Forced and Internal Components of Winter Air Temperature Trends over North America During the Past 50 Years: Mechanisms and Implications

Clara Deser\textsuperscript{1*}, Laurent Terray\textsuperscript{2} and Adam S. Phillips\textsuperscript{1}

\textsuperscript{1} Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO 80305

\textsuperscript{2} Sciences de L’Univers au CERFACS, URA-1875, CERFACS/CNRS

Toulouse, France

Submitted to \textit{Journal of Climate}

April 24, 2015

*Corresponding author: Clara Deser (cdesar@ucar.edu)
Abstract

This study elucidates the physical mechanisms underlying internal and forced components of winter surface air temperature (SAT) trends over North America during the past 50 years (1963-2012) using a combined observational and modeling framework. The modeling framework consists of 30 simulations with the Community Earth System Model at 1° latitude/longitude resolution, each of which is subject to identical scenario of historical radiative forcing but starts from a slightly different atmospheric state. Hence, any spread within the ensemble results from unpredictable internal variability superimposed upon the forced climate change signal. A dynamical-adjustment technique based on constructed atmospheric circulation analogues is used to assess the contributions of dynamical and thermodynamic processes to the forced and internal components of the SAT trends. Forced trends are primarily a result of thermodynamic effects, whereas internal trends are mainly dynamically-induced. Removing the effects of internal atmospheric circulation variability narrows the model spread and brings the observed trends closer to the model’s forced response. It also enhances the signal-to-noise ratio and advances the “time of emergence” of the forced SAT component. Internal circulation trends are estimated to account for 38% of the observed warming over North America, and more than 50% locally over parts of Canada and the United States. The approach outlined here can be applied more generally to improve physical understanding and interpretation of observed and simulated climate anomalies worldwide, and may help to reconcile the diversity of SAT trends across the Coupled Model Intercomparison Project version 5 (CMIP5) models.
1. Introduction

With emerging public awareness of human-induced climate change, a major challenge is to understand and communicate the causes of recent observed climate trends, particularly at local and regional scales. Such trends are often interpreted in the context of rising anthropogenic emissions of greenhouse gases (GHG) and sulfate aerosols associated with the burning of fossil fuels. However, internally-generated variability may also contribute to regional climate changes over periods of several decades and longer (e.g., van Oldenborgh et al., 2009; Hoerling et al., 2010; Kelley et al., 2011; Ting et al., 2011; Meehl et al., 2013; Wallace et al., 2014; Johnson and Mantua, 2014; Abatzoglou et al., 2014). Distinguishing between anthropogenic and internal influences on time scales less than 50 years and spatial scales smaller than continental remains an outstanding issue (Solomon et al., 2007; IPCC, 2013).

Global coupled climate model (GCM) simulations provide estimates of the climatic impacts of anthropogenic (e.g., GHG and sulfate aerosols) and natural (volcanic and solar) radiative forcings (IPCC 2013). However, comparison of observed trends with those simulated by models is not straightforward due to that the chronological sequences of internal variability need not match (e.g., van Oldenborgh et al., 2009; Deser et al., 2012a; Wallace et al., 2014). Systematic model biases, incomplete radiative forcing specifications, and observational uncertainty further complicate direct comparison between models and observations (e.g., Raisanen, 2006; Hegerl, et al., 2007; van Oldenborgh et al., 2009).

Isolating the effects of anthropogenic climate change from those of internal multi-decadal variability is relatively straightforward in climate models provided there are
enough simulations to define the forced response (Deser et al., 2012b). That is, by averaging across ensemble members from a particular model, the random sequences of internally-generated variability in the individual realizations can be sufficiently muted to reveal the model’s response to external forcing. Once the externally-forced response is obtained, it can be subtracted from each run to find the contribution from internal variability. Most models however, including those participating in the Coupled Model Intercomparison Project phases 3 and 5 (CMIP3 and CMIP5; Taylor et al. 2012), contain too few realizations to adequately estimate the forced response on local/regional scales (Deser et al., 2012a). Note that while it is common practice to average single runs from multiple models to obtain a robust estimate of anthropogenic climate change (e.g., the “multi-model mean”; IPCC, 2013), this approach does not allow the forced and unforced components of the response to be isolated in any given model.

The use of large “initial condition” GCM ensembles to identify the relative roles of forced and internal variability in determining future climate trends has led to important insights regarding uncertainty in climate change projections at local/regional scales (e.g., Deser et al., 2012a; Hu and Deser, 2013; Fischer et al., 2013; Deser et al., 2014; Wettstein and Deser, 2014; Wallace et al., 2014; Hawkins et al., 2015; Thompson et al., 2015). Such ensembles typically include 30-40 members, each of which is subject to the identical scenario of radiative forcing but starts from a slightly different atmospheric state. The degree to which the different ensemble members diverge over time is indicative of the importance of unpredictable, internal climate variability (Deser et al., 2012b). Such ensembles are designed to sample the range of possible trend outcomes resulting from the superposition of internal climate variability and anthropogenic climate
change. In addition, the large number of integrations allows for a robust determination of
the forced response at local/regional scales.

Here we apply a new 30-member initial-condition ensemble conducted with the
National Center for Atmospheric Research (NCAR) Community Earth System Model
version 1 (CESM1; Kay et al., 2014) covering the period 1920-2100 to the understanding
of observed climate trends, with a particular focus on surface air temperature (SAT) over
North America during the past 50 years. Configured at a spatial resolution of 1°
latitude/longitude (approximately 85 km at 40°N), this ensemble provides important
context for the interpretation of nature’s one realization: specifically, the relative
importance of internal variability and external radiative forcing at local/regional scales.

In addition to apportioning trends into forced and internal contributions, we assess
the physical mechanisms underlying these components, in particular the roles of
dynamics (atmospheric circulation changes in the absence of radiatively-induced changes
in SAT) and thermodynamics (changes in SST, sea ice and land surface properties in the
absence of atmospheric circulation changes). Previous work has highlighted the
importance of dynamics to observed trends in Northern Hemisphere SAT (e.g., Wallace
et al., 1995; Hurrell, 1996; Thompson et al., 2009; Cattiaux et al., 2010; Smoliak et al.,
2015). However, none distinguished between internal and forced dynamical
contributions, a unique contribution of our study enabled by the inclusion of the CESM1
large ensemble. While the focus of this work is on winter SAT trends over North
America during the past 50 years, our methodology is generic and can be used to improve
physical understanding and interpretation of observed and simulated climate trends of
any length worldwide.
The remainder of this study is organized as follows. Section 2 contains a description of the model experiments, observational data sets, and methods for determining dynamical and thermodynamic contributions to forced and internal SAT trend components. Results are presented in Section 3 and discussed in Section 4. Implications for model evaluation and interpretation of the CMIP5 archive are included in Section 4. A summary is provided in Section 5.

2. Data and Methods

a. The CESM1 Large Ensemble (CESM-LE)

We make use of a new set of simulations conducted with the Community Earth System Model version 1/Community Atmospheric Model version 5 (CESM1/CAM5) at 1° spatial resolution, hereafter referred to as the CESM1 Large Ensemble or CESM-LE for short. The CESM-LE consists of 30 simulations for the period 1920-2100, each subject to the identical external radiative forcing but beginning from slightly different atmospheric initial conditions. Following the CMIP5 design protocol, historical natural and anthropogenic radiative forcing was applied for 1920-2005 and Representative Concentration Pathway 8.5 (RCP8.5) radiative forcing was used for 2006-2100 (Taylor et al. 2012). The first ensemble member was initialized with January 1 conditions taken from a randomly selected year of a 1700-year pre-industrial (1850) control integration (PiCTL) of the same model. This first ensemble member was then integrated forward from 1850 to 2080. Ensemble members 2-30 were all started on January 1, 1920 with initial conditions taken from the first ensemble member. A small (order $10^{-14}$ K) round-off difference to the initial air temperature field was added to each ensemble member.
This infinitesimal perturbation serves to create spread among the ensemble members as internally-generated modes of climate variability grow over time. A full description of the CESM-LE is given in Kay et al. (2014).

b. The CMIP5 ensemble

We also make use of the Coupled Model Intercomparison Project version 5 (CMIP5). Specifically, we use a single run from each of the 38 models that submitted both historical and RCP8.5 simulations to the CMIP5 archive (see Table 9.A.1 in IPCC, 2013). If multiple runs exist for a given model, we use the first one.

c. Observational data sets

We make use of three monthly mean surface air temperature (SAT) data sets: 1) Merged Land–Ocean Surface Temperature analysis (MLOST) version 3.5 (Vose et al. 2012) on a 5° latitude x 5° longitude grid; 2) GISS Surface Temperature Analysis (GISTEMP) on a 2° latitude x 2° longitude grid and smoothed with a 250km spatial filter (Hansen et al., 2010); and 3) Cowtan and Way (2014) on a 5° latitude x 5° longitude grid. We also use two monthly mean sea level pressure (SLP) data sets: 1) Twentieth Century Reanalysis version 2 (Compo et al., 2011) on a 2° latitude x 2° longitude grid, and 2) HadSLP2 on a 5° latitude x 5° longitude grid (Allan and Ansell, 2006). Results are similar for each combination of SAT and SLP datasets; for conciseness, we present results based on MLOST SAT and 20CR SLP.

d. Methods
For each data set and model simulation, we compute monthly anomalies by subtracting the long-term (1963-2012) monthly means from the corresponding month of each year. We then form 3-month winter (December-February) averages from the monthly anomalies. Finally, we compute linear trends over the 50-year period 1963-2012 using least-squares regression analysis. When computing pattern correlations and rms differences between model simulations and observations, we re-grid the model output to the observational grid.

e. Dynamical adjustment using the constructed circulation analogue approach

The objective is to empirically determine the component of SAT variability due solely to atmospheric circulation changes (e.g., the dynamically-induced contribution), with all other factors including ocean and land surface conditions held fixed at climatological values. Here we use a variation of the constructed circulation analogue method (van den Dool 2003) to find the dynamical contribution to SAT anomalies in both the CESM-LE and observations. Circulation analogues were pioneered by Lorenz (1969) and van den Dool (1994, 2003) as a statistical approach to weather prediction, and have been applied more recently to downscale climate projections from coarse resolution models (e.g Zorita et al., 1995) as well as to infer the contribution of dynamics to observed European SAT trends (Cattiaux et al., 2010). A detailed description of our methodology is provided in the Appendix; only a brief summary is given here.

Application to the CESM-LE
For each month and year of each of the 30 CESM-LE simulations, we find a set of $Na$ closest SLP analogues within the 1700-year CESM PiCTL, ranked according to their Euclidean distance from the target CESM-LE SLP field. For example, the SLP analogues for January 1920 of run 1 are found by searching all of the January SLP fields in the PiCTL. The advantage of using the model’s control run to obtain the circulation analogues is two-fold: one, there is no effect of forced climate change on either the analogues or the associated SAT anomalies; two, the number of samples from which to draw analogues is very large. We then randomly subsample $Ns$ of the $Na$ analogues, and compute their optimal linear combination that best fits the target CESM-LE SLP field. The dynamically-induced SAT anomaly field is then defined as the corresponding optimal linear combination of PiCTL SAT anomalies associated with the $Ns$ SLP analogues. We then repeat this random sampling procedure $Nr$ times. Finally, we average the $Nr$ optimal sets of SLP analogues and associated SAT anomalies to obtain a “best estimate” of the circulation-induced component of SAT anomalies in the absence of climate change. The repeated subsampling of optimal linear combinations of analogues ensures robustness of the results (see Appendix). Note that any local thermodynamic feedbacks, due for example to snow cover or soil moisture anomalies, common to all circulation analogues will be included in this “best estimate” of dynamically-induced SAT; however, any variations in feedbacks among the analogues will be omitted, as will any changes in feedbacks associated with climate change. We use the domain $[20^\circ-90^\circ N$ and $180^\circ-350^\circ E]$ for the SLP analogues following Wallace et al. (2012) and Deser et al. (2014), but there is little sensitivity to the precise region used (not shown). The results shown here are based on the parameter values $Na = 150$, $Ns = 100$, and $Nr = 50$. 


However, there is little sensitivity to the precise choice of these values as long as they lie within the parameter space that has converged (see Appendix).

Note that by using so many circulation “analogs” and by weighting them according to their degree of resemblance with the target SLP field, our technique can be viewed in some sense as a type of linear regression rather than a strict circulation analog approach. An advantage of our method compared to that of Wallace et al. (2012) and Smoliak et al. (2015) is that we are able to evaluate how much of the target SLP field is contained in our dynamical estimate. Finally, despite the different algorithms, our results are very similar to those of Wallace et al. (2012) and Smoliak et al. (2015), lending confidence to both approaches (Fig. S1).

Application to observations

To dynamically adjust the observed SAT, we use SLP analogues from 20CR over the period 1899-2012 leaving out the targeted month/year, and then construct their optimal linear set following the procedures described above. For the results shown here, we use $N_a = 80$, $N_s = 50$, and $N_r = 100$. The results are not sensitive to the precise choices of these parameters (see Appendix). Since we do not have an observational “control run” from which to draw the SLP analogues and their associated effects on SAT, we high-pass filter the SAT data before computing their dynamically-induced component to minimize any potential effects from forced, thermodynamically-induced SAT; a similar strategy was employed in Smoliak et al. (2015). Note that the dynamically-induced component is then subtracted from the full (unfiltered) SAT field. We use the perfect-model framework of the CESM-LE to test how well this approach works and to
determine the sensitivity to our choice of high-pass filter. We find that subtracting a quadratic trend or a spline trend over the full period 1920-2012 produce the closest estimates of the true unforced dynamically-induced SAT (within 8% of the true rms amplitude); 10-year or 20-year high pass filters perform less well (within 16% and 23% of the true rms amplitude: see Appendix). For simplicity, the results shown here are based on quadratic trend removal.

3. Results

a. Raw trends

Figure 1 shows maps of 50-year (1963-2012) winter SAT trends (°C per 50 years) over North America from each of the 30 members of the CESM-LE (labeled “1” … “30”). The simulated SAT trends vary widely in both pattern and magnitude across the individual model realizations, despite that they were subject to the identical radiative (e.g., GHG) forcing. Many ensemble members even show cooling over large regions. For example, runs 19 and 28 exhibit contrasting patterns, with cooling (warming) over northwestern (southeastern) North America in run 26 and opposite-signed trends in run 19. There are many other examples of individual runs that show trends of opposite polarity or disparate amplitude at a given location.

The observed SAT trends (Fig. 1, lower right panel labeled “OBS”) are positive everywhere, with the largest warming (> 4 °C) over northwestern Canada and smaller amplitude warming (< 1 °C) over the western U.S. and far eastern Canada. Some members (e.g., runs 7, 12, 24) resemble the observed trend pattern while others are nearly orthogonal (e.g., runs 14, 21, 28). Viewed in the context of the 30-member CESM-LE,
the observed SAT trend distribution lies within the range of simulated outcomes, each of which is consistent with the response to radiative forcing. As discussed further below, ensemble member 7 provides a particularly close match to the observed trends in both pattern and magnitude.

The degree of resemblance between the observed and simulated trend maps can be quantified in terms of centered pattern correlation (e.g., the area-average SAT trend is removed before computing the pattern correlation) and rms difference (Fig. 2, gray dots). Pattern correlations range from -0.7 to +0.7 and rms differences range from 0.6 to 1.8 °C per 50 years, with a close linear relationship between the two (e.g., lowest rms corresponding to highest pattern correlation and vice versa). In other words, there are numerous possible trend configurations due to the superposition of forced and internally-generated components, and no single CESM-LE run need match nature’s one realization (although given enough ensemble members, one would expect some to be realistic if the model is credible).

The externally-forced component of the simulated SAT trends can be obtained by averaging the trends from all 30 members together (Fig. 1, lower panel labeled “EM”). This component shows a generally poleward-amplified warming pattern, with magnitudes \( \sim 2-3 \, ^\circ\)C per 50 years along the Arctic border and \( \sim 0.5-1.5 \, ^\circ\)C per 50 years farther south. In addition to poleward amplification, the forced trend in the CESM-LE exhibits zonal contrasts across Canada, with smaller warming in the west compared to the east, and a local maximum over the U.S. Rocky Mountains. Areas of amplified warming may be due to local positive feedbacks associated with reductions in snow and sea ice.
The observed SAT trend distribution shows similarities and differences with the CESM-LE forced response. While both exhibit warming over the entire continent, the magnitude and pattern of warming are somewhat distinct. The observed trend pattern shows some evidence of amplified warming along the Arctic coast, similar to the model’s forced response, although missing data over the Canadian Archipelago precludes a more definitive comparison. In particular, the maximum warming (> 4 °C per 50 years) occurs over western Canada in nature compared to (< 3 °C per 50 years) over northeastern Canada and far western Alaska in the CESM ensemble mean. The pattern correlation between observations and the CESM-LE ensemble mean is only 0.12 with an rms difference of 0.9 °C per 50 years (recall that individual ensemble members exhibit much larger pattern correlations and smaller rms differences than the ensemble-mean; Fig. 2).

The diversity of SAT trends within the CESM-LE is accompanied by a variety of atmospheric circulation (SLP) trend patterns and polarities (Fig. 3). Opposite-signed trends of similar magnitude are evident over the North Pacific in individual ensemble members: for example, negative SLP trends in runs 7, 13, 17, 19 and 26 compared to positive trends in runs 8, 10, 14, 20, 21, and 30, with maximum amplitudes approximately 4-8 hPa per 50 years. By contrast, the ensemble-mean SLP trends are near zero everywhere, indicating that SLP trends in any given realization are almost entirely due to internal variability. This point will become especially important when we turn to the diagnosis of dynamical vs. thermodynamic contributions to the forced and internal components of the SAT trends.

The observed SLP trend pattern shows negative values over the northeast Pacific extending into western Canada, and weaker positive values to the south. The observed
SLP trend pattern bears some resemblance to individual ensemble members, notably runs 12 and 17, and a lack of resemblance to others (e.g., runs 8, 10, 14 and 21). Consistent with this visual impression, the pattern correlations between the simulated and observed SLP trend maps range from -0.6 to +0.8 (Fig. 4). Interestingly, there is a close relationship between the pattern correlations of the simulated SLP and SAT trends with observations (Fig. 4). That is, ensemble members that show large positive (negative) SLP trend pattern correlations with observations also show large positive (negative) SAT trend pattern correlations with observations. This linear dependence of the SLP and SAT trend pattern correlations suggests that the circulation plays an important role in the diversity of SAT trends across the CESM-LE, and by extension in the observed SAT trends.

b. Dynamical adjustment

Removing the effects of internal atmospheric circulation variability via the constructed analogue technique greatly reduces the diversity of SAT trends across the CESM-LE (Fig. 5). Not only are the patterns and magnitudes more similar among the dynamically-adjusted SAT trend maps, their amplitudes are generally smaller than their un-adjusted counterparts (recall Fig. 1). Notably, regions of cooling are largely eliminated and areas of extreme warming (> 4 °C 50 yrs⁻¹) are reduced. The remaining spread across the dynamically-adjusted SAT trend maps is largely attributable to internal thermodynamic processes, as the circulation analogue technique removes more than 90% of the internal SLP trend variance (Fig. S2). Indeed, the SLP trend maps based on the PiCTL constructed analogues are very similar to their CESM-LE targets, with pattern
correlations > 0.96 and rms differences < 0.11 hPa 50 yrs\(^{-1}\) for all ensemble members (see Fig. S3). In observations, the dynamically-adjusted SAT trend distribution (bottom right panel of Fig. 5) shows less spatial heterogeneity and a more dominant expression of poleward amplification compared to the raw SAT trend map. In particular, the warming across western Canada is considerably reduced, and the east-west contrast across the United States is largely alleviated.

Dynamical adjustment also improves the resemblance between the observed SAT trend distribution and those of the individual CESM-LE ensemble members. In particular, the rms differences are considerably lower after applying dynamical adjustment to the internal component of circulation trends, with values ranging from 0.32 to 0.71 °C 50 yrs\(^{-1}\) and an average of 0.49 °C 50 yrs\(^{-1}\) (blue circles in Fig. 2) compared to 0.62 to 1.75 °C 50 yrs\(^{-1}\) and an average of 1.15 °C 50 yrs\(^{-1}\) for the raw trends (gray circles in Fig. 2). The pattern correlations are also improved for the dynamically adjusted SAT trends, ranging from 0.01 to 0.72 with a mean of 0.37 compared to a range of -0.70 to 0.66 with a mean of 0.02 for the raw trends (Fig. 2; note that the spatial-mean trend is excluded when computing pattern correlations).

Dynamical adjustment also improves the resemblance between the observed trend pattern and the \textit{forced} SAT trend pattern obtained from the average of the CESM-LE members (Fig. 5, lower right panels labeled “OBS” and “EM, respectively). In particular, the pattern correlation between the CESM-LE ensemble mean trend and the observed trend increases from 0.12 to 0.54, and the rms difference decreases from 0.90 to 0.34 °C 50 yrs\(^{-1}\), after applying dynamical adjustment. Thus, dynamical adjustment provides an improved estimate of the anthropogenic signal in observed 50-year trends.
The raw and dynamically-adjusted SAT trends for a given location or region may be summarized in terms of probability distributions based on the set of 30 CESM-LE runs. These distributions also provide context for the single estimate from observations. Figure 6 shows results for three regions of contrasting spatial scale: North America, Western Canada (140° - 90°W), and the grid box containing Fairbanks, Alaska. Compared to the raw trend distributions (gray bars), the dynamically-adjusted ones (blue bars) are considerably narrower and more Gaussian in character. Not surprisingly, the relative degree of narrowing of the trend distributions depends on spatial scale, with the greatest reductions for Fairbanks and the smallest for North America. For example, Fairbanks shows a broad range of SAT trend values (-1.8 to +6.1 °C 50 yrs⁻¹) with 75% of the simulations in the range (-0.8 to +3.5 °C 50 yrs⁻¹). After removing the influence of internal atmospheric circulation variability, the distribution narrows to +0.1 to +2.5 °C 50 yrs⁻¹ and exhibits a more Gaussian character with ~75% of simulations within the range (+0.8 to +2.2 °C 50 yrs⁻¹).

For all three regions, the observed trends (raw and dynamically-adjusted) lie within their respective model distributions, with the raw values (solid green lines) near the upper end of the model range and the dynamically-adjusted values (dashed green lines) close to the center of the model range (e.g., the model’s forced response; Fig. 6). Note that the only model information used in the calculation of the observed dynamically-adjusted SAT trends is the CESM-LE ensemble-mean SLP trend which is nearly zero (recall Fig. 3). For North America, the observed warming trend is reduced from 2.2 to 1.4 °C 50 yrs⁻¹ after dynamical adjustment, a decrease of 36%. Put another way, internal circulation variability has augmented the warming trend due to radiative
forcing (plus a small contribution from internal thermodynamics; see below) by 57%.
The effects are larger for Western Canada, where dynamical adjustment reduces the
warming trend by 50% [from 3.6 to 1.8 °C 50 yrs⁻¹]; e.g., internal circulation variability
has doubled the forced (plus internal thermodynamic) warming trend. The effect of
dynamical adjustment for Fairbanks is smaller than that for North America and Western
Canada but still substantial, with a reduction in warming of 29% (from 2.8 to 2.0 °C 50
yrs⁻¹).

To highlight the importance of internal circulation effects on the observed SAT
trend pattern, we show a map of the contribution of internal dynamics expressed as a
fraction of the total SAT trend at each location (Fig. 7). Unforced circulation trends
contribute more than 40% of the total warming over much of western Canada and the
eastern United States, and up to 70% in some locations. In other areas such as the western
U.S., parts of Alaska and far northeastern Canada, internal dynamics slightly offsets the
warming trend.

A more general view of the impact of dynamical adjustment on the spread of SAT
trends in the CESM-LE can be assessed by comparing maps of SAT trend variance based
on the raw and dynamically-adjusted fields (Fig. 8). The raw trend variance is largest
over western Canada and Alaska, and along a narrow band extending southeastward to
the Great Lakes and over the U.S. Rocky Mountains, with maximum values ~ 3-5 [°C 50
yrs⁻¹]² (Fig. 8a). After dynamical adjustment, the SAT trend variance reduces to < 0.5 [°C
50 yrs⁻¹]² over much of the continent, with slightly higher values 0.5 – 1 [°C 50 yrs⁻¹]²
over Alaska and central Canada (Fig. 8b). In terms of percentages, dynamical adjustment
reduces the variance in SAT trends by > 80% over most of the western and southeastern
portions of the continent; smaller but still substantial reductions of 50-60% occur around Hudson Bay, the Great Lakes, central Rockies and Mexican Highlands (Fig. 8c). Note that the reduction in SAT trend diversity after dynamical adjustment is ~20% less when using standard deviation (as opposed to variance) as a metric of spread, but it is still appreciable (generally > 50% except over the midsection of the continent; Fig. S4).

Another useful metric of the utility of dynamical adjustment is in terms of signal-to-noise ratio (SNR), where the signal is defined as the forced (ensemble-mean) SAT trend and the noise as the standard deviation of SAT trends across the 30 ensemble members (Fig. 9). For the raw trends, SNR values are close to unity over most of the continent, with values < 1 over the northwestern portion of the continent, and only limited areas with values > 2 (Mexico and east of Hudson Bay; Fig. 9a). On average, the SNR over North America is 1.3, indicating that the forced climate change signal and internal variability are comparable for winter SAT trends over the past 50 years. Removing the effects of internal atmospheric circulation variability makes a notable impact on SNR (Fig. 9b). Dynamically-adjusted SAT trends show SNR values generally above two over much of the continent, roughly twice as large as those based on raw SAT trends (compare Figs. 9a and b). The highest values (> 3) occur along the Arctic border, Florida, the southwestern United States and central Mexico, with lower values (1.5 – 2) over western Canada and the north-central United States. On average, the SNR over North America is 2.3 for the dynamically-adjusted trends (compared to 1.3 for the raw trends). The ratio of dynamically-adjusted SNR to raw SNR is shown in Fig. 9c. Dynamical adjustment more than doubles the SNR over the northwestern and
southeastern portions of the continent, and increases it by at least 50% elsewhere (Fig. 9c).

c. Decomposition of forced and internal SAT trends into dynamic and thermodynamic components

As discussed in the Introduction, SAT trends from any single model run can be partitioned into forced and internal components, each of which is made up of dynamical and thermodynamic contributions. This decomposition is illustrated in Fig. 10 for member 7 of the CESM-LE, the run with an SAT trend pattern most similar to observations (recall Fig. 1). The figure panels are arranged as follows. The top row shows the total SAT trends and their constituent internal and forced parts, the latter obtained from the CESM-LE ensemble mean and the former obtained by subtracting the forced component from the total. The middle row shows the contribution of dynamics to the total, internal and forced components. The total dynamical contribution is obtained directly by applying the constructed analogue technique to the total SLP trend field for that particular run; the forced dynamical contribution is obtained by averaging the total dynamical contributions for all 30 ensemble members; the internal dynamical contribution is obtained by subtracting the forced dynamics contribution from the total dynamical contribution. Finally, the bottom row shows the contribution of thermodynamics to the total, internal and forced components, obtained as residuals from the total minus dynamical contributions. Strictly speaking, the residuals also contain dynamical contributions not accounted for in the circulation analogue technique (due to both inadequate sampling as well as methodological uncertainty). However, as discussed
in the Appendix, these errors are generally very small given the length of control run used
to identify the constructed analogues as well as the repeated random sampling and
averaging of the optimal linear combinations of constructed analogues.

As already described, run 7 features strong warming over the northern two-thirds
of the continent, with maximum values exceeding 4 °C 50 yrs⁻¹ over large regions of
Canada and the northeast United States (Fig. 10a). Atmospheric circulation trends
contribute much of the warming over western Canada by virtue of anomalous southerly
flow associated with a deepened Aleutian Low (recall Fig. 4; Fig. 10b). Thermodynamic
processes are responsible for the enhanced warming in the vicinity of Hudson Bay, the
Canadian archipelago, and northern Alaska (Fig. 10c), in association with diminished
snow cover and sea ice cover (not shown). The total dynamical component (Fig. 10b) is
mainly a result of internal circulation trends (Fig. 10e), with forced circulation trends
contributing mainly to warming along the Arctic border (Fig. 10h). On the other hand, the
total thermodynamic contribution (Fig. 10c) is largely forced (Fig. 10i); the internal
thermodynamic component (Fig. 10f) augments the forced warming over eastern Canada,
the Great Lakes region and Alaska, and partially offsets the forced warming elsewhere. In
summary, both dynamics and thermodynamics contribute to the total SAT trends: further,
the dynamical contribution is almost entirely internal whereas the thermodynamic
contribution is both internal and forced.

The same decomposition is performed for the observed SAT trends in Fig. 11.
Strictly speaking, the only partitioning that can be made based purely on observations is
the separation of the raw trends into dynamical and thermodynamic components (left
column). Using the model’s forced SLP response in conjunction with observed SLP-SAT
relationships enables separation of the dynamical contribution into internal and forced components (middle row). [Note that we have performed a separate dynamical adjustment for the total and internal SLP components, where the internal component is obtained by subtracting the ensemble-mean of the CESM-LE at each time step.] The remaining contributions require use of the model’s forced SAT (e.g., panels d, f, g, i). Note that the model’s forced SAT trends have been regridded to the 5°x5° mesh of the MLOST observations in Fig. 11g.

Dynamics contributes to warming (2-3°C 50 yrs⁻¹) across western Canada and weak cooling over northeastern Canada, western Alaska and the northwestern United States (Fig. 11b), while thermodynamics accounts for a broad pattern of warming over the entire continent, with largest amplitudes over Canada (Fig. 11c). These patterns are remarkably similar to those from CESM-LE run 7 (Figs. 10b and c), although they are based on completely independent data sets. This, in turn, lends confidence to the robustness of the decomposition and interpretation of the results in both the model and observations. The circulation-induced component of the observed SAT trends (Fig. 11b) is almost entirely internal (Fig. 11e), similar to that found for run 7 of the CESM-LE except along the Arctic border where the model shows modest warming from forced circulation changes (Fig. 10h). It is difficult to assess whether the differences along the Arctic border are a result of model shortcomings or observational data constraints, since this is a region of limited data coverage. Further, this region may be subject to thermodynamic influences from dynamically-forced changes in sea ice cover that are implicitly included in the circulation analogues from the model’s control run, but may not be adequately sampled in the short observational record. Finally, it is evident that the
thermodynamic component of the observed SAT trends (Fig. 11c) is mainly forced (Fig. 11i), with secondary contributions from internal processes that augment the warming over western Canada and Alaska and cool much of the U.S. (Fig. 11f), in line with the results from run 7. In summary, observed DJF SAT trends over the past 50 years are largely the result of internal dynamics and forced thermodynamics, with secondary contributions from internal thermodynamics.

A complementary, temporal view of the relative contributions of internal and forced dynamics and thermodynamics to North American SAT anomalies during 1920-2012 in run 7 of the CESM-LE is shown in Fig. 12 (observational results are shown in Fig. S5). The top pair of curves contrasts the raw (black) and dynamically-adjusted (magenta) SAT records. While the raw time series displays prominent interannual fluctuations superimposed upon a long-term warming trend beginning in the early 1960s, the dynamically-adjusted record exhibits more muted variability and a smaller but more monotonic rise in SAT starting in the late 1970s (similar results are obtained when both forced and internal dynamics are removed; not shown). Thus, not only does dynamical adjustment alter the timing and amplitude of the long-term warming trend, it also provides for a more stable estimate of these parameters due to the reduction in interannual variance (noise in this context). The second pair of curves contrasts the forced (red) and internal (blue) components of North American SAT. The forced time series bears a close resemblance to the dynamically-adjusted record, with even less interannual variability, while the internal time series is dominated by high-frequency fluctuations. The third pair of curves compares the thermodynamic (brown) and dynamical (orange) contributions to forced SAT. It is clear that thermodynamics dominates, with dynamics
making a small but non-negligible contribution to the forced component of warming in recent decades. The final pair of curves compares the thermodynamic (green) and dynamical (cyan) contributions to the internal SAT record. Both are primarily high frequency in character, with larger amplitudes for the dynamical component than the thermodynamic one. In summary, dynamical adjustment leads to an improved estimate of the forced component of the North American SAT time series, and provides a more stable estimate (e.g., less subject to sampling fluctuations) of the timing and amplitude of low-frequency trends.

e. Time-of-emergence

Many studies have examined the “time of emergence” (TOE) of the anthropogenic climate change signal, estimated by evaluating when the forced response first exceeds a given amplitude of internal variability (typically, one or two standard deviations; e.g., Mahlstein et al., 2011; Diffenbaugh and Scherer, 2011; Deser et al., 2012b). Here we assess the effect of dynamical adjustment on TOE. Figure 13a shows a map of TOE from the CESM-LE based on 10-yr running means of DJF SAT anomalies relative to 1920-1949. To compute TOE, we find the year when the forced (ensemble-mean) 10-year running mean SAT anomaly first exceeds and remains above one standard deviation of the internal variability (computed across the 30 SAT anomalies at each time step after applying a 10-year running mean). We also compute TOE based on dynamically-adjusted data (internal dynamics removed; Fig. 13b). This “adjusted” TOE will generally be earlier than the raw TOE due to the smaller amplitude of internal variability in the dynamically-adjusted SAT (recall Fig. 12).
The map of TOE based on the raw data (Fig. 13a) shows that the forced signal in DJF SAT has not yet emerged above the internal variability over a wide swath of the western portion of the continent (gray shading), and has only recently emerged (since 2005) over much of the inter-mountain west, the central portion of North America, and far eastern Canada. The earliest TOE values (1980s) occur along the Arctic border and the southeastern United States. Dynamical adjustment advances the TOE considerably at all locations, and nearly all areas are now “emergent” (Fig. 13b). Specifically, the dynamically-adjusted TOE values are in the 1970s over Alaska and northern Canada, the 1980s over much of the eastern half of the continent, the 1990s over the western U.S., and after 2005 over the mid-section of the U.S. and along the western portion of the Canadian-US border (Fig. 13b). This advancement of TOE is entirely due to the reduction of interannual variance in the dynamically-adjusted data compared to the raw data, since the forced signal is the same in both calculations.

3. Discussion

The utility of dynamical adjustment for attributing SAT trends was demonstrated in the seminal studies of Wallace et al. (2012) and Smoliak et al. (2015) using a variety of techniques. These studies focused primarily on the period 1965-2000, an interval characterized by a prominent trend in the Northern Annular Mode with attendant effects on winter SAT over North America and Eurasia. Similar to our results, both find that dynamics account for approximately 40% of the total cold season SAT trend over land poleward of 40°N during 1965-2000. As shown in Smoliak et al. (2015), the spatial pattern of the dynamical contribution to SAT trends over North America during 1965-

A key result of our study is to demonstrate and make use of the fact that the forced component of SLP anomalies in any given year (and in trends over the last 50 years) is small compared to the internal component. Note that the forced component of SLP trends is also near zero in the CMIP5 multi-model archive (Fig. S6). By using the model’s estimate of the forced SLP response in conjunction with the observed relationships between SLP and SAT anomalies, we were able to evaluate the internal and forced dynamical contributions to observed SAT trends. Further, by incorporating the model’s forced SAT response, we were able to decompose the remaining portion of the observed SAT trends into internal and forced thermodynamic components.

Our methodology, based on a combination of observations, a large initial-condition ensemble of historical simulations, and a technique for estimating the dynamical contribution, has general applicability beyond the specific purpose of this study. In particular, it can be used to inform attribution of observed climate anomalies on a near real-time basis, as well as improve the physical understanding of differences in climate changes simulated by different models. As with any empirical method, inherent uncertainties due to limited sampling of circulation statistics in the short (~ 100 year) observational record must be taken into account. This problem is ameliorated to a great extent in models for which lengthy control simulations exist. There is also the caveat that
as models improve and resolve finer spatial scales, their forced SLP responses may change.

Although dynamical adjustment greatly reduces the diversity of SAT trends across the 30-member CESM-LE, differences remain, especially in the vicinity of Hudson Bay, the Great Lakes and Alaska (Fig. 8b). For example, dynamically-adjusted warming trends over much of Canada exceed 4 °C in run 4 but are < 1.5 °C in run 27 (Fig. 5). What processes contribute to the remaining spread in dynamically-adjusted SAT trends? The pattern of residual SAT trend variance (Fig. 14a) suggests that snow cover and sea ice changes may play a role. Indeed, a similar pattern is obtained by regressing trends in dynamically-adjusted SAT onto trends in sea ice concentration averaged over Hudson Bay (Fig. 14b). Our physical interpretation of this regression pattern is that internal trends in Hudson Bay sea ice (possibly initiated by atmospheric forcing) feedback onto SAT locally and over adjacent land areas via thermodynamic processes. In addition, associated changes in snow cover west of the Canadian and U.S. Great Lakes may amplify these thermodynamic SAT feedbacks. We speculate that similar physical mechanisms are at work over Alaska.

As discussed in Hawkins and Sutton (2009) and other studies, the CMIP5 archive contains three sources of uncertainty: model response uncertainty due to structural differences amongst models, radiative forcing uncertainty, and uncertainty due to internal variability. How large is the contribution of structural model uncertainty vs. internal variability uncertainty to the spread of SAT trends within CMIP5 for a given radiative forcing scenario, and does internal atmospheric circulation variability play a role? This question is difficult to address without a sufficient number of ensemble members to
define the forced response in each model, although results based on a previous 40-
member initial-condition ensemble with Community Climate System Model version 3
(CCSM3) suggested an important role for internal variability in the spread within CMIP3
(Deser et al., 2012a). However, we can gain some insight into these questions by
comparing the SLP and SAT trends simulated by the CMIP5 models with those simulated
by the CESM-LE. Qualitatively, the diversity of SAT and SLP trend patterns and
magnitudes within the set of CMIP5 model runs is reminiscent of that within the CESM-
LE, although there appear to be fewer cases of strong negative SAT trends (Fig. S6).
Particularly noteworthy is the wide range of circulation trends whose multi-model mean
(panel labeled “EM” in Fig. S6) is near zero, just as in the CESM-LE. A more
quantitative comparison is provided in Fig. 15a, which shows pattern correlations and
rms differences between the SAT trends from each of the 38 CMIP5 models against
observations. The set of CMIP5 model runs shows a similar range of pattern correlations
(albeit with fewer large negative values) and rms differences as the CESM-LE,
suggesting that internal variability may be an important source of SAT trend spread
within the CMIP5 archive (compare red and gray dots in Fig. 15a). The range of pattern
correlations between the observed and simulated SLP trends is also comparable across
the CMIP5 models and the CESM-LE ensemble, and importantly, the SLP and SAT
pattern correlations show a similar relationship in the two model ensembles (Fig. 15b).
Taken together, these results are suggestive of an important role for internal dynamics (as
opposed to model response uncertainty) in the diversity of North American SAT trends
within the CMIP5 archive. Further investigation of the role of internal variability vs.
model uncertainty in the CMIP5 archive is clearly warranted.
Another important implication of our results pertains to model evaluation. A common metric for climate model assessment is the skill with which observed trends during recent decades are simulated. However, as this study highlights, a single simulation with a credible model need not match the observed climate trajectory at local/regional scales if the contribution from internal variability is comparable or larger than that from external forcing. Conversely, a single integration with a model that is lacking in realism may show fortuitous agreement with the observed trends. The methodology outlined in this paper, including application of dynamical adjustment in conjunction with an estimate of the forced circulation response, can be used to provide a more informative and reliable assessment of a model’s ability to simulate observed climate trends.

Thompson et al. (2015) have recently argued that climate model evaluation would be well served to focus on two simple statistical properties of the unforced interannual variability: standard deviation ($\sigma$) and autocorrelation. These two parameters (especially $\sigma$) dictate the confidence intervals that can be placed on trends of any length due to internal variability, provided the data are normally distributed and stationary in time (see discussion in Thompson et al., 2015). In this context, it is worth noting that the CESM-LE simulates generally realistic magnitudes (within 10% - 20% of observations in most regions) and spatial patterns of SAT and SLP $\sigma$ for both unfiltered and 8-year low-pass filtered data (Fig. S7).

4. Summary
By combining observations with a 30-member initial-condition ensemble of CESM1 coupled model simulations (the CESM-LE) and making use of “dynamical adjustment”, we have provided insight into the mechanisms of internal and forced components of winter SAT trends over North America during the past 50 years. Our results show that at local/regional scales, the simulated SAT trends are strongly influenced by internal variability, with an average signal-to-noise ratio of only 1.3. The importance of internal variability is evidenced by the variety of spatial patterns, amplitudes and polarities of trends among the 30 simulations, each of which is subject to the identical scenario of historical radiative forcing. The range of model solutions spans the single realization of the real world, providing important context for the interpretation of observed long-term trends.

The internally-generated component of SAT trends within the CESM-LE is largely dynamically-induced, whereas the forced component is primarily thermodynamically controlled, either directly via radiative effects from increased GHGs or indirectly via changes in SSTs, sea ice and snow cover. This follows from the fact that the simulated SLP trends are almost entirely a result of internal climate variability, with a negligible forced component. For the real world, we estimate that internal circulation trends account for approximately 38% of the observed wintertime warming over North America during the past 50 years, and more than 50% locally over parts of Canada and the United States. Removing the effects of internal atmospheric circulation variability via a constructed analogue technique narrows the spread within the CESM-LE, thereby enhancing the signal-to-noise ratio of the simulated SAT trends (by a factor of two, on average) and advancing the “time of emergence” of the forced SAT component (by
approximately a decade in many locations). Dynamical adjustment also brings the observed trends closer to the model’s forced response, both in terms of pattern and amplitude, facilitating their interpretation.

The methodological framework proposed in this study provides a general template for improving physical understanding and interpretation of observed and simulated climate trends worldwide. Application to other seasons, regions, time periods and parameters will be pursued in future work.

Acknowledgements

We have benefited from useful discussions with Drs. Julien Boé, Brian V. Smoliak, David W. J. Thompson, and John M. Wallace during the course of this work. We also acknowledge the efforts of all those who contributed to producing the CESM-LE, and thank the Climate and Global Dynamics Division at NCAR for supporting L. Terray’s sabbatical visit. NCAR is sponsored by the National Science Foundation.

Appendix

Constructed analogue technique for dynamical adjustment

Here we provide additional details on our constructed analogue methodology for dynamical adjustment. As described in the main text, we first determine the Na closest analogues (Xc) from the PiCTL for each time step of a particular CESM-LE run (Xh), and then randomly draw Ns of the Na analogues. We then estimate a constructed SLP analogue Xca as a linear combination of these Ns analogues according to:

\[
X_{ca} = \sum_{i=1}^{Ns} w_i X_{ci}
\]
where \( X_c \) is a matrix of column vectors comprising the selected \( N_s \) closest analogues of the column vector \( X_h \), and \( \beta \) is a column vector of the fitted regression coefficients that are the linear proportions of the contributions of each column of \( X_c \) to the constructed analogue \( X_{ca} \). The dimensions of \( X_h \) are \( m \times 1 \), where \( m \) is the number of grid points contained in the SLP pattern. The dimensions of \( X_c \) are \( m \times N_s \) and that of \( \beta \) are \( N_s \times 1 \). The \( \beta \) coefficients can be estimated by using the Moore-Penrose pseudo-inverse of \( X_c \):

\[
\beta = \left[ \left( X_c^T \cdot X_c \right)^{-1} \cdot X_c^T \right] \cdot X_h
\]

In practice, we calculate \( \beta \) using a singular value decomposition of \( X_c \). The \( \beta \) coefficients are then applied to the \( N_s \) SAT patterns from the PiCTL simulation corresponding to the selected SLP analogues to reconstruct the internally-generated dynamically-induced component of SAT in the CESM-LE run.

For each month, we repeat the drawing of the \( N_s \) analogues \( N_r \) times. We then end up with \( N_r \) different samples of dynamically-induced SAT. Note that these differ mainly because of thermodynamically-induced internal variability. We then take the mean of the \( N_r \) samples as the final dynamically-induced SAT. The spread among the \( N_r \) samples gives a lower bound for the spread due to thermodynamically-induced internal variability. We repeat all the steps for each month of the CESM-LE simulation to obtain a
complete reconstruction of dynamically-induced SAT that run. Finally, we repeat the entire algorithm for each member of the CESM-LE.

Figure A1 shows the sensitivity of the results to the choice of closest $Ns$ analogues, for both observations and run 7 of the CESM-LE. The SLP trends obtained from the constructed analog methodology closely match the actual SLP trends in both observations and CESM-LE run 7 for a wide range of $Ns$ values (5-50 for observations and 10-100 for the model run; note that the observations were only computed for $Ns \leq 50$). Similarly, only minor differences in the total estimated dynamical contribution to SAT trends are found between $Ns=50$ and $Ns=5$ for observations (Fig. A1 b and c, respectively) and between $Ns=100$ and $Ns=10$ for the model run (Fig. A1 e and f, respectively), although the dynamical contribution to simulated SAT trends is slightly greater when more closest analogs are used.

The sensitivity of the results to $Nr$, quantified in terms of rms error, is shown in Fig. A2 for both SAT and SLP. The shading indicates the range across all 30 CESM-LE simulations, and the colored curve denotes the average over the ensemble. It is clear that the results converge for $Nr > 20$, and that there is less spread for SLP compared to SAT across all $Nr$.

As discussed in Section 2e of the main text, we have used the perfect-model framework of the CESM-LE to determine which high-pass filter to use for dynamically adjusting the observations. Using run 7 from the CESM-LE, we can subtract the forced SLP and SAT responses (obtained from the CESM-LE ensemble mean) to obtain the unforced “residual” component of variability. We can then find circulation analogues and associated SAT anomalies from this residual record, and apply our dynamical adjustment.
procedure to obtain the “true” unforced dynamically-induced component in run 7. We can then compare this “true” dynamical estimate to that obtained without knowledge of the forced response (as is the case when dealing with observations). The latter is found by high-pass filtering the SAT record over the period 1920-2012 and then applying our constructed circulation analog protocol. Figure A3 shows the dynamical SAT trend contribution for run 7 of the CESM-LE based on a) the “true” estimate, and based on 4 different choices of high-pass filters: b) quadratic trend removal, c) spline trend removal, d) 20-year high-pass filter, and e) 10-year high-pass filter. All four choices yield similar patterns of dynamical contribution, but the quadratic and spline trend removals are the closest to the “true” estimate, while the two high-pass filters underestimate the true dynamical contribution. Specifically, the rms of the “true” dynamical SAT trend distribution shown in Fig. A3a is overestimated by 2% (8%) using the spline (quadratic) trend removal, and underestimated by 16% (23%) using the 10-year (20-year) high-pass filter.

References


Hoerling, M. P., J. Eischeid, and J. Perlwitz, 2010: Regional precipitation trends:


Zorita, E., J. P. Hughes, D. P. Lettenmaier and H. von Storch. 1995: Stochastic characterization of regional circulation patterns for climate model diagnosis and
**Figure Captions**

**Figure 1.** Winter SAT trends (1963-2012; °C 50 yrs⁻¹) for each member of the CESM-LE (labeled 1-30), the CESM-LE ensemble-mean trend (labeled “EM”), and observations (MLOST; labeled “OBS”).

**Figure 2.** Pattern correlation (x-axis) vs. rms difference (y-axis; °C 50 yrs⁻¹) of DJF SAT trends (1963-2012) over North America for each member of the CESM-LE against observations (MLOST; 20CR). Gray dots are for total trends, and blue dots are for dynamically-adjusted trends (internal dynamics removed). See text for explanation.

**Figure 3.** As in Fig. 1 but with SLP trends superimposed (contour interval is 1 hPa 50 yrs⁻¹, negative values are dashed and the zero contour is thickened).

**Figure 4.** Pattern correlations between simulated and observed 1963-2012 winter SAT (x-axis) and SLP (y-axis) trends. Each dot represents a different CESM-LE run. Pattern correlations are based on the Pacific-North American domain for SLP and North America for SAT. SAT and SLP observation are from MLOST and 20CR, respectively.

**Figure 5.** As in Fig. 1 but internal circulation effects removed (see text).

**Figure 6.** Histograms of DJF SAT trends (1963-2012; °C 50 yrs⁻¹) for a) North America, b) Western Canada, and c) Fairbanks, Alaska. Gray (blue) bars denote raw (dynamically-adjusted) trends from the CESM-LE. Solid (dashed) green vertical lines denote raw...
(dynamically-adjusted) trends from observations. Note the different horizontal and vertical axis ranges in each panel.

**Figure 7.** DJF SAT trends (1963-2012). a) Total ('C 50 yrs⁻¹), b) contribution from internal dynamics ('C 50 yrs⁻¹), and c) fractional contribution from internal dynamics.

**Figure 8.** Variance of DJF SAT trends [1963-2012; ('C 50 yrs⁻¹)^2] across the CESM-LE (a) before and (b) after dynamical adjustment. The proportion of total trend variance (%) accounted for by dynamical adjustment is shown in panel c.

**Figure 9.** Signal-to-noise ratio (SNR) for winter SAT trends (1963-2012) from the CESM-LE based on: a) raw and b) dynamically-adjusted data. The ratios of the SNR values in b) divided those in a) are shown in panel c). Signal is defined as the ensemble-mean trend, and noise is defined as the standard deviation of the trends across the 30-members.

**Figure 10.** Decomposition of DJF SAT trends (1963-2012; °C 50 yrs⁻¹) into internal, forced, dynamical and thermodynamic components for run 7 of the CESM-LE. See text for details.

**Figure 11.** As in Fig. 10 but for observations (MLOST).
Figure 12. Timeseries decomposition of DJF SAT anomalies (°C) averaged over North America from run 7 of the CESM-LE into internal, forced, dynamical and thermodynamic components. The top panel shows the raw (black) and dynamically-adjusted (magenta; internal dynamics removed) components. The second panel from the top shows the free (blue) and forced (red) components. The second panel from the bottom shows the forced thermodynamics (brown) and forced dynamics (orange) components. The bottom panel shows the free thermodynamics (green) and free dynamics (cyan) components. Note the different vertical scales for each set of curves. See text for explanation.

Figure 13. Maps of time-of-emergence of forced winter SAT anomalies based on 10-year running means from the CESM-LE for a) raw and b) dynamically-adjusted data. See text for details. Grey areas denote grid boxes where the forced signal has not yet emerged by 2012.

Figure 14. a) Standard deviation of dynamically-adjusted DJF SAT trends [1963-2012; (°C 50 yrs⁻¹)] across the CESM-LE. b) Regression map of dynamically-adjusted SAT trends onto the Hudson Bay sea ice trend index. See text for details.

Figure 15. a) Pattern correlation vs. rms difference (°C 50 yrs⁻¹) of DJF SAT trends (1963-2012) over North America for each member of the CESM-LE (gray dots) and each member of the CMIP5 archive (red dots) against observations (MLOST). b) Pattern correlations between simulated and observed 1963-2012 winter SAT and SLP trends.
Gray and red dots are for the CESM-LE and CMIP5 simulations, respectively. Pattern correlations are based on the Pacific-North American domain for SLP and North America for SAT.

Figure A1. Total estimated dynamical contribution to SLP and SAT trends (1963-2012) as a function of the number of closest $N$s analogues used. Observational results are shown in the top row, and model results based on run 7 of the CESM-LE are shown in the bottom row. Panels a) and d) show the actual SLP trends, and panels b) – f) show the constructed analog estimates (SAT in color shading; °C 50 yrs$^{-1}$ and SLP in contours; contour interval = 1hPa 50 yrs$^{-1}$) based on different choices of $N$s as stated in the panel titles.

Figure A2. RMS Error (RMSE) as a function of Nr (“Number of averaged iterations”) for a) SAT (°C 50 yrs$^{-1}$) and b) SLP (hPa 50 yrs$^{-1}$) DJF trends (1963-2012) from the CESM-LE. Shading indicates the range across an ensemble of 200 averaging iteration steps that differ by a random initial sequence of Nr SAT and SLP trends. The colored curve denotes the average over the ensemble.

Figure A3. Dynamical contribution to SAT trends (1963-2012; °C 50 yrs$^{-1}$) in run 7 of the CESM-LE. Panel a) shows the “true” estimate, and the remaining panels show the estimates based on b) quadratic trend removal, c) spline trend removal, d) 20-year high-pass filter, and e) 10-year high-pass filter. See text for explanation.
Figure 1. Winter SAT trends (1963-2012; °C 50 yrs⁻¹) for each member of the CESM-LE (labeled 1-30), the CESM-LE ensemble-mean trend (labeled “EM”), and observations (MLOST; labeled “OBS”).
Figure 2. Pattern correlation (x-axis) vs. rms difference (y-axis; °C 50 yrs⁻¹) of DJF SAT trends (1963-2012) over North America for each member of the CESM-LE against observations (MLOST; 20CR). Gray dots are for total trends, and blue dots are for dynamically-adjusted trends (internal dynamics removed). See text for explanation.
Figure 3. As in Fig. 1 but with SLP trends superimposed (contour interval is 1 hPa 50 yrs⁻¹, negative values are dashed and the zero contour is thickened).
Figure 4. Pattern correlations between simulated and observed 1963-2012 winter SAT (x-axis) and SLP (y-axis) trends. Each dot represents a different CESM-LE run. Pattern correlations are based on the Pacific-North American domain for SLP and North America for SAT. SAT and SLP observation are from MLOST and 20CR, respectively.
**Figure 5.** As in Fig. 1 but internal circulation effects removed (see text).
Figure 6. Histograms of DJF SAT trends (1963-2012; °C 50 yrs⁻¹) for a) North America, b) Western Canada, and c) Fairbanks, Alaska. Gray (blue) bars denote raw (dynamically-adjusted) trends from the CESM-LE. Solid (dashed) green vertical lines denote raw (dynamically-adjusted) trends from observations. Note the different horizontal and vertical axis ranges in each panel.
Figure 7. DJF SAT trends (1963-2012), a) Total (°C 50 yrs⁻¹), b) contribution from internal dynamics (°C 50 yrs⁻¹), and c) fractional contribution from internal dynamics.
Figure 8. Variance of DJF SAT trends [1963-2012; (°C 50 yrs⁻¹)²] across the CESM-LE (a) before and (b) after dynamical adjustment. The proportion of total trend variance (%) accounted for by dynamical adjustment is shown in panel c.
Figure 9. Signal-to-noise ratio (SNR) for winter SAT trends (1963-2012) from the CESM-LE based on: a) raw and b) dynamically-adjusted data. The ratios of the SNR values in b) divided those in a) are shown in panel c). Signal is defined as the ensemble-mean trend, and noise is defined as the standard deviation of the trends across the 30-members.
Figure 10. Decomposition of DJF SAT trends (1963-2012; °C 50 yrs⁻¹) into internal, forced, dynamical and thermodynamic components for run 7 of the CESM-LE. See text for details.
Figure 11. As in Fig. 10 but for observations (MLOST).
Figure 12. Timeseries decomposition of DJF SAT anomalies (°C) averaged over North America from run 7 of the CESM-LE into internal, forced, dynamical and thermodynamic components. The top panel shows the raw (black) and dynamically-adjusted (magenta; internal dynamics removed) components. The second panel from the top shows the free (blue) and forced (red) components. The second panel from the bottom shows the forced thermodynamics (brown) and forced dynamics (orange) components. The bottom panel shows the free thermodynamics (green) and free dynamics (cyan) components. Note the different vertical scales for each set of curves. See text for explanation.
Figure 13. Maps of time-of-emergence of forced winter SAT anomalies based on 10-year running means from the CESM-LE for a) raw and b) dynamically-adjusted data. See text for details. Grey areas denote grid boxes where the forced signal has not yet emerged by 2012.
**Figure 14.** a) Standard deviation of dynamically-adjusted DJF SAT trends [1963-2012; (°C 50 yrs⁻¹)] across the CESM-LE. b) Regression map of dynamically-adjusted SAT trends onto the Hudson's Bay sea ice trend index. See text for details.
Figure 15. a) Pattern correlation vs. rms difference (°C 50 yrs⁻¹) of DJF SAT trends (1963-2012) over North America for each member of the CESM-LE (gray dots) and each member of the CMIP5 archive (red dots) against observations (MLOST). b) Pattern correlations between simulated and observed 1963-2012 winter SAT and SLP trends. Gray and red dots are for the CESM-LE and CMIP5 simulations, respectively. Pattern correlations are based on the Pacific-North American domain for SLP and North America for SAT.
Figure A1. Total estimated dynamical contribution to SLP and SAT trends (1963-2012) as a function of the number of closest $N_s$ analogues used. Observational results are shown in the top row, and model results based on run 7 of the CESM–LE are shown in the bottom row. Panels a) and d) show the actual SLP trends, and panels b) – f) show the constructed analog estimates (SAT in color shading; °C 50 yrs$^{-1}$ and SLP in contours; contour interval = 1hPa 50 yrs$^{-1}$) based on different choices of $N_s$ as stated in the panel titles.
Figure A2. RMS Error (RMSE) as a function of Nr (“Number of averaged iterations”) for a) SAT (°C 50 yrs-1) and b) SLP (hPa 50 yrs-1) DJF trends (1963-2012) from the CESM-LE. Shading indicates the range across an ensemble of 200 averaging iteration steps that differ by a random initial sequence of Nr SAT and SLP trends. The colored curve denotes the average over the ensemble.
Figure A3. Dynamical contribution to SAT trends (1963-2012; °C 50 yrs⁻¹) in run 7 of the CESM-LE. Panel a) shows the “true” estimate, and the remaining panels show the estimates based on b) quadratic trend removal, c) spline trend removal, d) 20-year high-pass filter, and e) 10-year high-pass filter. See text for explanation.