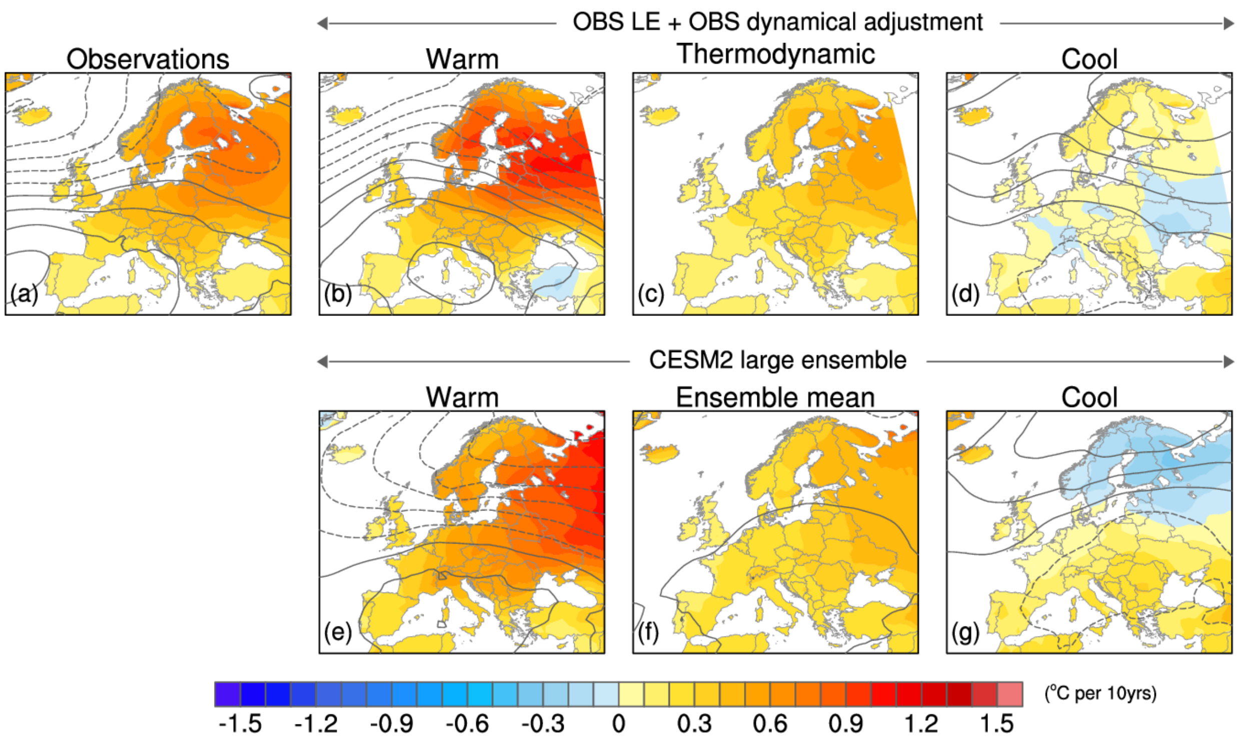

Figure 14. As in Fig. 10 but for the period 1962-2021. The top row is based on the Observational Large Ensemble combined with the residual thermodynamic component of observed trends. The bottom row is based on the 100-member CESM2 Large Ensemble. See text for details.

Precipitation trends in the "wet" and "dry" end-members are also similar between the model and observationally-based results (Fig. 15). The "wet" members show widespread increases in precipitation over southern and central Europe (maximum values of 0.2-0.4 mm d^{-1} per decade) and drying over the northern UK and parts of Scandinavia (Figs. 15 b and e). Largely opposite patterns prevail in the "dry" members (Figs. 15 d and g). The contrasting precipitation trends in the “wet” and “dry” end-members are associated with opposite flow configurations, with regions of drying corresponding to high pressure and vice versa.

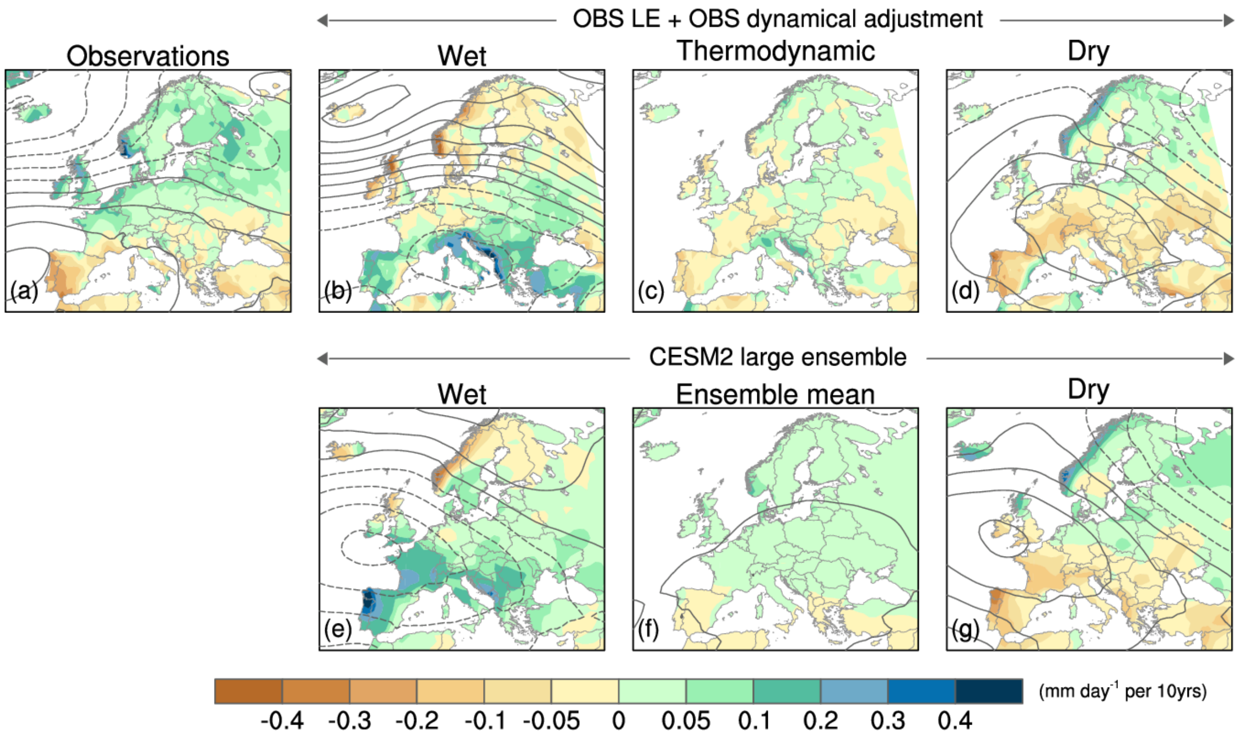

Figure 15. As in Fig. 14 but for precipitation (mm d^{-1} per decade).

4. Summary and open questions

Disentangling the effects of internal variability and anthropogenic forcing on regional climate trends remains a long-standing issue in climate sciences. Recent advances in climate modeling and physical understanding have led to new insights on this topic, and provided an improved source of information on the future risks of weather extremes associated with human-induced climate change. Here, we have highlighted new findings for European winter climate based on the following complementary tools: Earth System Model Large Ensemble simulations; an observationally-based Large Ensemble; and an empirical approach for removing the influence of atmospheric circulation variability from observed temperature and precipitation data, termed "dynamical adjustment".

The new 100-member CESM2 Large Ensemble shows that internal climate variability imparts considerable uncertainty to past and future 50-year trends in winter temperature and precipitation over Europe. Such uncertainty is irreducible due to the lack of predictability of the simulated internal variability on decadal time scales. A novel synthetic Large Ensemble constructed from the statistical characteristics of internal variability in the observational record exhibits quantitatively similar levels of uncertainty in past 50-year trends as the CESM2 LE, reinforcing the credibility of the model's internally-generated trends. Additionally, the results of our "dynamical adjustment" procedure applied to observations shows good agreement between the observed thermodynamic-residual trend component and the model's forced thermodynamic trend, further underscoring the realism of CESM2. Finally, we have combined internal variability of trends from an Observational Large Ensemble with an observational estimate of the forced trend (the thermodynamic-residual component obtained from "dynamical adjustment") to show what the observed range of past trends in European temperature and precipitation could have been. Because it does not rely on climate model information, this observationally-based range of trend outcomes provides a powerful test of the range of simulated trends in a model Large Ensemble. To the best of our knowledge, this is the first time that such a synthesis of the two purely observational methods has been undertaken.

Many outstanding questions remain regarding the relative influences of internal climate variability and anthropogenic forcing on regional climate change in models and the real world. Fortunately, promising new tools are being developed to help address these challenges. For example, innovative machine learning methods may be able to improve upon existing techniques for constructing Observational Large Ensembles. Such methods have shown good results as statistical emulators of model-based LEs, but their application to the observational record remains to be pursued (Beusch et al. 2019). Similarly, neural network approaches to dynamical adjustment may offer increased skill compared to conventional methods (Davenport and Diffenbaugh, 2021), but have yet to be applied with the aim of separating forced and internal components of observed trends. Complementary physically-based approaches such as Linear Inverse Modeling and Low-Frequency Pattern Analysis mentioned in Section 1d also offer promise for estimating the forced response in observations without reliance on climate models and should be pursued more widely.

We have relied on the fact that the CESM2 LE (like other models of its class; see Deser et al. 2020a and references therein) simulates a negligible forced atmospheric circulation trend over the past 50-60 years to interpret our observed dynamical adjustment results (i.e., we have equated the observed dynamically-induced trend with the internal component, and the observed thermodynamic-residual trend with the forced component). If the model is erroneous in this regard, then our interpretation of our decomposition of observed trends into "internal dynamical" and "forced thermodynamic" components is flawed. Recent work suggests that large-scale extra-tropical atmospheric circulation variability simulated by climate models may be less predictable on seasonal-to-decadal timescales than that in the real world, implying that models underestimate the signal-to-noise ratio of predictable components (Scaife et al. 2014; Eade et al. 2014; Scaife and Smith, 2018). But the underlying mechanisms for this underestimation, and whether this so-called "signal-to-noise paradox" found in initial-value predictability studies applies to models' atmospheric circulation response to anthropogenic forcing, remain unknown at this time.

Emerging efforts to develop higher-resolution (km scale) global coupled climate models may provide the key to addressing this elusive challenge (Slingo et al. 2022).

Data and code availability statement

All data used in this study are publicly available as follows:

CESM2 Large Ensemble: https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm2le.output.html

GPCC precipitation: https://www.dwd.de/EN/ourservices/gpcc/gpcc.html

BEST temperature: http://berkeleyearth.org/data/

and ERA5 SLP: https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5

Code used to create the Observational Large Ensemble and Dynamical Adjustment results are publicly available at: https://github.com/karenamckinnon/observational_large_ensemble/ and https://github.com/terrayl/Dynamico, respectively.

Author contributions

CD led the overall effort and wrote the manuscript. ASP performed some of the calculations and prepared the figures.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

We acknowledge the efforts of all those who contributed to producing the model simulations and observational data sets used in this study. We thank L. Terray for providing the dynamical adjustment results and K. McKinnon for providing the observational large ensemble results. We sincerely appreciate the comments and suggestions from the Reviewers, which led to improved clarity of presentation. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

References

Andrews, T., Bodas-Salcedo, A., Gregory, J. M., Dong, Y., Armour, K. C., Paynter, D., Lin, P., Modak, A., Mauritsen, T., Cole, J. N. S., Medeiros, B., Benedict, J. J., Douville, H., Roehrig, R., Koshiro, T., Kawai, H., Ogura, T., Dufresne, J. -L., Allan, R. P., and Liu, C.: On the effect of historical SST patterns on radiative feedback, J. Geophys. Res.-Atmos., 127, e2022JD036675, https://doi.org/10.1029/2022JD036675, 2022.

Barnes, E. A., Hurrell, J. W., and Uphoff, I. E.: Viewing forced climate patterns through an AI lens, Geophys. Res. Lett., 46, 13389–13398, https://doi.org/10.1029/2019GL084944, 2019.

Beusch, L., Gudmundsson, L., and Seneviratne, S. I.: Emulating Earth System Model temperatures: from global mean temperature trajectories to grid-point level realizations on land, Earth Syst. Dyn. Discuss., https://doi.org/10.5194/esd-2019-34, 2019.

Bódai, T., G. Drótos, M. Herein, F. Lunkeit, and V. Lucarini: The Forced Response of the El Niño–Southern Oscillation–Indian Monsoon Teleconnection in Ensembles of Earth System Models. *J. Climate*, 33, 2163–2182, https://doi.org/10.1175/JCLI-D-19-0341.1, 2020.

Bódai, T., J.-Y. Lee, and A. Sundaresan: Sources of Nonergodicity for Teleconnections as Cross-Correlations, *Geophys. Res. Lett.*, 49, 8, e2021GL096587, doi: 10.1029/2021GL096587, 2022.

Bonfils, C. J. W., Santer, B. D., Fyfe, J. C., Marvel, K., Phillips, T. J., and Zimmerman, S. R. H.: Human influence on joint changes in temperature, rainfall and continental aridity. Nat. Clim. Change, 10, 726-731, https://doi.org/10.1038/s41558-020-0821-1, 2020.

Branstator, G. and Teng, H.: Two limits of initial-value decadal predictability in a CGCM, J. Climate, 23, 6292-6311, https://doi.org/10.1175/2010JCLI3678.1, 2010.

Capotondi, A., Deser, C., Phillips, A., Okumura, Y. and Larson, S.: ENSO and Pacific DecadalVvariability in the Community Earth System Model Version 2, J. Adv. Model. Earth Sy., 12, e2019MS002022, https://doi.org/10.1029/2019MS002022, 2020.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø., Ross, T.F., Trigo, R.M., Wang, X.L., Woodruff, S.D. and Worley, S.J.: The twentieth century reanalysis project, Q. J. Roy. Meteor. Soc., 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Davenport, F. V. and Diffenbaugh, N. S.: Using machine learning to analyze physical causes of climate change: A case study of U.S. Midwest extreme precipitation, Geophys. Res. Lett., 48, e2021GL093787. https://doi.org/10.1029/2021GL093787, 2021.

Deser, C., Phillips, A., Bourdette, V., and Teng, H. Y.: Uncertainty in climate change projections: The role of internal variability, Clim. Dynam., 38, 527–546. https://doi.org/10.1007/s00382-010-0977-x, 2012.

Deser, C., Phillips, A., Alexander, M. A., and Smoliak, B. V.: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. J. Climate, 27, 2271–2296, https://doi.org/10.1175/JCLI-D-13-00451.1, 2014.

Deser, C., Terray, L., and Phillips, A. S.: Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications, J. Climate, 29, 2237–2258. https://doi.org/10.1175/JCLI-D-15-0304.1, 2016.

Deser, C., Simpson, I. R., McKinnon K. A., and Phillips, A. S.: The Northern Hemisphere extra-tropical atmospheric circulation response to ENSO: How well do we know it and how do we evaluate models accordingly? J. Climate, 30, 5059-5082, https://doi.org/10.1175/JCLI-D-16-0844.1, 2017.

Deser, C., Hurrell, J. W., and Phillips, A.S.: The role of the North Atlantic Oscillation in European Climate Projections, Clim. Dynam., 49, 3141–3157, https://doi.org/10.1007/s00382-016-3502-z, 2017.

Deser, C., Simpson, I. R., Phillips, A. S., and McKinnon, K.A.: How well do we know ENSO's climate impacts over North America, and how do we evaluate models accordingly? J. Climate, 30, 4991-5014, https://doi.org/10.1175/JCLI-D-17-0783.1, 2018.

Deser, C.: Certain uncertainty: The role of internal climate variability in projections of regional climate change and risk management. Earths Future, 8, e2020EF001854, https://doi.org/10.1029/2020EF001854, 2020.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A, Simpson, I. R., and Ting, M.: Insights from earth system model initial-condition large ensembles and future prospects. Nat. Clim. Change, 10, 277-286, https://doi.org/10.1038/s41558-020-0731-2, 2020a.

Deser, C., Phillips, A. S., Simpson, I. R., Rosenbloom, N., Coleman, D., Lehner, F., Pendergrass, A., DiNezio, P., and Stevenson, S.: Isolating the Evolving Contributions of Anthropogenic Aerosols and Greenhouse Gases: A New CESM1 Large Ensemble Community Resource. J. Climate, 33, 7835-7858, https://doi.org/10.1175/JCLI-D-20-0123.1, 2020b.

Deser, C. and Phillips, A.S.: Defining the internal component of Atlantic Multidecadal Variability in a changing climate. Geophys. Res. Lett., 48, e2021GL095023, https://doi.org/10.1029/2021GL095023, 2021.

DiNezio, P. N., Deser, C., Okumura, Y., and Karspeck, A.: Predictability of 2-year La Niña events in a coupled general circulation model, Clim. Dyn. 49, 4237–4261, 2017.

Dong, Y, Armour, K.C., Zelinka, M., Proistosescu, C., Battisti, D., Zhou, C., and Andrews, T.: Inter-model spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models, J. Climate, https://doi.org/10.1175/JCLI-D-19-1011.1, 2020.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N.: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?, Geophys. Res. Lett., 41, 5620– 5628, https://doi.org/10.1002/2014GL061146, 2014.

Fasullo, J. T. and R.S. Nerem: "Altimeter-era emergence of the patterns of forced sea-level rise in climate models and implications for the future. *Proc. Natl Acad. Sci.* https://doi.org/10.1073/pnas.1813233115, 2018.

Fasullo, J., Phillips, A. S., and Deser, C.: Evaluation of leading modes of climate variability in the CMIP Archives, J. Climate, 33, 5527–5545, https://doi.org/10.1175/JCLI-D-19-1024.1, 2020.

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al.: High climate sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophys. Res. Lett.*, 46, 8329– 8337. https://doi.org/10.1029/2019GL083978, 2019.

Gordon, E. M. and Barnes, E.A.: Incorporating uncertainty into a regression neural network enables identification of decadal state-dependent predictability, Geophys. Res. Lett., e2022GL098635, https://doi.org/10.1029/2022GL098635, 2022.

Gould, S. J.: Wonderful Life: The burgess shale and the nature of history, W. W. Norton & Co., 978-0-393-30700-9, 1989.

Griffies, S. M. and Bryan, K.: Predictability of North Atlantic multidecadal climate variability, Science, 275, 181–184, https://doi.org/10.1126/science.275.5297.181, 1997.

Guo, R. X., Deser, C., Terray, L., and Lehner, F.: Human influence on terrestrial precipitation trends revealed by dynamical adjustment, Geophys. Res. Lett., 46, 3426-3434, https://doi.org/10.1029/2018GL081316, 2019.

Hegerl, G.C., Zwiers, F. W. , Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J. A., Nicholls, N., Penner, J. E., and Stott, P. A.: Understanding and attributing climate change. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Hurrell J. W., Kushnir, Y., Ottersen G., and Visbeck M. (eds): The North Atlantic Oscillation: climate significance and environmental impact, Geophys. Monogr. Ser, 134, AGU, Washington, D. C, 2003.

James, I. N. and James, P. M.: Spatial structure of ultra-low-frequency variability of the flow in a simple atmospheric circulation model. Quart. J. Roy. Meteor. Soc., 118, 1211-1233, https://doi.org/10.1002/qj.49711850810, 1992.

Jin, E.K., Kinter, J.L., and Wang, B: Current status of ENSO prediction skill in coupled ocean–atmosphere models, Clim. Dynam. 31, 647–664, https://doi.org/10.1007/s00382-008-0397-3, 2008.

Kay, J., 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., Polvani, L. and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability, B. Am. Meteorol. Soc., 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

Klavans, J.M., Cane, M.A., Clement, A.C., and Murphy, L. N.: NAO predictability from external forcing in the late 20th century, Npj Clim. Atmos. Sci., 4, 22 (2021), https://doi.org/10.1038/s41612-021-00177-8, 2021.

Lehner, F., Schurer, A. P., Hegerl, G. C., Deser, C., and Frölicher, T. L.: The importance of ENSO phase during volcanic eruptions for detection and attribution, Geophys. Res. Lett. 43, 2851–2858, https://doi.org/10.1002/2016GL067935, 2016.

Lehner, F., Deser, C., and Terray, L.: Towards a new estimate of "time of emergence" of anthropogenic warming: insights from dynamical adjustment and a large initial-condition model ensemble. J. Climate, 30, 7739-7756, http://doi.org/10.1175/JCLI-D-16-0792.1, 2017.

Lehner, F., Deser, C., Simpson, I. R., and Terray, L.: Attributing the US Southwest's recent shift into drier conditions, Geophys. Res. Lett., 45, 6251–6261, https://doi.org/10.1029/2018GL078312, 2018.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E., Brunner, L., Knutti, R. and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. Earth Syst. Dynam. Discuss., Special Issue on Large Ensembles, 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Leith, C. E.: The standard error of time-average estimates of climatic means, J. Appl. Meteorol. Clim., 12(6), 1066-1069, https://doi.org/10.1175/1520-0450(1973)012%3C1066:TSEOTA%3E2.0.CO;2, 1973.

Lorenz, E. N.: Deterministic nonperiodic flow, J. Atmos. Sci., 20, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963.

Madden, R. A.: Estimates of the natural variability of time-averaged sea-level pressure, Mon. Weather. Rev., 104, 942-952, https://doi.org/10.1175/1520-0493(1976)104%3C0942:EOTNVO%3E2.0.CO;2, 1975.

Maher, N., Matei, D., Milinski, S., and Marotzke, J.: ENSO change in climate projections: Forced response or internal variability?, Geophys. Res. Lett., 45, 11390–11398, https://doi.org/10.1029/2018GL079764, 2018.

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., Raddatz, T., Stevens, B., and Marotzke, J.: The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability, J. Adv. Model. Earth Sy., 11, 2050–2069, https://doi.org/10.1029/2019MS001639, 2019.

McGraw, M. C., Barnes, E. A., and Deser, C.: Reconciling the observed and modeled southern hemisphere circulation response to volcanic eruptions, Geophys. Res. Lett., 43, 7259–7266, https://doi.org/10.1002/2016GL069835, 2016.

McKenna, C. M., and Maycock, A. C.: Sources of uncertainty in multimodel large ensemble projections of the winter North Atlantic Oscillation, Geophys. Res. Lett., 48, e2021GL093258, https://doi.org/10.1029/2021GL093258, 2021.

McKinnon, K. A., A. Poppick, Dunn-Sigouin, E. and C. Deser: An 'Observational Large Ensemble' to compare observed and modeled temperature trend uncertainty due to internal variability. *J. Climate*, **90**, 7585-7598, doi:10.1175/JCLI-D-16-0905.1, 2017.

McKinnon, K. A and Deser, C.: Internal variability and regional climate trends in an Observational Large Ensemble, J. Climate, 31, 6783–6802, https://doi.org/10.1175/JCLI-D-17-0901.1, 2018.

McKinnon, K. A. and Deser, C.: The inherent uncertainty of precipitation variability, trends, and extremes due to internal variability, with implications for Western US water resources, J. Climate, 34, 9605-9622, https://doi.org/10.1175/JCLI-D-21-0251.1, 2021.

Meehl, G., Hu, A. and Teng, H: Initialized decadal prediction for transition to positive phase of the Interdecadal Pacific Oscillation, Nat. Commun., 7, 11718 (2016), https://doi.org/10.1038/ncomms11718, 2016.

Meehl, G.A., J.H. Richter, Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M. G., England, M. H., Fyfe, J. C., Han, W., Kim, H., Kirtman, B. P., Kushnir, Y., Lovenduski, N. S., Mann, M. E., Merryfield, W. J., Nieves, V., Pegion, K., Rosenbloom, N., Sanchez, S. C., Scaife, A. A., Smith, D., Subramanian, A. C., Sun, L., Thompson, D., Ummenhofer, C. C., and Xie, S. -P.: Initialized Earth system prediction from subseasonal to decadal timescales, Nat. Rev. Earth Environ., 2, 340–357 (2021), https://doi.org/10.1038/s43017-021-00155-x, 2021.

Merrifield, A., Lehner, F., Xie, S. -P., and Deser, C.: Removing circulation effects to assess Central US land-atmosphere interactions in the CESM Large Ensemble, Geophys. Res. Lett., 44, 9938-9946, https://doi.org/10.1002/2017GL074831, 2017.

Milinski, S., Maher, N., and Olonscheck, D.: How large does a large ensemble need to be?, Earth Syst. Dyn. Discuss., 11, 885-901, https://doi.org/10.5194/esd-11-885-2020, 2019.

Newman, M.: Interannual to decadal predictability of tropical and North Pacific sea surface temperatures, J. Climate, 20, 2333–2356, https://doi.org/10.1175/JCLI4165.1, 2007.

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., Mantua, N. J., Miller, A. J., Minobe, S., Nakamura, H., Schneider, N., Vimont, D. J., Phillips, A. S., Scott, J. D., and Smith, C. A.: The Pacific decadal oscillation, revisited, J. Climate, 29, 4399–4427, https://doi.org/10.1175/ JCLI-D-15-0508.1, 2016.

O'Brien, J. P. and Deser, C.: Quantifying and understanding forced changes to unforced modes of atmospheric circulation variability over the North Pacific in a coupled model large ensemble, *J. Climate*, 36, 17-35, doi: https://doi.org/10.1175/JCLI-D-22-0101.1, 2023.

Olivarez, H. C., Lovenduski, N. S., Brady, R. X., Fay, A. R., Gehlen, M., Gregor, L., Landschützer, P., McKinley, G. A., McKinnon, K. A., and Munro, D. R.: Alternate histories: Synthetic large ensembles of sea-air CO_2 flux, Global Biogeochem. Cy., 36, e2021GB007174, https://doi.org/10.1029/2021GB007174, 2022.

Persad, G. G., and Caldeira, K.: Divergent global-scale temperature effects from identical aerosols emitted in different regions. Nat. Commun., 9, 3289, https://doi.org/10.1038/s41467- 018-05838-6, 2018.

Rodgers, K. B., Lee, S. -S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., Edwards, J., Kim, J. -E., Simpson, I., Stein, K., Stuecker, M. F., Yamaguchi, R., Bodai, T., Chung, E. -S., Huang, L., Kim, W., Lamarque, J. -F., Lombardozzi, D., Wieder, W. R., and Yeager, S. G.: Ubiquity of human-induced changes in climate variability. Earth Sys. Dyn., 12, 1393–1411, https://doi.org/10.5194/esd-12-1393-2021, 2021.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., and Mosher, S.: Berkeley Earth temperature averaging process, Geoinf. Geostat. Overview, 1:2, https://doi.org/10.4172/2327-4581.1000103, 2013.

Santer, B., Fyfe, J. C., Solomon, S., Painter, J. F., Bonfils, C., Pallotta, G., and Zelinka, M. D.: Quantifying stochastic uncertainty in detection time of human-caused climate signals, Proc. Natl. Acad. Sci., 116, 19821–19827, https://doi.org/10.1073/pnas.1904586116, 2019.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J. and Williams, A: Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41, 2514–2519, https://doi.org/10.1002/2014GL059637, 2014.

Scaife, A. A. and Smith, D.: A signal-to-noise paradox in climate science. Npj Clim. Atmos. Sci., 1, 28 (2018), https://doi.org/10.1038/s41612-018-0038-4, 2018.

Schneider, D. P., Deser, C., and Fan, T.: Comparing the impacts of tropical SST variability and polar stratospheric ozone loss on the Southern Ocean westerly winds, J. Climate, 28, 9350-9372, https://doi.org/10.1175/JCLI-D-15-0090.1, 2015.

Schneider, U., Fuchs, T., Meyer-Christoffer, A., and Rudolf, B.: Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation, 1-12.

Shepherd, T.: Atmospheric circulation as a source of uncertainty in climate change projections, Nature Geosci., 7, 703–708 (2014), https://doi.org/10.1038/ngeo2253, 2014.

Sippel, S. Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer, E., and Knutti, R.: Uncovering the forced climate response from a single ensemble member using statistical learning, J. Climate, 32, 5677-5699, https://doi.org/10.1175/JCLI-D-18-0882.1, 2019.

Slingo, J., P. Bates, P. Bauer, S. Belcher, T. Palmer, G. Stephens, B. Stevens, T. Stocker, G. Teutsch: Ambitious partnership needed for reliable climate prediction. *Nature Climate Change*, doi:10.1038/s41558-022-01384-8, 2022.

Smith, D., Eade, R., Scaife, A. A., Caron, L. -P., Danabasoglu, G., DelSole, T. M., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Hermanson, L., Kharin, V., Kimoto, M., Merryfield, W. J., Mochizuki, T., Müller, W. A., Pohlmann, H., Yeager, S., and Yang, X: Robust skill of decadal climate predictions, Npj Clim. Atmos. Sci., 2, 13 (2019), https://doi.org/10.1038/s41612-019-0071-y, 2019.

Smith, D. M. et al. North Atlantic Climate far more predictable than models imply, Nature, 583, 796–800, 2020.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L. F., Caron, L.P., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F. J., Dunstone, N.J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V., Kimoto, M., Merryfield, W. J., Mignot, J., Mochizuki, T., Modali, K., Monerie, P. A., Müller, W. A., Nicolí, D., Ortega, P., Pankatz, K., Pohlmann, H., Robson, J., Ruggieri, P., Sospedra-Alfonso, R., Swingedouw, D., Wang, Y., Wild, S., Yeager, S., Yang, X., and Zhang, L.: North Atlantic climate far more predictable than models imply, Nature, 583, 796-800, https://doi.org/10.1038/s41586-020-2525-0, 2020.

Smoliak, B. V., Wallace, J. M., Lin, P., and Fu, Q.: Dynamical adjustment of the Northern Hemisphere surface air temperature field: Methodology and application to observations, J. Climate, 28, 1613–1629. https://doi.org/10.1175/JCLI-D-14-00111.1, 2015.

Stevenson, S., Fox-Kemper, B., Jochum, M., Neale, R., Deser, C., and Meehl, G.: Will there be a significant change to El Nino in the 21st Century?, J. Climate, 25, 2129-2145, https://doi.org/10.1175/JCLI-D-11-00252.1, 2012.

Suarez-Gutierrez, L., Milinski, S., and Maher, N.: Exploiting large ensembles for a better yet simpler climate model evaluation, Clim. Dynam., 57, 2557–2580 (2021), https://doi.org/10.1007/s00382-021-05821-w, 2021.

Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E., and Jahn A.: Influence of internal variability on Arctic sea-ice trends. Nat. Clim. Change, 5, 86–89, https://doi.org/10.1038/nclimate2483, 2015.

Tebaldi, C., Dorheim, K., Wehner, M., Leung, R.: Extreme metrics from large ensembles: investigating the effects of ensemble size on their estimates, Earth Syst. Dynam., 12 (4), 1427-1501, https://doi.org/10.5194/esd-12-1427-2021, 2021.

Tél, T., Bódai, T., Drótos, G., Haszpra, T., Herein, M., Kaszás, B., and Vincze, M.: The theory of parallel climate realizations, J. Stat, Phys., 179, 1496–1530, https://doi.org/10.1007/s10955-019-02445-7, 2020.

Terray, L.: A dynamical adjustment perspective on extreme event attribution, Weather Clim. Dynam., 2, 971–989, https://doi.org/10.5194/wcd-2-971-2021, 2021.

Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E., and Phillips, A. S.: Quantifying the role of internal climate variability in future climate trends, J. Climate, 28, 6443–6456, https://doi.org/10.1175/JCLI-D-14-00830.1, 2015.

Trenary, L. and DelSole, T.: Does the Atlantic Multidecadal Oscillation Get Its Predictability from the Atlantic Meridional Overturning Circulation?, J. Climate, 29, 5267-5280, https://doi.org/10.1175/JCLI-D-16-0030.1, 2016.

Wallace, J.M., Deser, C., Smoliak, B. V., and Phillips, A. S.: Attribution of climate change in the presence of internal variability, In Climate Change: Multidecadal and Beyond (Eds: C.P. Chang, M. Ghil, M. Latif, and J. M. Wallace), World Scientific Series on Asia-Pacific Weather and Climate, 6, 1-29, https://doi.org/10.1142/9789814579933_0001, 2013.

Wang, C., Deser, C., Yu, J. -Y., DiNezio, P., and Clement, A.: El Nino and Southern Oscillation (ENSO): A Review. Coral Reefs of the Eastern Pacific, P. Glymn, D. Manzello and I. Enochs, Eds., Springer Science Publisher, 4, 85-106, https://doi.org/10.1007/978-94-017-7499-4_4, 2017.

Wills, R. C. J., Battisti, D. S., Armour, K. C., Schneider, T., and Deser, C.: Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations, J. Climate, 33, 8693-8719, https://doi.org/10.1175/JCLI-D-19-0855.1, 2020.

Wittenberg, A. T.: Are historical records sufficient to constrain ENSO simulations?, Geophys. Res. Lett., 36, L12702, https://doi.org/10.1029/2009GL038710, 2009.

Wu, X., Okumura, Y. M., Deser, C., and DiNezio, P. N.: Two-year dynamical predictions of ENSO event duration during 1954–2015, *J. Climate*, 34(10), 4069-4087, https://doi.org/10.1175/JCLI-D-20-0619.1, 2021.

Yeager, S. Danabasoglu, D., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., and Lovenduski, N. S.: Predicting near-term changes in the Earth System: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model, *Bull. Am. Meteorol*. Soc. 99, 1867–1886, https://doi.org/10.1175/BAMS-D-17-0098.1, 2018.

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., Amrhein, D. E., and Little, C. M.: A review of the role of the Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and associated climate impacts, Rev. Geophys., 57, 316–375, https://doi.org/10.1029/2019RG000644, 2019.