

## MIT Open Access Articles

*Quantifying the Lead Time Required for a Linear Trend to Emerge from Natural Climate Variability*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Li, Jingyuan, et al. "Quantifying the Lead Time Required for a Linear Trend to Emerge from Natural Climate Variability." *Journal of Climate*, vol. 30, no. 24, Dec. 2017, pp. 10179–91. © 2017 American Meteorological Society.

**As Published:** <http://dx.doi.org/10.1175/JCLI-D-16-0280.1>

**Publisher:** American Meteorological Society

**Persistent URL:** <http://hdl.handle.net/1721.1/117019>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



# Quantifying the Lead Time Required for a Linear Trend to Emerge from Natural Climate Variability

JINGYUAN LI, DAVID W. J. THOMPSON, AND ELIZABETH A. BARNES

*Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado*

SUSAN SOLOMON

*Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Manuscript received 5 April 2016, in final form 13 August 2017)

## ABSTRACT

This study introduces a simple analytic expression for calculating the lead time required for a linear trend to emerge in a Gaussian first-order autoregressive process. The expression is derived from the standard error of the regression and is tested using the NCAR Community Earth System Model Large Ensemble of climate change simulations. It is shown to provide a robust estimate of the point in time when the forced signal of climate change has emerged from the natural variability of the climate system with a predetermined level of statistical confidence. The expression provides a novel analytic tool for estimating the time of emergence of anthropogenic climate change and its associated regional climate impacts from either observed or modeled estimates of natural variability and trends.

## 1. Introduction

The time of emergence (TOE) is defined as the point in time when the signal of climate change emerges from the underlying noise of background natural variability (Madden and Ramanathan 1980; Santer et al. 1995; Weatherhead et al. 1998; Christensen et al. 2007; Giorgi and Bi 2009; Mahlstein et al. 2011; Deser et al. 2012b; Hawkins and Sutton 2012; Zappa et al. 2015). It is helpful for anticipating when the impacts of climate change are likely to have significant effects across societies and ecosystems, and can inform risk assessments, mitigation policies, and climate adaptation planning.

As noted in the IPCC Fifth Assessment Report (AR5) (Kirtman et al. 2013, p. 983), there is “no single metric” for estimating TOE. In most cases, TOE is estimated as the first lead time when the anthropogenic signal in climate change exceeds (and then remains continuously above) a predetermined factor of the amplitude of the natural variability. In this case, the TOE for a time series  $x(t)$  is expressed as

$$n_{\text{TOE}} = \frac{ks_e}{b}, \quad (1)$$

where  $n_{\text{TOE}}$  is the time of emergence (the number of time steps at which anthropogenic climate change is judged to have emerged from the background natural variability; this is expressed relative to the start of the time series and is referred to subsequently as the lead time),  $k$  is the required ratio of the forced signal to the natural variability (generally between 1 and 3),  $s_e$  is the amplitude of the internal (unforced) variability, and  $b$  is the linear trend per time step. Note that in this study, the term *natural variability* refers to the sum of internal climate variability as a result of stochastic dynamic processes and external variability as a result of natural forcings, such as volcanic eruptions and solar variability.

Most previous studies of TOE are based on empirical estimates of the first lead time at which Eq. (1) (or a closely related variant) is satisfied. The differences lie in the methodologies used to determine  $b$ ,  $s_e$ , and  $k$ . For example, Giorgi and Bi (2009), Mahlstein et al. (2011), Diffenbaugh and Scherer (2011), and Zappa et al. (2015) all define  $b$  as the change in the climate state averaged over an ensemble of climate change simulations, where the forced signal is smoothed with different averaging periods. Weatherhead et al. (1998) estimate  $b$  using a generalized least squares regression model. Hawkins and Sutton (2012) define  $b$  as the linear projection of

*Corresponding author:* Jingyuan Li, jingyuan.li@colostate.edu

regional temperatures onto smoothed values of simulated global-mean temperatures. [Giorgi and Bi \(2009\)](#), [Mahlstein et al. \(2011\)](#), and [Hawkins and Sutton \(2012\)](#) estimate  $b$  from ordinary least squares (OLS) linear regression and prescribe a signal-to-noise ratio  $k$  that is an integer factor of the natural variability. [Santer et al. \(1995\)](#) also employ signal-to-noise ratios, but they consider them as a function of the time scale of both the signal and the noise. [Christensen et al. \(2007\)](#), [Deser et al. \(2012b\)](#) and [Zappa et al. \(2015\)](#) consider various “epoch differences” and a value of  $k$  derived from the  $t$  statistic for the difference of means. [Mahlstein et al. \(2012\)](#) also consider differences between epochs when assessing the time of emergence and apply a Kolmogorov–Smirnov test to assess whether two sample epochs are significantly different (i.e., rather than a  $t$  statistic).

The existing literature on TOE provides valuable insight into the point in time when anthropogenic climate change will emerge from natural climate variability on regional spatial scales. But it also has several shortcomings. The methodologies used to estimate the trend [ $b$  in Eq. (1)], the amplitude of the natural variability [ $s_e$  in Eq. (1)], and the predetermined signal-to-noise ratio [ $k$  in Eq. (1)] vary widely from one study to the next, which makes it difficult to reproduce and compare different estimates of the TOE. Times of emergence defined on the basis of a fixed signal-to-noise ratio (e.g., [Hawkins and Sutton 2012](#)) do not correspond to a particular level of statistical significance. Several existing methods require smoothing the data using a wide range of methodologies. Furthermore, many of the methods are based on empirical rather than analytical techniques.

The purposes of this study are twofold. First, we introduce a simple and novel expression for estimating the lead time required for a linear trend to emerge from natural variability at a predetermined level of statistical confidence. The expression is developed from the standard error of the regression, which is widely used in climate research but as far as we know has not been exploited for the explicit purpose of calculating TOE. Second, we will test the resulting expression in a large ensemble of climate change simulations. The results demonstrate the robustness of the assumptions that underlie the expression and make clear its utility for assessing the emergence of linear trends in climate data. The expression is developed in [section 2](#). The application of the expression to climate trends is explored in [section 3](#), and its advantages relative to other methods of calculating TOE are considered in [section 4](#). Conclusions are provided in [section 5](#).

## 2. An analytic expression for the lead time required for a linear trend to emerge from natural variability

Consider the case of a first-order autoregressive [AR(1)] time series of length  $N$  with a linear trend  $b$  imposed upon it such that

$$x(n_t) = bn_t + \alpha x(n_t - 1) + \varepsilon(n_t),$$

where  $n_t = 1, 2, \dots, N$  is the number of time steps,  $x(0) = 0$  by assumption, and  $\varepsilon$  is white noise (independent Gaussian noise with a mean of zero and a variance of  $\sigma_\varepsilon^2$ ). The parameter  $\alpha$  is between 0 and 1, and represents the memory in the time series  $x(n_t)$  from one time step to the next.

Here we estimate  $b$  using simple linear regression, where  $\hat{b}$  is the regression estimator of the trend. The parameter  $\alpha$  can be estimated as the lag-1 autocorrelation of the time series ( $r_1$ ). The confidence interval in the total change in  $x(n_t)$  over time  $n_t$  can thus be expressed as

$$\text{CI} = \hat{b}n_t \pm e,$$

where  $n_t$  is again the number of time steps and  $e$  is the uncertainty in the change in  $x(n_t)$  over time  $n_t$  given by  $\hat{b}n_t$ . The margin of error ( $e$ ) is given in units of  $\Delta x/(n_t \Delta t)$ , where  $\Delta t$  is the time step. The trend  $\hat{b}$  is given in units of  $\Delta x/\Delta t$  so that  $\hat{b}n_t$  is the change over the length of the record and has the same units as  $e$ .

Following [Thompson et al. \(2015\)](#), if detrended values of  $x(n_t)$  are well modeled as an AR(1) process, then the margins of error on the linear trend in  $x(n_t)$  can be expressed as

$$e = t_c n_t s \gamma(n_t, r_1) g(n_t), \quad (2)$$

where

$$\gamma(n_t, r_1) \equiv \left\{ (n_t - 2) \left/ \left[ n_t \left( \frac{1 - r_1}{1 + r_1} \right) - 2 \right] \right. \right\}^{1/2}$$

and

$$g(n_t) \equiv \sqrt{\frac{12}{n_t^3 - n_t}}.$$

In Eq. (2),  $t_c$  is the  $t$  statistic for the desired confidence level and  $s$  is the standard deviation of the residuals of the regression [i.e., of detrended values of  $x(t)$ ]. The expressions for  $\gamma(n_t, r_1)$  and  $g(n_t)$  account for 1) the effects of persistence on the estimate of  $s$ , where  $r_1$  is the lag-1 autocorrelation of the residuals; and 2) the standard deviation of the time axis, respectively. Note that

Eq. (2) is simply the standard error of the regression for the case where 1) the predictor is time, and 2) detrended values of the predictand are well modeled as an AR(1) process (e.g., Santer et al. 2000; Thompson et al. 2015).

The trend in  $x(t)$  is statistically significant when it exceeds its margins of error. Setting  $e = \hat{b}n_t$  in Eq. (2) yields

$$\frac{(T_{\text{SIG}}^3 - T_{\text{SIG}})}{12} \times \frac{T_{\text{SIG}} \left( \frac{1-r_1}{1+r_1} \right) - 2}{(T_{\text{SIG}} - 2)} = \left( \frac{t_c s}{\hat{b}} \right)^2, \quad (3)$$

where  $T_{\text{SIG}}$  denotes the lead time when the trend in  $x(t)$  is statistically significant (in units of time steps). That is, given our parameter estimates of  $\hat{b}$ ,  $s$ , and  $r_1$ ,  $T_{\text{SIG}}$  is the estimated number of time steps required for the trend to be statistically significant at the desired confidence level. The value of  $T_{\text{SIG}}$  can be trivially calculated given  $\hat{b}$ ,  $s$ ,  $r_1$ , and  $t_c$ . It requires no subjective analysis choices (such as the length of the periods used in epoch differences) and no smoothing of the data.

The parameters  $\hat{b}$ ,  $s$ , and  $r_1$  are calculated using the entire length of the time series, while  $t_c$  is a function of  $n_r$ . Thus, to solve for  $T_{\text{SIG}}$ , Eq. (3) is calculated iteratively at each time step. For instance, a time series with a  $T_{\text{SIG}}$  of 10 time steps is calculated using the two-tailed 95% value of  $t_c = \pm 2.26$ , while a time series with a  $T_{\text{SIG}}$  of 50 time steps is calculating using the respective value of  $t_c = \pm 2.01$ . We use the entire length of the time series to determine the values of  $\hat{b}$ ,  $s$ , and  $r_1$ . This provides the most accurate estimates given that the linear trend and the Gaussian AR(1) distribution are consistent throughout the entire period, as calculations using short time series can produce erroneous values. A graphical representation of the calculation of  $T_{\text{SIG}}$  can be seen in Fig. 1, where  $T_{\text{SIG}}$  corresponds to the time step when the lower 95% error bound intersects zero (Fig. 1 will be discussed in more detail in the next section).

Equation (3) can be simplified greatly given two conditions. First, detrended values of  $x(n_t)$  are not serially correlated ( $r_1 \approx 0$ ). This condition holds for climate variability at most terrestrial locations on interannual time scales, since there is very little memory in the internal variability of land surface climate from one year to the next [see discussion in Thompson et al. (2015)]. Second, the trend length is at least  $\sim 15$  time steps. In this case,  $T_{\text{SIG}}^3 \gg T_{\text{SIG}}$  and the two-tailed  $t$  statistic for the 95% confidence level is  $t_c \sim 2$ . Applying both conditions yields a simplified version of Eq. (3) that is suitable for cases where the natural variability is not serially correlated from one time step to the next:

$$T_{95\%} \approx 3.6 \left( \frac{s}{\hat{b}} \right)^{2/3}, \quad (4)$$

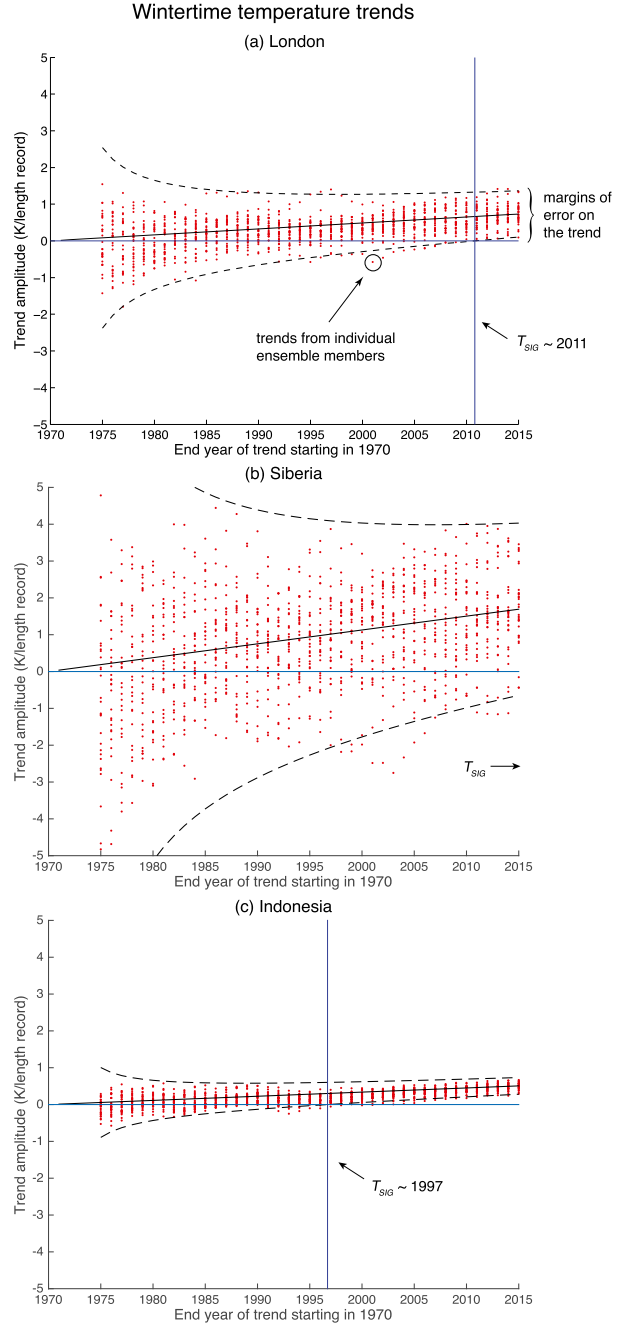


FIG. 1. Trend amplitudes for modeled surface temperature at the grid boxes collocated with (a) London, (b) central Siberia, and (c) Jakarta using CESM-LE output. Red dots indicate actual trends from all 30 ensemble members, while the dashed lines indicate the predicted ranges of trends from applying Eq. (2) to the statistics of the model natural variability. The blue vertical line indicates the lead time when the forced trend is statistically significant as per Eq. (3). See text for details.

where  $T_{95\%}$  is the lead time when the trend in  $x(t)$  is statistically significant at the 95% confidence level. Equation (4) places Eq. (3) in a “signal to noise” format that is similar but not identical to that used in many



previous studies and provides a useful back-of-the-envelope estimate for  $T_{\text{SIG}}$ . All analyses in this paper use the general expression of Eq. (3) for accuracy.

### 3. Application to climate trends

In this section, we test the robustness of Eq. (3) ( $T_{\text{SIG}}$ ) for assessing the point in time when the signature of anthropogenic warming emerges from the background noise of natural climate variability (i.e., achieves statistical significance) on regional scales. We perform the assessment for land surface temperature changes at individual grid boxes. To do so, we exploit a large ensemble of climate change simulations.

In a large ensemble of climate change simulations, each individual ensemble member may be viewed as a unique realization of “model reality.” Here we test the expression for  $T_{\text{SIG}}$  using output from the NCAR Community Earth System Model Large Ensemble (CESM-LE).

Details of the simulations are provided in Hurrell et al. (2013) and Kay et al. (2015). In short, the CESM-LE consists of 40 climate change simulations run using the same model configuration with the same external forcings. Differences in climate trends from one realization to the next are entirely due to the internal variability in the model. Note that internal variability is distinct from “natural” variability, since the latter includes both internal climate variability and external forcings, such as those resulting from volcanic eruptions and variations in solar output (see the appendix). Here, we use the original 30 CESM-LE simulations released in 2014. The runs are available from 1920 to 2100, with historical forcings used for the period 1920–2005 and RCP8.5 forcings used for the period 2006–2100. The analyses are based on seasonal-mean values of surface temperature for the Northern Hemisphere (NH) cold-season (October–March) and warm-season (April–September) months over the 1970–2015 period. There are three reasons for choosing this time period: 1) We wish to focus on the period with the largest global warming observed to date (Bindoff et al. 2013); 2) we wish to compare results derived from the CESM-LE with results derived from observations, which are relatively sparse before 1970; and 3) our analytic expression is based on a linear least squares fit to the forced signal, which is approximately linear over the selected period (the linear assumption is discussed in more detail in the final section). The simulated trends in global-mean surface temperature from the CESM-LE are not linear over the full simulation period 1920–2100; that is, the trends increase from roughly zero in the mid-twentieth century to roughly  $0.5 \text{ K decade}^{-1}$  in the latter part of

the twenty-first century (Kay et al. 2015; see Fig. 2 therein). However, they are approximately linear over the comparatively short 1970–2015 period examined here. We tested the linearity assumption by comparing residual temperature time series derived by subtracting a linear fit to the data with those derived by subtracting second- and third-order fits to the data. The higher-order fits do not significantly change the variance explained by the residual time series (see also the discussion in the appendix).

The expression for  $T_{\text{SIG}}$  is tested as follows. First, we calculate the “empirically derived  $T_{\text{SIG}}$ ” as the first time step when 29 out of 30 ensemble members exhibit trends of the same sign as that of the model forced signal in the current and all subsequent time steps. In the context of large ensembles, for a confidence level of 95%, the expression for  $T_{\text{SIG}}$  [Eq. (3)] should thus correspond to the lead time when 97.5% of all possible realizations of model reality exhibit trends of the same sign as the forced signal (for a two-tailed confidence interval). Given that the CESM-LE consists of only 30 members, 29 is the closest approximation to our 95% confidence level. Note that the empirically derived  $T_{\text{SIG}}$  does not correspond to a strict statistical quantity and is calculated primarily to explore the robustness of Eq. (3) in the context of a large ensemble of climate simulations. The additional requirement that 29 out of 30 ensemble members must also exhibit trends of the same sign as that of the model forced signal in *all subsequent* time steps is to control for any false positives in the  $T_{\text{SIG}}$  results (i.e., a  $T_{\text{SIG}}$  that has “emerged” but then falls below the 29/30 threshold at a future time step).

Second, we calculate the “analytically derived  $T_{\text{SIG}}$ ” at all grid boxes by solving Eq. (3) for  $T_{\text{SIG}}$  using 1) the ensemble-mean trends in temperature calculated over the period 1970–2015 ( $\bar{b}$ ); 2) the standard deviations of the residuals of the regression (i.e., the variability about the long-term trends;  $s$ ); and 3) the lag-1 autocorrelations of the residuals of the regression ( $r_1$ ). The ensemble-mean trends are assumed to reflect the forced signal in surface temperature. The standard deviation and lag-1 autocorrelation of the residuals are found by 1) detrending the seasonal-mean temperature time series in each of the ensemble members and at each grid box and 2) calculating the pooled standard deviations and ensemble-averaged lag-1 autocorrelations of the residual time series. The resulting values of  $s$  and  $r_1$  are assumed to reflect the amplitude and persistence of the model’s natural variability. We note that for short time series (fewer than 20 time steps) or time series with large memory, using detrended residuals from ordinary least squares can result in erroneous lag-1 autocorrelation values. In such cases, it is advisable to use a generalized least squares (GLS) estimator of the trend to calculate  $r_1$ .

In principle, the model's natural variability can be isolated using a variety of different methodologies. We have chosen to isolate the natural variability by removing a linear fit to the temperature time series in all ensemble members, since Eq. (3) is a function of the standard error of the residuals of the regression. In the appendix, we explore three alternative methods for isolating the natural variability: 1) by removing a second-order polynomial (rather than linear) fit to the gridbox temperature time series, thus retaining natural external forcings as a result of, say, volcanic eruptions and allowing for exponential changes in temperature; 2) by removing the gridbox *ensemble-mean* temperature time series from the gridbox temperature time series in all ensemble members, thus explicitly removing the signals of all forms of forced variability from the ensemble members and allowing for forced variability on a range of time scales; and 3) by using the last 1380 yr of the preindustrial control run, in which there are no natural external forcings (e.g., volcanoes or solar irradiance changes). In practice, all four methods yield very similar estimates of internal/natural climate variability and thus very similar estimates of the time of emergence (Figs. A1, A2).

Figure 1 illustrates the analytically and empirically derived values of  $T_{\text{SIG}}$  in NH wintertime surface temperatures at three grid boxes: one from Northern Hemisphere midlatitudes (at a grid box whose node is close to London, United Kingdom), one from a region of relatively high temperature variance (located in central Siberia), and one from a region of relatively low temperature variance (close to Jakarta, Indonesia). The sloping black lines in all three panels indicate the ensemble-mean trend over the 1970–2015 period at each location. As noted above, the ensemble-mean trend is interpreted as the “forced signal” of climate change. The small red dots indicate the trends in surface temperature from all 30 individual ensemble members, where the trends start in 1970 and end on the date indicated on the ordinate axis. The units on all trends are kelvin per length of the record. The dashed lines in all three panels indicate the analytically derived 95% margins of error on the forced signal, where the margins of error are derived from Eq. (2). Note that the close agreement of the 95% margins of error given by Eq. (2) (dashed lines) and the trends calculated from the large ensemble (red dots) attests to the robustness of Eq. (2) for estimating the role of natural variability in climate trends (see Thompson et al. 2015).

The analytically derived values of  $T_{\text{SIG}}$  are calculated at each terrestrial location by inserting the estimated forced signal and natural variability for each grid box into Eq. (3). For example, in the case of London, the

estimated forced signal is  $\hat{b} = 0.02 \text{ K yr}^{-1}$ , the amplitude of the natural variability is  $s = 0.6 \text{ K}$ , and the winter-to-winter autocorrelation is not significantly different from zero ( $r_1 \sim 0$ ). Inserting the above values into Eq. (3) yields  $T_{\text{SIG}} = 41 \text{ yr}$ , or 2011, which by definition is the lead time when the lower bound of the 95% confidence levels intersects zero (the intersection is marked by the vertical blue line in Fig. 1). Both the forced signal and natural variability vary from one location to the next in Fig. 1, but in general the latter dominates the variations in  $T_{\text{SIG}}$ . For example,  $T_{\text{SIG}}$  is longer over Siberia, where the interannual temperature variance is much larger ( $s = 2 \text{ K}$ ), but shorter over Indonesia, where the interannual temperature variance is relatively small ( $s = 0.2 \text{ K}$ ). The inverse relationship between regional temperature variance and the emergence of the forced signal has been noted extensively in previous studies (e.g., Christensen et al. 2007; Mahlstein et al. 2011). The key point in Fig. 1 is that the expression given in Eq. (3) for  $T_{\text{SIG}}$  clearly provides a simple and robust estimate of the first lead time at which effectively all realizations of model “reality” (as given by individual ensemble members) exhibit warming.

Figure 2 shows the results for a similar test at all terrestrial grid boxes during the NH winter and summer seasons. The top panels indicate the empirically derived values of  $T_{\text{SIG}}$  found by empirically calculating the lead time when 29 of the 30 ensemble members exhibit warming in the current and subsequent time steps. The bottom panels in Fig. 2 indicate the analytically derived values of  $T_{\text{SIG}}$  obtained from Eq. (3) (very similar results are derived for Eq. (4), since the lag-1 autocorrelation of seasonal-mean surface temperature is not significantly different from zero at most terrestrial grid boxes). Warm colors indicate relatively early signal detection times (e.g., times of emergence less than 2015). White denotes lead times that exceed the analysis period (TOE beyond 2015), while gray denotes oceans and any missing data.

The strong similarities between the top and bottom panels in Fig. 2 are important. They suggest that the lead time given by Eq. (3) provides a reliable estimate of the geographical pattern of detection time—the time at which virtually all possible realizations of model reality indicate trends of the same sign as the forced signal. They also support the assumptions that underlie Eq. (3), for example, that the natural variability is sufficiently Gaussian and that the forced signal is sufficiently linear to warrant use of the standard error of the regression. As noted in numerous previous studies (e.g., Christensen et al. 2007; Mahlstein et al. 2011; Hawkins and Sutton 2012), the forced signal in surface temperature emerges earliest in regions where the variance is smallest (i.e., the

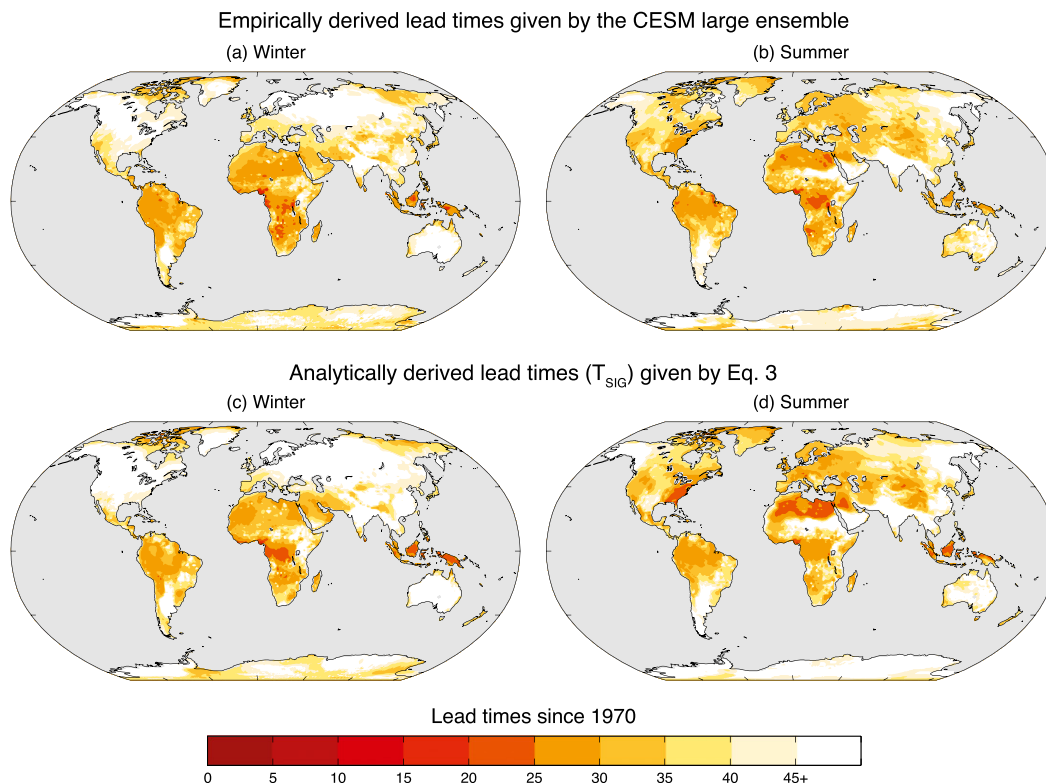


FIG. 2. Using CESM-LE output to test Eq. (3) for (left) winter and (right) summer. (top) Empirically derived lead times when the trends emerge from natural variability, calculated as the time step when 29 out of 30 ensemble members exhibit positive trends in the current and all subsequent time steps. (bottom) Analytically derived lead times ( $T_{\text{SIG}}$ ) derived by applying Eq. (3) to the model natural variability. Note that all lead times beyond the limit of the analysis period (45 yr or 2015) are white. Gray denotes ocean regions.

tropics during all seasons and the extratropics during the warm season months).

The top panels in Fig. 3 examine analogous results, but for the case where 1) the estimated forced signal ( $\hat{b}$ ) is again derived from the CESM-LE but 2) the natural variability ( $s$  and  $r_1$ ) is derived from observations of surface temperature from the HadCRUT4 dataset. The HadCRUT4 data are obtained from the Climatic Research Unit at the University of East Anglia and are analyzed on a  $5^\circ \times 5^\circ$  grid for the period January 1970–September 2015. The advantage of using observations to estimate the natural variability is that—by definition—they best reflect the variance of the “real world.” The disadvantages are that 1) the observed record may be too short to fully sample variability on decadal time scales and that 2) the observed record includes missing data and may include residual errors that influence estimates of the observed variability. As in the case of Fig. 2,  $\hat{b}$  is defined as the ensemble-mean trend from the CESM-LE over 1970–2015. In contrast to Fig. 2,  $s$  is found by 1) detrending the *observed* wintertime-mean surface temperature data at each grid box and 2) calculating the standard

deviation of the resulting time series. Note that the detrending methodology is identical to that applied to individual ensemble members (except for the pooling) in Fig. 2 and is discussed in the appendix.

Results based on the amplitude of observed natural variability are similar but not identical to those based on the natural variability displayed by the CESM-LE. Regions of strong agreement between the top panels in Fig. 3 and the bottom panels in Fig. 2 correspond to areas where the variability in the CESM-LE closely corresponds to that in the observations. Regions where the top panels in Fig. 3 and the bottom panels in Fig. 2 are notably different point to areas where differences between the simulated and observed natural variability lead to differences in the lead time when surface warming emerges in a statistically significant sense. These differences can be seen more clearly in the top panels of Fig. 4, which show the ratios of the amplitudes of natural variability derived from the CESM-LE to those derived from observations. For the most part, the CESM-LE overestimates the variance of surface temperature and thus underestimates the times of

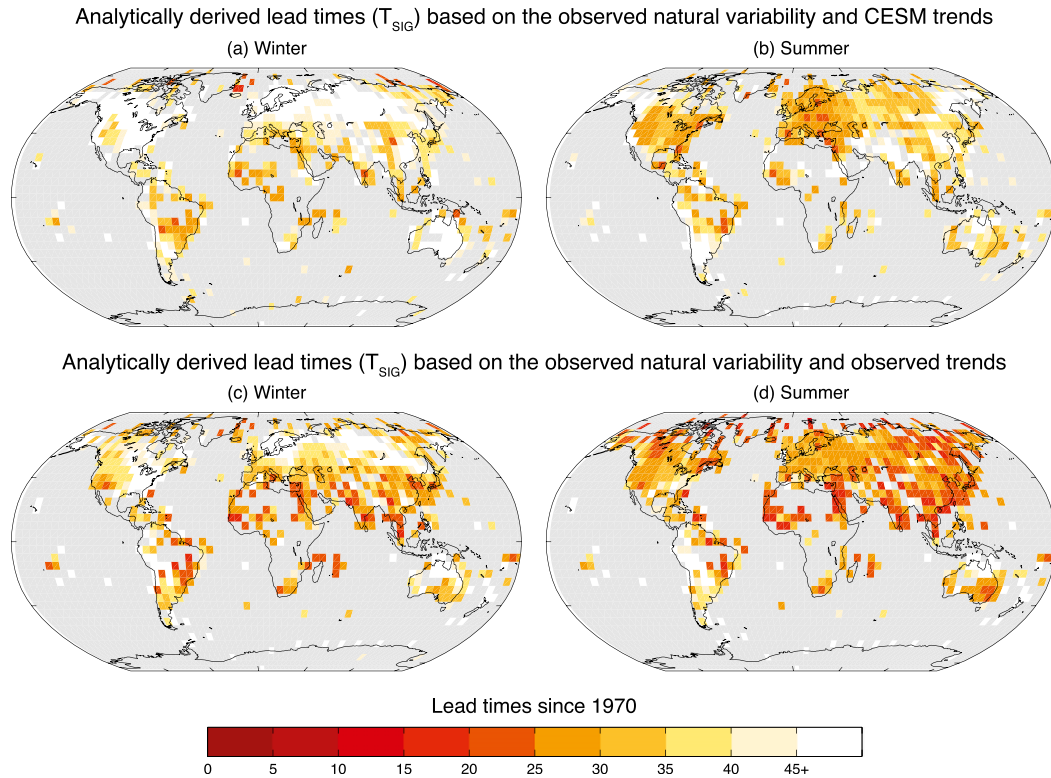


FIG. 3. As in the bottom panels in Fig. 2, but for lead times calculated by applying Eq. (3) to the (top) ensemble-mean trends from the CESM-LE and the observed natural variability and (bottom) linear trends from the observations and the observed natural variability. The observed natural variability is estimated in both panels as the standard deviation of the detrended data. The observations are used over the period 1970–2015. White denotes lead times beyond the limit of the analysis period (longer than 45 yr), while gray denotes oceans and any regions of missing data. See text for details.

emergence over much of the Northern Hemisphere midlatitudes.

The bottom panels in Fig. 3 show analogous results to those in the top panels, but in this case both the natural variability and the forced signal (the linear trends) are estimated from observations; that is,  $\hat{b}$  is defined as the linear trend calculated from observations over the period 1970–2015 and  $s$  is found in an identical manner to the top panel. The observed trends reflect only one realization of reality and are therefore noisier than the model ensemble-mean trends, particularly over regions of large temperature variance such as the Northern Hemisphere midlatitudes during winter (e.g., Deser et al. 2012a). Nevertheless, the resulting lead times are interesting in that they provide a purely observational estimate of the lead time when the observed warming emerges from the observed natural climate variability in a statistically significant sense.

The differences between the upper and lower panels in Fig. 3 arise solely from differences between trends from the CESM-LE ensemble mean and the observations.

The CESM-LE ensemble-mean trends from 1970 to 2015 are weaker than those derived from the observations over much of the tropical land areas, Europe, and East Asia during summer (see the bottom panels of Fig. 4). Hence, the purely observational lead times in these regions are shorter than those derived from the ensemble-mean trends.

Figure 5 explores whether the TOE estimates obtained solely from observations lie outside the range of TOE estimates derived from all individual ensemble members. To address this question, we calculated the TOE at all grid boxes and for all ensemble members using the individual ensemble member trends and detrended standard deviations as estimates of the forced signal and natural variability (i.e., we treated output from individual ensemble members as we treated the observations in the lower panel of Fig. 3). Interestingly, the observed TOE estimates given in the bottom panel of Fig. 3 lie within the 95% bounds on TOE estimates derived from individual ensemble members over 95% of all land areas (Fig. 5).



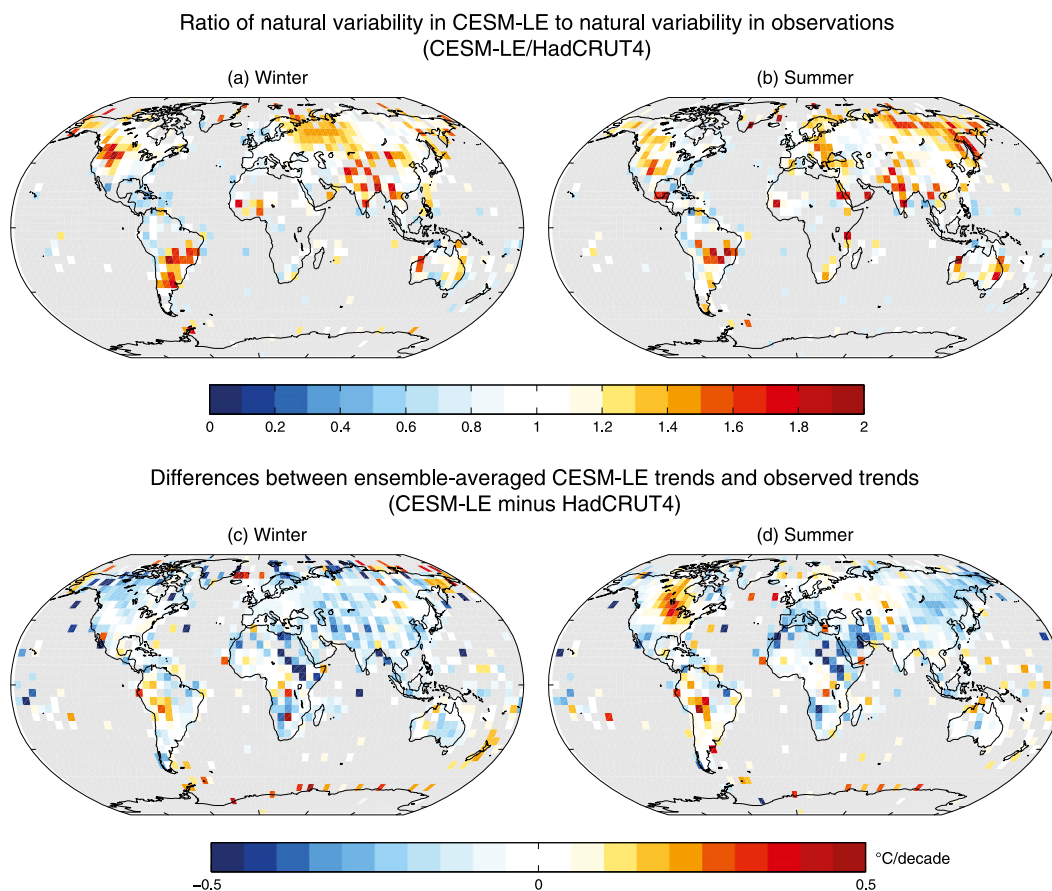


FIG. 4. (top) Ratio of amplitudes of natural variability from the CESM-LE to that from observations for (a) the winter season and (b) the summer season. Warm colors denote regions where the amplitude of natural variability is larger in the model for the 1970–2015 period, while cool colors denote regions where the amplitude of natural variability is larger in observations for the same period. (bottom) Differences between the 1970–2015 ensemble-averaged trends from the CESM-LE and trends from HadCRUT4 observations for (c) winter and (d) summer. The CESM-LE trends were used in calculating the lead times in the top panels of Fig. 3, while the observed trends were used in calculating the lead times in the bottom panels of Fig. 3. The predominance of cool colors for both seasons indicates that observed trends from 1970 to 2015 were larger than the simulated ensemble-mean trends over the same period. Gray denotes oceans and any regions of missing data.

#### 4. Discussion

The standard error of the regression is widely used in climate research. But to the best of our knowledge, it has not been explicitly used to develop an expression for the time of emergence of anthropogenic climate change. The resulting expression for  $T_{\text{SIG}}$  provides a novel and general “rule of thumb” for assessing the lead time when anthropogenic climate change will emerge from natural climate variability. The methodology has some disadvantages relative to existing methods; for example, it assumes that the natural variability is Gaussian, which is not required in existing metrics based on the Kolmogorov–Smirnov test (e.g., [Mahlstein et al. 2012](#)). However, it also has several key advantages:

- 1) The expression for  $T_{\text{SIG}}$  given by Eq. (3) [and Eq. (4) for the case where the data are not serially correlated] indicates the lead time when the forced signal of the trend has emerged in a statistically significant sense. Some previous studies explicitly consider TOE in the context of statistical significance (e.g., [Christensen et al. 2007](#); [Deser et al. 2012a](#); [Zappa et al. 2015](#)). But others consider it in the context of specific values of the natural variability. For example, consider the case of TOE defined as the first lead time when the forced signal exceeds 2 times the amplitude of the natural variability [e.g., one of the criteria outlined in [Hawkins and Sutton \(2012\)](#)]. At the grid box close to London, the TOE for  $k = 2$  in Eq. (1) occurs at a lead time of 74 yr, which is more than three decades longer than the point in time

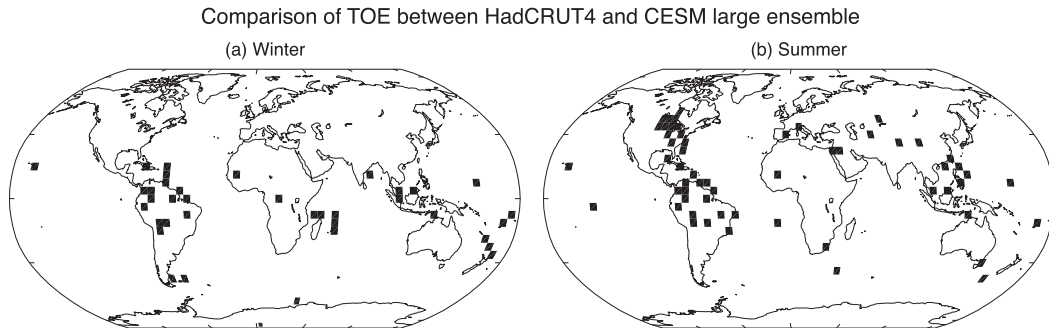


FIG. 5. Grid boxes where the TOE calculated from HadCRUT4 surface temperature observations fall outside the 95% bounds on TOE calculated for individual ensemble member trends and standard deviations. Only 5% of the observed TOE estimates lie outside the bounds given by the individual ensemble members.

when the trend is significant (Fig. 1). Similarly large differences are found throughout much of the extratropics (Fig. 6).

- 2) The expression for  $T_{\text{SIG}}$  exploits linear regression instead of epoch differences to estimate the linear trend. For example, Christensen et al. (2007), Deser et al. (2012b), and Zappa et al. (2015) all consider statistical significance when assessing the time of emergence, but they consider the differences in means between epochs of various lengths rather than linear trends. The distinction is important. Linear regression uses all of the data in a time series, while epoch differences take data only from the beginning and end of the time series. Additionally, the variance of the epoch difference estimator varies greatly depending on the length of the epoch used and is always larger than the variance of the linear trend estimator for AR(1) time series with lag-1 autocorrelations less than about 0.85 (Barnes and Barnes 2015). Thus, for all time series with a lag-1 autocorrelation less than 0.85, we believe the linear regression estimator to be preferable to epoch differences.
- 3) The expression for  $T_{\text{SIG}}$  is not subject to a multiple-testing problem. Many previous methods of calculating TOE have relied on stepping through continuous time steps and defining the TOE as the first time step when the criteria is met (e.g., the first time step when the forced signal exceeds 2 times the amplitude of the natural variability). This sequential testing increases the rate of type I errors (false positives in the results).
- 4) The expression for  $T_{\text{SIG}}$  can be solved analytically and requires no additional modifications to the data. Hence, the resulting estimate of TOE can be easily reproduced from one study to the next and readily compared across different model configurations and forcing scenarios.

## 5. Conclusions

The lead times at which anthropogenic warming and its related impacts emerge from the background noise of natural variability vary greatly from one location to the next. The expression derived in Eq. (3) provides a simple analytic tool for estimating the lead time when the regionally dependent impacts of climate change emerge

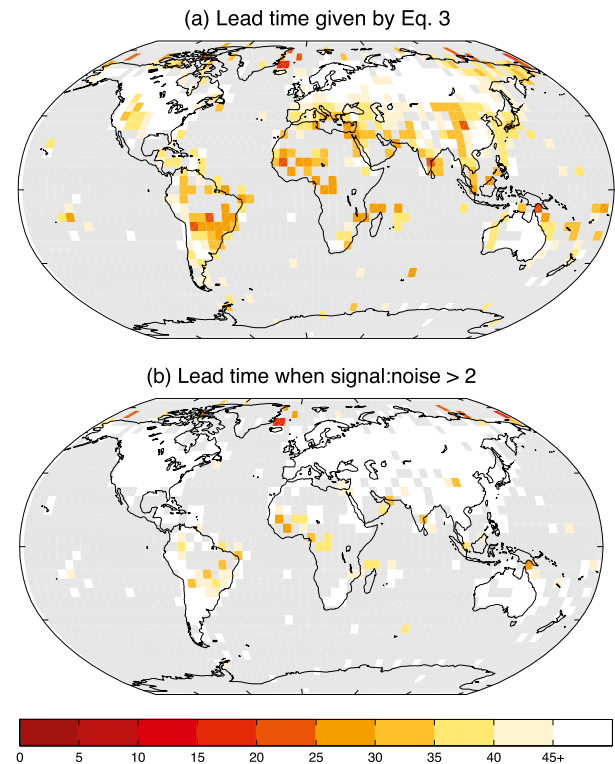


FIG. 6. Comparison between lead times calculated using (top) Eq. (3) and (bottom) Eq. (1), where  $k = 2$ . In both cases, the forced signal is given as the ensemble-mean temperature trends over 1970–2015, and the natural variability is given as the detrended observed interannual standard deviation. Gray denotes oceans and any regions of missing data. Note the top panel is reproduced from Fig. 3a.



from the natural variability in a statistically significant sense.

We have focused on the application of Eq. (3) to surface temperature, but the expression holds for any time series where the following three conditions are met: 1) the forced signal can be modeled as a linear trend; 2) the statistics of the natural variability (detrended values of the time series) are closely Gaussian; and 3) the standard deviation of the natural variability is stationary. These three assumptions derive from our use of the standard error of the regression. The bases for all three assumptions are discussed and justified in Thompson et al. (2015). The linear assumption warrants additional comment here.

In principle, the signature of anthropogenic forcing in the climate system is not necessarily linear. For example, atmospheric aerosols likely contributed to the slowdown of globally averaged warming during the mid-twentieth century (Bindoff et al. 2013), and the surface temperature trends of the next 50 yr are expected to be notably larger than those of the past 50 yr (Kirtman et al. 2013). However, in practice, the simulated response of surface temperature to greenhouse gas increases can be modeled as a linear trend on time scales shorter than roughly 50 yr, including the 1970–2015 period considered here (e.g., see Kay et al. 2015; cf. Fig. 2 therein).

The methodology outlined here is potentially useful in climate change research for four primary reasons. 1) It provides an *analytic* estimate of the lead time required for a trend to emerge and can thus be trivially calculated given (i) the amplitude and autocorrelation of the observed natural variability and (ii) the simulated forced signal. 2) It provides an estimate of the time required for a linear trend to emerge in a *statistically significant* sense, rather than as a (statistically arbitrary) factor of the internal variability. 3) It is not burdened by issues arising from sequential testing of the data, as is the case for many other TOE methods. 4) The expression requires no treatment of the data, which renders the resulting lead times easy to compare across different model configurations, different forcing scenarios, and different estimates of the natural variability.

*Acknowledgments.* We thank Benjamin D. Santer and two anonymous reviewers for their insightful and constructive comments on the manuscript. We also thank Daniel Cooley for the helpful discussions on the statistical methodology. JL and DWJT are supported by the NSF Climate Dynamics program. EAB is supported, in part, by the Climate and Large-Scale Dynamics program of the National Science Foundation under Grants AGS-1419818 and AGS-1545675. SS is supported, in part, by the Climate and Large-Scale Dynamics program of the National Science Foundation under Grant AGS-1419667.

## APPENDIX

### Estimating the Amplitude of the Natural (or Internal) Variability Used in Eqs. (2)–(4)

Figure A1 explores four different approaches for estimating  $s$  in Eqs. (2)–(4) using the CESM output: 1) removing a linear fit to the temperature time series in each of the ensemble members (as done in the main text; top panel); 2) removing a second-order polynomial (rather than linear) fit to the gridbox temperature time series in all ensemble members; 3) removing the gridbox ensemble-mean temperature time series from the gridbox temperature time series in all ensemble members; and 4) taking the last 1380 yr of the CESM preindustrial control run.

The four methods have various advantages and disadvantages. The advantages of method 1 are that (i) the residuals of the linear fit correspond directly to the residuals of the regression that form the basis for  $s$  in Eq. (3) and (ii) a similar method can be applied to observations in the absence of climate model output. The disadvantages are that (i) the anthropogenic signal is not necessarily best modeled as a linear trend; (ii) the linear fits include a component of the internal variability, since stochastic variability includes a trend component; and (iii) the linear fit does not account for externally forced variability resulting from, say, volcanic eruptions. Method 2 is similar to method 1, but it has the additional advantage that it allows for exponential changes in temperature. However, the residuals of the second-order polynomial fit do not—strictly speaking—correspond to the residuals of the regression that form the basis for  $s$  in Eq. (3). The residuals of method 3 also do not form the basis for  $s$  in Eq. (3), but removing the ensemble-mean time series arguably reflects the most robust method for removing the variability caused by all forms of external forcing in the CESM-LE, including anthropogenic forcings (e.g., as a result of increasing greenhouse gases) and external natural forcings (e.g., as a result of volcanoes). Method 4 is the simplest and most accurate for estimating pure internal variability, as the preindustrial control run does not include either anthropogenic forcings or natural external forcings such as volcanic eruptions (which are included in the forced simulations). Note that the first two methods include both internal climate variability and natural variability as a result of volcanic forcings and solar irradiance changes, whereas the latter two methods include only internal variability.

As shown in Fig. A1, all four methods yield very similar estimates of the nonanthropogenic amplitude of climate variability. But there are subtle differences. For example, the top panels in Fig. A2 show the ratios

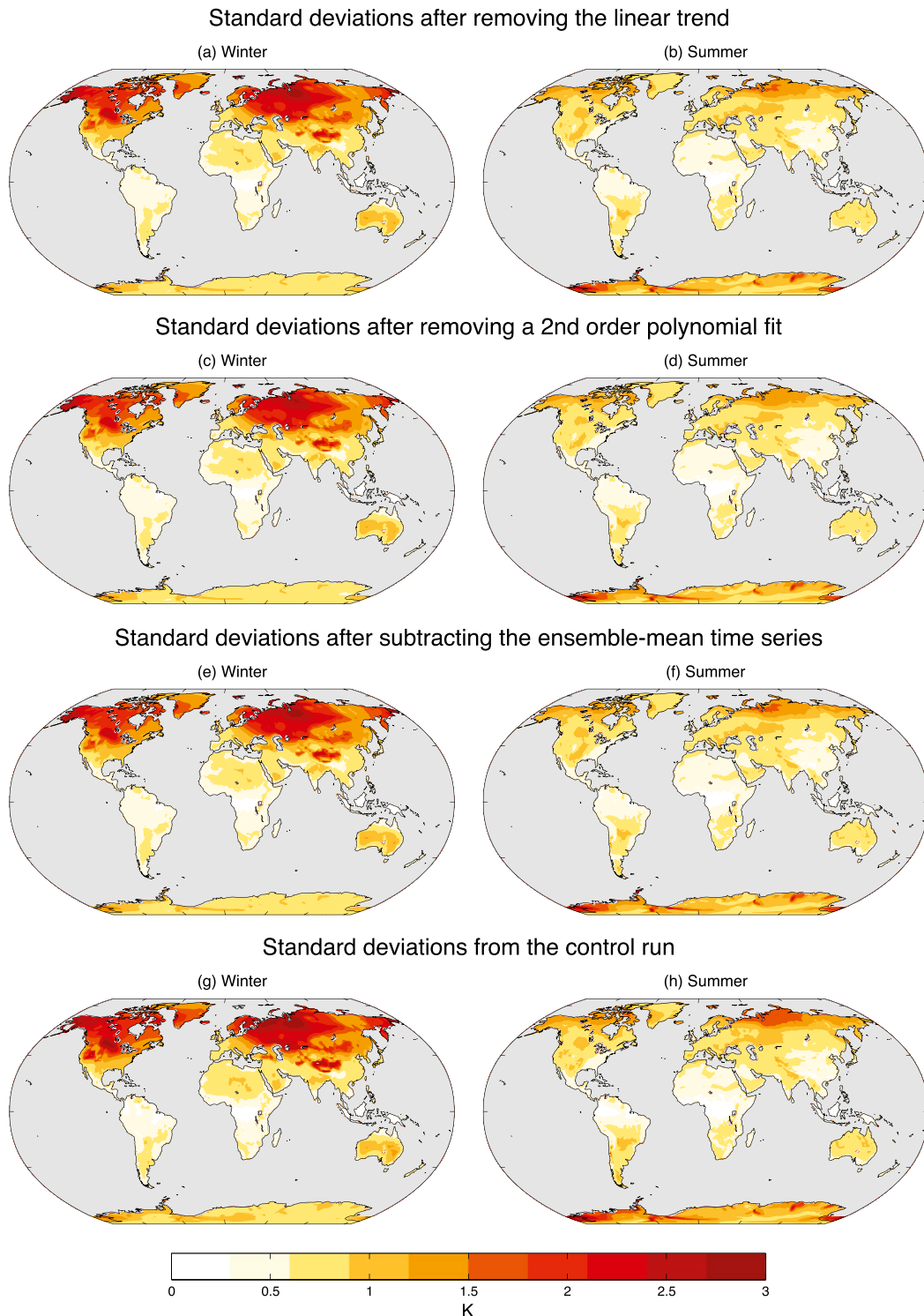


FIG. A1. Comparisons of the standard deviations calculated from the CESM-LE over the period 1970–2015 using four different methods to remove the long-term forced signal for (left) winter and (right) summer. (a),(b) Pooled standard deviations after removing the linear trend from all gridpoint time series in all ensemble members (as used for Figs. 2c,d, respectively, from the main text). (c),(d) Pooled standard deviations after removing a second-order polynomial fit from all gridpoint time series in all ensemble members. (e),(f) Pooled standard deviations after removing the ensemble-mean time series from all gridpoint time series in all ensemble members. (g),(h) Standard deviations from the last 1380 yr of the control run. Gray denotes ocean regions.

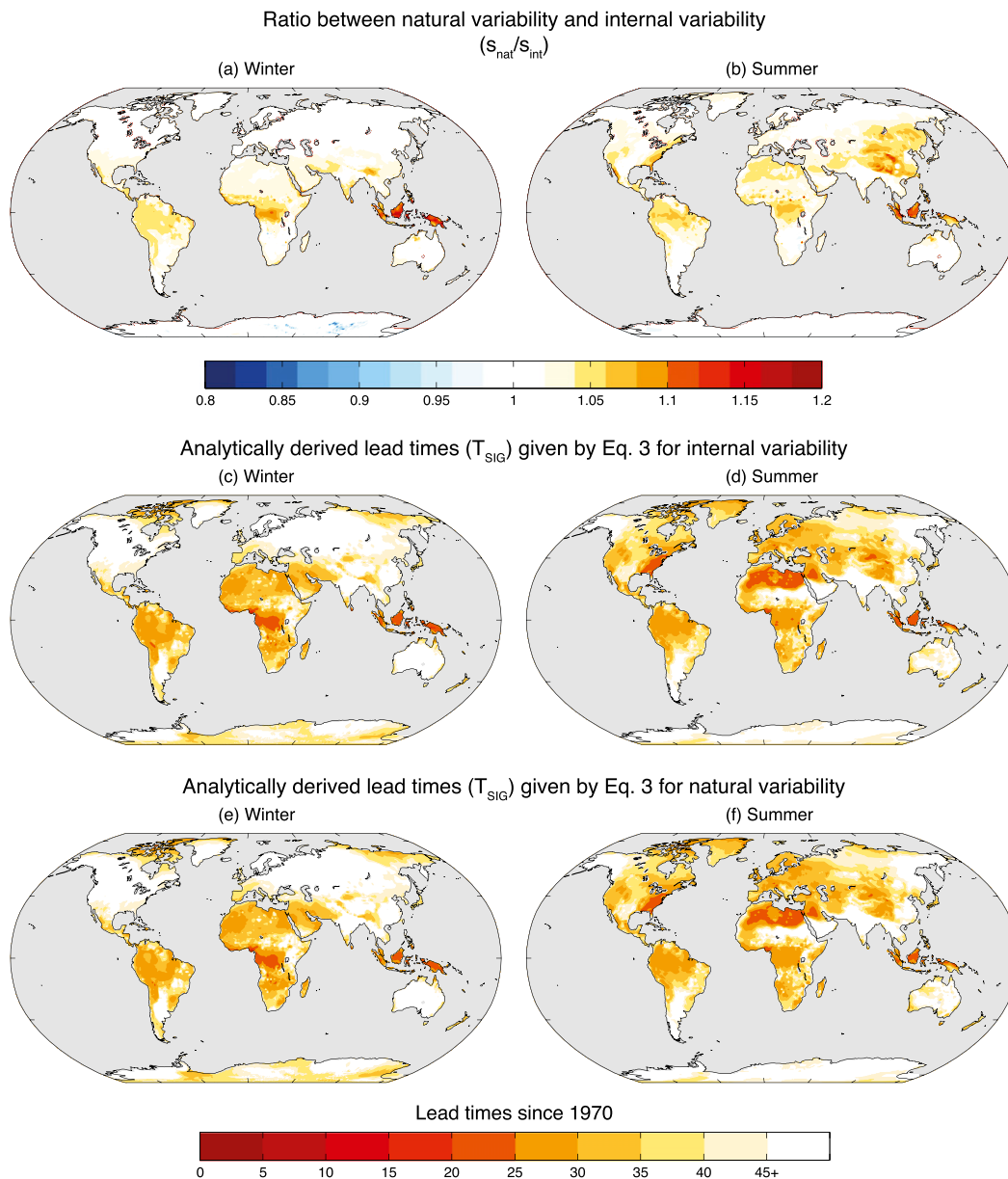


FIG. A2. Comparisons between natural variability and internal variability as calculated from the large ensemble for (left) winter and (right) summer. (a),(b) Ratio between the amplitude of natural variability [standard deviations as calculated in (a) and (b) in Fig. A1] and internal variability [standard deviations as calculated in (e) and (f) in Fig. A1]. (c),(d) Analytically derived lead times from Eq. (3) using the estimate of internal variability from Figs. A1e,f. (e),(f) Analytically derived lead times from Eq. (3) using the estimate of natural variability from Figs. A1a,b (reproduced from Figs. 2c,d, respectively). Gray denotes ocean regions.

between 1) natural variability calculated using method 1 (i.e., the method used in the main text) and 2) internal variability calculated using method 3. The dominance of warm colors (and lack of cool colors) demonstrates that the amplitude of natural variability is indeed slightly larger than that of internal variability as we expect, since natural variability includes both internal variability and natural external forcings as a result of volcanic eruptions,

etc. However, these additional sources of variability are neither large nor significant enough to show up in the standard deviations in Fig. A1. Importantly, the two methods yield nearly identical estimates of the time of emergence ( $T_{\text{SIG}}$ , bottom four panels in Fig. A2).

Last, we note that in the CESM-LE, the historical forcings with volcanic eruptions and solar irradiance changes extend only to the year 2005. After 2005, the

CESM-LE uses RCP8.5 forcing, which does not include volcanoes. Thus, the “natural variability” calculated here over the period 1970–2015 does not account for external natural forcings between 2006 and 2015. However, there have been no major volcanic eruptions during the period 2006–15, and plots comparing the amplitude of natural and internal variability calculated for the period 1970–2005 (not shown) are almost identical to those calculated for the period 1970–2015.

## REFERENCES

- Barnes, E. A., and R. J. Barnes, 2015: Estimating linear trends: Simple linear regression versus epoch differences. *J. Climate*, **28**, 9969–9976, doi:10.1175/JCLI-D-15-0032.1.
- Bindoff, N. L., 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.
- Christensen, J. H., and Coauthors, 2007: Regional climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 847–940.
- Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012a: Communication of the role of natural variability in future North America climate. *Nat. Climate Change*, **2**, 775–779, doi:10.1038/nclimate1562.
- , A. S. Phillips, V. Bourdette, and H. Teng, 2012b: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, doi:10.1007/s00382-010-0977-x.
- Diffenbaugh, N. S., and M. Scherer, 2011: Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. *Climatic Change*, **107**, 615–624, doi:10.1007/s10584-011-0112-y.
- Giorgi, F., and X. Bi, 2009: Time of emergence (TOE) of GHG-forced precipitation change hot-spots. *Geophys. Res. Lett.*, **39**, L06709, doi:10.1029/2009GL037593.
- Hawkins, E., and R. Sutton, 2012: Time of emergence of climate signals. *Geophys. Res. Lett.*, **39**, L01702, doi:10.1029/2011GL050087.
- Hurrell, J. W., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, doi:10.1175/BAMS-D-12-00121.1.
- 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.
- Kirtman, B., and Coauthors, 2013: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 953–1028.
- Madden, R., and V. Ramanathan, 1980: Detecting climate change due to increasing carbon dioxide. *Science*, **209**, 763–768, doi:10.1126/science.209.4458.763.
- 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.
- Santer, B. D., U. Mikolajewicz, W. Brüggemann, U. Cubasch, K. Hasselmann, H. Höck, E. Maier-Reimer, and T. M. L. Wigley, 1995: Ocean variability and its influence on the detectability of greenhouse warming signals. *J. Geophys. Res.*, **100**, 10 693–10 725, doi:10.1029/95JC00683.
- , T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor, 2000: Statistical significance of trends and trend differences in layer average atmospheric temperature time series. *J. Geophys. Res.*, **105**, 7337–7356, doi:10.1029/1999JD901105.
- Thompson, D. W. J., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the role of internal climate variability in future climate trends. *J. Climate*, **28**, 6443–6456, doi:10.1175/JCLI-D-14-00830.1.
- Weatherhead, E. C., and Coauthors, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, **103**, 17 149–17 161, doi:10.1029/98JD00995.
- Zappa, G., B. J. Hoskins, and T. G. Shepherd, 2015: Improving climate change detection through optimal seasonal averaging: The case of the North Atlantic jet and European precipitation. *J. Climate*, **28**, 6381–6397, doi:10.1175/JCLI-D-14-00823.1.