Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1029/2018JD029066

Key Points:
- Internal variability renders the detection of future changes in the seasonal cycle of surface temperature difficult over many regions
- Europe, North Africa, and Siberia, however, exhibit remarkably robust and easily detectable changes across climate model ensembles
- The robust changes over the three regions are primarily driven by changes in surface longwave and turbulent heat fluxes

Supporting Information:
- Supporting Information S1

Correspondence to:
V. Yettella,
vineel.yettella@colorado.edu

Citation:
Yettella, V., & England, M. R. (2018). The role of internal variability in twenty-first-century projections of the seasonal cycle of Northern Hemisphere surface temperature. *Journal of Geophysical Research: Atmospheres*, *123*. https://doi.org/10.1029/2018JD029066

Received 25 MAY 2018
Accepted 11 NOV 2018
Accepted article online 21 NOV 2018

The Role of Internal Variability in Twenty-First-Century Projections of the Seasonal Cycle of Northern Hemisphere Surface Temperature

Vineel Yettella[1,2] **and Mark R. England**[3]

[1]Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, Colorado, USA, [2]Cooperative Institute for Research in Environmental Sciences, Boulder, Colorado, USA, [3]Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, USA

Abstract The seasonal cycle is fundamental to the Earth's climate system, accounting for the vast majority of temperature variance. Understanding how the seasonal cycle will change in the future, and by when, is a key question with important implications. Here a 40-member initial condition climate model ensemble is used to investigate the influence of internal variability on the detection of changes in the amplitude and timing of the seasonal cycle of surface temperature over Northern Hemisphere land in response to increasing greenhouse gases. Internal variability renders the detection of these changes challenging; even by the mid-twenty-first century, small ensembles will be insufficient to separate the forced signals from internal variability over many continental regions in the Northern Hemisphere. Despite this, projected changes over Europe, North Africa, and Siberia are large and easily detectable, even in a single member. Specifically, amplitude increases over Europe and North Africa while it decreases over Siberia. On the other hand, the timing of the seasonal cycle is delayed over all three regions. It is found that these changes are remarkably robust across model ensembles from the Coupled Model Intercomparison Project phase 5 archive. To understand the mechanisms underlying these robust changes, a simple energy balance model is used to partition changes into contributions arising from changes in the physical parameters that control the seasonal cycle. It is found that future changes in the seasonal cycle over the three regions are most strongly controlled by changes in surface longwave and turbulent heat fluxes.

1. Introduction

The seasonal cycle of Earth's surface temperature is a fundamental aspect of climate variability; the annual frequency band contains more temperature variance than any of the well-known modes of circulation variability (e.g., El Niño–Southern Oscillation, Bjerknes (1969); Northern Annular Mode, Thompson and Wallace (2000); Pacific North American pattern, Wallace and Gutzler (1981)). Beyond being a fundamental feature of the climate system, the seasonal cycle in temperature has a number of first-order ecological and societal effects. For example, it has a modulating effect on biological cycles in plants (Schwartz et al., 2006), determines animal migration patterns (Carey, 2009), influences agricultural development (Lambert, 1971), can affect the estimation of climate trends and variability (Qian et al., 2011), and has even been argued to have played a crucial role in the growth and collapse of civilizations (Patterson et al., 2010).

Anthropogenic emissions, which are projected to increase for the foreseeable future (IPCC, 2013), are rapidly changing the Earth's climate, both in the mean (Hawkins & Sutton, 2012; Madden & Ramanathan, 1980) and in the variability about the mean (Huntingford et al., 2013; Yettella et al., 2018). Given the high societal and ecological relevance of the seasonal cycle of surface temperature, it is crucial to understand how increasing anthropogenic emissions might change its climatological characteristics in the coming century. A notable previous study that investigated changes in the annual cycle in a warmer world is that of Dwyer et al. (2012), hereafter referred to as D12. The authors analyzed output from the Coupled Model Intercomparison Project phase 3 (CMIP3; Meehl et al., 2007) suite of models and found that, in the global mean, the models robustly projected a delay in the phase and a decrease in the amplitude of the seasonal cycle of surface temperature. Using a simple energy balance model, they further showed that the changes are largely driven by sea ice loss in a warming climate. Despite the importance of the seasonal cycle, its response to increasing greenhouse gases has been investigated by only a few other model studies (e.g.,

Lynch et al., 2016; Mann & Park, 1996). As such, several fundamental questions about the future seasonal cycle response remain unaddressed.

This paper addresses two specific research goals. Our first and main goal is to understand the influence of internal variability on the ability to detect forced changes in the twenty-first century seasonal cycle of surface temperature over Northern Hemisphere (NH) continental landmasses. Unforced internal variability that arises solely from modes inherent to the climate system can give rise to year-to-year variations (Ault et al., 2011) and even multidecadal trends (Cornes et al., 2017; Stine & Huybers, 2012) in the seasonal cycle adding uncertainty to future projections. Quantifying this uncertainty and assessing its impact on the detection of forced changes is essential for several obvious and important reasons. First, the detection of a statistically significant change is often the first step in attributing change to a particular cause, for example, increasing greenhouse gases (Bindoff et al. 2013). Furthermore, detection is essential for informing policy makers and stake holders to enable near-term risk mitigation and long-term adaptation (IPCC, 2013). Finally, given that recent trends in the observed seasonal cycle have been partly attributed to human influence (Qian & Zhang, 2015; Stine et al., 2009), it is easy to envisage that anthropogenically forced changes in the seasonal cycle will become even more significant in the future. It is therefore vital to understand how internal variability will influence the detection of these changes. To the best of our knowledge previous studies have not investigated the role of internal variability on seasonal cycle projections at either the global or regional scales. A recent exception is Labe et al. (2017), who use a large ensemble of climate model simulations to investigate the role of internal variability in future projections of spring onsets over the continental United States and demonstrate a large projected increase in the likelihood of early springs.

We address the influence of internal variability on the detection of forced changes in the seasonal cycle using a coupled climate model ensemble. Using the ensemble, we quantify the uncertainty arising from internal variability by addressing two specific questions: (1) how soon can we detect forced changes in the seasonal cycle for a given number of ensemble members? and (2) how many ensemble members would be required to detect forced changes at a future point in time?

Our second goal pertains to addressing the mechanisms underlying future changes in the seasonal cycle, leveraging models that participated in the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor et al. 2012). For this purpose, we utilize the energy balance model developed by D12, and partition future changes in the seasonal cycle into contributions arising from changes in the physical parameters that control the seasonal cycle. Specifically, we quantify the role of effective heat capacity and surface energy fluxes in driving future changes across different CMIP5 models. Using the energy balance model, we also briefly explore biases in the simulation of the historical seasonal cycle within the CMIP5 models.

We organize the paper as follows: In section 2, we describe our methods and data. In section 3, we present the climatological structure of the seasonal cycle in the CESM-LE. In section 4, we present projected changes and assess the influence of internal variability on the detection of projected changes in the CESM-LE and in several CMIP5 ensembles. We find that internal variability exerts a profound influence on forced changes in seasonality over much of the NH. We also reveal areas where forced changes are strong compared to internal variability and are therefore easily detectable, even in a single realization of the climate system, and further show that these changes are robust across multiple ensembles from the CMIP5 archive. In section 5, we utilize the energy balance model to explore the mechanisms underlying future changes in the seasonal cycle, as well as historical biases in CMIP5 models. Finally, in section 6, we offer some concluding remarks.

2. Methods and Data

2.1. Definition of the Seasonal Cycle

Two methods have primarily been used in the literature to define the seasonal cycle of surface temperature (Stine et al., 2009). The first is based on the time of the year when temperatures reach a level of interest, a useful method when one is interested in phenomena that depend on temperature crossing a threshold. Threshold-based definitions confound the description of systematic changes in the seasonal cycle when a warming signal is present in the annual mean (see Stine et al., 2009, Figure S1). The alternative method of describing the seasonal cycle is based on a decomposition of the yearly temperature time series, whose annual mean has been removed, into orthogonal sinusoids or harmonics. As such, this method enables a

Table 1
CMIP5 Historical and RCP 8.5 Data Used in This Study

	Model	Ensemble Member
1	ACCESS1–0	r1i1p1
2	ACCESS1–3	r1i1p1
3	bcc-csm1–1	r1i1p1
4	bcc-csm1–1-m	r1i1p1
5	BNU-ESM	r1i1p1
6	CanESM2	r1i1p1
7	CCSM4	r1i1p1
8	CESM1-BGC	r1i1p1
9	CESM1-CAM5	r1i1p1
10	CNRM-CM5	r1i1p1
11	FGOALS-g2	r1i1p1
12	GFDL-CM3	r1i1p1
13	GFDL-ESM 2G	r1i1p1
14	GFDL-ESM 2 M	r1i1p1
15	HadGEM2-CC	r1i1p1
16	HadGEM2-ES	r1i1p1
17	inmcm4	r1i1p1
18	MIROC5	r1i1p1
19	MIROC-ESM	r1i1p1
20	MIROC-ESM-CHEM	r1i1p1
21	NorESM1-M	r1i1p1
22	NorESM1-ME	r1i1p1

description of the structure of the seasonal cycle that is unaffected by annual mean warming. Since the focus of this paper is on the detectability of systematic changes of the seasonal cycle in a warming climate, without regard to any threshold-based phenomena, we adopt the second definition.

The harmonic structure of the seasonal cycle exhibits a notable geographic dependence. For example, as noted by D12, the Sun passes overhead twice in the tropics, and as a result, the tropical seasonal cycle of insolation and in turn that of temperature has a strong semiannual character. In contrast, the extratropical seasonal cycle is predominantly annual in nature. The annual and semiannual sinusoids together capture the vast majority of the variance in the seasonal cycle of surface temperature virtually everywhere on the globe. In this paper, we limit our focus to the annual sinusoid and to the regions where >85% of the variance in the seasonal cycle is explained by the annual sinusoid. Our choice therefore excludes large portions of the tropics. We further limit the scope of this study to the NH.

Adopting the methods of Stine et al. (2009), we compute the annual sinusoid (*annual cycle* hereafter) for year t_0 using the Fourier transform:

$$Y_x(t_0) = \frac{2}{12}\Sigma_{t=0.5}^{11.5} e^{2\pi i t/12} x(t+t_0) \quad (1)$$

where $x(t + t_0)$, $t = 0.5, \ldots, 11.5$ are the 12-monthly average values of either the surface (skin) temperature or solar insolation. Monthly average values resolve the annual cycle adequately for the purposes of this study (Mckinnon et al., 2013; Stine & Huybers, 2012). We compute the phase of the annual cycle as $\phi_x = \tan^{-1}(Im(Y_x)/\, Re\,(Y_x))$ and the amplitude as $A_x = \mid Y_x \mid$. To enable a standardized comparison of the timing of the annual cycle across different latitudes, we reference its phase, ϕ_T, and amplitude A_T, to the phase, ϕ_S, and amplitude, A_S, of the local solar insolation, and define gain $G_T = \frac{A_T}{A_S}$ and lag $\lambda_T = \phi_T - \phi_s$. We hereafter quantify changes in the annual cycle in terms of changes in its gain and lag (see methods of Stine et al. (2009)). We note that previous studies have used slightly varying methods to define the annual cycle. For example, Qian & Zhang (2015), rather than investigate gain, examine the amplitude in terms of the range of seasonal mean temperatures within a year; we expect our studies to be broadly comparable, although caution should be taken for results at lower latitudes where the correspondence of gain and amplitude is weaker.

2.2. Data

We assess the influence of internal variability using an initial condition ensemble: the CESM large ensemble (CESM-LE; Kay et al., 2015). The CESM-LE consists of 40 simulations of a single climate model (CESM1-CAM5; Hurrell et al., 2013) that are all run under the same forcing: historical (Lamarque et al., 2010) from 1920 to 2005 and Representative Concentration Pathway 8.5 (RCP 8.5; Meinshausen et al., 2011) from 2006 to 2100. The ensemble members are initialized identically at 1920 except for small differences (order of 10^{-14} K) in the air temperature field. Chaos leads to growth in these initial differences eventually leading to spread among the ensemble members. As such, each member represents an independent realization of the climate system and the spread represents uncertainty arising from internal variability alone. To assess the robustness of the CESM-LE projections, we analyze output from an initial-condition ensemble based on a different model: the CanESM2 (Canadian Earth System Model 2; Arora et al., 2011) large ensemble (Fyfe et al., 2017) that consists of 50 ensemble members and was run under the same forcing as the CESM-LE, albeit at a lower resolution of 5° × 5°. We also briefly examine the CESM medium ensemble (CESM-ME; Sanderson et al., 2018) which consists of 15 members that are run under RCP 4.5 (Thomson et al., 2011) conditions from 2006 to 2080 to understand the sensitivity of our results to the chosen emission scenario.

We explore the mechanisms underlying projected changes in the annual cycle in 22 CMIP5 historical and RCP 8.5 model simulations (Table 1; Taylor et al., 2012) using a simple energy balance model (section 5). This analysis requires temperature, shortwave flux fields at the surface, and solar insolation which are all obtained at

2.5° resolution from the CMIP5 Next Generation (CMIP5-ng) database. Monthly data from the first ensemble member (r1i1p1) of each CMIP5 model are utilized for the computation of the annual cycle. We note that all models used in this study prescribe solar insolation based on a reconstruction of total solar irradiance by Wang et al. (2005). Exceptions are the ACCESS1–0 and ACCESS1–3 models that used a reconstruction by Lean (2000).

In this study, we also explore model biases by comparing model results with observational data. We derive the observational gain and lag using the Berkeley Earth Surface Temperature (BEST; Rohde et al., 2013) data set available at 1° resolution and historical solar irradiance data based on the Wang et al. (2005) reconstruction. We bilinearly interpolate BEST data to the lower resolution model grids to facilitate comparison. Shortwave fluxes at the surface are obtained from National Centers for Environmental Prediction–Department of Energy Reanalysis II (Kanamitsu et al., 2002).

2.3. Detection of Forced Changes in the Annual Cycle

The primary goal of this study is to understand the influence of internal variability on the detection of future changes in the annual cycle; for this purpose, we use the statistical approach outlined in Deser et al. (2012). In each ensemble member of the CESM-LE, we quantify future amplitude and lag changes in terms of decadal epoch differences within that ensemble member. We estimate forced change as the ensemble mean epoch difference $\overline{X}$ and internal variability as the ensemble standard deviation of the epoch differences $\widehat{\sigma}$. We consider the forced change $\overline{X}$ to be statistically significant relative to $\widehat{\sigma}$ at 95% significance if the following condition, based on the standard error of the mean, is met:

$$\left|\frac{\overline{X}}{\widehat{\sigma}}\right| \geq \frac{2}{\sqrt{N-1}} \tag{2}$$

where N is the number of ensemble members. The minimum number of ensemble members (N_{min}) required to detect a significant forced epoch difference at 95% significance is computed as (Sardeshmukh et al., 2000)

$$N_{\text{min}} = \frac{8}{\left(\frac{\overline{X}}{\widehat{\sigma}}\right)^2} \tag{3}$$

3. Climatological Structure

We begin by presenting the climatological structure of the annual cycle in the CESM-LE in the NH over both land and ocean. Specifically, we look at the long-term (1960–2005) mean gain and lag within member 1 (Figures 1a and 1c) of the CESM-LE and the standard deviation across the ensemble members (Figures 1b and 1d). The standard deviations of both gain and lag (Figures 1b and 1d) are very small compared to the long-term means (Figures 1a and 1c), indicating little variability across the ensemble members. Two striking spatial features in the annual cycle response to solar insolation are observed. First, gain is generally larger over land compared to that over the oceans, while lag is larger over the oceans, reaching more than 75 days in the lower midlatitudes. These differences in the annual cycle response are consistent with the vastly different effective heat capacities of land and ocean: the larger effective heat capacity of the ocean damps the amplitude of the temperature response to the oscillatory forcing of solar insolation and delays its phase compared to that over land. This dependence of annual cycle characteristics on the effective heat capacity will be made more quantitative in section 5. Second, there is a notable zonal gradient in gain and lag, following the direction of the background circulation that carries the tempering influence of the oceans onto land. Specifically, gain generally increases as one moves west to east across the continents while lag exhibits the opposite tendency. A similar spatial structure is also seen in the variability of gain and lag (Figures 1b and 1d)—the variability of gain increases as one moves to the east across continents while the variability of lag decreases.

In Figure 2, we document the long-term mean gain and lag in observations (Figures 2a and 2d) and associated biases in the CESM-LE (Figures 2b and 2e). On comparing Figures 1a and 1c with Figures 2a and 2d

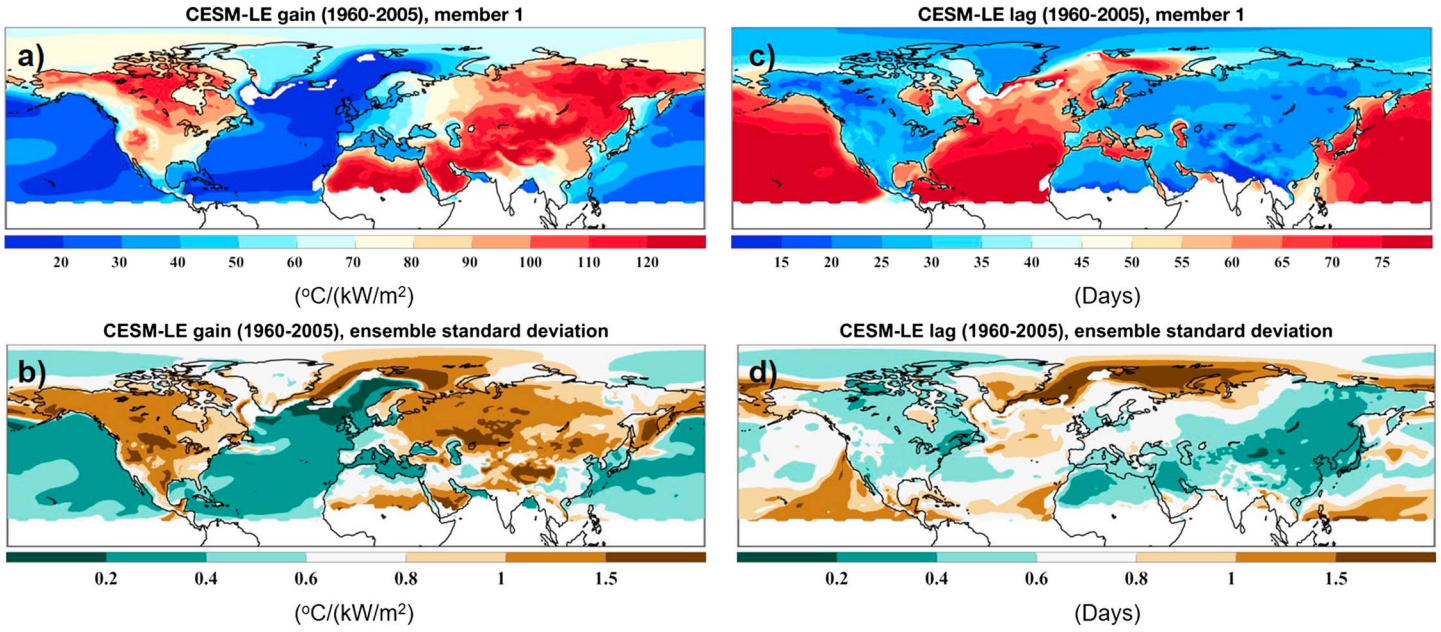

Figure 1. (a) The 46-year (1960–2005) mean gain in member 1 of the CESM-LE. (b) Standard deviation across CESM-LE members of 46-year (1960–2005) mean gain. (c) Same as in (a). (d) Same as in (b) but for lag. Regions where less than 85% of the variance in an average demeaned year (averaged over 1960–2005) is explained by the annual cycle, in one or more ensemble members, have been excluded.

and also with results from previous observational (Stine et al., 2009; Stine & Huybers, 2012) and model studies (D12), we find that the CESM-LE is able to reproduce the broad spatial structure of the annual cycle. However, after computing the fractional differences in gain and lag between the model and observations (Figures 2b and 2e), we find that the model exhibits large biases over many regions. For example, the CESM-LE overestimates gain by more than 30% over the Rocky Mountains and North Africa, and lag by more than 50% over Mexico. When aggregated over land, the CESM-LE underestimates lag by 8.5% while overestimating gain by 14% compared to BEST observations. Interestingly, the spatial pattern of biases seen in Figure 2b, most notably the biases over the sea ice edge, persist to some extent in the CMIP5 models (not shown) and also in the CMIP3-based results of D12.

While the mean state is important, it is crucial to assess model fidelity in simulating the variability of the historical annual cycle. To gain a qualitative understanding of how representative the simulated internal variability is of the real climate system, we show in Figures 2c and 2f the ratio of the interannual standard deviations of detrended gain and lag in member 1 of the CESM-LE to those in the BEST observations. Over most land regions, modeled variability is statistically indistinguishable from observed variability. Exceptions are the Tibetan Plateau, the Rocky Mountains, and Western Russia, where the modeled variability in gain is almost 1.5 to 2 times greater than observed variability and is statistically distinguishable at the 95% confidence level. It is important to note that if future internal variability over these regions is also overestimated by the CESM-LE, then the estimates of the time of detection of forced changes will be biased to later years and the number of ensemble members will be biased to a higher number (see equations (2) and (3)).

4. Future Changes and the Influence of Internal Variability

Because the impacts of changes in the seasonal cycle on society are largest over land, we restrict our analyses to the large continental landmasses. We examine changes in the gain (Figure 3a) and lag (Figure 3d). To reduce the noise of internal variability and bring out robust forced signals, we focus on the far-future (2091–2100) ensemble mean change relative to this past decade (2008–2017). By the end of the twenty-first century, the gain of the annual cycle increases over Europe, Northern Africa, vast portions of the United States, and Greenland. The increases seen over Europe are particularly striking with the amplitude difference

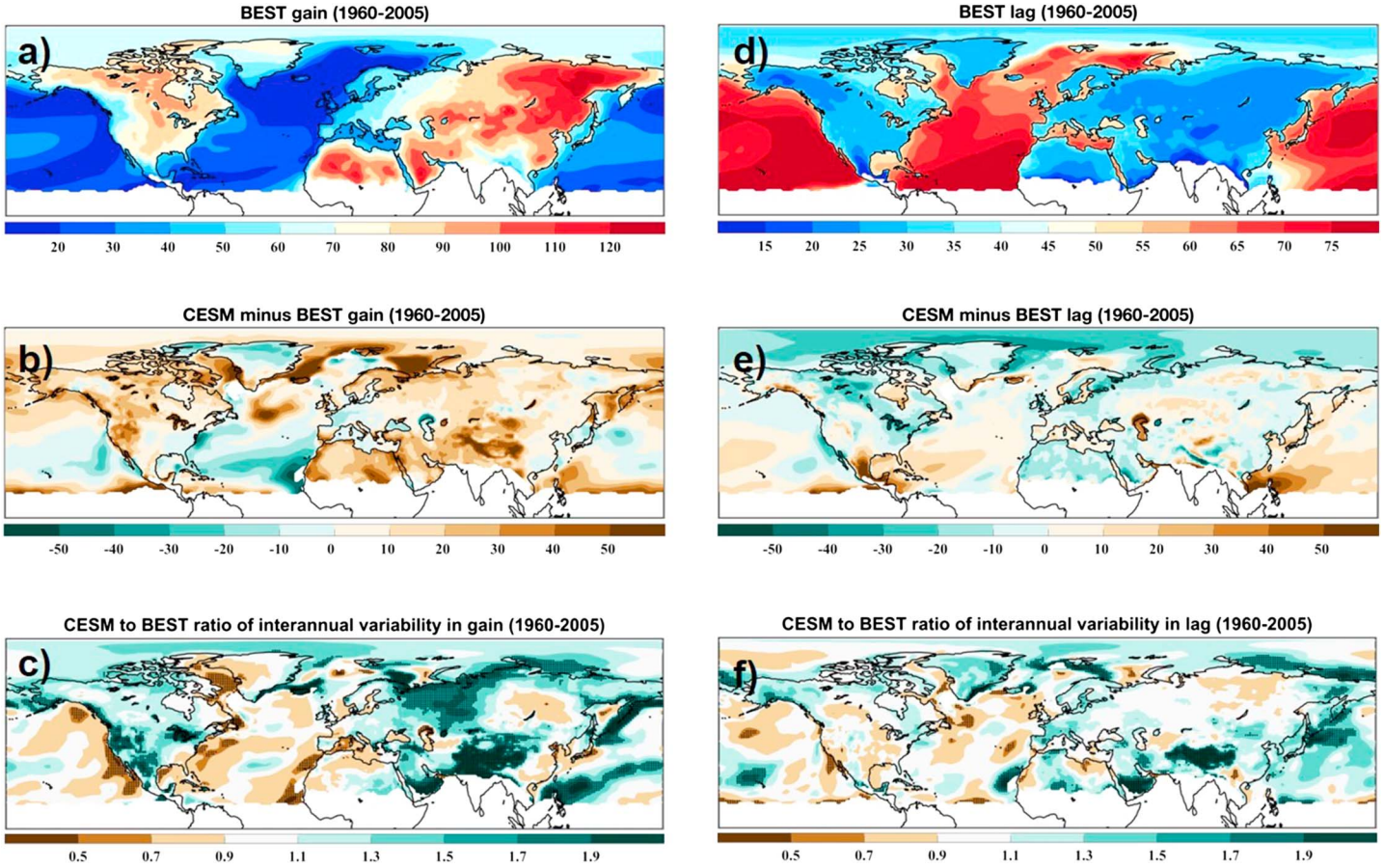

Figure 2. (a) The 46-year (1960–2005) mean gain in BEST observations. (b) The 46-year (1960–2005) mean gain percentage difference between ensemble mean of the CESM-LE and BEST observations. (c) Ratio of interannual standard deviation of gain (1960–2005) in member 1 of the CESM-LE to that in BEST observations. (d) Same as in (a). (e) Same as in (b). (f) Same as in (c) but for lag. Regions where less than 85% of the variance in an average demeaned year (averaged over 1960–2005) is explained by the annual cycle, in one or more ensemble members, or in observations, have been excluded. Stippling indicates regions where the interannual variance in member 1 of the CESM-LE is different from that in BEST observations according to a two-tailed *F* test at the 95% confidence level.

reaching than 12 °C kW m^2 (roughly 2.5 °C in terms of amplitude). There are also regions where the gain decreases, specifically, over Siberia, Alaska, and Canada. The largest reductions in gain, up to −10 °C kW m^2 (roughly −3 °C in terms of amplitude), occur over Siberia and Alaska. The spatial pattern of change in lag (Figure 3d) is very different to that of gain. Lag increases (i.e., the timing of the annual cycle is delayed) by almost two days over most of the NH. Exceptions, however, are lag reductions over the Iberian Peninsula, Mexico, and central Asia. Interestingly, this is in contrast with Stine & Huybers (2012), who suggest that the historical annual cycle has shifted toward earlier seasons. It is argued that anomalous atmospheric conditions were responsible for this shift; however, it is unclear that these changes in atmospheric conditions are a forced response to recent anthropogenic emissions. Therefore, the forced changes in lag seen in Figure 3 do not necessarily contradict these previous results.

We repeat the above analysis with the CanESM2, a different large ensemble consisting of 50 ensemble members (see section 2.2) and with 22 CMIP5 models. We find that the CanESM2 large ensemble and the CMIP5 multimodel ensemble project very similar gain (Figures 3b and 3c) and lag (Figures 3e and 3f) changes over most regions by the end of the twenty-first century. While the magnitude of change in the CMIP5 projections is smaller, likely due to averaging across models with compensating changes, the direction and spatial pattern of change is consistent with those in the CESM-LE and CanESM2 and with the CMIP3 results of D12. It is interesting to note that future changes in the amplitude seem to be a continuation of trends in the historical record (Qian & Zhang, 2015). The striking similarities

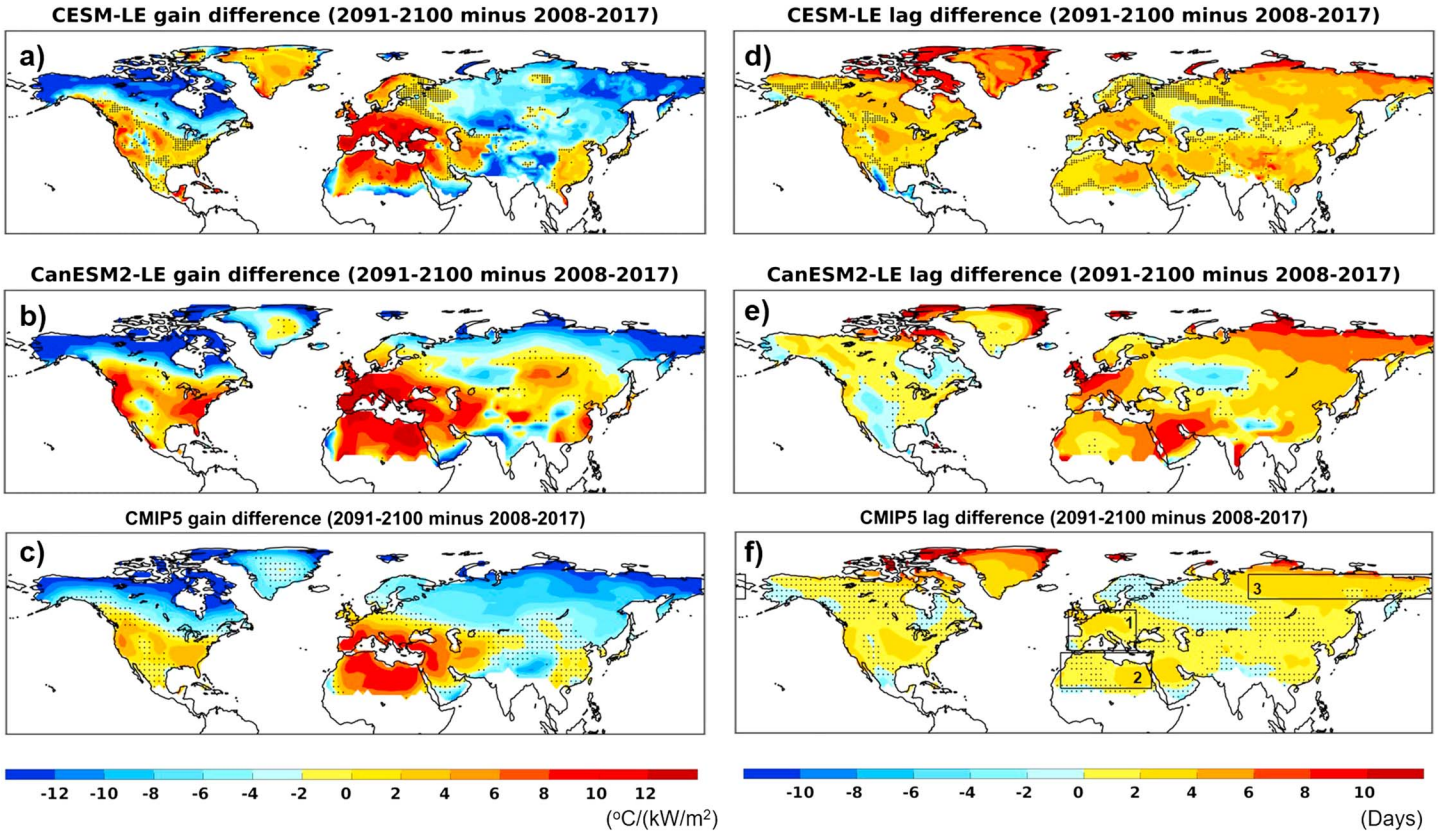

Figure 3. Ensemble mean epoch difference (2091–2100 minus 2008–2017) for gain over land in (a) 40-member CESM-LE, (b) 50-member CanESM2-LE, and (c) CMIP5 multimodel ensemble (Table 1). (d) Same as in (a). (e) Same as in (b). (f) Same as in (c) but for lag. Stippling in (a), (b), (d), and (e) indicates changes that are *not* statistically significant at 95% confidence (under a two-tailed *t* test). Stippling in (c) and (f) indicates that the multimodel mean change has the same sign as at least 75% of the models. Regions where less than 85% of the variance in an average demeaned year (averaged over 1920–2100) is explained by the annual cycle, in one or more ensemble members, have been excluded. In (f), bounding boxes numbered 1, 2, and 3 correspond to the definitions of Europe, North Africa, and Siberia used in this study.

between anthropogenically forced changes in the annual cycle in two different large climate model ensembles with different model physics and in the CMIP5 multimodel mean increase confidence in the projections.

4.1. How Soon Can Forced Changes in the Annual Cycle Be Detected?

We now address the first goal of this study: quantifying the influence of internal variability. We begin by assessing the time of emergence of forced changes in gain and lag. Following the strategy of Deser et al. (2012), we use equation (2) to compute the decade of emergence of statistically significant forced changes for two ensemble sizes: small (N = 5) and large (N = 40). Forced changes are computed as decadal-mean ensemble-mean differences for future decades relative to the past 10 years (2008–2017). With a small ensemble, the detection of forced changes in the gain (Figure 4a) and lag (Figure 4c) is not possible until midtwenty-first century over most regions of the NH. Particularly striking are the continental United States and Western Russia where the detection of changes, in either the gain or lag, is not possible even by the end of the twenty-first century. Changes in the annual cycle that are detectable with a small ensemble within the next 20 years occur only over a handful of regions—Europe, Eastern Canada, and Eastern Siberia being the most prominent examples.

Increasing the ensemble size to 40 leads to earlier detection times for both gain (Figure 4b) and lag (Figure 4d) consistent with the expectation that a larger ensemble size would lead to stronger signal-to-noise ratios. Indeed, with 40 ensemble members, forced changes in the gain and lag are detectable within the next 30 years over the majority of the Northern Hemispheric land areas. However, there are also a few regions

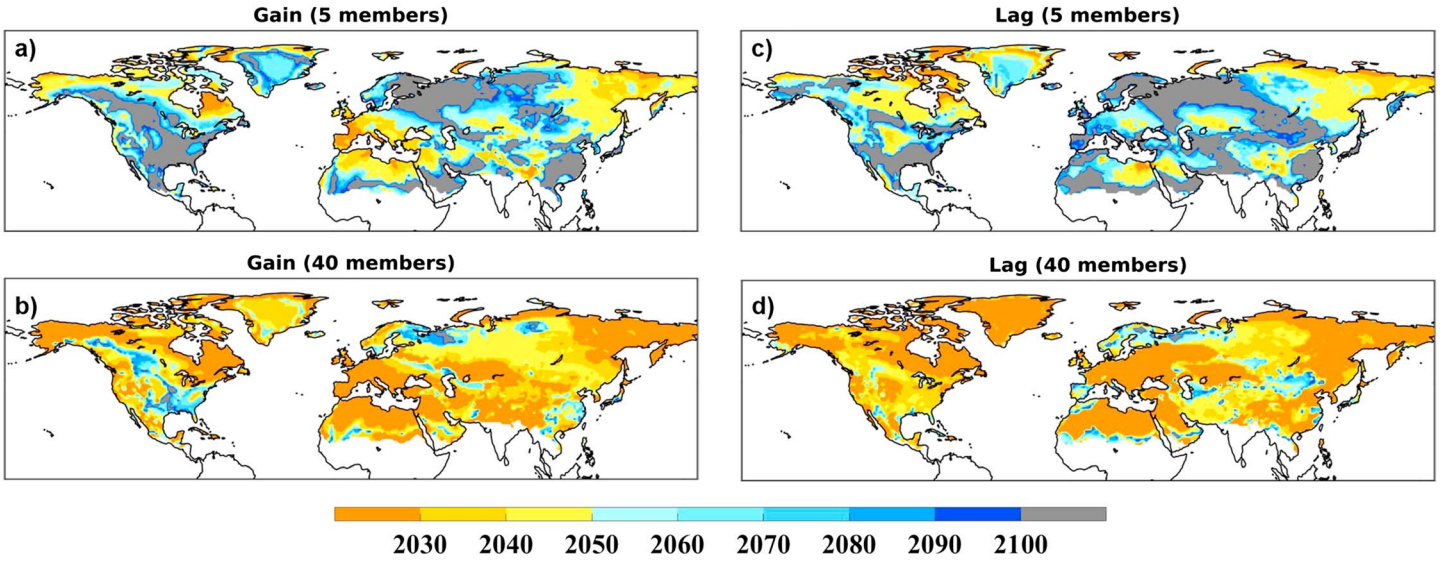

Figure 4. Decade when ensemble mean gain change over land relative to the period 2008–2017 first becomes detectable at 95% confidence (under a two-tailed *t* test) for an ensemble size of (a) 5 and (c) 40. (b) Same as in (a). (d) Same as in (c) but for lag. Year refers to midpoint of decade. Gray indicates regions where detection is not possible by 2100. Excluded regions are the same as in Figure 3.

where the changes are not detectable by midcentury even with 40 ensemble members. For example, changes in gain over the eastern United States and in lag over the Iberian Peninsula become detectable with a large ensemble only in the second half of the twenty-first century.

4.2. How Many Ensemble Members Are Necessary to Detect Forced Changes?

An alternate and useful perspective of quantifying the influence of internal variability is to assess the minimum number of ensemble members (N_{min}) required to detect forced signals. We compute N_{min} for three future epochs in relation to 2008–2017: near-future (2026–2035), midtwenty-first century (2046–2055), and the far-future (2091–2100). The effect of internal variability is strongest in the near-future with changes in gain (Figure 5a) and lag (Figure 5d) requiring more than 40 ensemble members for detection over most regions. Moving toward 2046–2055, the forced signals strengthen leading to smaller ensemble size requirements for detection (Figures 5b and 5e). However, even by midtwenty-first century, there are vast regions where detection of forced signals from the noise of internal variability is not possible even with a large ensemble of 40 ensemble members. For example, detecting midtwenty-first century changes in gain over most of the United States and changes in lag over Western Russia requires more than 40 ensemble members. The effect of internal variability on detection is evident even as we move to the end of the twenty-first century. While far-future forced changes in gain and lag are strong enough to be detectable over many regions using fewer than 10 ensemble members (Figures 5c and 5f), there are regions, for example, Western Russia, where forced changes in either the gain or lag are not detectable even with 40 members.

While Figure 5 demonstrates the strong confounding influence internal variability exerts on simulated changes in the structure of the annual cycle over many regions in the NH, there are also regions in the CESM-LE where the forced signals are relatively strong compared to internal variability and therefore are easier to detect. For example, forced signals in gain and lag over Europe, Northern Africa, and Siberia are detectable with as few as three ensemble members by midtwenty-first century (Figures 5b and 5e). Consistent with Figure 3, the regions with the largest projected changes tend to be the regions where detection with a smaller ensemble size is possible. Since the results in Figures 4 and 5 are derived from equivalent mathematical expressions (equations (2) and (3)), it should not be surprising that the two figures lead to similar conclusions.

4.3. Detecting Forced Twenty-First-Century Seasonal Cycle Changes in Future Observational Records

While it is useful to view the climate system as an ensemble of independent trajectories, recall that only one of these trajectories is actually realized in the future. Assessing the possibility of detection of forced changes in a single realization, and transferring that knowledge to the future observational record, is therefore of

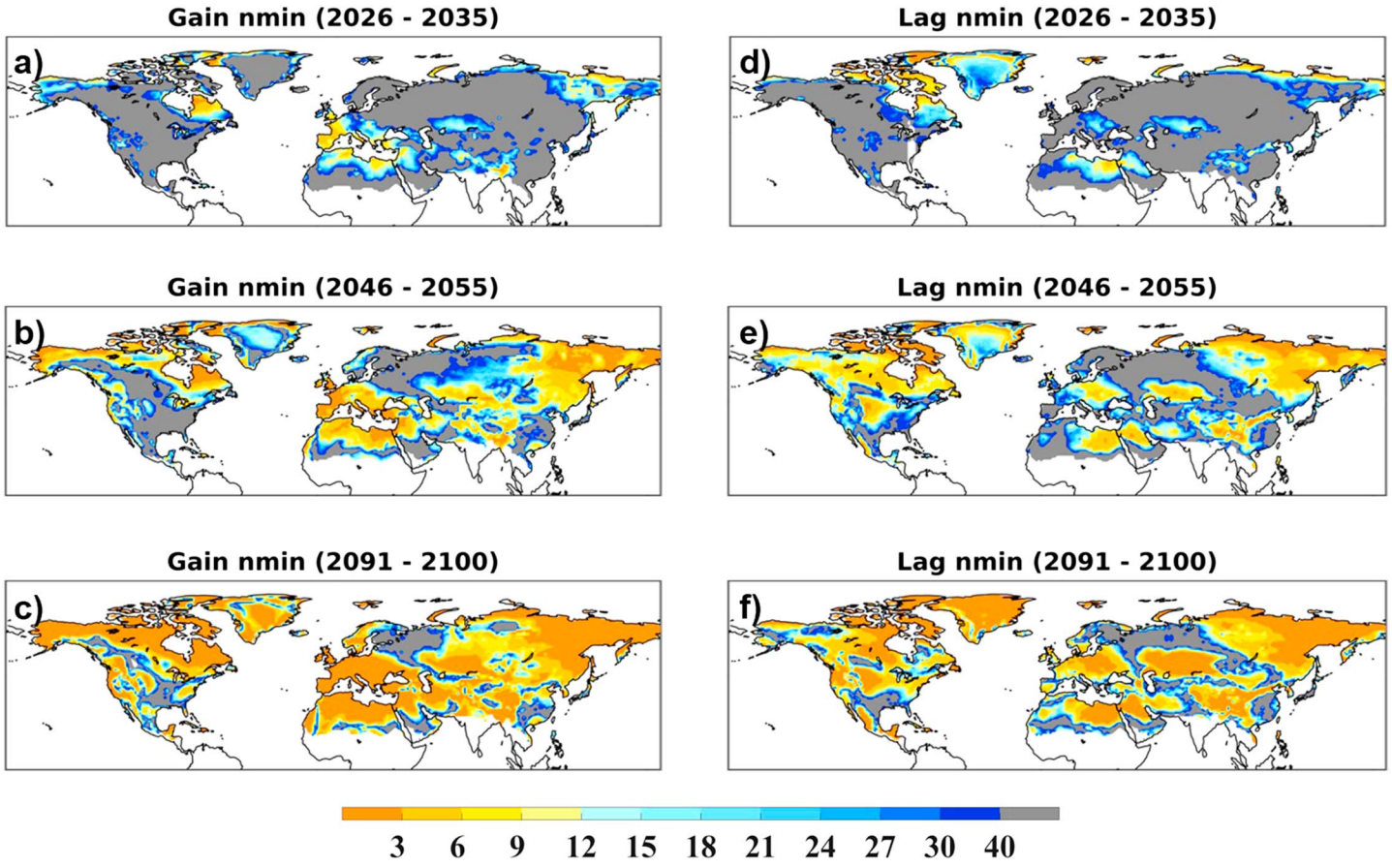

Figure 5. Minimum number of ensemble members required to detect change in gain over land at 95% confidence (under a two-tailed *t* test) for the periods (a) 2026–2035, (b) 2046–2055, and (c) 2091–2100, relative to the period 2008–2017. (d) Same as in (a). (e) Same as in (b). (f) Same as in (c) but for lag. Gray indicates regions where even 40 members are not sufficient to detect a significant change. Excluded regions are the same as in Figure 3.

obvious practical interest. This however requires an approach that is slightly different from the ensemble-based methods applied thus far (see equations (2) and (3)). A metric that is particularly useful here is the signal-to-noise ratio (S/N; e.g., Hawkins and Sutton 2012) that quantifies the signals of climate change relative to a baseline noise level in a single realization. Adopting the S/N ratio approach, we assess in this section the prospect of detecting changes in a single realization of the future climate system. We define the noise as the interannual variability of the historical (1960–2005) gain and lag. The signals are obtained by first removing the historical mean from gain and lag and linearly regressing them onto time at each grid point. The year at which S/N exceeds a ratio of one is considered the time of emergence (ToE) of the signal, as done in previous studies (Hawkins & Sutton, 2012). The ToE can be interpreted as the year in which one can first identify a robust change in the seasonal cycle in the future observed record from internal variability.

We apply the signal-to-noise ratio approach to member 1 of the CESM-LE and show the results for gain in Figure 6a. While the signal does not emerge from noise over most land areas in the NH, we are able to detect anthropogenic influence in three previously identified regions (Europe, North Africa, Siberia) as well as Northern Canada, as soon as the coming decade. We further apply the procedure to the remaining ensemble members individually and display the 10th and 90th percentiles of the ToE across the members (Figures 6b and 6c) with the goal of qualitatively assessing the range of possible ToE. At the 10th percentile, changes emerge by 2025 over the three regions, whereas at the 90th percentile, changes emerge by the midtwenty-first century. These percentiles give crude estimates of the lower and upper bounds of the ToE. This indicates a high likelihood that forced changes in the annual cycle over the three regions will be detectable even in a single realization, and thus the future observed record, by the midtwenty-first century. We do

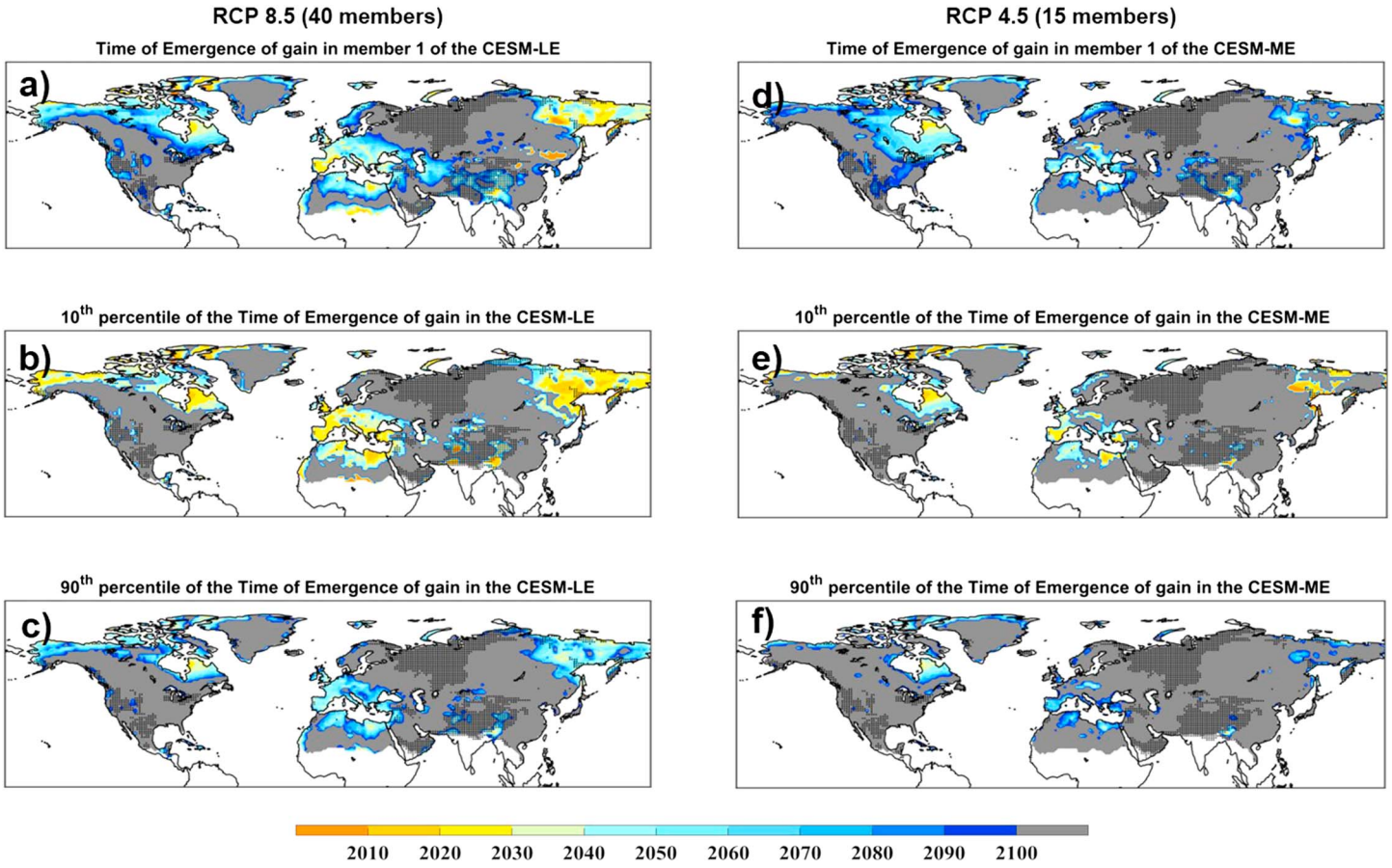

Figure 6. (a) Time of emergence (ToE) of gain in member 1 of the CESM-LE. (b) The 10th percentile of the ToE of gain across 40 members in the CESM-LE. (c) Same as in (b) but for the 90th percentile. (d) Same as in (a). (e) Same as in (b). (f) Same as in (c) but for the 15-member CESM-ME. ToE is defined as the year when the signal-to-noise ratio exceeds 1 for the first time. Gray indicates regions where the signal-to-noise ratio does not exceed 1 by 2100. Excluded regions are the same as in Figure 3. Stippling indicates regions where the interannual variance in member 1 of the CESM-LE is different from that in BEST observations according to a two-tailed *F* test at the 95% confidence level.

not include the corresponding lag figure because the internal variability swamps the forced response and so changes in lag are not detectable by the end of the century.

For the ToE estimated from the CESM-LE to be relevant to the real world, one assumption is that the variability simulated by the model is not considerably different to that found in observations. In general, the CESM-LE overestimates the interannual variability of gain over NH land (Figure 2c), and therefore, the ToE are likely to be overestimates when compared to future observational records. It is conceivable therefore that, in the three regions which this study identifies, the changes in observations are detectable at an earlier year. Even accounting for these differences between the model and observations, it seems unlikely that changes in the annual cycle will be detectable from observations over much of the NH (gray regions in Figure 6) by the end of this century using our methodology. In some regions, mostly over mountainous terrain, the model significantly overestimates variability (Figure 2), and we indicate these areas in Figure 6 with stippling. These are regions in which the ToE estimated from the CESM-LE is likely biased and so we cannot make robust conclusions over these regions. One should note, however, that these regions have no overlap with the three key regions identified in this study (Europe, North Africa, and Siberia).

This study focuses on scenario RCP 8.5 which is used in the CESM-LE projections and is considered to be a "business-as-usual" scenario. Figures 6a–6c indicate that forced changes in the annual cycle in the identified regions will be detectable even in a single realization of the climate system under RCP 8.5 conditions, as soon as the coming decade. If, however, significant efforts are made to reduce future anthropogenic emissions and warming is limited to a further 1.5° or 2° increase (Mitchell et al., 2017), then RCP 4.5 is

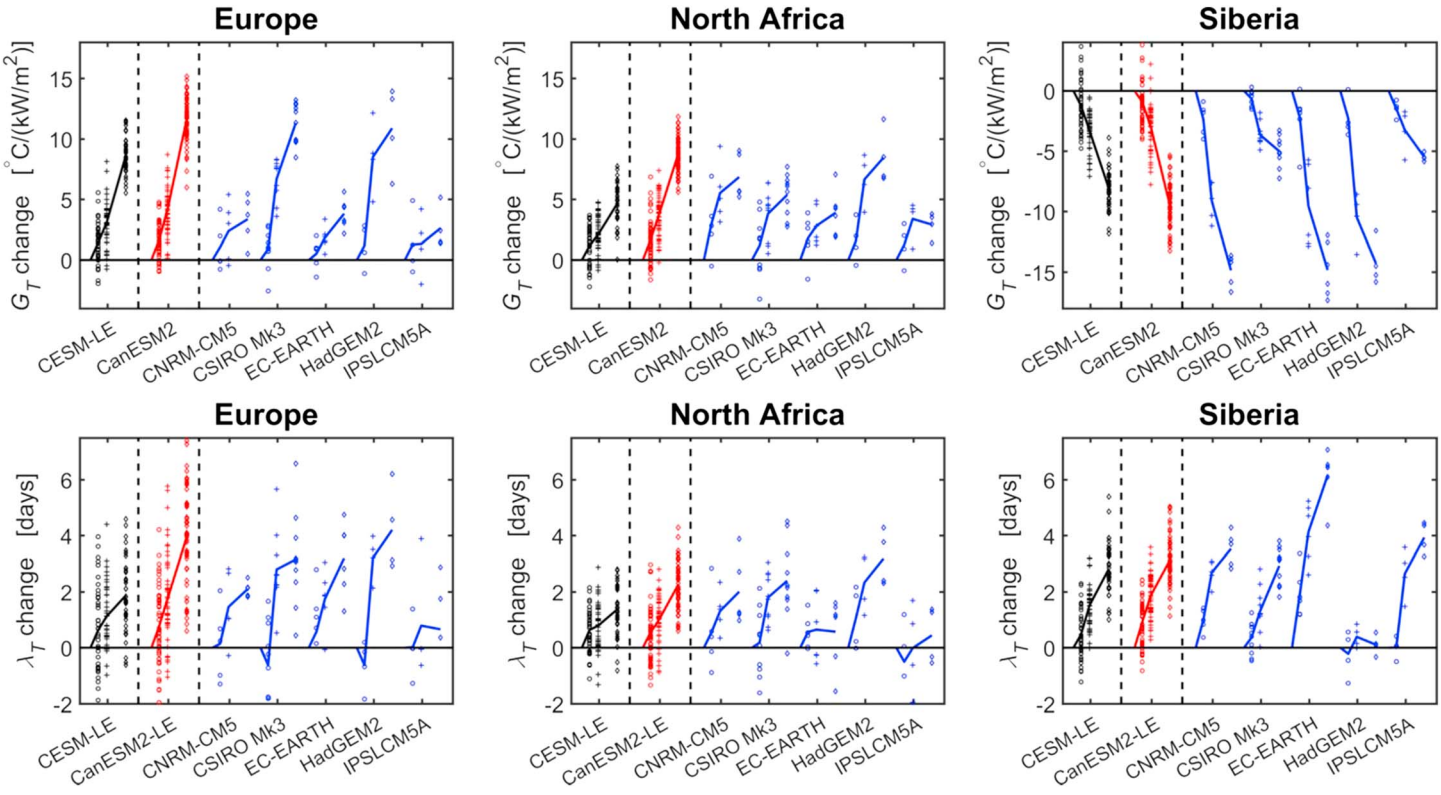

Figure 7. Change in (top row) gain and (bottom row) lag for area averaged regions over (left column) Europe, (middle column) North Africa, and (right column) Siberia over the coming century from RCP8.5 integrations. See section 4.3 and Figure 3f for how the three regions are defined. For each region, changes are shown for the two large ensembles (CESM-LE in black and CanESM2 in red) and the five small ensembles from the CMIP5 archive (blue). Each change is relative to the decadal average of the last 10 years, 2008–2017. Circles represent projected changes by the near future (2026–2035—present), crosses represent projected changes by the midcentury (2046–2055—present), and diamonds represent projected changes by the far future (2091–2100). The lines join the ensemble average for each set of changes. The values are spaced proportionally to time on the *x* axis.

a more realistic scenario to investigate. Figures 6d–6f show the estimated ToE in gain using the 15 members of the CESM-ME which is run with RCP 4.5 forcing. As one would expect, the signal is larger under RCP 8.5 and so the ToE occurs much earlier. Under RCP 4.5, very few regions, except from Western Europe, exhibit any detectable change in gain by the end of this century. Although these results are not unexpected, we note that the CESM-ME only has 15 members compared to the 40 members of the CESM-LE and so there is additional uncertainty that arises from the use of a smaller sample. Overall, this shows that our findings are sensitive to the choice of emission scenario. An exception is ToE over Canada which seems to be insensitive to the choice of emission scenario.

4.4. How Robust Are Projections of the Twenty-First-Century Annual Cycle?

Figures 4–6 indicate the existence of land areas where forced changes in the seasonal cycle are strong compared to the noise of internal variability and are detectable even with small ensembles in the near future. We next test this hypothesis using multiple CMIP5 ensembles. Specifically, we use five CMIP5 models that each contributed at least four ensemble members under the RCP 8.5 scenario and quantify forced signals in each model in terms of the gain and lag differences over land. We limit our focus to averages over land areas in three regions where the forced changes are particularly strong compared to internal variability, and therefore may be detected with few ensemble members (see Figure 3f for region definitions): Europe (37°N–55°N, 8°W–27°E), Northern Africa (20°N–36°N, 12°W–35°E), and Siberia (60°N–71°N, 85°E–185°E). We assess forced changes for three future time periods: near-future (2026–2035), midtwenty-first century (2046–2055), and far-future (2091–2100) relative to 2008–2017. We present the forced changes (estimated as the ensemble mean) in Figure 7, along with changes in individual members to convey ensemble spread. All model ensembles project gain increases over Europe and North Africa,

and decreases over Siberia, and lag increases over all three regions during the time periods under investigation, although with differing rates. Consistent with the finding from the CESM-LE that the forced signals over the three regions will be detectable with fewer than 10 ensemble members, the forced signals projected by almost all models rise above the noise of internal variability and become statistically significant by the midtwenty-first century. We note that these findings are insensitive to small variations in the geographical limits used to define the three regions.

Having shown that the direction of forced changes in the structure of the annual cycle over the three regions is highly robust and easily detectable across the models, it is useful to make note of the magnitude of these changes. Forced changes in lag over the three regions are rather small ranging from two to six days across the model ensembles by the end of the twenty-first century. In contrast, forced changes in amplitude (not shown) are large and are comparable to annual mean warming, reaching almost 50% of the projected mean temperature increase over Europe.

The direction of change in gain over the three regions in Figure 7 matches that of the historical trend in gain (Stine et al., 2009). In combination with annual mean warming, the historical trend in gain has contributed to differing warming rates of summers and winters (e.g., Qian & Zhang, 2015). Specifically, increases in gain over Europe and Northern Africa have manifested as summers warming faster than winters, and decreases in gain over Siberia as winters warming faster than summers. The highly robust results in Figure 7 suggest that this observed differential in summer and winter warming rates over the three regions will continue into the twenty-first century.

5. Mechanisms

5.1. Energy Balance Model

We next turn to our second goal of understanding the mechanisms underlying future changes in the annual cycle using multiple models from the CMIP5 archive. We find it useful to leverage the simple surface energy balance model proposed by D12:

$$C_{\text{eff}}\frac{dT}{dt} = F[t, T] \tag{4}$$

where C_{eff} is the effective heat capacity of the surface, T is the temperature, F is the net energy flux into the surface, and t is the time. While C_{eff}, in general, exhibits seasonality due to seasonal changes in surface properties like soil moisture (Carson & Moses, 1963), D12 report that results based on equation (4) are insensitive to seasonal changes in C_{eff}. We therefore assume C_{eff} to be constant within a given year. It must be noted that C_{eff} is the heat capacity of a layer of material that the atmosphere can thermally influence on the annual time scale (typically the top 1 or 2 m of soil; Carson and Moses 1963) and *not* the heat capacity of some substance per unit mass or of a fixed mass of a substance.

To bring out factors affecting the annual cycle in surface temperature, we follow D12 and partition the net energy flux into the surface layer as $F[t, T] = Q(t) - \beta T$. $Q(t)$ represents net shortwave flux at the surface, computed as downwelling minus upwelling shortwave radiation at the surface, and is assumed to be linearly independent of T.$-\beta T$ represents the sum of longwave and turbulent heat fluxes at the surface layer. β is a constant and may be interpreted as a damping factor that controls the extent to which the surface longwave and turbulent heat fluxes influence the temperature response to solar forcing. Increases in β may be interpreted as these fluxes, in some combination, becoming more effective at maintaining the surface temperature at equilibrium, and vice versa for decreases in β.

Diagnostic relationships between the gain and lag of surface temperature and the controlling parameters C_{eff}, β, and Q may be derived by applying a Fourier transform to equation (4):

$$\lambda_T = \lambda_Q + \arctan\left(\frac{\omega C_{\text{eff}}}{\beta}\right) \tag{5}$$

$$G_T = \frac{G_Q}{\sqrt{\beta^2 + \omega^2 C_{\text{eff}}^2}} \tag{6}$$

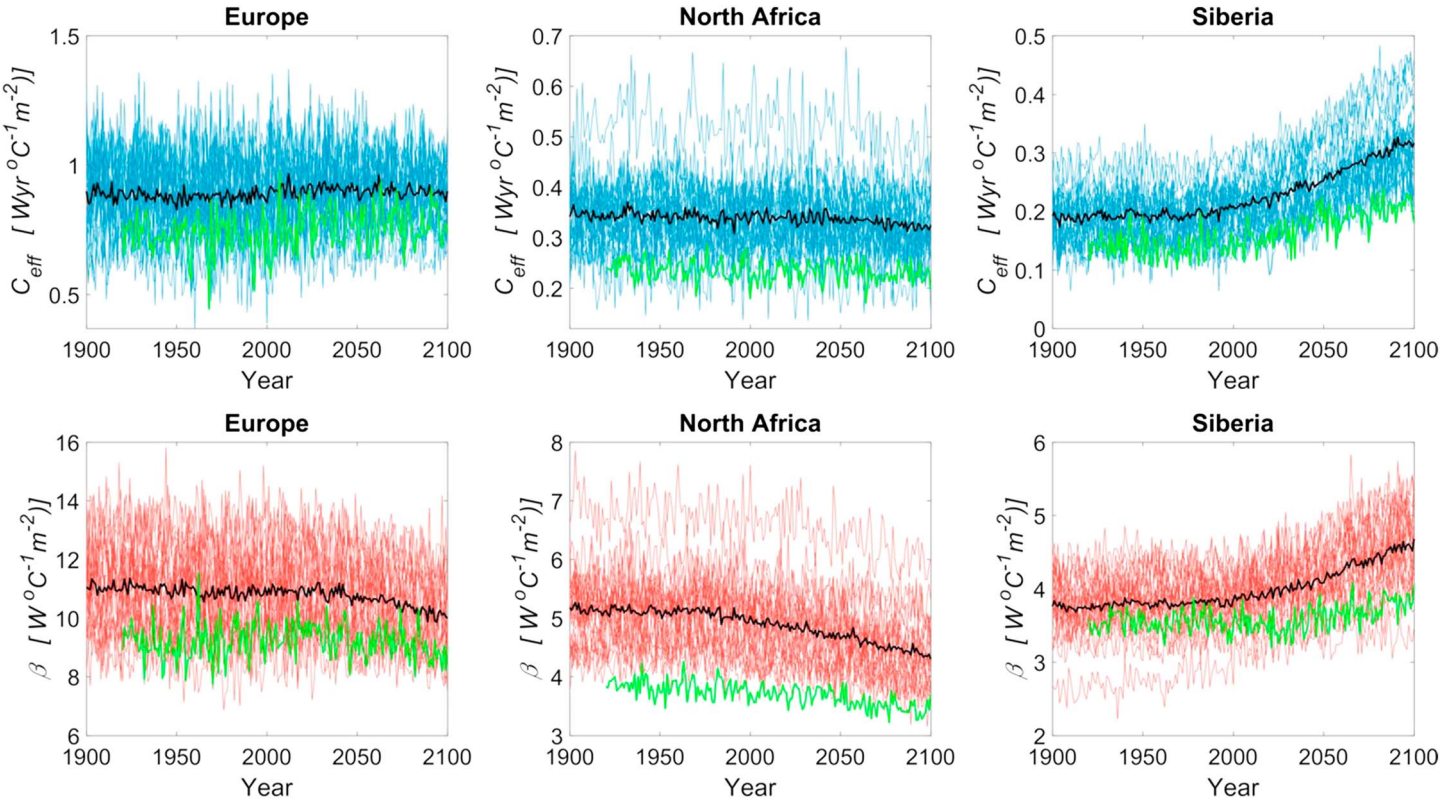

Figure 8. Time series of (top row) effective heat capacity and (bottom row) β for area-averaged regions over (left column) Europe, (middle column) North Africa, and (right column) Siberia in 22 CMIP5 models (red; see Table 1 for list of models) and member 1 of the CESM-LE (green). Thick black lines represent CMIP5 multimodel means.

where G_Q and λ_Q are the gain and lag of Q calculated using the methods in section 2.1 and $\omega = 2\pi$/year. In practice, C_{eff} and β can be obtained by inverting equations (5) and (6):

$$C_{\text{eff}} = \frac{\sin(\lambda_T - \lambda_Q)}{\omega\left(G_T / G_Q\right)} \tag{7}$$

$$\beta = \frac{\cos(\lambda_T - \lambda_Q)}{G_T / G_Q} \tag{8}$$

Equations (5) and (6) delineate the relationship between the gain and lag of the annual cycle of surface temperature, and the physical parameters that control them. In particular, the equations show that C_{eff} has a direct relationship with λ_T and an inverse relationship with G_T. On the other hand, the equations reveal an inverse relationship of β with both λ_T and G_T. Since β parameterizes the combined influence of longwave and turbulent heat fluxes, it is not straightforward to physically interpret the effect of β on the annual cycle. The effect of C_{eff}, on the other hand, can be understood more readily: a larger effective heat capacity is expected to result in a smaller and more delayed temperature response to solar forcing. As a simple check, we plot the historical (1960–2005) gain and lag for NH land in 22 CMIP5 models, computed using the definitions in section 2.1, against C_{eff} (equation (7)) in Figure S1 in the supporting information. In line with physical intuition, C_{eff} exhibits a strong negative relationship ($r = -0.50$, $p = 0.02$) with gain (Figure S1) and a strong positive relationship ($r = 0.68$, $p = 0.00$) with lag (Figure S2), with most models (20 out of 22) underestimating C_{eff} compared to observations.

5.2. Application of Energy Balance Model to Forced Changes

We use equations (5)–(8)) to understand future changes in the annual cycle over the three regions where we earlier identified highly robust changes across a range of ensembles (Figure 7). In Figure 8, we plot C_{eff} and β

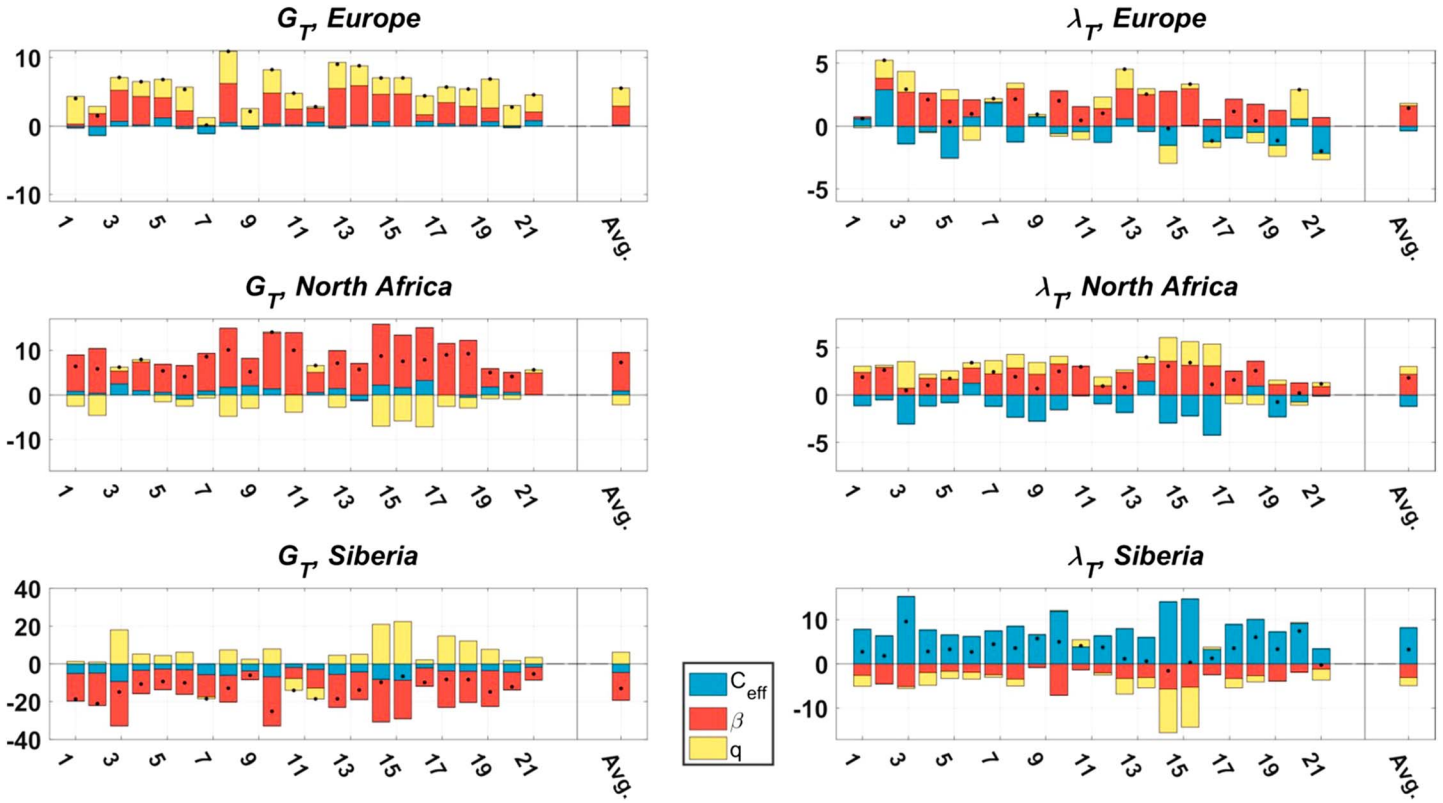

Figure 9. Changes in (left column) gain and (right column) lag for area-averaged regions over (top row) Europe, (middle row) North Africa, and (bottom row) Siberia partitioned into contributions from changes in effective heat capacity (blue), β (red), and net solar insolation at the surface (yellow), across 22 CMIP5 models (see Table 1 for list of models). Black dots represent sum of contributions. Changes are calculated for the far-future (2091–2100) relative to the present (2008–2017).

(computed using equations (7) and (8)) as a function of time for the three regions. Although there is considerable intermodel spread in all regions, the most pronounced changes in C_{eff} and β occur over Siberia where, in the multimodel mean, C_{eff} nearly doubles and β increases by about 25% relative to the early twentieth century. Equation (6) suggests that increases in C_{eff} and β should act together to reduce gain. Consistent with this, we previously noted a large projected reduction in gain over Siberia (Figure 7). In contrast, equation (5) suggests that increases in C_{eff} should act to increase lag while increases in β should have the opposite effect. Despite the 25% increase in β and a 20% decrease in λ_Q (not shown) that contribute to reduce lag, the larger 90% increase in C_{eff} has the greater influence leading to an increase in lag over Siberia (Figure 7).

Over Europe and North Africa, changes in C_{eff} and β are smaller and less robust than those over Siberia. β decreases by 12% over Europe and 15% over North Africa, while changes in C_{eff} are even smaller: C_{eff} decreases by 4% over Europe and 8% over North Africa. Ascertaining the combined influence of these small changes on changes in the gain and lag of the annual cycle is not straightforward since fractional changes in λ_Q and G_Q, while small, are of magnitudes comparable to the fractional changes in C_{eff} and β over these regions, and their influence as such needs to be taken into consideration (see equations (5) and (6)). We address this problem next by explicitly partitioning changes in gain and lag over the three regions into contributions arising from changes in the four controlling parameters C_{eff}, β, G_Q, and λ_Q.

The partitioning is achieved by linearizing equations (5) and (6):

$$\Delta\lambda_T = \Delta\lambda_Q + \frac{R}{1+R^2}\frac{\Delta C_{\text{eff}}}{C_{\text{eff}}} - \frac{R}{1+R^2}\frac{\Delta\beta}{\beta} \tag{9}$$

$$\Delta G_T = \frac{G_T}{G_Q}\Delta G_Q - G_T \frac{R^2}{1+R^2}\frac{\Delta C_{\text{eff}}}{C_{\text{eff}}} - G_T \frac{1}{1+R^2}\frac{\Delta \beta}{\beta} \quad (10)$$

where $R = {}^{\omega C_{\text{eff}}}/_{\beta}$. The first terms in equations (9) and (10) represent contributions from changes in the annual harmonic of Q, the second terms, from changes in C_{eff}, and the third terms, from changes in β. This linearization was found to provide a good approximation; differences between the sum of the contributions and the full changes in the gain and lag were found to be small. Using equations (9) and (10), we partition future changes in the annual cycle in the three regions on an individual model basis and plot the results in Figure 9. We also list the multimodel mean values in Table S1. While there is considerable intermodel spread in the magnitude of future changes in gain and lag, the direction of change is more or less the same across the models in all three regions and is consistent with the results from Figure 7. Changes in β have the largest influence on changes in gain and lag in the multimodel mean except over Siberia, where increases in C_{eff} dominate increases in lag. Interestingly, despite the proximity of the two regions, the gain increases over Europe and North Africa arise in different ways. β and C_{eff} act in opposition over North Africa with β dominating and resulting in an increase in gain while they both contribute to an increase in gain over Europe. It is also seen that G_Q has a considerable influence on gain, albeit smaller than that of β. In contrast, changes in λ_Q affect lag to a negligible degree. Overall, Figure 9 suggests that β is the most important parameter for understanding the highly robust future changes in the annual cycle revealed in Figure 7, except for the change in phase over Siberia.

It has previously been suggested that changes in seasonality over the NH are most strongly influenced by changes in atmospheric circulation (Stine & Huybers, 2012). We note that atmospheric circulation changes affect the annual cycle primarily by impacting surface fluxes, and therefore, the effects of these circulation changes are implicitly captured by changes in β. It is important to acknowledge, however, that changes in atmospheric circulation would predict the opposite response in gain over Europe (Stine & Huybers, 2012), and it therefore seems likely that this is not the only factor causing these changes.

Given the abstract nature of the definitions of C_{eff} and β, how should our results from the energy balance model be interpreted physically? If we look at the projected changes in Siberia, we hypothesize that the increase in β is related to the loss of Arctic sea ice northward of this region which allows for more heat to be fluxed from the newly exposed ocean. It has been shown that although Arctic sea ice loss peaks in September, the resultant heat flux to the atmosphere is maximum in the winter (e.g., see Sun et al. 2015, Figure 2 and England et al. 2018, Figure 2) which would lead to a reduction in the gain. The energy balance model allows us to weigh the contribution of different processes. For example, this suggests that the possible effect of sea ice loss is far more important than the effect of the increase in C_{eff} over Siberia (Figure 9), which could be a result of increased soil moisture from snowmelt, and changes in Q arising from changes in albedo due to a reduction in the snow season length. Another example that demonstrates the use of the energy balance model is the different signs of the contributions of Q over Europe and North Africa (Figure 9). The results clearly show that changes in these two regions are not governed by identical processes; we are unsure of the reason behind this, but we think it could be related to changes in cloud cover (e.g., Tselioudis et al., 2016).

To explore the influence of internal variability on the physical parameters, we next partition future changes in the annual cycle within the CESM-LE and document the results in Figure 10. The magnitude of the ensemble mean change in gain and lag in the three regions is comparable to that of the CMIP5 multimodel mean (Figure 9). While internal variability leads to spread in the magnitude of changes across the ensemble members, the direction of the changes is, in general, consistent across the members and matches that in Figure 9. It is interesting to note that the spread of the contributions across the ensemble members, while small compared to the ensemble mean contributions (Table S2), is roughly half that of the CMIP5 multimodel spread (compare standard deviations in Tables S1 and S2). In other words, to the extent that the CESM-LE represents the variability of the real climate system, internal variability is an important consideration not only for detection but also for understanding the relative roles of the factors that contribute to future changes in the annual cycle.

5.3. Application of Energy Balance Model to Historical Biases

Motivated by the utility of the energy balance model in understanding future changes in the annual cycle, we finally apply the model to explore biases in the simulation of the historical (1960–2005) annual cycle by the CMIP5 ensemble. Specifically, we use equations (9) and (10) to partition CMIP5 model biases

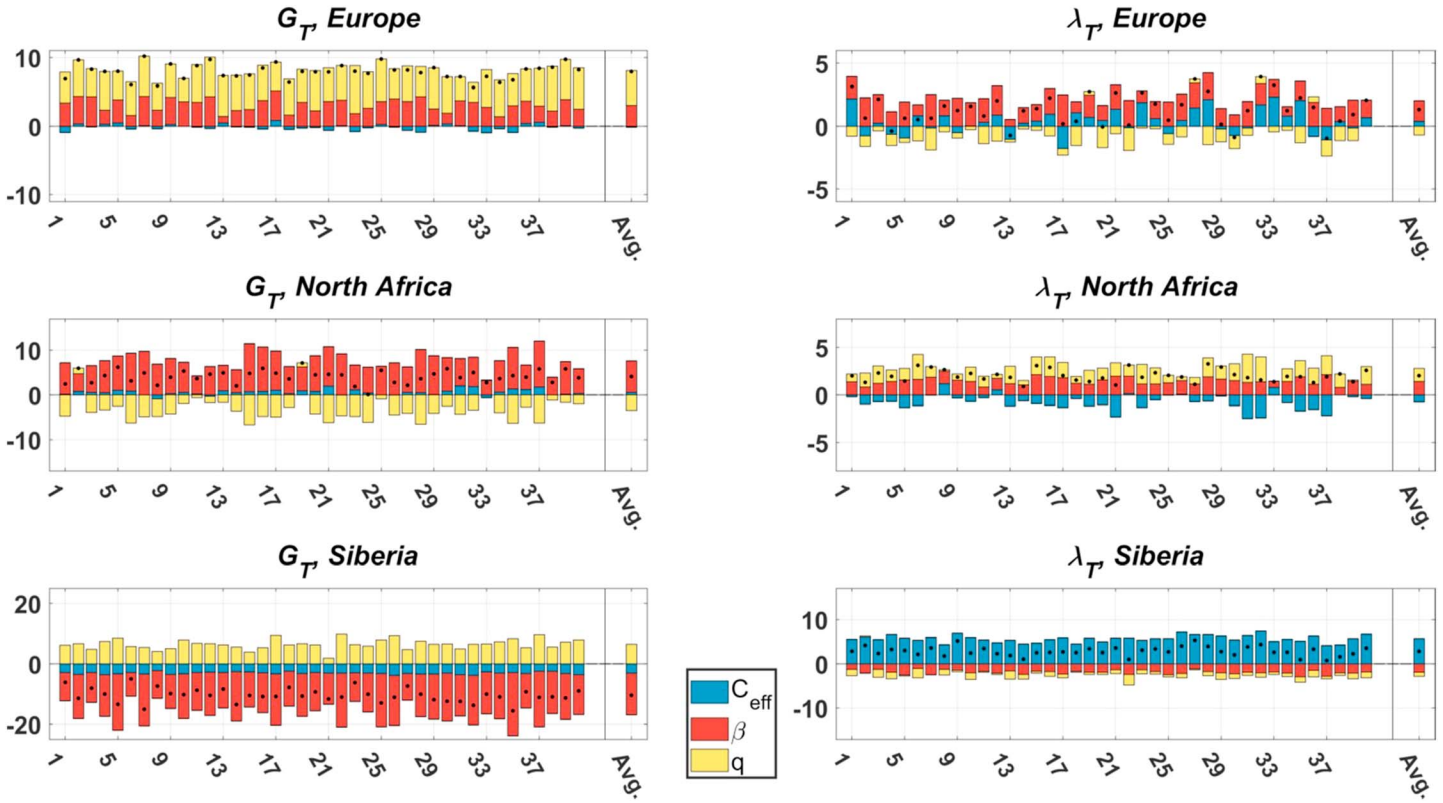

Figure 10. Changes in (left column) gain and (right column) lag for area-averaged regions over (top row) Europe, (middle row) North Africa, and (bottom row) Siberia partitioned into contributions from changes in effective heat capacity (blue), β (red), and net solar insolation at the surface (yellow), across 40 CESM-LE member. Black dots represent sum of contributions. Changes are calculated for the far-future (2091–2100) relative to the present (2008–2017).

in the simulation of the historical annual cycle with respect to BEST observations into contributions from biases in the controlling parameters C_{eff}, β, G_Q, and λ_Q. We present these results in Figure 11. Despite spread in their magnitude, the sign of the contributions of the biases is largely consistent across the models, suggesting a common origin to the biases in the models. In the multimodel mean, gain is slightly underestimated over Europe and Siberia while it is slightly overestimated over North America. Lag on the other hand is underestimated in Europe and North Africa and overestimated in Siberia. The total multimodel mean biases in gain in the three regions, and in lag over Siberia, are relatively small. It is seen that these small changes are in fact the result of cancellations between large and opposing contributions from biases in the controlling parameters.

It is important to note that biases in the controlling parameters are in many cases larger than future changes in the parameters (compare Figure 11 with Figure 9), which could introduce uncertainty to the interpretation of our results. However, on performing a simple correlation analysis (not shown) between historical biases and future changes in C_{eff}, β, G_Q, and λ_Q across the 22 CMIP5 models, we found no obvious correlations that could serve as indicators of links between historical biases and spread in future projections. We finally note that the impact of internal variability on these biases in the CESM-LE was minimal (not shown).

6. Discussion and Summary

Consistent with previous studies that documented the influence of internal variability on observed trends in seasonality (e.g., Stine & Huybers, 2012), our results show that future projections of the annual cycle will also be strongly modulated by internal variability. In particular, our study suggests that the detection of systematic changes in the annual cycle due to increasing greenhouse gases will be challenging from the observational record, and even with small ensembles, over many NH land regions in the coming decades.

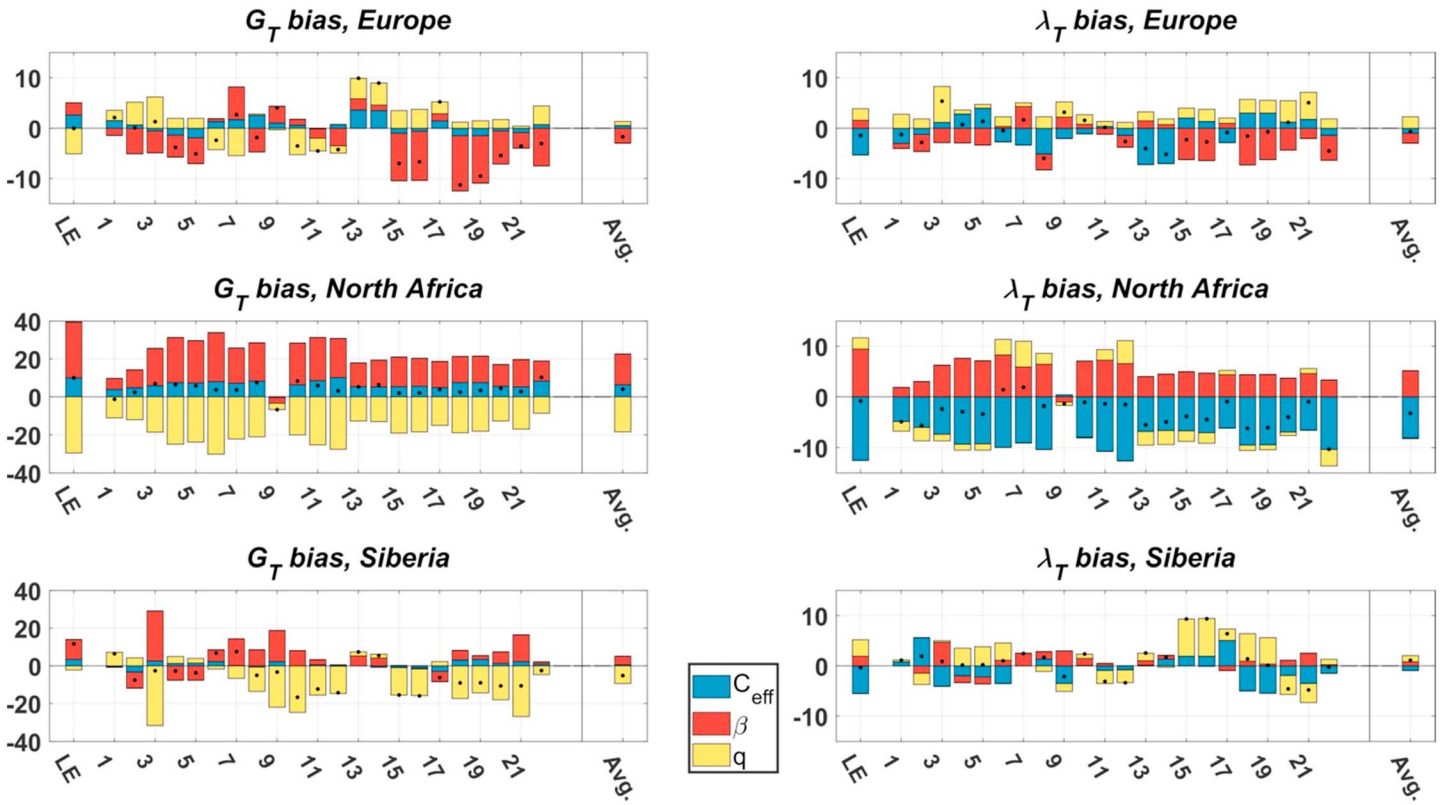

Figure 11. Biases in (left column) gain and (right column) lag for area-averaged regions over (top row) Europe, (middle row) North Africa, and (bottom row) Siberia partitioned into contributions from changes in effective heat capacity (blue), β (red), and net solar insolation at the surface (yellow), across 22 CMIP5 models (see Table 1 for list of models) and member 1 of the CESM-LE (indicated by the label "LE" on the *x* axes). Black dots represent sum of contributions. Biases are calculated as 46-year (1960–2005) mean differences between models and BEST observations.

Yet despite the noise from internal variability, our investigation did find regional exceptions where the forced annual cycle signals are relatively strong compared to internal variability. Indeed, we found regions where changes in the seasonal cycle are detectable with ensembles consisting of approximately five members by the midtwenty-first century. Specifically, our multimodel analysis of annual cycle projections over three such regions (Europe, Northern Africa, Siberia) leveraging five CMIP5 ensembles revealed a strong consensus across the model ensembles on the direction of change as well as the strength of the forced response in relation to internal variability (Figure 7). The direction of change over these regions may therefore be considered robust with a high degree of confidence regardless of the observation that the rate of change varies to some extent across models.

It should be noted that our results could be sensitive to our choice of comparing decadal epoch differences. The effect of internal variability will likely be more pronounced, and consequently, ensemble size requirements will be greater for detecting forced changes at the interannual time scale. Conversely, choosing a longer averaging period, for example, 20 years, would lead to a reduction in temporal variance, and consequently, detection would require less ensemble members. Additionally, we note that fitting a linear trend is a more robust approach than comparing epoch differences; however, we chose to examine decadal averages so that we did not need to assume linearity in time (Barnes & Barnes, 2015). Our choice of decadal averaging was motivated by our goal to investigate changes over a range of future time periods (near-, medium-, and far-future).

Likewise, averaging across spatial domains, in general, results in a reduction of variability which may help to discern a signal. Qian and Zhang (2015) are able to attribute a reduction in the seasonal cycle amplitude over the observational period to increased anthropogenic emissions, most clearly when averaged over all NH land. Averaging over larger scales, however, can also mask the climate change signal; we have identified competing responses in gain in Siberia and Europe which would counteract each other. The

regions we have chosen to investigate further are identified because they contain strong uniform signals. Additionally, we also note that spatial resolution could also play an important role. For example, it is likely that the biases which are found over mountainous regions (e.g., Figures 2b and 2c) arise because the resolutions of the models smooth out the effect of the topographic features notwithstanding possible defects in surface parameterization (Rhoades et al., 2016). An alternative approach to that pursued in this study which could address the concerns about model biases and resolution is to estimate internal variability from the observational record on temporal and spatial scales of interest (e.g., Mckinnon et al., 2017; Thompson et al., 2015).

We finally list two limitations of this study. First, we found large and persistent CMIP5 model biases in the physical parameters controlling the historical annual cycle. As mentioned earlier, how these biases imprint upon future projections is not clear to us and is worthy of further investigation. Second, it is important to note that the time of emergence of climate change signals in a climate model ensemble is a function of the variability simulated by the ensemble (e.g., Mckinnon et al., 2017). In this study we have attempted to assess the simulated variability by comparing the interannual variability from CESM with observations, but the longer-term variability is much more difficult to assess. If the variability in the ensemble is higher than that present in the real climate system would lead to a later detection while variability that is lower would lead to an earlier detection. Regardless, this paper underscores the importance of internal variability for future projections of the annual cycle. Further, our study alerts policymakers and stakeholders in the three regions to the remarkable robustness of changes in the annual cycle, and suggests that observational studies aimed at detecting forced changes in the annual cycle are best served by focusing on these regions.

Acknowledgments
We thank Jennifer Kay for helpful comments related to this work. We are grateful to the Advanced Climate Dynamics Courses summer school 2017, "The Dynamics of the Seasonal Cycle," that motivated this study. We also thank the Yellowstone CESM CSL for providing computing resources, and the scientists and software engineers who build and maintain CESM. We acknowledge Environment and Climate Change Canada's Canadian Centre for Climate Modeling and Analysis for executing and making available the CanESM2 large-ensemble simulations used in this study, and the Canadian Sea Ice and Snow Evolution Network for proposing the simulations. We thank Urs Beyerle for kindly providing access to the CMIP5-ng database. V.Y. is supported by the CIRES Graduate Research Award program and startup funds awarded to Jennifer Kay by the University of Colorado CIRES. M.R.E. is funded by the Columbia University Bakhmeteff Research Fellowship in fluid mechanics. The data sets necessary to reproduce our results are freely available at Zenodo: http://doi.org/10.5281/zenodo.1477091.

References

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., et al. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, *38*, L05805. https://doi.org/10.1029/2010GL046270

Ault, T. R., Macalady, A. K., Pederson, G. T., Betancourt, J. L., & Schwartz, M. D. (2011). Northern Hemisphere modes of variability and the timing of spring in western North America. *Journal of Climate*, *24*(15), 4003–4014. https://doi.org/10.1175/2011JCLI4069.1

Barnes, E. A., & Barnes, R. J. (2015). Estimating linear trends: Simple linear regression versus epoch differences. *Journal of Climate*, *28*(24), 9969–9976. https://doi.org/10.1175/JCLI-D-15-0032.1

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review*, *97*(3), 163–172. https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2

Carey, C. (2009). The impacts of climate change on the annual cycles of birds. *Philosophical Transactions of the Royal Society B*, *364*(1534), 3321–3330. https://doi.org/10.1098/rstb.2009.0182

Carson, J. E., & Moses, H. (1963). The annual and diurnal heat-exchange cycles in upper layers of soil. *Journal of Applied Meteorology*, *2*, 397–406. https://doi.org/10.1175/1520-0450(1963)002%3C0397:TAADHE%3E2.0.CO%5Cn2

Cornes, R. C., Jones, P. D., & Qian, C. (2017). Twentieth-century trends in the annual cycle of temperature across the Northern Hemisphere. *Journal of Climate*, *30*(15), 5755–5773. https://doi.org/10.1175/JCLI-D-16-0315.1

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

Dwyer, J. G., Biasutti, M., & Sobel, A. H. (2012). Projected changes in the seasonal cycle of surface temperature. *Journal of Climate*, *25*(18), 6359–6374. https://doi.org/10.1175/JCLI-D-11-00741.1

England, M., Polvani, L., & Sun, L. (2018). Contrasting the Antarctic and Arctic atmospheric responses to projected sea ice loss in the late twenty-first century. *Journal of Climate*, *31*(16), 6353–6370. https://doi.org/10.1175/JCLI-D-17-0666.1

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., et al. (2017). Large near-term projected snowpack loss over the western United States. *Nature Communications*, *8*, 14996. https://doi.org/10.1038/ncomms14996

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*, L01702. https://doi.org/10.1029/2011GL050087

Huntingford, C., Jones, P. D., Livina, V. N., Lenton, T. M., & Cox, P. M. (2013). No increase in global temperature variability despite changing regional patterns. *Nature*, *500*(7462), 327–330. https://doi.org/10.1038/nature12310

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The Community Earth System Model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1

IPCC (2013). Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In *Intergov. Panel Clim. Chang. Work. Gr. I Contrib. to IPCC Fifth Assess. Rep. (AR5)* (Vol. 1535). Cambridge, New York: Cambridge University Press. https://doi.org/10.1029/2000JD000115

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). NCEP-DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*, *83*(11), 1631–1644. https://doi.org/10.1175/BAMS-83-11-1631

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Labe, Z., Ault, T., & Zurita-Milla, R. (2017). Identifying anomalously early spring onsets in the CESM large ensemble project. *Climate Dynamics*, *48*(11-12), 3949–3966. https://doi.org/10.1007/s00382-016-3313-2

Lamarque, J. F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., et al. (2010). Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmospheric Chemistry and Physics*, *10*(15), 7017–7039. https://doi.org/10.5194/acp-10-7017-2010

Lambert, L. D. (1971). The role of climate in the economic development of nations. *Land Economics*, *47*(4), 339. https://doi.org/10.2307/3145070

Lean, J. (2000). Evolution of the Sun's spectral irradiance since the Maunder minimum. *Geophysical Research Letters*, *27*, 2425–2428.

Lynch, C., Seth, A., & Thibeault, J. (2016). Recent and projected annual cycles of temperature and precipitation in the Northeast United States from CMIP5. *Journal of Climate*, *29*(1), 347–365. https://doi.org/10.1175/JCLI-D-14-00781.1

Madden, R. A., & Ramanathan, V. (1980). Detecting climate change due to increasing carbon dioxide. *Science*, *209*, 763–768.

Mann, M., & Park, J. (1996). Greenhouse warming and changes in the seasonal cycle of temperature: Model versus observations. *Geophysical Research Letters*, *23*(10), 1111–1114. https://doi.org/10.1029/96GL01066

Mckinnon, K. A., Poppick, A., Dunn-Sigouin, E., & Deser, C. (2017). An "observational large ensemble" to compare observed and modeled temperature trend uncertainty due to internal variability. *Journal of Climate*, *30*(19), 7585–7598. https://doi.org/10.1175/JCLI-D-16-0905.1

Mckinnon, K. A., Stine, A. R., & Huybers, P. (2013). The spatial structure of the annual cycle in surface temperature: Amplitude, phase, and Lagrangian history. *Journal of Climate*, *26*(20), 7852–7862. https://doi.org/10.1175/JCLI-D-13-00021.1

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., et al. (2007). The WCRP CMIP3 multimodel dataset: A new era in climatic change research. *Bulletin of the American Meteorological Society*, *88*(9), 1383–1394. https://doi.org/10.1175/BAMS-88-9-1383

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1-2), 213–241. https://doi.org/10.1007/s10584-011-0156-z

Mitchell, D., AchutaRao, K., Allen, M., Bethke, I., Beyerle, U., Ciavarella, A., et al. (2017). Half a degree additional warming, prognosis and projected impacts (HAPPI): Background and experimental design. *Geoscientific Model Development*, *10*(2), 571–583. https://doi.org/10.5194/gmd-10-571-2017

Patterson, W. P., Dietrich, K. A., Holmden, C., & Andrews, J. T. (2010). Two millennia of North Atlantic seasonality and implications for Norse colonies. *Proceedings of the National Academy of Sciences*, *107*(12), 5306–5310. https://doi.org/10.1073/pnas.0902522107

Qian, C., Fu, C., & Wu, Z. (2011). Changes in the amplitude of the temperature annual cycle in China and their implication for climate change research. *Journal of Climate*, *24*(20), 5292–5302. https://doi.org/10.1175/JCLI-D-11-00006.1

Qian, C., & Zhang, X. (2015). Human influences on changes in the temperature seasonality in mid- to high-latitude land areas. *Journal of Climate*, *28*(15), 5908–5921. https://doi.org/10.1175/JCLI-D-14-00821.1

Rhoades, A. M., Huang, X., Ullrich, P. A., & Zarzycki, C. M. (2016). Characterizing Sierra Nevada snowpack using variable-resolution CESM. *Journal of Applied Meteorology and Climatology*, *55*(1), 173–196. https://doi.org/10.1175/JAMC-D-15-0156.1

Rohde, R., Muller, R. A., Jacobsen, R., Perlmutter, S., & Rosenfeld, A. (2013). Berkeley Earth temperature averaging process. *Geoinformatics & Geostatistics : An Overview*, *01*(02), 1–39. https://doi.org/10.4172/2327-4581.1000103

Sanderson, B. M., Oleson, K. W., Strand, W. G., Lehner, F., & O'Neill, B. C. (2018). A new ensemble of GCM simulations to assess avoided impacts in a climate mitigation scenario. *Climatic Change*, *146*(3-4), 303–318. https://doi.org/10.1007/s10584-015-1567-z

Sardeshmukh, P. D., Compo, G. P., & Penland, C. (2000). Changes of probability associated with El Niño. *Journal of Climate*, *13*(24), 4268–4286. https://doi.org/10.1175/1520-0442(2000)013<4268:COPAWE>2.0.CO;2

Schwartz, M. D., Ahas, R., & Aasa, A. (2006). Onset of spring starting earlier across the Northern Hemisphere. *Global Change Biology*, *12*(2), 343–351. https://doi.org/10.1111/j.1365-2486.2005.01097.x

Stine, A. R., & Huybers, P. (2012). Changes in the seasonal cycle of temperature and atmospheric circulation. *Journal of Climate*, *25*(21), 7362–7380. https://doi.org/10.1175/JCLI-D-11-00470.1

Stine, A. R., Huybers, P., & Fung, I. Y. (2009). Changes in the phase of the annual cycle of surface temperature. *Nature*, *457*(7228), 435–440. https://doi.org/10.1038/nature07675

Sun, L., Deser, C., & Tomas, R. A. (2015). Mechanisms of stratospheric and tropospheric circulation response to projected Arctic Sea ice loss. *Journal of Climate*, *28*(19), 7824–7845. https://doi.org/10.1175/JCLI-D-15-0169.1

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E., & Phillips, A. S. (2015). Quantifying the role of internal climate variability in future climate trends. *Journal of Climate*, *28*(16), 6443–6456. https://doi.org/10.1175/JCLI-D-14-00830.1

Thompson, D. W. J., & Wallace, J. M. (2000). Annular mode in the extratropical circulation. Part I: Month-to-month variability. *Journal of Climate*, *13*, 1000–1016. https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., et al. (2011). RCP4.5: A pathway for stabilization of radiative forcing by 2100. *Climatic Change*, *109*(1-2), 77–94. https://doi.org/10.1007/s10584-011-0151-4

Tselioudis, G., Lipat, B. R., Konsta, D., Grise, K. M., & Polvani, L. M. (2016). Midlatitude cloud shifts, their primary link to the Hadley cell, and their diverse radiative effects. *Geophysical Research Letters*, *43*, 4594–4601. https://doi.org/10.1002/2016GL068242

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review*, *109*(4), 784–812. https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2

Wang, Y. -M., Lean, J. L., & Sheeley, Jr, N. R. (2005). Modeling the Sun's magnetic field and irradiance since 1713. *The Astrophysical Journal*, *625*(1), 522–538. https://doi.org/10.1086/429689

Yettella, V., Weiss, J. B., Kay, J. E., & Pendergrass, A. G. (2018). An ensemble covariance framework for quantifying forced climate variability and its time of emergence. *Journal of Climate*, *31*(10), 4117–4133. https://doi.org/10.1175/JCLI-D-17-0719.1