

# Toward a New Estimate of “Time of Emergence” of Anthropogenic Warming: Insights from Dynamical Adjustment and a Large Initial-Condition Model Ensemble

FLAVIO LEHNER AND CLARA DESER

*Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

LAURENT TERRAY

*CECI, Université de Toulouse, CERFACS/CNRS, Toulouse, France*

(Manuscript received 8 November 2016, in final form 9 June 2017)

## ABSTRACT

Time of emergence of anthropogenic climate change is a crucial metric in risk assessments surrounding future climate predictions. However, internal climate variability impairs the ability to make accurate statements about when climate change emerges from a background reference state. None of the existing efforts to explore uncertainties in time of emergence has explicitly explored the role of internal atmospheric circulation variability. Here a dynamical adjustment method based on constructed circulation analogs is used to provide new estimates of time of emergence of anthropogenic warming over North America and Europe from both a local and spatially aggregated perspective. After removing the effects of internal atmospheric circulation variability, the emergence of anthropogenic warming occurs on average two decades earlier in winter and one decade earlier in summer over North America and Europe. Dynamical adjustment increases the percentage of land area over which warming has emerged by about 30% and 15% in winter (10% and 5% in summer) over North America and Europe, respectively. Using a large ensemble of simulations with a climate model, evidence is provided that thermodynamic factors related to variations in snow cover, sea ice, and soil moisture are important drivers of the remaining uncertainty in time of emergence. Model biases in variability lead to an underestimation (13%–22% over North America and <5% over Europe) of the land fraction emerged by 2010 in summer, indicating that the forced warming signal emerges earlier in observations than suggested by models. The results herein illustrate opportunities for future detection and attribution studies to improve physical understanding by explicitly accounting for internal atmospheric circulation variability.

## 1. Introduction

It is well established that anthropogenic increases in greenhouse gases (GHGs) have caused the globe as a whole to warm over the past 50 years beyond the level of natural variability (IPCC 2013). Consequently, concerns have been raised regarding the vulnerability of ecosystems to the rising temperatures (Scholze et al. 2006), in particular that a shift to a different mean state might pose an additional stress for ecosystems that have adapted to a particular reference climate over hundreds

or thousands of years (Rockström et al. 2009). Time of emergence (ToE) has been introduced as a framework to depict the timing of anthropogenic climate change and to investigate whether such changes are potentially beyond the known adaptability of ecosystems (Giorgi and Bi 2009). It is based on the simple notion that a new climate state, often defined by temperature at a given location, can be said to have “emerged” when it deviates significantly from a prior reference state for a given length of time, taking into account natural variability. Various versions of the concept of ToE have been used to detect shifts that have already occurred or are projected to occur in temperature (Mahlstein et al. 2011, 2012), precipitation (Giorgi and Bi 2009), climate extremes (Scherer and Diffenbaugh 2014; King et al. 2015, 2016; Bador et al. 2016), and the carbon cycle (Keller et al. 2014), as well as to estimate the future time horizon

---

Supplemental information related to this paper is available at the Journals Online website: <https://dx.doi.org/10.1175/JCLI-D-16-0792.s1>.

---

Corresponding author: Flavio Lehner, [flehner@ucar.edu](mailto:flehner@ucar.edu)

DOI: 10.1175/JCLI-D-16-0792.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

of emerging climate change as a motivation to inform human exposure to climate change (Lehner and Stocker 2015; Harrington et al. 2016) or to protect potentially endangered ecosystems (Beaumont et al. 2011).

Similar to other frameworks used to detect climate change, such as classical detection and attribution, ToE is subject to uncertainties from model structural differences, choice of emissions scenario, and internal climate variability (Deser et al. 2012; Hawkins et al. 2014). Further uncertainties arise from the choice of ToE metric, including differences in the definition of ToE thresholds (e.g., standard deviation or maximum value), the statistical test that defines robust emergence (e.g., significant epoch differences or simple threshold exceedance), or the temporal filtering of data to estimate the forced signal (e.g., linear trend, running mean, spline, or model ensemble mean). Generally, it is assumed that emergence from a reference climate occurs as a result of radiative forcing from anthropogenic increases in GHGs. However, emergence might not necessarily be synonymous with anthropogenic climate change if internal multidecadal variability is substantial. This is particularly problematic if ToE is calculated from observations, for which the anthropogenically forced response is difficult to quantify precisely. Large initial-condition ensembles of historical simulations with climate models, for which ToE of human-induced warming can be well estimated, provide a useful tool for investigating the contribution of internal multidecadal variability to ToE uncertainty. Recent work using large initial-condition model ensembles has shown that unforced changes in atmospheric circulation can advance or delay the emergence of anthropogenically forced trends in regional temperature and precipitation by up to several decades, especially in the extratropics (Deser et al. 2012; Hawkins et al. 2016). Hence, accounting for this internal variability could reduce the uncertainty on estimates of ToE at local and regional scales, provide information about the forced response that drives ToE, and facilitate differentiation between predictable and unpredictable components of ongoing climate change. In this context, it is worth mentioning that unpredictable internal variability represents an irreducible uncertainty for estimates of ToE, in contrast to model structural uncertainties and forcing uncertainties, which, in principle, are reducible (Hawkins and Sutton 2009).

Past studies investigating ToE of surface air temperature or precipitation have focused on a number of different aspects and technical issues: Mahlstein et al. (2011) detected and contrasted regions of early and late ToE; Giorgi and Bi (2009) investigated the impact of different emissions scenarios and different climate models on ToE; and Hawkins and Sutton (2012)

demonstrated how the magnitudes of both warming and internal variability differ across climate models and how this in turn affects estimates of ToE. While all of these studies explicitly mention the important influence of internal variability on estimates of ToE, to our knowledge none has investigated its physical origins or has tried to quantify its impact. This is particularly important for the Northern Hemisphere midlatitudes, where large uncertainties in ToE due to the influence of internal atmospheric circulation variability and the uncertain future of the storm track (Shaw et al. 2016) coincide with high population density, and hence an inherent desire for more accurate climate projections. While Deser et al. (2012) explicitly showed the effects of internal variability on ToE, their results pertained to future climate projections only, not to the historical period and not to observations.

The aim of this paper is to provide new estimates of ToE for winter and summer temperatures over North America and Europe after accounting for the influence of internal atmospheric circulation variability, using both gridded observational data for the period 1920–2015 and a large ensemble of simulations with a fully coupled Earth system model for the period 1920–2100. The circulation influence is removed with a dynamical adjustment technique based on constructed circulation analogs (Deser et al. 2016). While ToE calculated from dynamically adjusted temperatures may not necessarily reflect the actual ToE that ecosystems experience in the real world, it allows for separation between potentially unpredictable components of ToE (atmospheric circulation) from potentially predictable ones (thermodynamic processes, oceanic influence, etc.) and provides physical insight into the processes governing ToE. We calculate ToE in different ways in order to isolate additional sources of uncertainty including imperfect knowledge of the forced response, length of record, degree of temporal filtering, model biases in internal variability, and observational uncertainty; however, by using a single model, we do not sample the effect of model structural uncertainty on ToE (Hawkins and Sutton 2012). We also contrast local ToE (at the gridbox level) with spatially aggregated ToE to illustrate how the two perspectives differ in the conclusions they allow (Fischer et al. 2013; Fischer and Knutti 2014).

The remainder of the paper is structured as follows. Section 2 introduces the observational datasets and model simulations used, the dynamical adjustment technique applied, and methods used to calculate ToE. Section 3 illustrates the effect of dynamical adjustment on ToE, discusses the influence of model biases in variability, and investigates potential drivers of the

residual uncertainty. Section 4 provides a summary and discussion.

## 2. Data and methods

### a. Observational datasets and model simulations

We use monthly mean surface air temperature (SAT) and sea level pressure (SLP) from the fully coupled Community Earth System Model (CESM), version 1, large ensemble (CESM-LE; Kay et al. 2015). The CESM-LE consists of 40 simulations for the period 1920–2100 at a spatial resolution of approximately  $1^\circ$  in both latitude and longitude. Each simulation begins from slightly different atmospheric but identical ocean, land, and sea ice initial conditions. In accordance with protocols from phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012), historical natural and anthropogenic radiative forcings and land use were applied from 1920 to 2005, and representative concentration pathway 8.5 (RCP8.5) radiative forcing and land use thereafter. Because of its many simulations, the CESM-LE allows us to sample the model's internal variability as well as its forced response in a robust manner. We also make use of an 1800-yr CESM control simulation under preindustrial (1850) radiative forcing conditions (Kay et al. 2015).

For observationally derived datasets, we use monthly mean SAT from the Berkeley Earth Surface Temperature dataset (BEST; Rohde et al. 2013) at  $1^\circ \times 1^\circ$  resolution and the ensemble mean SLP from the Twentieth Century Reanalysis (20CR; Compo et al. 2011) at  $1^\circ \times 1^\circ$  resolution (after bilinearly interpolating it from the original  $2.5^\circ \times 2.5^\circ$  resolution) for the period 1920–2015. The 20CR, which currently ends in 2012, is extended from January 2013 to December 2015 by adding the monthly SLP anomalies from ERA-Interim (Dee et al. 2011;  $1^\circ \times 1^\circ$  resolution) to the 1920–2012 SLP climatology of 20CR. 20CR and ERA-Interim show very good agreement during the time of overlap (1979–2012; not shown), justifying this approach to construct a continuous SLP dataset. In addition, we use two SAT datasets, the Merged Land–Ocean Surface Temperature (MLOST) analysis (Vose et al. 2012) and the University of Delaware dataset (Willmott and Matsuura 1995), to test observational uncertainty.

### b. Dynamical adjustment

Empirical dynamical adjustment techniques have a long history and are motivated by the need to remove circulation-induced SAT variability that obscures the radiatively forced response of the climate system to increasing greenhouse gases (Hurrell 1996; Thompson

et al. 2009; Wallace et al. 2012; Smoliak et al. 2015). Although the circulation itself may respond to GHG forcing, this response is generally much weaker than internally generated variability in the extratropics, including that on multidecadal time scales (Deser et al. 2012). To estimate the contribution of atmospheric circulation to SAT changes, we apply a dynamical adjustment methodology based on constructed circulation analogs using SLP (Deser et al. 2016). The approach is summarized below; full details are available in Deser et al. (2016).

For observations, we consider the 96-yr period 1920–2015 and apply the method to North America ( $20^\circ$ – $90^\circ$ N,  $180^\circ$ – $10^\circ$ W) and Europe ( $25^\circ$ – $90^\circ$ N,  $60^\circ$ W– $40^\circ$ E) separately. For a given “target” month and year (e.g., January 2012), we rank the remaining 95 January SLP fields according to their similarity with the target SLP pattern using a Euclidean distance metric. From the 80 SLP fields with the smallest Euclidean distances, we randomly select 50 and compute the optimal linear combination that best fits the SLP pattern of the target month. Applying the same set of linear coefficients to the accompanying SAT fields, we obtain the associated optimal linear combination of SAT. The entire process is repeated 100 times (using random selection with replacement), and the resulting set of 100 optimal linear combinations is then averaged to arrive at a best estimate of the target SLP field and its associated dynamically induced SAT field. Deser et al. (2016) illustrate the importance of this iterative random selection process and the sensitivity of the results to the choice of parameters. Also, while some of the chosen fields might not be that similar to the target pattern, the linear combination will assign them small weights. In practice, our method gives very similar results to the partial least squares regression used by Smoliak et al. 2015 (Deser et al. 2016). Our procedure is applied to each month of each year, so that eventually the dynamically induced component of monthly SAT anomalies at each grid box over North America and Europe during 1920–2015 is obtained. Prior to applying the procedure, the SAT time series at each grid point for each month (all Januaries, and so on) are detrended with a quadratic fit to remove the ostensible global warming signal [see the appendix of Deser et al. (2016) for details on how this detrending approach was chosen]. This step is necessary, since otherwise months picked from the end of the record will contribute higher SAT anomalies simply because of the anthropogenically forced warmer background climate, even if the SLP patterns are the same. Note that the detrending is only for the purpose of obtaining the dynamical SAT contribution, SLP is not detrended and can have a (forced) trend, which in turn can project onto the

dynamical SAT contribution; the thermodynamic contribution, or dynamically adjusted SAT, is then obtained as a residual between the raw (not detrended) SAT and the dynamical SAT contribution.

For the CESM-LE simulations, we obtain the constructed circulation analogs and their corresponding SAT anomalies from the CESM 1800-yr preindustrial control simulation. This is done to increase the pool of samples from which analogs are selected (e.g., 1800 vs 96) and to allow a direct estimate of the dynamical contribution to SAT in the absence of climate change, alleviating the need for the detrending step described above. We leverage the larger sample size by randomly selecting 100 from the closest 150 SLP “analogues” (out of 1800 samples) and repeat this entire procedure 100 times. The larger number of circulation analogs and their closer resemblance to the target SLP field results in a more certain estimate of the dynamical contribution to SAT in the CESM-LE compared to observations as discussed further below. In addition, we used analogs from the historical period itself to investigate whether this leads to significantly different results in dynamically adjusted SAT and subsequent ToE calculations; no significant differences were found (see Fig. S1 in the supplemental material).

### c. Time of emergence

We define the time of emergence at each grid box as the year when the 10-yr-running-mean SAT first exceeds (and remains above) the average SAT during the reference period 1920–49 by at least two standard deviations based on the variability of the 10-yr-running-mean SAT during the reference period. ToE is computed separately for each season, and for both the raw and dynamically adjusted SAT fields. The 10-yr running mean serves to reduce the noise from interannual variability while retaining key features of decadal-scale fluctuations. We have tested other running-mean lengths and found that 10 years strikes a good balance between retaining decadal-scale variability while at the same time not being unduly susceptible to errors from higher-frequency noise, especially for dynamically adjusted SAT (see Fig. S2 and accompanying discussion in the supplemental material). The reference period was chosen because of the 1920 start date of the CESM-LE and the reduced coverage and reliability of SLP measurement in the Pacific–North American sector prior to that time (Raible et al. 2014; Krueger et al. 2013). It is acknowledged that the period 1920–49 can only serve as a quasi-natural state from which to base detection of ToE, since arguably anthropogenic forcing might already have had an influence (King et al. 2016), but it is the best we can do given the data constraints. This

analysis is carried out separately for the raw and dynamically adjusted data during the period 1920–2015. Finally, we note that there is no unique definition of ToE [see, e.g., alternative approaches in Mahlstein et al. (2012)] and that the results may vary according to the definition of both the signal and the noise. We shall assess the sensitivity of our ToE estimates to both factors in a controlled fashion below.

Estimates of ToE based on a single realization of climate (as is the case for observations) are inherently uncertain, as they cannot adequately sample the full range of trends in atmospheric circulation or thermodynamic processes related to conditions on land and ocean. Therefore, there is no guarantee that a particular SAT time series that has already emerged will not fall below the ToE threshold again in the future, even when the time series is low-pass filtered (Hawkins et al. 2014). This issue challenges the robustness of ToE estimates from observations, which might show a “false” ToE. Note, however, that a false ToE might still have significant impacts on affected systems, depending on how long SAT remains above the threshold; it simply indicates that temperature has not completely shifted into uncharted territory, where it will not drop below the threshold any longer. We can quantify the likelihood of a false ToE, at least within the framework of the CESM-LE, by considering the future portion of each simulation in our calculations. That is, we can use the full period 1920–2100 to calculate the probability that SAT drops below the ToE threshold in any year after 2010 (note that 2010 indicates the midpoint of the last 10-yr period available: 2006–15).

For the model simulations, we calculate ToE in four ways (calculations 1–4) for both the raw and dynamically adjusted SAT:

- 1) For each ensemble member, we disregard data after 2015 (for direct comparison to observations).
- 2) We proceed as in calculation 1, but use the full period of simulation until 2100. By comparing with the results from calculation 1, we are able to assess the rate of false positives resulting from not knowing the future SAT evolution.
- 3) For the forced ToE, we use the ensemble mean SAT time series at each grid point, but calculate ToE based on the ToE thresholds of all individual ensemble members using data until 2100. Recall that the ensemble mean represents the model’s forced response, which is what we aim to detect. Hence, this way of calculating ToE represents an estimate of the ToE of the anthropogenically forced response and ignores uncertainties in SAT evolution due to internal variability.

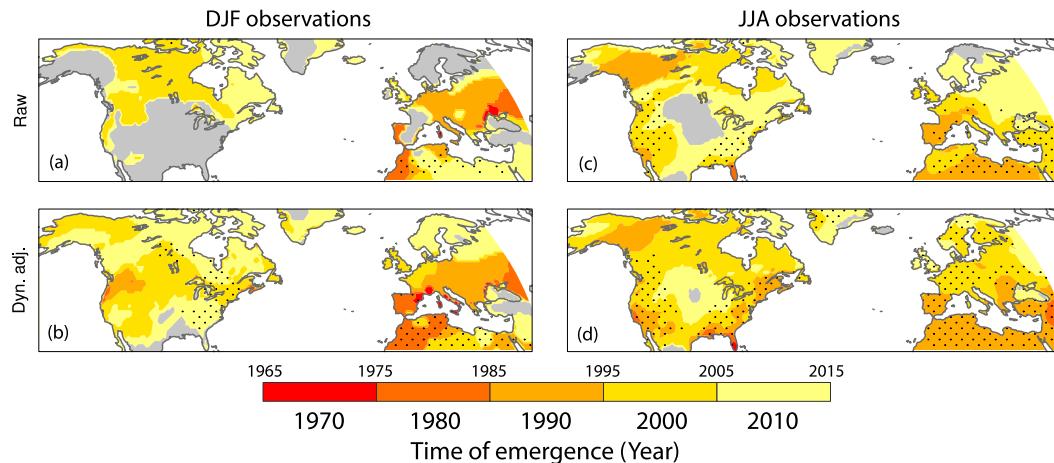


FIG. 1. ToE of observed 10-yr-running-mean time series of SAT anomalies relative to the reference period 1920–49 based on DJF (a) raw and (b) dynamically adjusted data and JJA (c) raw and (d) dynamically adjusted data. The color bar shows the central year of the 10-yr-running-mean period (e.g., 1990 refers to 1986–95). Gray shading indicates that emergence has not yet occurred. Stippling indicates a <10% chance of false emergence.

4) We proceed as in calculation 3, but use the observed ToE threshold at each grid point (after bilinearly interpolating the observed ToE threshold to the model grid). This represents the ToE of the model’s forced response given the observed variability. This measure allows us to make inferences about how model biases in SAT variability affect our results.

### 3. Results

#### a. Time of emergence in observations

Figure 1 shows maps of observed ToE based on raw and dynamically adjusted SAT in December–February (DJF) and June–August (JJA). The year of emergence represents the midpoint of the 10-yr-running-mean period (e.g., 1990 refers to 1986–95); note that 2010 is the latest possible year of emergence since it represents the last complete 10-yr running mean in the observational record (2006–15). Stippling indicates regions where there is a less than 10% chance, according to CESM-LE, of a false positive in ToE. We describe results for North America first, followed by those for Europe.

For raw DJF SAT over North America, most of Canada emerged in 1990 whereas most of the United States has not yet emerged (Fig. 1a). In contrast, after dynamical adjustment, nearly the entire continent is considered emerged by 2010 in DJF (Fig. 1b). However, there is a greater than 10% chance that these ToE estimates may represent false positives (indicated by areas without stippling). Although dynamical adjustment advances ToE over the United States, it delays it over parts of north-central Canada. This indicates that dynamically

induced decadal SAT variability can augment or offset an underlying signal of anthropogenic climate change depending on location (see also Deser et al. 2016). For raw SAT in JJA, the entire continent except for the midsection and Alaska emerged by 2000 or 2010, with earlier emergence over western Canada and Southern California (1990) and Florida (1980) (Fig. 1c). Dynamical adjustment has a smaller impact in JJA compared to DJF, but does advance ToE over the eastern portion of the continent by 5–10 yr and fills in the unemerged area in the center of the continent (Fig. 1d). Unlike DJF, about one-third of North America shows robust emergence of both raw and dynamically adjusted SAT in JJA, defined as a less than 10% chance of false positive (stippling in Figs. 1c,d). It is worth noting that the reference period includes the exceptionally warm 1930s over the central United States (“Dust Bowl”) in observations, and hence might prompt relatively later ToE in that specific region.

Unlike North America, large parts of the European continent and North Africa are emerged by 2010 in both seasons based on raw SAT (Figs. 1a,c). The earliest emergence occurs in DJF over eastern Europe (1970–80), and in JJA over most of the Mediterranean region (1990). Consequently, in the dynamically adjusted data, the emerged area increases only moderately (Figs. 1b,d): in DJF, parts of France and Spain (1980–90), as well as Scandinavia (2010) are newly emerged, and in JJA eastern Europe’s ToE advances by about a decade. However, only a fraction of the allegedly emerged areas is likely to be robust. Most spatial patterns described here are robust across the two alternative SAT datasets, with the exception of the eastern United States in DJF

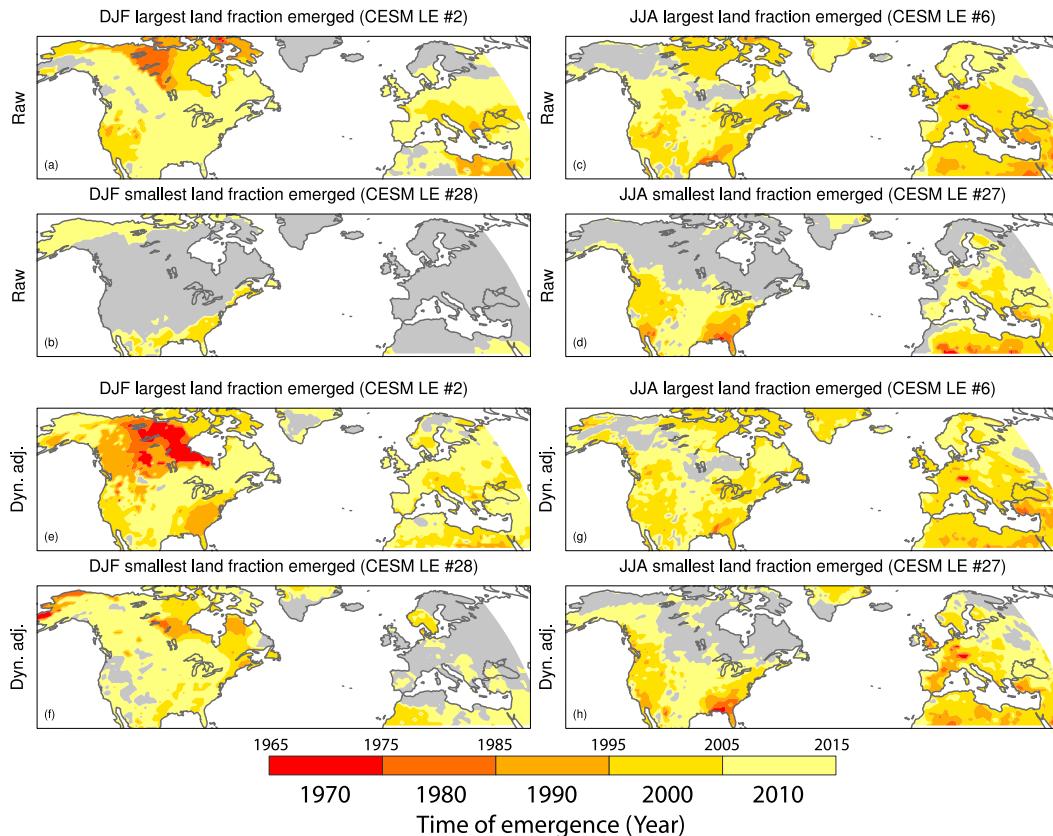


FIG. 2. As in Fig. 1, but for selected members of the CESM-LE based on (a)–(d) raw and (e)–(h) dynamically adjusted data. See text for details.

after dynamical adjustment, where BEST shows emergence, while the other two datasets do not (Fig. S3 in the supplemental material).

We have further tested the effect of the dynamical adjustment on ToE in observations using the metric of [Mahlstein et al. \(2011\)](#) and [Mahlstein et al. \(2012\)](#). Their metric detects emergence when consecutive 30-yr periods are, and remain, significantly warmer than a particular 30-yr reference period [see [Mahlstein et al. \(2011\)](#) for details]. This metric, which is more conservative than ours because they have not applied any low-pass filtering to the data, shows very few emerged grid cells in the raw data (Fig. S4 in the supplemental material). Interestingly, some regions that are emerged in the raw data are not emerged after dynamical adjustment, suggesting that even this conservative metric might show a false ToE due to internal variability.

### b. Time of emergence in CESM-LE

To put the observational results in perspective, we turn to the CESM-LE. We first examine ToE in four individual ensemble members (Fig. 2; without assessing the rate of false positives; i.e., no stippling): those with the largest and

smallest fractions of land area emerged by 2010 in winter (members 2 and 28, respectively) and in summer (members 6 and 27, respectively). The dynamically adjusted results are based on calculation 1 outlined in [section 2c](#) (i.e., treated just like the observations ending in 2015). CESM-LE member 2 in DJF shows almost all of North America and Europe emerged by 2010 (Fig. 2a), while member 28 shows almost no emergence by 2010 (Fig. 2b). These differences are somewhat reconciled when both members are dynamically adjusted, although regional differences remain (Figs. 2e,f). In JJA, differences between the two contrasting members are smaller than those in DJF, especially when dynamically adjusted (Figs. 2c,d). The spatial patterns of the dynamically adjusted ToE in JJA over the United States resemble the observations, in particular with regard to earlier emergence over the western and southeastern United States (recall Fig. 1d). These examples from the CESM-LE serve to illustrate the large range of possible ToE outcomes due solely to internal variability, and that uncertainties may remain even after dynamical adjustment, particularly in DJF.

For a more general examination of ToE within the CESM-LE, Fig. 3 shows the median ToE across all

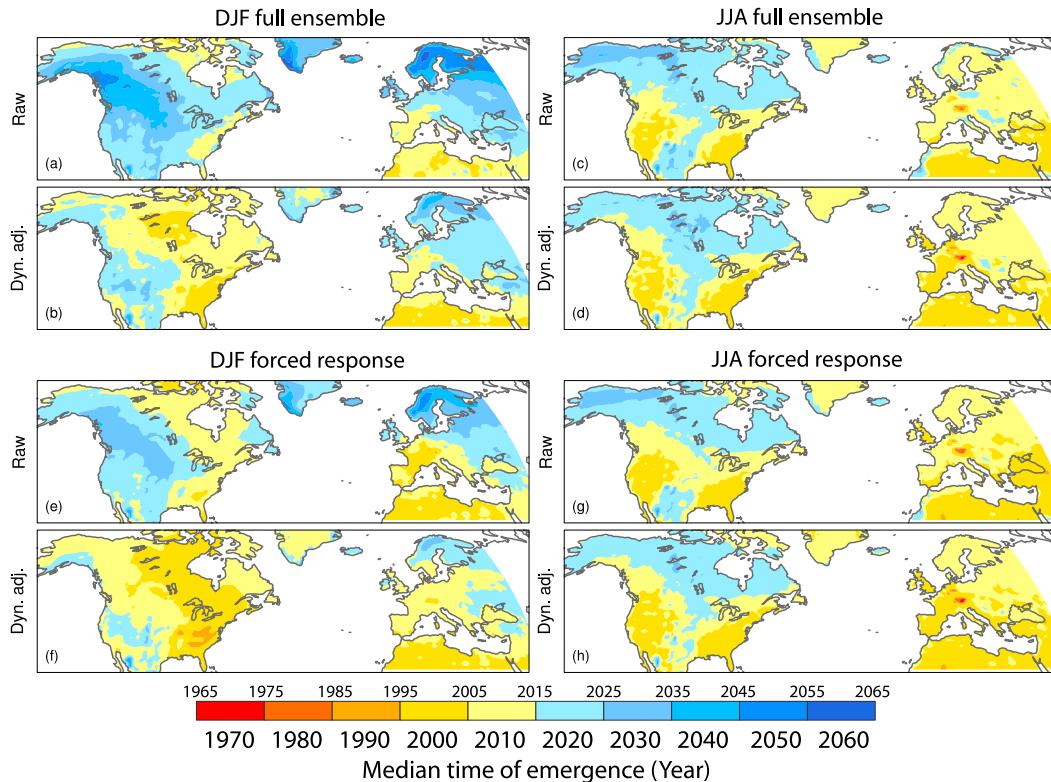


FIG. 3. Median ToE of 10-yr-running-mean SAT anomalies relative to the reference period 1920–49 in the CESM-LE based on DJF (a) raw and (b) dynamically adjusted data and JJA (c) raw and (d) dynamically adjusted data. (e),(g) As in (a),(c), but using the forced response with ToE thresholds based on the raw data. (f),(h) As in (e),(g), but with ToE thresholds based on dynamically adjusted data. See text for details. Note that the same range of colors as in Figs. 1 and 2 is used for twentieth-century ToE values; blue shading indicates twenty-first-century ToE values.

ensemble members based on raw SAT using calculation 2 in section 2c. Note that here we have included the future segment of the simulations to estimate ToE, which is not possible with observations. In DJF, a large swath of western North America and all of northern Europe are projected to emerge, on average, as late as the 2040s or 2050s (Fig. 3a). After dynamically adjusting DJF SAT, the median ToE over North America is advanced almost everywhere by 1–2 decades, in particular the sector reaching from Alaska down to the middle of the continent (Fig. 3b). Over Europe, the median ToE in DJF advances 20 or more years mainly at high latitudes, while the areas bordering the Mediterranean Sea are relatively unchanged. The distinctive northward gradient in the median value of dynamically adjusted ToE in DJF over Europe in CESM-LE resembles observations (Fig. 1) and previous observation-based studies (Mahlstein et al. 2012). The pattern of the median dynamically adjusted ToE over North America does not resemble observations.

In JJA, the median value of raw ToE in the CESM-LE is earliest (2000) over the eastern and western United

States, in good agreement with observations; however, the values over Canada and the midsection of the United States are much later (2020), unlike observations (Fig. 3c). Europe shows a relatively uniform median raw ToE in JJA, with values around 2000 in the south and 2010 in the north. Dynamically adjusting JJA SAT has only a small effect, advancing median ToE by about a decade at most (Fig. 3d). We defer discussion of the remaining panels in Fig. 3 to section 3d.

The median values shown in Fig. 3 belie the wide range of ToE across the individual members, with some showing values in the 2080s and 2090s over particular areas (not shown; see also Hawkins et al. 2014). Given such late ToE close to the end of the simulation period, the possibility of a false emergence cannot be excluded, especially if GHG concentrations were to stabilize or decline after 2100. Section 3f will address the rate of false positives in ToE in more detail.

### c. Land fraction emerged by 2010

To further compare model results and observations, we use the fraction of land area over which warming has

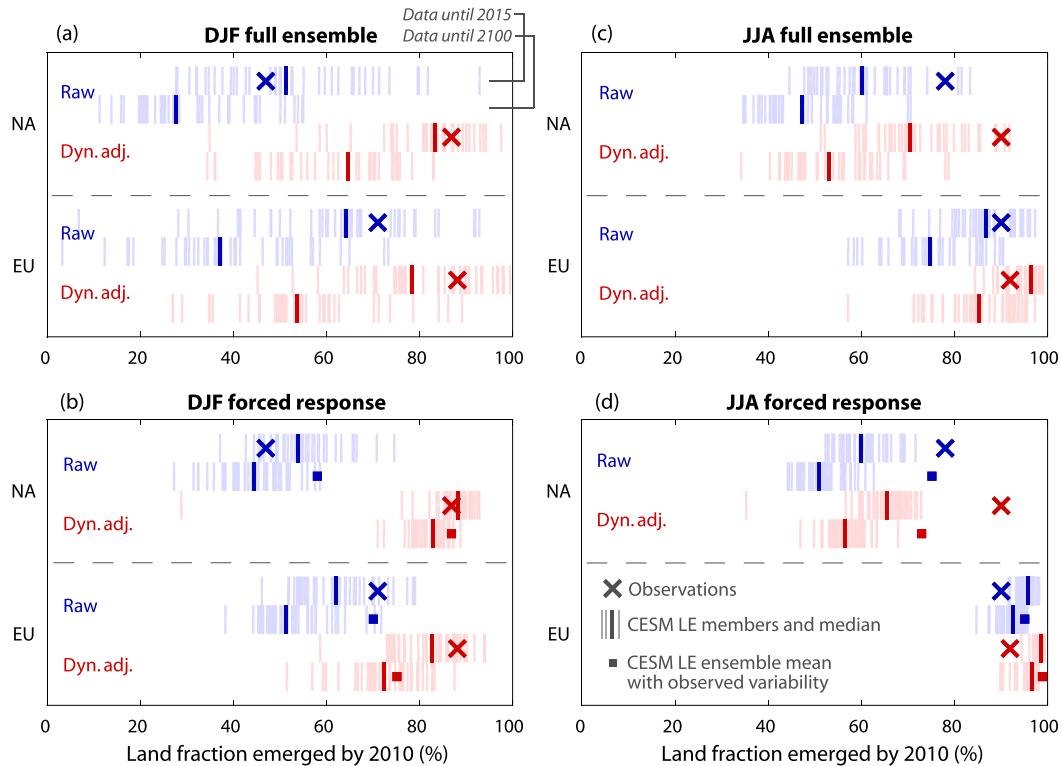


FIG. 4. Land fraction of North America (NA) and Europe (EU) over which warming has emerged by 2010 in observations and CESM-LE for (a),(b) DJF and (c),(d) JJA. Blue denotes raw data, and red denotes dynamically adjusted data. Crosses denote observations (BEST), thin vertical lines denote the individual CESM-LE ensemble members, and thick vertical lines denote the median of the CESM-LE ensemble members. ToE is computed based on (a),(c) the full CESM-LE and (b),(d) the CESM-LE ensemble mean plus ToE thresholds from individual members. In (b) and (d), filled squares denote ToE calculated using the CESM-LE ensemble mean and observed ToE thresholds. For the CESM-LE, the top and bottom parts of each panel show results based on the period 1920–2015 and 1920–2100, respectively [see annotation in (a)].

emerged by 2010 [hereinafter land fraction emerged by 2010 (LFE2010)] as a spatially aggregated metric (Fig. 4). For this comparison we use calculations 1 and 2 from section 2c. To recap, calculation 1 is based on the simulations up to year 2015, while calculation 2 is based on their full length (up to year 2100). In other words, calculation 1 is much more likely to show false emergence than calculation 2, with implications for observations.

In DJF (Fig. 4a) the observed LFE2010 for North America and Europe (47% and 72%, respectively) lies within the very large range of the individual model ensemble members (28%–93% and 7%–93%, respectively, with median values of 51% and 64%, respectively). This also holds true after dynamical adjustment, for which LFE2010 in observations increases to 87% and 88% for North America and Europe, respectively, and in the model increases to 35%–97% (median value of 83%) and 45%–99% (median value of 78%), respectively. However, these high values may be

misleading, since they could represent false detections of ToE because the records end in 2015. Indeed, if one considers the full length of the simulations, the median values for North America and Europe in the raw data drop to 28% and 37%, respectively, and in the dynamically adjusted data to 65% and 54%, respectively. That is, LFE2010 is reduced by 23% on average over both continents. A similar result occurs in JJA (Fig. 4c), where the LFE2010 is reduced over North America and Europe. However, in contrast to DJF, JJA does not exhibit large gains in LFE2010 as a result of dynamical adjustment, generally less than 10% for both North America and Europe.

In summary, for six out of the eight cases where we can compare observations and simulations directly (i.e., calculation 1 for North America and Europe in DJF and JJA for raw and dynamically adjusted SAT), the observed value of LFE2010 exceeds the CESM-LE median

value. Only for raw DJF over North America and dynamically adjusted JJA over Europe is LFE2010 lower than the CESM-LE median. However, in all cases the observed value lies within the range of the individual ensemble members, indicating that potential model biases in temperature trends or variability are not necessarily needed to explain the tendency toward lower LFE2010 in the model compared to observations. The role of model biases will be discussed in [section 3e](#).

#### *d. Time of emergence of the forced response in CESM-LE*

The metrics presented so far were intended for comparison with observations, for which the forced response is per se unknown. Within CESM-LE, however, we can start to disentangle uncertainties in ToE arising from internal variability since the forced response can be estimated from the ensemble mean. Accordingly, we have computed ToE in each ensemble member using the CESM-LE ensemble-mean 10-yr-running-mean SAT time series (i.e., the forced SAT response) and the ToE thresholds from the individual ensemble members (calculation 3 in [section 2c](#)). This constitutes an estimate of the ToE of the forced response [referred to as the forced ToE (FToE)] in the presence of uncertainty in the internal variability threshold. The median values of FToE computed from all ensemble members are shown in [Figs. 3e–h](#). Their spatial patterns are similar to those of the median values of the total ToE discussed previously. However, quantitatively, the median FToE values tend to be about 10 years earlier than the median values of the total ToE, for both seasons, both continents, and regardless of whether they are based on raw or dynamically adjusted data. This advance in FToE over total ToE is because the ensemble-mean SAT time series is smoother than the SAT time series in any individual simulation, decreasing the likelihood of dropping below the ToE threshold once it has been exceeded. Expressed in terms of LFE2010, FToE shows larger values and a smaller uncertainty range than total ToE for the same reason ([Figs. 4c,d](#)). The range of LFE2010 values based on FToE encompasses the observations in DJF, but not in JJA when it lies below the observed value for North America and above it for Europe. This indicates that observed LFE2010 in JJA over North America (Europe) is significantly larger (smaller) in observations than the forced response of the model would suggest.

#### *e. Role of model biases in variability*

Next we investigate the effect of model biases in interannual SAT variability upon the estimates of ToE and LFE2010. To get a sense of where CESM might

have too much or too little SAT variability, we plot maps of the standard deviation of 10-yr-running-mean SAT anomalies from observations alongside the mean standard deviation across the CESM-LE members ([Fig. 5](#)). Areas enclosed by light and dark blue contours with hatching indicate regions where the observed value lies outside the entire CESM-LE. It is evident that CESM has too much variability over the western United States and eastern Europe in DJF and over northwestern North America in JJA, as well as too little variability over Newfoundland and along the U.S. Gulf Coast in DJF, and in the southeastern United States in JJA ([Figs. 5c,g](#)). In raw JJA, there are a number of regions, such as around Hudson Bay, southwestern Canada and the United States, or Spain, where CESM's variability is not outside observations, but the mean from CESM-LE at least suggests an overestimation of variability. Interestingly, dynamical adjustment reconciles most of these biases ([Figs. 5d,h](#)). Further, it is worth emphasizing that observational uncertainties in SAT variability exist ([Lehner et al. 2017](#)) and thus model fidelity may depend on the observational dataset being used to benchmark the model.

To investigate the role of these model biases in SAT variability, we estimate ToE and LFE2010 using the ensemble-mean SAT time series from the CESM-LE and the observed ToE threshold (calculation 4 in [section 2c](#); squares in [Figs. 4c,d](#)). By using the observed ToE threshold, we control for potential biases in simulated SAT variability, and hence any significant discrepancy between this estimate of ToE and FToE must arise from differences in the ToE threshold (i.e., the SAT variability during the reference period).

In DJF, no effect from model bias in SAT variability is apparent on LFE2010, as all values (indicated by the squares) lie within the range of the model's LFE2010, both for the raw data and the dynamically adjusted SAT over both continents ([Fig. 4b](#)). In other words, using the observed ToE threshold does not produce an LFE2010 that is outside of the range produced by the model. In JJA, however, this is not the case over North America for either the raw or dynamically adjusted data, with a substantially larger LFE2010 when the observed ToE threshold is used (about 22% in raw data and 13% in dynamically adjusted data; compare squares and thick solid vertical lines in [Fig. 4d](#)). This suggests that over North America in JJA, the model's slight but widespread overestimation of SAT variability has a discernible effect on the estimated FToE. Recent research suggests an overestimated land–atmosphere coupling might be responsible for this bias, although the observational constraint on this metric remains weak ([Merrifield and Xie 2016](#)). In JJA over Europe, the

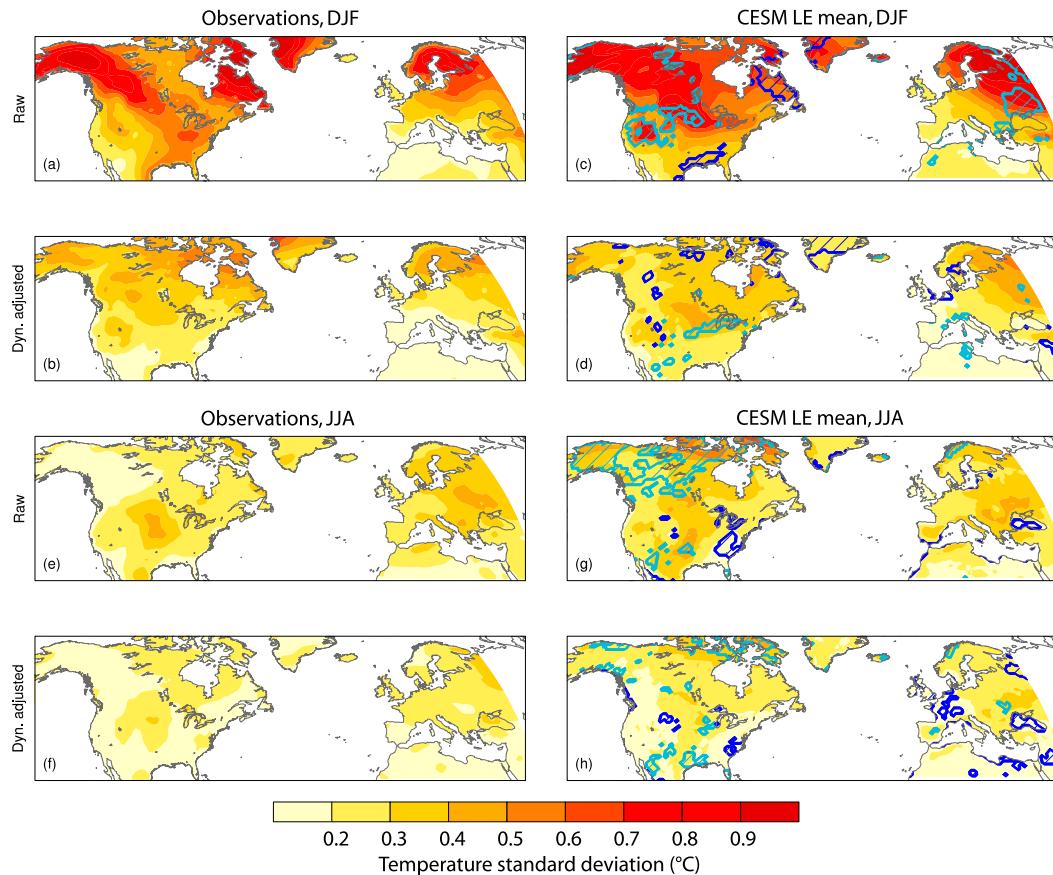


FIG. 5. Standard deviation  $\sigma$  of 10-yr-running-mean time series of SAT anomalies from (a),(b),(e),(f) observations (BEST) and (c),(d),(g),(h) CESM-LE for (a)–(d) DJF and (e)–(h) JJA. For CESM, the mean  $\sigma$  across all 40 members is shown. The time series have been quadratically detrended before calculating  $\sigma$ . Light blue (blue) contours and hatching enclose regions where the  $\sigma$  of all CESM-LE members is above (below) the observed  $\sigma$ .

results also suggest a potential role for model bias in SAT variability on LFE2010, as the squares lie at the upper end or just beyond the model range (Fig. 4d).

#### f. The 5%–95% range of time of emergence

The 5%–95% range of uncertainty in ToE across the CESM-LE is shown in Fig. 6. Over North America in DJF, the largest ToE uncertainty ( $>25$  yr) is concentrated in a corridor from Alaska to the U.S. East Coast (Fig. 6a). Most of central to northern Europe shows a similar range. These areas of largest uncertainty in ToE are collocated with high variability in the midlatitude westerlies and associated storm track (Strong and Davis 2007), which is likely responsible for the uncertainty in ToE of unadjusted DJF SAT. Position and variability of storm tracks, in turn, might be partially driven by interactions with the midlatitude ocean (e.g., Booth et al. 2012). After dynamical adjustment, the 5%–95% range of DJF ToE is greatly diminished over

both continents, with some areas showing reductions of more than 10 years (Fig. 6b). The remaining uncertainty in dynamically adjusted ToE is largest across the northern portion of the eastern United States, central Canada, and Scandinavia, with values of 2–3 decades (Fig. 6b).

In JJA, the largest 5%–95% ToE ranges based on raw SAT occur over the southern United States and into Mexico as well as eastern Mediterranean countries (Fig. 6c). However, the maximum values are considerably smaller than those in DJF. The uncertainty range in JJA ToE is only slightly reduced through dynamical adjustment (cf. Figs. 6c and 6d), in line with the expectation that the large-scale circulation exerts a weaker influence on SAT in summer compared to winter. It is interesting to note that the regions of greatest residual ToE uncertainty (after removing the effects of circulation variability) in JJA have been identified as locations with potentially strong land surface coupling, for

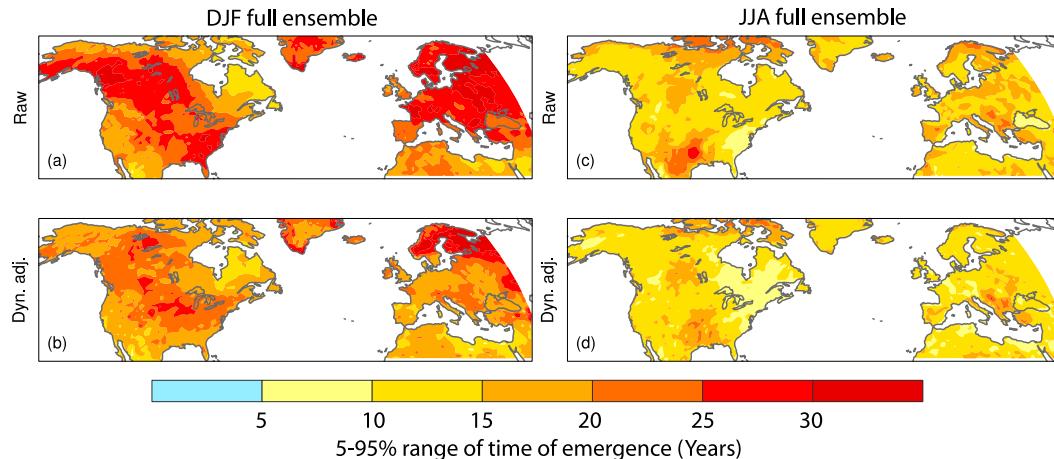


FIG. 6. The 5%–95% range of time of emergence in CESM-LE for (a),(b) DJF and (c),(d) JJA based on (a),(c) raw and (b),(d) dynamically adjusted data.

example the southern United States and eastern Mediterranean region (Seneviratne et al. 2013).

*g. Potential thermodynamic sources of residual ToE uncertainty*

After removing circulation-induced uncertainty in ToE, we expect any remaining uncertainty to be related primarily to thermodynamic processes, ocean-induced internal variability, and residual dynamical contributions not captured by the dynamical adjustment procedure. Here we provide a simple initial investigation into physically plausible sources of this uncertainty, focusing on the role of snow cover and sea ice in winter and soil moisture in summer. We will later remove these uncertainties empirically from the dynamically adjusted SAT fields to provide an empirical upper bound estimate of their impact on the overall uncertainty in ToE.

Figure 7a shows maps of the local correlation coefficient between lagged seasonal mean time series of snow cover (fraction of a grid cell covered by snow) and DJF dynamic SAT for lags  $-3$  through  $+2$  months, where negative lags indicate snow cover leading SAT and positive lags indicate snow cover lagging SAT [i.e., lag  $-3$  months indicates September–November (SON) snow cover is being correlated with DJF SAT]. Here, dynamic refers to the dynamical contribution to SAT, that is, the time series that is subtracted from the raw SAT to obtain the dynamically adjusted SAT. For this calculation, the CESM-LE ensemble mean time series of both variables are removed from each ensemble member and then the residual time series from all members are concatenated before computing the correlations. All correlations shown are significant at 95% confidence. Figure 7b shows the analogous correlations, but using dynamically adjusted SAT in place of dynamic

SAT. By comparing the correlation maps for dynamic and dynamically adjusted SAT at the various lags, we aim to gain insight into the role of thermodynamic interaction between snow cover and SAT variability. Similar patterns are obtained using regression coefficients in place of correlations (not shown).

At lag  $-3$  months, the correlation maps show that SON snow cover in the northern United States and southern Canada is significantly negatively correlated only with dynamically adjusted DJF SAT, while the correlation with dynamic DJF SAT is insignificant in the region of interest (Figs. 7a,b). Note that a negative correlation indicates below normal snow cover is associated with above normal SAT, and vice versa. At lag  $-2$  and  $-1$  month, significant correlations are apparent with both SAT fields, but are larger and/or more widespread for dynamically adjusted SAT than dynamic SAT. At lags from  $+1$  to  $+2$  months, the opposite is true. The simultaneous correlation maps show the strongest correlations with both SAT fields, with maximum amplitudes exceeding 0.6. Taken together, the distinctions in the lead-lag correlation maps with dynamic and dynamically adjusted SAT provide some evidence, albeit circumstantial, that once the effects of atmospheric circulation variability are removed from SAT, a physically plausible thermodynamic relationship (indicative of two-way feedbacks) between snow cover and SAT variability is revealed. Note that the correlations between snow cover and dynamic SAT likely indicate that atmospheric circulation variability is impacting both quantities.

These results, in turn, suggest that a plausible mechanism underlying the residual uncertainty in the 5%–95% range of dynamically adjusted ToE in these locations (recall Fig. 6b) may be associated with snow

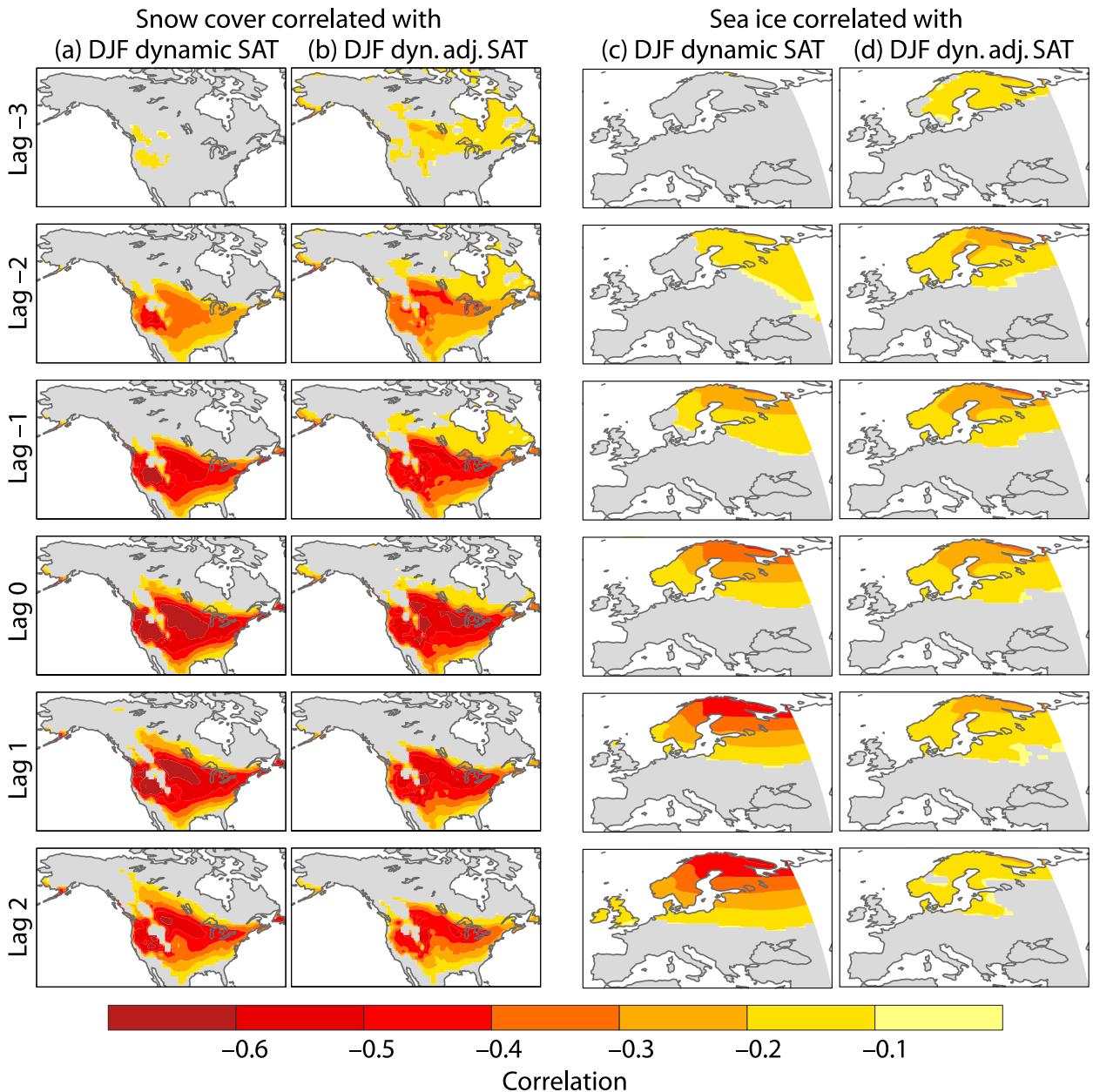


FIG. 7. Pointwise correlation across all CESM-LE members between lagged seasonal mean time series of (a) snow cover and DJF dynamical SAT (i.e., dynamic) and (b) snow cover and DJF dynamically adjusted SAT (i.e., thermodynamic). (c),(d) As in (a),(b), but for Barents Sea sea ice and SAT. The lag is given in months, for example, lag  $-3$  corresponds to the correlation between SON and DJF. Only correlations significant at the 95% level taking into account serial autocorrelation are shown. (b) As in (a), but for sea ice in the Barents Sea and dynamically adjusted SAT. The CESM-LE ensemble mean of each variable has been subtracted from each member prior to the analysis (see text for details).

cover variability. The residual ToE uncertainty in northern Canada (Fig. 6b) is likely not driven by snow cover variations, since this region is essentially always snow covered during DJF. Other processes, such as remote thermodynamic influences from decadal-scale variability in Arctic sea ice, may be important for this region [not shown, but see related results in Deser et al. (2016)].

The largest residual uncertainty in ToE over Europe in DJF is located in northern Scandinavia (Fig. 6b). Sea ice concentration in the Barents Sea can influence northern Europe SAT in winter through dynamic but also direct thermodynamic processes (Lehner et al. 2013; Screen et al. 2014; Mori et al. 2014; Sun et al. 2015; Sorokina et al. 2016) and could account in part for this

residual uncertainty. To investigate the influence of sea ice concentration (SIC) in the Barents Sea on dynamic and dynamically adjusted SAT, we construct time series of DJF SIC averaged over the Barents Sea for each ensemble member, and correlate them with dynamic and dynamically adjusted SAT at each grid cell over Europe, again removing the CESM-LE ensemble mean from each variable beforehand and concatenating the time series before correlating. Although weaker than that found for snow cover over North America, the Barents Sea SIC correlation is significantly negative over northern Scandinavia, and, importantly, the correlations with dynamically adjusted SAT (Fig. 7d) at lags of  $-3$  months, and also  $-2$  and  $-1$  month (i.e., SIC leading SAT), tend to be stronger than the ones with dynamic SAT (Fig. 7c). This suggests that Barents SIC exerts a thermodynamic influence on SAT. Also, the pattern resembles that of the residual uncertainty range of ToE (Fig. 7b), suggesting that residual ToE uncertainty over Europe in DJF may result in part from the thermodynamic influence of internal variability of Barents Sea SIC.

In JJA, the largest residual uncertainties in ToE occur in the U.S. southern Great Plains and the eastern Mediterranean region, regions identified with strong land–atmosphere coupling (Seneviratne et al. 2013). Hence, thermodynamic soil moisture–driven SAT variability across the CESM-LE ensemble members could in part be responsible for the remaining uncertainty there. We can illustrate this relationship by correlating at each grid box the JJA time series of column-integrated soil moisture with that of dynamic and dynamically adjusted SAT, after removing the CESM-LE ensemble mean from all variables and concatenating the time series (Fig. 8). The results show that dynamically adjusted SAT is significantly negatively correlated with soil moisture in the U.S. Great Plains and the Mediterranean region at all lags (Fig. 8b), and importantly, shows larger amplitude correlations than those between soil moisture and dynamic SAT when soil moisture leads (i.e., lags from  $-3$  to  $-1$  month), supporting the hypothesis that thermodynamic effects of summer soil moisture, likely with a memory from winter–spring precipitation totals and corresponding summer latent heat flux, are important contributors of residual uncertainty in ToE.

As a consequence of these analyses, one would expect the uncertainty range in ToE over North America and Europe to be reduced when the wintertime influences of snow cover and sea ice, and the summertime effects of soil moisture, are accounted for. To test this at the example of  $-1$ -month lag, we multiply the time series of snow cover, sea ice, or soil moisture with the respective

regression coefficients derived from  $-1$ -month-lag regressions (only for grid points with significant regression coefficients), yielding an empirical estimate of the individual linear contributions of snow cover, sea ice, and soil moisture to the dynamically adjusted SAT. After removal of this contribution, ToE and its 5%–95% range are calculated again. The reason we perform this calculation for only those grid boxes with significant regression coefficients is that the regression involving a bounded quantity such as snow cover can yield unrealistically high or low coefficients, which would result in overestimation of their contribution to SAT variability. Note that this regression approach does not account for possible collinearity of the dynamic and the dynamically adjusted SAT variability that would reflect the two-way feedback between, for example, snow cover and SAT but rather assumes a one-way influence of the thermodynamic driver onto the residual SAT variability. Thus, such an empirical estimate should be interpreted as a possible upper bound on the contribution of snow cover, sea ice, or soil moisture to the dynamically adjusted SAT variability, and hence residual ToE uncertainty. Further dissecting this collinearity and more comprehensively quantifying the feedback contributions from thermodynamic drivers are beyond the scope of this study.

The effect that this additional removal of thermodynamic SAT variability has on the 5%–95% range of ToE is illustrated in Fig. 9. In DJF over North America, removing the influence of snow cover yields an additional reduction in ToE uncertainty over the midsection of North America, with largest values (up to 8 yr) to the west of the Great Lakes and smaller reductions of 1–4 yr elsewhere (Fig. 8a). Over Europe in DJF, where we only focus on the influence of Barents Sea sea ice, there is a significant reduction in uncertainty ( $>10$  yr) over northern Scandinavia but none elsewhere (Fig. 8a). In JJA, the main regions of residual ToE uncertainty over North America (central Canada and the U.S. Great Plains) and Europe (the eastern Mediterranean region) see reductions in ToE uncertainty of approximately 1–5 yr after linearly removing the local thermodynamic influence of soil moisture (Fig. 8b). Overall, our analysis shows that these individual thermodynamic drivers of SAT variability in both winter and summer might account for part of the residual ToE uncertainty, but that additional factors need to be taken into account for a full explanation.

#### 4. Discussion

The utility of dynamical adjustment methods for understanding the role of the atmospheric circulation on

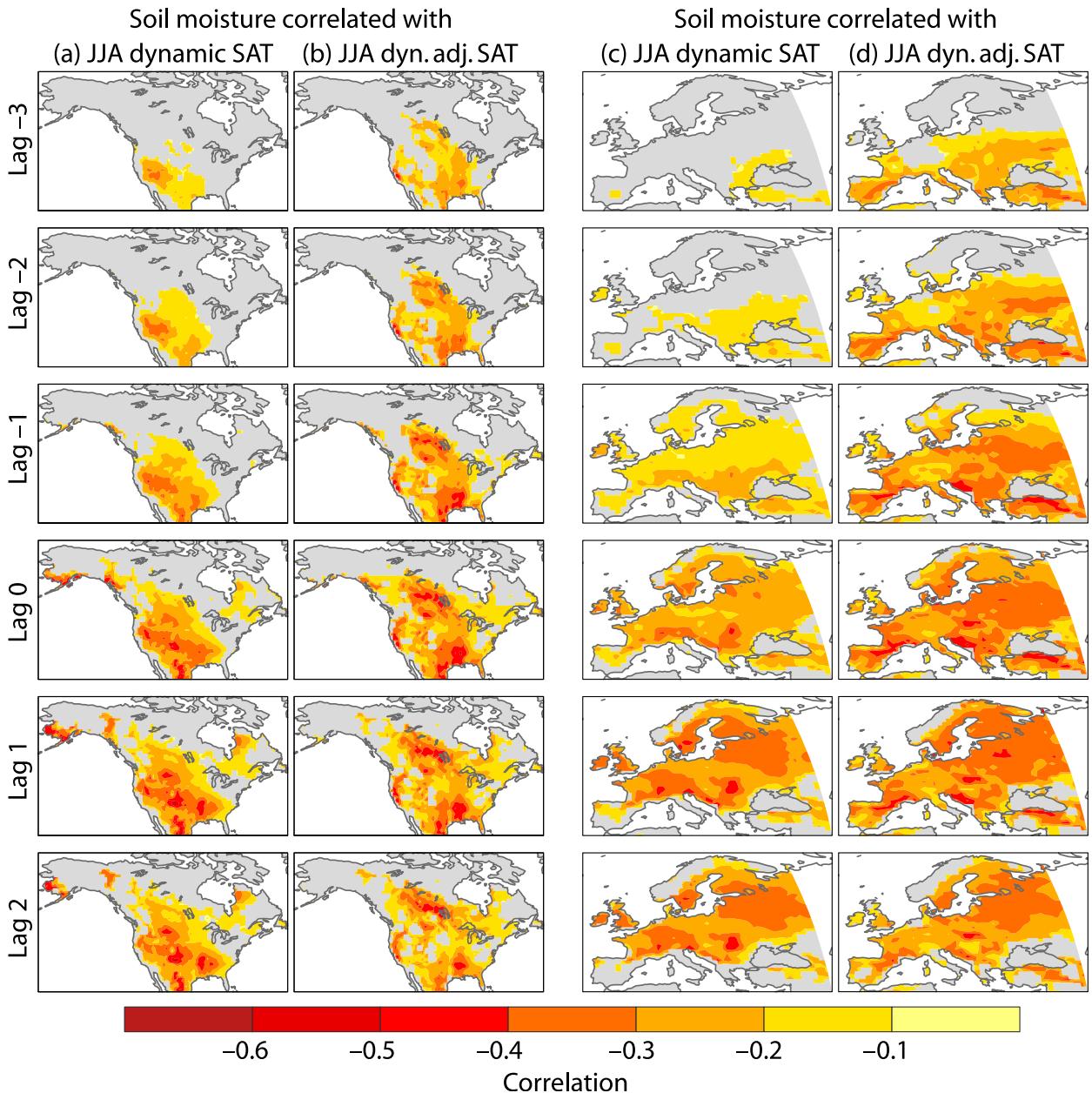


FIG. 8. As in Fig. 7, but for soil moisture and JJA.

SAT variability and trends has been demonstrated in recent studies (Wallace et al. 2012; Smoliak et al. 2015; Deser et al. 2016). Here, we applied such a method to observations and CESM simulations to study the role of atmospheric circulation variability in estimates of ToE, as an example of a commonly used climate change detection metric. Consistent with previous studies, we find that atmospheric dynamics has a particularly large influence in winter on terrestrial SAT variability over the mid-to-high latitudes, with the result of advancing ToE by multiple decades, or delaying it by a decade,

depending on whether it contributes to warming or cooling. We find that when internal atmospheric circulation variability is not accounted for in observationally based estimates of ToE, even conservative metrics [such as those adopted in Mahlstein et al. (2012)] might show a nonpermanent ToE in certain regions. We also find that in some regions there exist notable differences in ToE between observational datasets. Using the CESM-LE, we showed that the uncertainty in detecting a robust emergence in observations is almost as large, in terms of number of years but also in terms of area emerged, as the

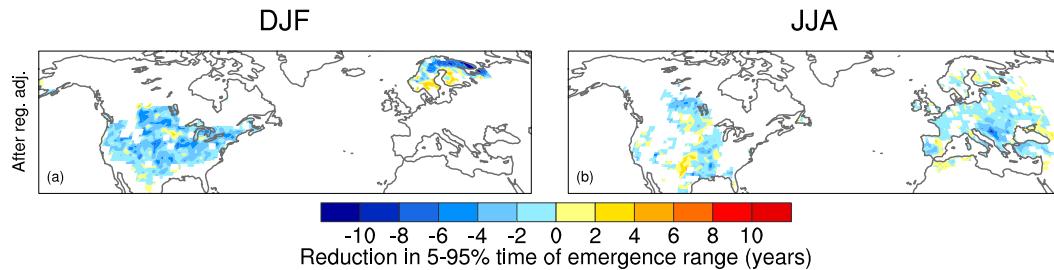


FIG. 9. Reduction in 5%–95% range of ToE in CESM-LE for (a) DJF and (b) JJA after removing an estimate of thermodynamic influences from dynamically adjusted ToE (see text for details).

advancement of ToE due to removing the influence of atmospheric circulation. However, this result also shows that dynamical adjustment can partially compensate for this unavoidable uncertainty in observed ToE arising from not knowing the future. Also, in some places the reduction in uncertainty from dynamical adjustment ( $>20$  yr) is almost as large as model structural uncertainty discussed in other studies ( $>30$  yr; Hawkins and Sutton 2012). Note that by using a large ensemble with a single model, we did not sample model structural uncertainty in ToE here. This would be a crucial part of a more comprehensive investigation of ToE and has to be considered in the interpretation of the results here.

Further, one needs to carefully consider what exactly “emergence” means, since ToE has not necessarily been used synonymously with the detection of anthropogenically forced warming in observational studies. Using the ensemble mean of CESM-LE to estimate the forced response to increased GHGs, we demonstrated that robust ToE of forced warming for a given location in North America or Europe can occur as late as the second half of the twenty-first century. Indeed, depending on the models and methods used, robust ToE may occur even later (Hawkins et al. 2014). This seems to contradict studies suggesting that detection and attribution of anthropogenic warming has already occurred for some continental averages (Bindoff et al. 2013). However, our aggregated metric of the fraction of land with emerged warming indicates that approximately 75% of North America and 95% of Europe have emerged from their 1920–49 reference climate, highlighting that spatially aggregated metrics show higher signal-to-noise ratios of anthropogenic climate change than local measures (Fischer and Knutti 2014). While this reconciles findings from traditional detection and attribution studies with results from ToE studies such as the one here, it also illustrates how different climate change risk assessments may require different approaches. For example, an insurance company with a diverse portfolio might be interested in spatially aggregated risk changes (Mills 2005), while the manager of a

protected ecosystem might have more localized concerns (McLeod et al. 2009).

Finally, we have explored the role of thermodynamic processes in determining the remaining uncertainty in ToE after the influence of internal atmospheric circulation variability has been removed. While our results are suggestive of a thermodynamic influence from snow cover, sea ice, and soil moisture on ToE, targeted model sensitivity experiments with controlled lower boundary conditions are needed to properly determine the underlying cause of the spread in future projections of SAT and ToE (Seneviratne et al. 2013). This is particularly critical for regions where a strong model sensitivity exists for surface–atmosphere coupling (Boé and Terray 2014).

## 5. Summary

This study has investigated time of emergence (ToE) of anthropogenic warming over North America and Europe in observations and a 40-member ensemble of historical and future simulations with CESM. In addition to quantifying for the first time the role of internal atmospheric circulation variability on observed ToE, we have highlighted the utility of the CESM-LE for exploring various factors influencing the determination of ToE and its uncertainty. In particular, knowledge of the anthropogenically forced response in the CESM-LE (given by the ensemble mean) allowed us to explore uncertainties associated with various ways of estimating ToE in cases where the forced response has to be estimated empirically (such as for the single “realization” of the real world). We have further demonstrated the utility of dynamical adjustment for reducing uncertainty in ToE estimates and in revealing the anthropogenically forced response in both nature and the CESM-LE. Finally, we have investigated ToE and its uncertainty on both local scales and from an aggregated perspective in terms of the land fraction emerged.

The main findings are summarized as follows. For observations in winter, most of Canada emerged in the

1990s whereas most of the United States has not yet emerged by 2010. In summer, large parts of North America are emerged by 2010 and as early as the 1980s over western Canada and Florida. In Europe, a considerable region is emerged by 2010 in both seasons. In winter, we find that atmospheric circulation advanced ToE over Canada by one decade and delayed it over the United States by two decades. After removing the influence of atmospheric circulation, most of North America showed emergence between the 1990s and 2000s. Over Europe in winter, the difference between unadjusted and dynamically adjusted ToE indicates that the influence of atmospheric circulation in determining observed ToE has been less compared to North America. In summer, both continents show widespread emergence in the early 1990s with generally smaller influence of atmospheric circulation on ToE than in winter. Along with these findings, the land fraction for which human-induced warming has emerged by 2010 increased by approximately 10%–30% after dynamical adjustment, depending on season and region.

In a longer-term perspective based on the CESM-LE simulations, all areas in North America and Europe in both winter and summer show robust emergence of anthropogenic warming by approximately the 2040s–2060s under RCP8.5. Similar to observations, removing the influence of atmospheric circulation in CESM-LE tends to advance ToE by about two decades in winter and about one in summer, leading to a median emergence around the 2020s–2040s.

Accounting for circulation variability via dynamical adjustment also reduces the uncertainty in the range of ToE, especially in winter over the main regions of jet stream variability, according to the CESM-LE. Residual uncertainty beyond that explained by atmospheric circulation variability may be due to decadal-scale variations in snow cover and sea ice concentration in winter. In particular, snow cover variability was found to explain parts of the residual uncertainty over the interior United States and Canada, while Barents Sea sea ice variability explains part of the residual uncertainty over northern Europe.

In summer, the reduced importance of the large-scale atmospheric circulation for temperature variability results in dynamical adjustment having only a small effect on ToE. Similarly, the uncertainty in ToE is generally smaller in summer than in winter, with the largest residual uncertainty coinciding with areas of strong land–atmosphere coupling that, according to CESM-LE, are associated with thermodynamic influences from soil moisture variability.

Finally, we have quantified the influence of potential model biases on the model-based ToE. For most

regions, CESM-LE shows SAT variability in agreement with observations. There are a few regions where CESM-LE's SAT variability is clearly outside of observations, with overestimated variability over eastern Europe and the western United States in winter and underestimated variability at the U.S. East Coast in summer. However, it is the slight but widespread overestimation of SAT variability in summer over North America (and to a lesser extent over Europe) that significantly affects the model-based estimate of ToE and leads to an underestimated summer ToE in the model compared to observations: 22% less land fraction emerged by 2010 over North America for raw data, and 13% less for dynamically adjusted data (less than 5% over Europe for both cases). Hence, CESM might underestimate how early the forced warming signal in summer emerges in observations.

While this study has not investigated the sensitivity of the results to model structural differences or choice of emissions scenario, our results indicate that such sensitivities should be addressed within the framework of dynamical adjustment. In particular, accounting for the effects of internal atmospheric circulation variability on ToE of anthropogenic warming provides an opportunity to increase signal-to-noise ratios for detection and attribution of forced climate change and enables assessment of the relative contributions of dynamic and thermodynamic processes. Our approach can be applied to other detection and attribution methods, such as fingerprinting, or to other impact targets, for example forced changes in heat waves, precipitation, and other facets of climate change.

*Acknowledgments.* We are very grateful to Anna Merrifield and Reto Knutti for helpful discussion and thank three anonymous reviewers for constructive feedback. We also acknowledge the efforts of all those who contributed to producing the CESM-LE. The National Center for Atmospheric Research is sponsored by the National Science Foundation. F. L. is supported by an Early Postdoc Mobility fellowship from the Swiss National Science Foundation and a Postdoc Applying Climate Expertise (PACE) fellowship cosponsored by NOAA and the Bureau of Reclamation.

## REFERENCES

- Bador, M., L. Terray, and J. Boé, 2016: Emergence of human influence on summer record-breaking temperatures over Europe. *Geophys. Res. Lett.*, **43**, 404–412, doi:10.1002/2015GL066560.
- Beaumont, L. J., A. Pitman, S. Perkins, N. E. Zimmermann, N. G. Yoccoz, and W. Thuiller, 2011: Impacts of climate change on the world's most exceptional ecoregions. *Proc. Natl. Acad. Sci. USA*, **108**, 2306–2311, doi:10.1073/pnas.1007217108.

- Bindoff, N., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952.
- Boé, J., and L. Terray, 2014: Land–sea contrast, soil–atmosphere and cloud–temperature interactions: Interplays and roles in future summer European climate change. *Climate Dyn.*, **42**, 683–699, doi:10.1007/s00382-013-1868-8.
- Booth, J. F., L. Thompson, J. Patoux, and K. A. Kelly, 2012: Sensitivity of midlatitude storm intensification to perturbations in the sea surface temperature near the Gulf Stream. *Mon. Wea. Rev.*, **140**, 1241–1256, doi:10.1175/MWR-D-11-00195.1.
- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, doi:10.1002/qj.776.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, doi:10.1007/s00382-010-0977-x.
- , L. Terray, and A. S. Phillips, 2016: 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, doi:10.1175/JCLI-D-15-0304.1.
- Fischer, E. M., and R. Knutti, 2014: Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.*, **41**, 547–554, doi:10.1002/2013GL058499.
- , U. Beyerle, and R. Knutti, 2013: Robust spatially aggregated projections of climate extremes. *Nat. Climate Change*, **3**, 1033–1038, doi:10.1038/nclimate2051.
- Giorgi, F., and X. Bi, 2009: Time of emergence (TOE) of GHG-forced precipitation change hot-spots. *Geophys. Res. Lett.*, **36**, L06709, doi:10.1029/2009GL037593.
- Harrington, L. J., D. J. Frame, E. M. Fischer, E. Hawkins, M. Joshi, and C. D. Jones, 2016: Poorest countries experience earlier anthropogenic emergence of daily temperature extremes. *Environ. Res. Lett.*, **11**, 055007, doi:10.1088/1748-9326/11/5/055007.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, doi:10.1175/2009BAMS2607.1.
- , and —, 2012: Time of emergence of climate signals. *Geophys. Res. Lett.*, **39**, L01702, doi:10.1029/2011GL050087.
- , and Coauthors, 2014: Uncertainties in the timing of unprecedented climates. *Nature*, **511**, E3–E5, doi:10.1038/nature13523.
- , R. S. Smith, J. M. Gregory, and D. A. Stainforth, 2016: Irreducible uncertainty in near-term climate projections. *Climate Dyn.*, **46**, 3807–3819, doi:10.1007/s00382-015-2806-8.
- Hurrell, J. W., 1996: Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. *Geophys. Res. Lett.*, **23**, 665–668, doi:10.1029/96GL00459.
- IPCC, 2013: Summary for policymakers. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 3–29.
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, doi:10.1175/BAMS-D-13-00255.1.
- Keller, K. M., F. Joos, and C. C. Raible, 2014: Time of emergence of trends in ocean biogeochemistry. *Biogeosciences*, **11**, 3647–3659, doi:10.5194/bg-11-3647-2014.
- King, A. D., and Coauthors, 2015: The timing of anthropogenic emergence in simulated climate extremes. *Environ. Res. Lett.*, **10**, 094015, doi:10.1088/1748-9326/10/9/094015.
- , M. T. Black, S.-K. Min, E. M. Fischer, D. M. Mitchell, L. J. Harrington, and S. E. Perkins-Kirkpatrick, 2016: Emergence of heat extremes attributable to anthropogenic influences. *Geophys. Res. Lett.*, **43**, 3438–3443, doi:10.1002/2015GL067448.
- Krueger, O., F. Schenk, F. Feser, and R. Weisse, 2013: Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. *J. Climate*, **26**, 868–874, doi:10.1175/JCLI-D-12-00309.1.
- Lehner, F., and T. F. Stocker, 2015: From local perception to global perspective. *Nat. Climate Change*, **5**, 731–734, doi:10.1038/nclimate2660.
- , A. Born, C. C. Raible, and T. F. Stocker, 2013: Amplified inception of European Little Ice Age by sea ice–ocean–atmosphere feedbacks. *J. Climate*, **26**, 7586–7602, doi:10.1175/JCLI-D-12-00690.1.
- , C. Deser, and B. M. Sanderson, 2017: Future risk of record-breaking summer temperatures and its mitigation. *Climatic Change*, doi:10.1007/s10584-016-1616-2, in press.
- Mahlstein, I., R. Knutti, S. Solomon, and R. W. Portmann, 2011: Early onset of significant local warming in low latitude countries. *Environ. Res. Lett.*, **6**, 034009, doi:10.1088/1748-9326/6/3/034009.
- , G. Hegerl, and S. Solomon, 2012: Emerging local warming signals in observational data. *Geophys. Res. Lett.*, **39**, L21711, doi:10.1029/2012GL053952.
- McLeod, E., R. Salm, A. Green, and J. Almany, 2009: Designing marine protected area networks to address the impacts of climate change. *Front. Ecol. Environ.*, **7**, 362–370, doi:10.1890/070211.
- Merrifield, A. L., and S. P. Xie, 2016: Summer U.S. surface air temperature variability: Controlling factors and AMIP simulation biases. *J. Climate*, **29**, 5123–5139, doi:10.1175/JCLI-D-15-0705.1.
- Mills, E., 2005: Insurance in a climate of change. *Science*, **309**, 1040–1044, doi:10.1126/science.1112121.
- Mori, M., M. Watanabe, H. Shioyama, J. Inoue, and M. Kimoto, 2014: Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nat. Geosci.*, **7**, 869–873, doi:10.1038/ngeo2277.
- Raible, C. C., F. Lehner, J. F. González-Rouco, and L. Fernández-Donado, 2014: Changing correlation structures of the Northern Hemisphere atmospheric circulation from 1000 to 2100 AD. *Climate Past*, **10**, 537–550, doi:10.5194/cp-10-537-2014.
- Rockström, J., and Coauthors, 2009: A safe operating space for humanity. *Nature*, **461**, 472–475, doi:10.1038/461472a.
- Rohde, R., and Coauthors, 2013: Berkeley Earth temperature averaging process. *Geoinfo. Geostat. Overview*, **1** (2), doi:10.4172/2327-4581.1000103.
- Scherer, M., and N. S. Diffenbaugh, 2014: Transient twenty-first century changes in daily-scale temperature extremes in the United States. *Climate Dyn.*, **42**, 1383–1404, doi:10.1007/s00382-013-1829-2.
- Scholze, M., W. Knorr, N. W. Arnell, and I. C. Prentice, 2006: A climate-change risk analysis for world ecosystems. *Proc. Natl. Acad. Sci. USA*, **103**, 13 116–13 120, doi:10.1073/pnas.0601816103.
- Screen, J. A., C. Deser, I. Simmonds, and R. Tomas, 2014: Atmospheric impacts of Arctic sea-ice loss, 1979–2009: Separating

- forced change from atmospheric internal variability. *Climate Dyn.*, **43**, 333–344, doi:10.1007/s00382-013-1830-9.
- Seneviratne, S. I., and Coauthors, 2013: Impact of soil moisture–climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. *Geophys. Res. Lett.*, **40**, 5212–5217, doi:10.1002/grl.50956.
- Shaw, T. A., and Coauthors, 2016: Storm track processes and the opposing influences of climate change. *Nat. Geosci.*, **9**, 656–664, doi:10.1038/ngeo2783.
- Smoliak, B. V., J. M. Wallace, P. Lin, and Q. Fu, 2015: Dynamical adjustment of the Northern Hemisphere surface air temperature field: Methodology and application to observations. *J. Climate*, **28**, 1613–1629, doi:10.1175/JCLI-D-14-00111.1.
- Sorokina, S. A., C. Li, J. J. Wettstein, and N. G. Kvamstø, 2016: Observed atmospheric coupling between Barents Sea ice and the warm-Arctic cold-Siberian anomaly pattern. *J. Climate*, **29**, 495–511, doi:10.1175/JCLI-D-15-0046.1.
- Strong, C., and R. E. Davis, 2007: Winter jet stream trends over the Northern Hemisphere. *Quart. J. Roy. Meteor. Soc.*, **133**, 2109–2115, doi:10.1002/qj.171.
- Sun, L., C. Deser, and R. A. Tomas, 2015: Mechanisms of stratospheric and tropospheric circulation response to projected Arctic sea ice loss. *J. Climate*, **28**, 7824–7845, doi:10.1175/JCLI-D-15-0169.1.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Thompson, D. W. J., J. M. Wallace, P. D. Jones, and J. J. Kennedy, 2009: Identifying signatures of natural climate variability in time series of global-mean surface temperature: Methodology and insights. *J. Climate*, **22**, 6120–6141, doi:10.1175/2009JCLI3089.1.
- Vose, R. S., and Coauthors, 2012: NOAA’s merged land–ocean surface temperature analysis. *Bull. Amer. Meteor. Soc.*, **93**, 1677–1685, doi:10.1175/BAMS-D-11-00241.1.
- Wallace, J. M., Q. Fu, B. V. Smoliak, P. Lin, and C. M. Johanson, 2012: Simulated versus observed patterns of warming over the extratropical Northern Hemisphere continents during the cold season. *Proc. Natl. Acad. Sci. USA*, **109**, 14 337–14 342, doi:10.1073/pnas.1204875109.
- Willmott, C. J., and K. Matsuura, 1995: Smart interpolation of annually averaged air temperature in the United States. *J. Appl. Meteor.*, **34**, 2577–2586, doi:10.1175/1520-0450(1995)034<2577:SIOAAA>2.0.CO;2.