Title

Improvements in Wintertime Surface Temperature Variability in the Community Earth System Model Version 2 (CESM2) Related to the Representation of Snow Density

Journal

Journal of Advances in Modeling Earth Systems, 14(4)

ISSN

1942-2466

Authors

Simpson, Isla R
Lawrence, David M
Swenson, Sean C
et al.

Publication Date

2022-04-01

DOI

10.1029/2021ms002880

Peer reviewed

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2021MS002880

Key Points:

- Northern Hemisphere wintertime surface temperature variability was overestimated in Community Earth System Model Version 1 and is improved in Community Earth System Model Version 2 (CESM2)
- This improvement has arisen through an increase in snow density and associated snow conductance
- A reduced ensemble spread and reduced decline in daily variance is also found in future projections of surface temperature with CESM2

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

I. R. Simpson,
islas@ucar.edu

Citation:

Simpson, I. R., Lawrence, D. M., Swenson, S. C., Hannay, C., McKinnon, K. A., & Truesdale, J. E. (2022). Improvements in wintertime surface temperature variability in the Community Earth System Model Version 2 (CESM2) related to the representation of snow density. *Journal of Advances in Modeling Earth Systems*, *14*, e2021MS002880. https://doi.org/10.1029/2021MS002880

Received 21 OCT 2021
Accepted 9 MAR 2022

Improvements in Wintertime Surface Temperature Variability in the Community Earth System Model Version 2 (CESM2) Related to the Representation of Snow Density

Isla R. Simpson[1], David M. Lawrence[1], Sean C. Swenson[1], Cecile Hannay[1], Karen A. McKinnon[2], and John E. Truesdale[1]

[1]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO, USA, [2]Department of Statistics and Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA

Abstract The Community Earth System Model (CESM) is widely used for the prediction and understanding of climate variability and change. Accurate simulation of the behavior of near surface air temperature (T_{2m}) is critical in such a model for addressing societally relevant problems. However, previous versions of CESM suffered from an overestimation of wintertime T_{2m} variability in Northern Hemisphere (NH) land regions. Here, it is shown that the latest version of CESM (CESM2) exhibits a much improved representation of wintertime T_{2m} variability compared to its predecessor and it now compares well with observations. A series of targeted experiments reveal that an important contributor to this improvement is the local effects of changes to the representation of snow density within the land surface component. Increased snow densities in CESM2 lead to enhanced conductance of the snow layer. As a result, larger heat fluxes across the snow layer are induced in the presence of T_{2m} anomalies, leading to a greater dampening of surface and near surface atmospheric temperature anomalies. The implications for future projections with CESM2 are also considered through comparison of the CESM1 and CESM2 large ensembles. Aligned with the reduction in surface temperature variability, compared to CESM1, CESM2 exhibits reduced ensemble spread in future projections of NH winter mean temperature and a smaller decline in daily wintertime T_{2m} variability under climate change. Overall, this improvement has increased the accuracy of CESM2 as a tool for the study of wintertime T_{2m} variability and change.

Plain Language Summary A societally relevant quantity that is predicted by Earth System Models is near surface air temperature (T_{2m}). Accurate simulation of this quantity requires accurate representation of atmospheric circulation, boundary layer processes and land-atmosphere interaction. Here we show substantial improvements in the representation of wintertime T_{2m} variability in the latest version of the Community Earth System Model and isolate the roles of different aspects of the model development. Increases in the density of snow in the new model are shown to reduce T_{2m} variability. Snow density governs the heat fluxes induced across the snow layer via conductance and increased snow density in the model has increased the ability of heat fluxes across the snow layer to dampen T_{2m} variability. The implications of this change for future climate projections are explored and it is found that CESM2 exhibits a reduced uncertainty in future projections of mean T_{2m} over Northern Hemisphere land regions, as might be expected given the reduced sampling uncertainty in T_{2m} due to reduced internal variability. Furthermore, wintertime T_{2m} variability is expected to decrease under climate change and this decrease is larger in the older version of the model that exhibited greater T_{2m} variability in its historical simulation.

1. Introduction

Numerical models of the coupled ocean-atmosphere-land system are routinely used to provide projections of near surface weather and climate on timescales ranging from short term weather forecasts, to seasonal-to-decadal climate predictions, to projections of long term climate change over the coming century. Given its relevance to humans and ecosystems that reside at the Earth's surface, one of the primary fields of interest from such models is near surface air temperature, at 2 m above the ground (referred to as T_{2m} hereafter). While T_{2m} is one of the most easily, and therefore most widely, measured meteorological quantities, it is not necessarily straightforward to simulate it with fidelity given its dependence on a multitude of factors including large scale meteorological influences, the representation of unresolved processes in the boundary layer, and interactions with the land surface.

Interactions with the land surface are known to play an important role in summertime T_{2m} variability when drier soils can exacerbate heat extremes by altering the proportion of incoming energy that is partitioned into latent heating (e.g., Durre et al., 2000; Fischer et al., 2007; Seneviratne et al., 2010; Vargas Zeppetello et al., 2020). In the wintertime, at extratropical latitudes where surface temperature variability is greatest, the advective influences of the large scale circulation are widely considered to dominate (e.g., Holmes et al., 2016). Indeed, a number of studies have examined the leading order characteristics of wintertime temperature variability by considering free tropospheric levels (e.g., 850 hPa), therefore assuming a relatively small influence of land-atmosphere coupling in shaping temperature distributions near the surface (Linz et al., 2019; Schneider et al., 2015). While many of the features of the wintertime surface temperature distribution can indeed be explained this way, it does not preclude the possibility that the details of land-atmosphere coupling may also be important in ultimately governing how temperature variability behaves near the surface during the winter. Indeed, a number of studies have demonstrated the importance of temporal variability in snow cover or snow cover characteristics for winter and spring surface temperature variability via albedo effects and hydrologic effects on soil moisture (Diro et al., 2018; Dutra et al., 2011; Fischer et al., 2011; Xu & Dirmeyer, 2011).

The topic of this study is the representation of daily average wintertime T_{2m} variability in the Community Earth System Model (CESM). The second generation version of CESM (CESM2) was released in 2018 (Danabasoglu et al., 2019). Compared to its predecessor, CESM1 (Hurrell et al., 2013), CESM2 includes major upgrades to most of its components. In particular, as the atmospheric component (the Community Atmosphere Model, CAM) transitioned from CAM5 to CAM6, almost every physical parameterization, except radiation, was upgraded (Bogenschutz et al., 2018; Danabasoglu et al., 2019). Similarly, as the land component (the Community Land Model, CLM) transitioned from CLM4 to CLM5, the representation of many existing processes was updated and the representation of many new processes was introduced (Lawrence et al., 2019). As will be shown below, CESM1 substantially overestimated wintertime daily T_{2m} variability in the Northern Hemisphere (NH) high latitudes and this bias has now been alleviated in CESM2. This is not a bias that developers had targeted during the development process. Rather, this improvement has emerged as a result of upgrades that had been implemented with other motivations in mind. The aim of this study, therefore, is to document this change, provide an explanation of how this improvement has arisen and discuss the implications of this improvement for future climate projections.

As will be shown, changes in the representation of snow density in CLM5 are of central importance to this change in T_{2m} variability, so we briefly describe CLM5's upgrades in snow density and density evolution but refer readers to van Kampenhout et al. (2017) for a more complete description. van Kampenhout et al. (2017) implemented the changes described below with the primary motivation of improving the representation of perennial snow and firn over the Greenland ice sheet, and elsewhere, where high winds and extreme cold temperatures prevail. However, the new parameterizations of snow density and densification have been implemented globally within the model and here we illustrate their collateral benefit on the representation of surface temperature variability over seasonally snow covered continental regions of the NH. Snow density in CLM is governed by both the density of fresh snow as it lands on the ground and the subsequent density evolution (or densification) in response to environmental factors. The representation of both fresh snow density and densification have changed in CLM5, as now described.

1. **Fresh snow density.** In CLM4, the density of fresh snow varied as a function of temperature only, with higher densities at warmer temperatures (Figure 1a of van Kampenhout et al. (2017)). This temperature dependence obeyed a functional form that was derived from a single high elevation measurement site in Alta, Utah (Anderson, 1976). Given substantial differences between conditions at this site and conditions on the Greenland ice sheet, van Kampenhout et al. (2017) upgraded this parameterization in two ways. First, they added a linear dependence of fresh snow density on wind speed (Figure 1b of van Kampenhout et al. (2017)), motivated by the fact that windy conditions lead to enhanced breaking of snow crystals, a smaller effective snow grain size (Sato et al., 2008) and, therefore, an enhanced ability of snow crystals to pack tightly into a denser snow layer. Second, they modified the functional form of the temperature dependence of snow density to account for the fact that the crystal size gets smaller at very cold temperatures, leading to an inversion of density at the lowest temperatures. The overall dependence of fresh snow density on temperature and wind speed resulting from both these changes can be seen in Figure 1c of van Kampenhout et al. (2017) and demonstrates that there is now a greatly increased likelihood for fresh snow to be denser than it was in CLM4

2. **Densification.** Various processes lead to increased density of snow once it is on the ground and a number of changes to these processes have been made in CLM5. First, destructive metamorphism allows the snow to densify as water molecules move along the snow crystals through sublimation and condensation. In CLM, the rate of densification due to this process is dependent on temperature and then tapers off exponentially when the snow density reaches some specified upper limit. From CLM4 to CLM5, the value of this upper limit has increased from 100 kgm^{-3} to 175 kgm^{-3}, which would allow this densification process to remain important up to higher densities. Second, snow can densify due to compaction associated with pressure from the snow layers above. In CLM5, a change has been made to the formulation of viscosity that is used to determine the rate of compaction by this process, which ultimately will reduce the densification due to compaction by overburden pressure - a change that was necessary to improve the representation of firn on the Greenland ice sheet (Section 3.4.2 of van Kampenhout et al. (2017)). Finally, in CLM5, a parameterization of the densification effects of drifting snow has been introduced. Drifting snow leads to higher densities as it causes the snow crystals to break, allowing them to pack more densely. A representation of this process has been introduced in a parameterized way in CLM5, with the densification being dependent on both the mobility of the snow and wind speed (Vionnet et al., 2012)

van Kampenhout et al. (2017) showed that, over Greenland, enhanced near surface snow densities associated with these changes lead to improvements in sub-surface melt rates, in association with enhanced conductance and an improved vertical redistribution of melt. In addition, the changes to the representation of snow compaction by overburden pressure result in an improved representation of firn. Here, we will show that these changes, in particular the density of fresh snow, have also led to improvements in surface temperature variability over seasonally snow covered NH land regions and we discuss the mechanisms behind this impact.

The simulations and observation-based datasets used in this study are described in Section 2. In Section 3, the changes in T_{2m} variability between CESM1 and CESM2 are described and the role for changes in the representation of snow density is identified. In Section 4, a mechanistic understanding of the snow density influence is provided and in Section 5 the implications for future climate projections are assessed before discussion is provided in Section 6 and conclusions are drawn in Section 7.

2. Model Simulations, Observation Based Datasets and Methods

2.1. Model Simulations

We use of a range of pre-existing CESM simulations as well as newly performed simulations aimed at isolating the ultimate cause of the T_{2m} variability change between CESM1 and CESM2. These are listed in Table 1 and use prescribed historical and future forcings of either the Phase 5 or Phase 6 era of the Coupled Model Intercomparison Project (CMIP) and are performed with a 1.25 ° longitude × 1 ° latitude horizontal resolution and 32 layers in the vertical with a model top at ∼40 km. Each of these simulations vary in length and all of our analyses of the historical period will use years 1979–2014. The exception is some of the sensitivity experiments that only extend to 2005 and for these we use 1979–2005.

2.1.1. Coupled Simulations

To explore the change in present day temperature variability and future projections between CESM1 and CESM2 with fully coupled simulations we use the CESM1 large ensemble (LENS1, Kay et al. (2014)) and the CESM2 large ensemble (LENS2, Rodgers et al. (2021)). LENS1 is a 40-member ensemble of coupled simulations using CESM1, initialized from 1920 and run under CMIP5 historical forcings to 2005 and forcings of the Representative Concentration Pathway 8.5 (RCP8.5), thereafter. LENS2 is a 100-member ensemble of coupled simulations using CESM2, initialized from 1850 and run under CMIP6 historical forcings to 2014 and forcings of the Shared Socioeconomic Pathway 3–7.0 (SSP3-7.0), thereafter. In LENS1, the ensemble members are initialized from the same state but with a random noise purturbation applied to the air temperature field to introduce ensemble spread (micro initialization). In LENS2, a mixture of micro and macro initializations are used, where "macro" refers to the initialization of members from different dates from the coupled pre-industrial control simulation. As detailed in Rodgers et al. (2021), some bug fixes and a change to the biomass burning emissions were introduced between the first and second 50 members of LENS2. However, since our focus here is primarily on sub-seasonal

Table 1
Description of Community Earth System Model Simulations (*) Only Two of These Members are Used to Force the Single Column Atmospheric Model (SCAM) Simulations and Only One Is Used for the Detailed Surface Energy Budget Analysis (**) Only One of These Is Used for the Detailed Surface Energy Budget Analysis

Name	# members	Length	Description
LENS1	40	1920–2100	CESM1 large ensemble (Kay et al., 2014). Coupled historical + RCP8.5 (after 2005)
LENS2	89	1850–2100	CESM2 large ensemble (Rodgers et al., 2021). Coupled historical + SSP3-7.0 (after 2014)
GOGA1	10	1880–2017	CESM1. Historical + RCP8.5 (after 2005), ERSSTv4 SSTs
GOGA2	10	1850–2020	CESM2. Historical + SSP3-7.0 (after 2014), ERSSTv5 SSTs
Sensitivity Experiments (all within the CESM2 codebase, using Hurrell et al. (2008) SSTs and BGC in CLM turned off)			
CAM6_CLM5	3*	1979–2014	Similar to GOGA2 but with different SSTs and land BGC turned off that is, these are standard "out-of-the-box" AMIP runs with CESM2.
CAM5_CLM4	1	1979–2005	As CAM6_CLM5 but reverting CAM6 back to CAM5 and CLM5 back to CLM4
CAM6_CLM4	1	1979–2005	As CAM6_CLM5 but reverting CLM5 back to CLM4
CAM5_CLM5	1	1979–2005	As CAM6_CLM5 but reverting CAM6 back to CAM5
SNWDENS	2**	1979–2005	As CAM6_CLM5 but reverting the snow density and densification settings in CLM5 back to those in CLM4
Single column model (SCAM) experiments (all within the CESM2 codebase)			
SCAM6_CLM5_CLM5F	2	1979–2014	Large scale forcing from CAM6_CLM5 and CLM5 as the land component
SCAM6_SNWDENS_SNWDENSF	2	1979–2014	Large scale forcing from SNWDENS and CLM5 with reverted snow densisty and densification settings as the land component
SCAM6_SNWDENS_CLM5F	2	1979–2014	Large scale forcing from CAM6_CLM5 with reverted snow density and densification settings as the land component

variability, we do not expect these changes to have a substantial impact and we consider both sets together. Furthermore, at the time of writing, only 89 members were available, so 89 members are used.

2.1.2. Prescribed SST Simulations

To isolate whether there is a role for the ocean and sea ice in T_{2m} variability changes between CESM1 and CESM2, we compare the coupled simulations with two 10-member ensembles run with prescribed Sea Surface Temperatures (SSTs) taken from observations (GOGA simulations, referring to Global Ocean Global Atmosphere). GOGA1 uses CESM1 and extends from 1880 to 2017 with CMIP5 historical forcing prior to 2006 and RCP8.5 forcing thereafter. The observed SSTs used in this ensemble are taken from ERSSTv4 (Huang et al., 2014). GOGA2 uses CESM2 and extends from 1850 to 2020 with CMIP6 historical forcing prior to 2015 and SSP3-7.0 forcing thereafter. The observed SSTs used in this ensemble are taken from ERSSTv5 (Huang et al., 2017).

Table 2
A Summary of the Namelist Setting That are Relevant to Snow Density and Densification in Community Earth System Model Version 1 (CESM1) (Left), Community Earth System Model Version 2 (CESM2) (Middle) and a Description of What They Represent (Right)

CESM1	CESM2	Description
lotmp_snowdensity_method = 'TruncatedAnderson1976'	lotmp_snowdensity_method = 'Slater2017'	Controls the density of fresh snow
wind_dependent_snow_density = .false.	wind_dependent_snow_density = .true	
upplim_destruct_metamorph = 100.d00	upplim_destruct_metamorph = 175.d00	Controls the upper density limit for destructive metamorphism
overburden_compress_tfactor = 0.08d00	snow_overburden_compaction_method = 'Vionnet2012'	Controls viscosity used in overburden compaction and the drifting snow parameterization
snow_overburden_compaction_method = 'Anderson, 1976'		

2.1.3. Sensitivity Experiments

The above simulations will reveal that the ocean and sea-ice are not involved in the T_{2m} variability change between CESM1 and CESM2 (see Section 3), so we will continue our investigations using uncoupled simulations with prescribed observation-based SSTs as listed in the middle section of Table 1. The following sensitivity experiments are designed to isolate the relative roles of the atmosphere component (CAM), the land component (CLM) and ultimately which land surface parameterization changes are responsible for the land influence. In the transition from CESM1 to CESM2, CAM has evolved from CAM5 to CAM6 and CLM has evolved from CLM4 to CLM5. We make use of a 3 member ensemble of simulations from 1979 to 2014 run with CESM2 using the "FHIST" component set and refer to this as CAM6_CLM5. These are "out-of-the-box" simulations with prescribed SSTs based on the Hurrell et al. (2008) data set and are essentially the same as GOGA2, but differ in the prescribed SSTs and also in land biogeochemistry (BGC), which was turned on in GOGA2 and is off in CAM6_CLM5. We use these rather than GOGA2 as our CESM2 baseline as the SSTs and BGC settings are consistent with those in the following sensitivity experiments.

CAM6_CLM5 will then be compared with a series of experiments in which various components or parameterizations are reverted back to those in CESM1 (to the extent possible) within the CESM2 codebase. In CAM5_CLM4 the atmosphere component is CAM5 and the land component is CLM4 that is, both have been reverted back to their CESM1 components. In CAM6_CLM4, the atmosphere is CAM6 but the land is CLM4 that is, only the land component is reverted back. In CAM5_CLM5, the atmosphere has been reverted to CAM5 but the land is still CLM5 that is, only the atmosphere component is reverted back and CMIP5 forcings are used. Finally, we perform simulations using both CESM2 components (CAM6 and CLM5) but we revert only the snow density and densification settings back to those of CLM4. Specifically, this involves the namelist changes listed in Table 2 and we refer to these simulations as SNWDENS.

2.1.4. Single Column Atmospheric Model Experiments

To isolate the relative roles of local versus non-local influences of the snow density, we perform a series of experiments with the Single Column Atmospheric Model (SCAM) (Gettelman et al., 2019), all of which use CAM6 atmospheric physics (summarized in the bottom portion of Table 1). SCAM solves only for the column physics in the atmosphere and coupling with the land surface at a single grid point, while the influence of the large scale atmospheric circulation is prescribed by forcing data. In our case, this forcing data is taken from 2 members from each of CAM6_CLM5 and SNWDENS at three NH snow-covered locations, two in Canada and one in Russia: Saskatoon (254°E, 52°N); Toronto (280°E, 44°N); and Siderovsk (83°E, 66°N). These locations were chosen at random to sample locations in the NH high latitudes where T_{2m} variability has changed between CESM1 and CESM2. The SCAM experiments, therefore, consist of two members from 1979 to 2014 and we will refer to them as SCAM6_X_YF where X denotes the configuration of CLM being used (either CLM5 for using default CLM5 or SNWDENS with the snow settings reverted) and Y denotes the configuration of CLM in the CESM experiment that was used to generate the forcing data (either CLM5 if the forcing data comes from the CAM6_CLM5 simulations or SNWDENS if the forcing data comes from the SNWDENS simulations). SCAM

uses the Eulerian dynamical core for the vertical advection calculation, which is different from the Finite Volume dynamical core used in the 3-dimensional simulations with CAM. To initialize the land in SCAM, the same initial state that was used for the CAM simulations was interpolated onto the relevant grid point on the Eulerian grid. In order to prevent drift, aerosol species (mass and number) in these SCAM simulations are relaxed toward the seasonally varying climatology from the run that was used to provide the large scale forcing and temperature is relaxed toward that of the initial state. To limit the impact of this relaxation on T_{2m} we increase the relaxation timescale from 2 days at the model top to 60 days at the surface.

2.1.5. Coupled Model Intercomparison Project, Phase 6 (CMIP6) Simulations

We compare our CESM results with those from 20 models from CMIP6 that had daily T_{2m} data available at the time of writing. These models are listed in the titles of Figure 13 and we use the data from the first member of each model from "historical" simulations for the period 1979–2014.

2.2. Observation Based Datasets

We will compare daily T_{2m} variability in CESM with that in three different observation-based products over the period 1979–2014: the Berkeley Earth Surface Temperature daily product (BEST, Rohde et al., 2013); station data from the Integrated Surface Database (ISD, Smith et al., 2011); and ERA5 reanalysis (Hersbach et al., 2020). BEST has been derived from station-based observations using the process described in Rohde et al. (2013) but the daily product is still considered "experimental" at the time of writing. For the ISD station data we use daily average T_{2m} for stations that have more than 20 years of record within the period 1979–2014 and, of those, we only use the stations that have values for more than 80% of days. The majority of stations in ISD are located at airports and we have chosen this data set since it provides daily averages that are obtained from high time resolution sampling. ERA5 is the latest reanalysis product from the European Center for Medium Range Weather forecasts and it assimilates screen level temperature as well as snow depth and density (Hersbach et al., 2020). There is no perfect method to compare models with observations: station based data have the potential to be representing different spatial scales than the coarser model grid cells, while gridded products such as ERA5 and BEST may have errors related to the underlying methods used in producing them. We, therefore, use all three datasets to confirm the robustness of our conclusions as well as the validity of comparisons with each of these products. Temperature at 850 hPa ($T850$) from ERA5 reanalysis is also used.

In Section 6 we provide a comparison of the behavior of surface fluxes between the model and both ERA5 and the FLUXNET2015 (Pastorello et al., 2020) data set. The ERA5 surface fluxes are determined from the forecast integration and as a result, have the potential to be subject to deficiencies in the underlying forecast model of ERA5. We, therefore, also diagnose the daily average T_{2m} analysis increments for ERA5 as the difference between the analysis and the forecast. In practise this is computed as the average of two values: the difference between the analysis at time 06 hr and step 12 of the forecast that was initialized at 18 hr on the previous day; and the difference between the analysis at time 18 hr and step 12 of the forecast that was initialized at 06 hr of the same day. As such, these increments represent the extent to which the underlying forecast model of ERA5 drifts from reality over the course of a 12 hr forecast window.

For FLUXNET2015 we use: air temperature (FLUXNET2015 variable TA_F); sensible heat flux (FLUXNET2015 variable H_F_MDS); and Net Radiation (FLUXNET2015 variable NETRAD). We use the daily values which are the daily averages of half hourly measurements and we only use days where a given variable has a quality control flag of greater than 0.5 indicating that more than half of the values contributing to the daily average are either measured, or good quality gap filled data. Stations are only shown if they provide all three of these variables, have a record length of 10 years or more, and if the DJF averaged snow fraction at the location according to CAM6_CLM5 is greater than 0.5, which leaves 10 stations with record lengths ranging from 10 to 17 years.

2.3. Methods

The majority of our analysis will be performed using sub-seasonal anomalies, defined as follows. At each grid point, the seasonally varying daily climatology is determined by averaging over the years considered. The seasonal cycle is then defined as the first four harmonics of this seasonally varying climatology and this is subtracted from the daily values to produce the anomalies from the seasonal cycle. To further isolate the high frequency

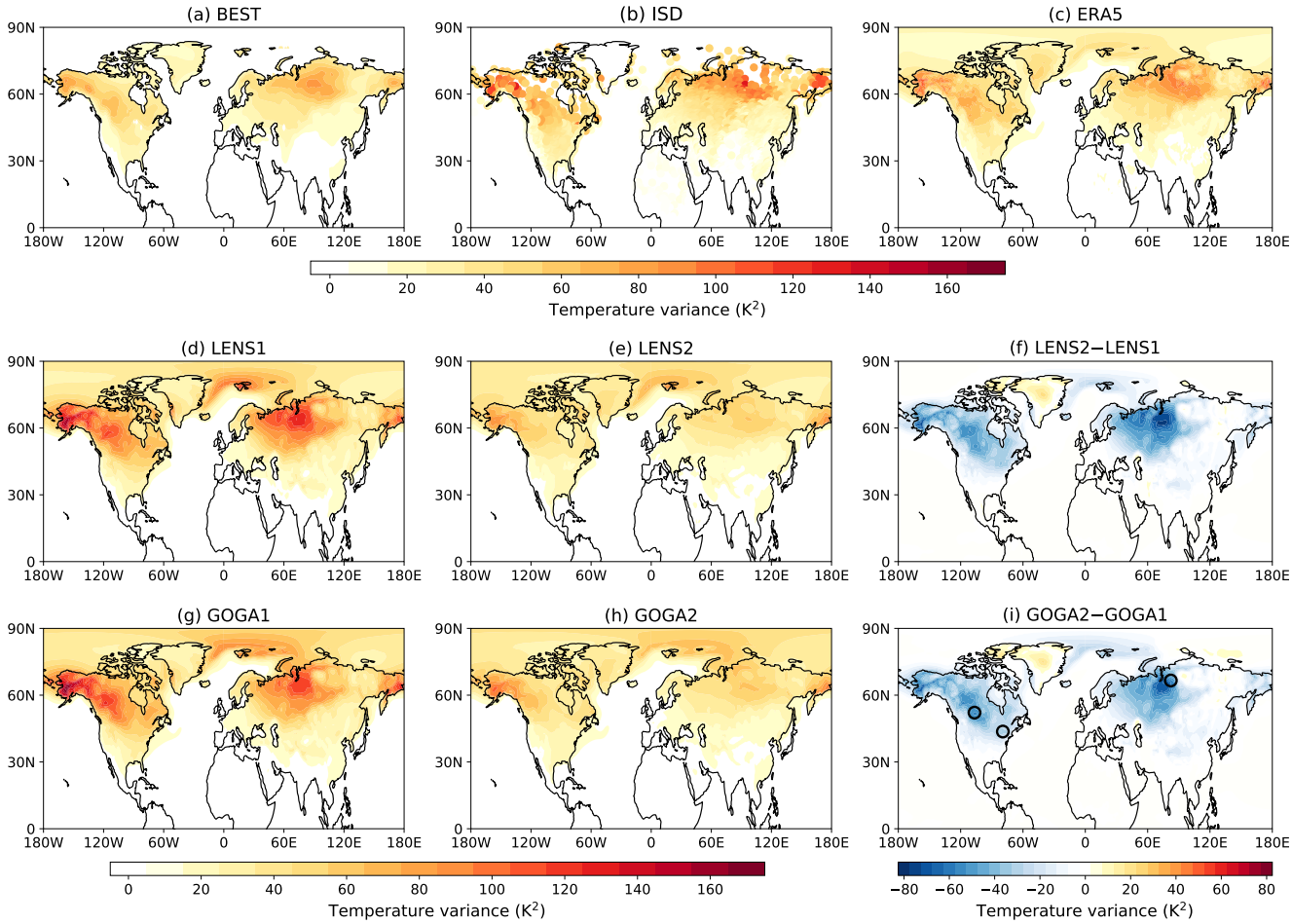

Figure 1. Sub-seasonal daily average T_{2m} variance during December-January-February, 1979–2014. (a)–(c) show observational datasets: Berkeley Earth Surface Temperature daily product; Integrated Surface Database; and ERA5. (d)–(f) show the Community Earth System Model Version 1 large ensemble (LENS1), the Community Earth System Model Version 2 large ensemble (LENS2) and the difference between them. (g)–(i) show the CESM1 GOGA ensemble (GOGA1), the CESM2 GOGA ensemble (GOGA2) and the difference between them. The black circles in (i) indicate the three locations used throughout this study: Saskatoon; Toronto; and Siderovsk.

variability from interannual variability or long term trends, when considering the December-January-February (DJF) season, we remove the mean anomaly from each DJF season to produce daily sub-seasonal anomalies. From now on, unless stated otherwise, we use T_{2m} to refer to these sub-seasonal anomalies.

3. The Change in Temperature Variability Between CESM1 and CESM2 and Its Cause

We begin by comparing the NH sub-seasonal DJF variance of T_{2m}, over the 1979–2014 period, between the observations, CESM1 and CESM2 in Figure 1. It can be seen that BEST (Figure 1a), ISD (Figure 1b) and ERA5 (Figure 1c) compare well in their observed estimates of T_{2m} variance, although there is a tendency for BEST to show slightly reduced variance compared to the other products, particularly over central Russia and Alaska. Comparing Figure 1d with Figures 1a–1c makes clear that CESM1 had too much variance in T_{2m} compared to observations (see also the difference plots in Figure S1 in Supporting Information S1 and in Figure 13 to be discussed in Section 6). Furthermore, a comparison of Figure 1e with Figure 1d demonstrates that CESM2 has greatly reduced T_{2m} variance compared to CESM1, with the difference between them shown in Figure 1f. This reduction in variance is also consistently present in both the daily minimum and maximum T_{2m} (Figure S2 in Supporting Information S1), so the reduction in variance from CESM1 to CESM2 is occurring throughout the day. In general, CESM2 is more comparable to the observations, although it does now appear to have too little

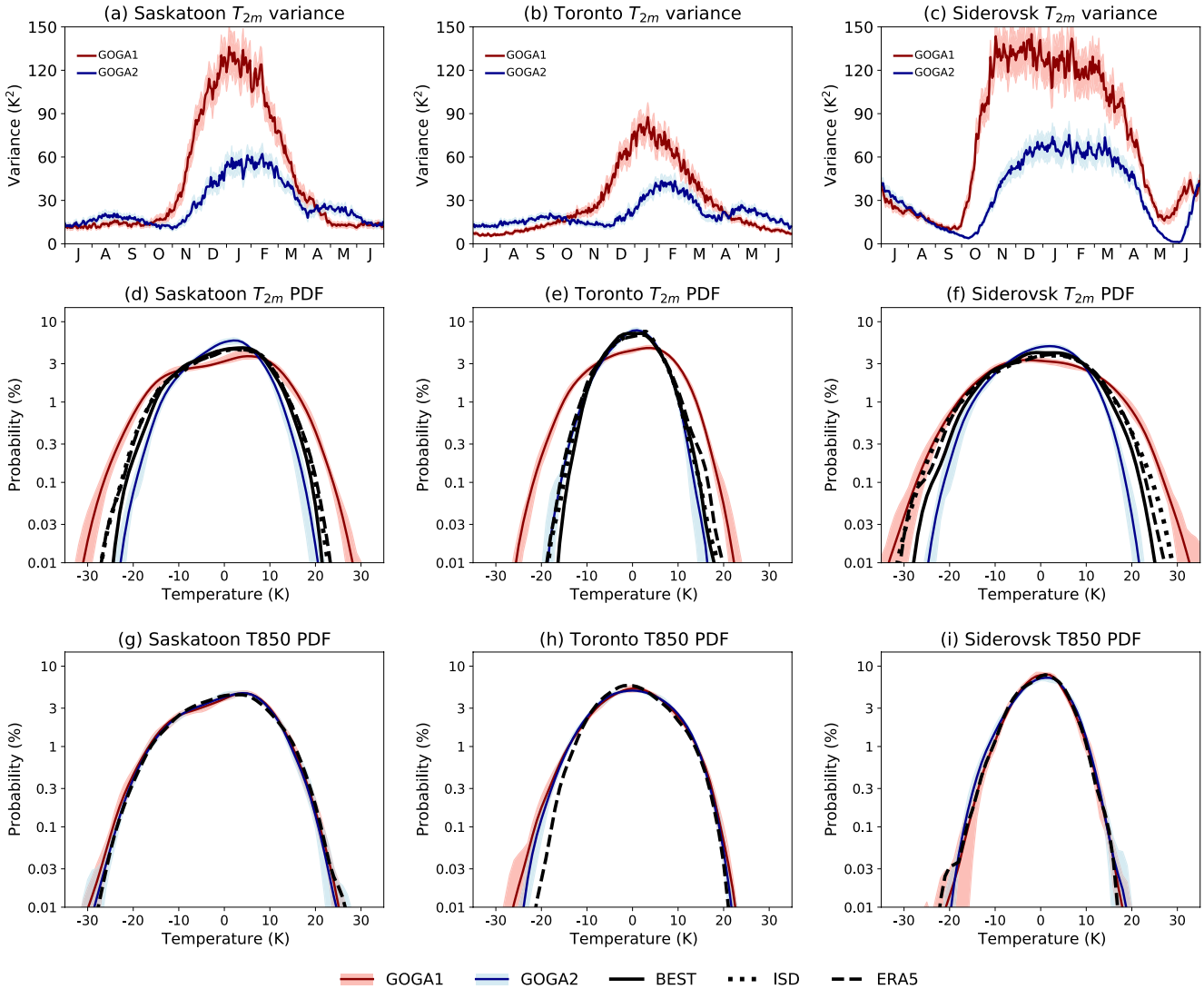

Figure 2. (a)–(c) Sub-seasonal variance of daily average T_{2m} calculated across years and ensemble members for 1979–2014 of the GOGA1 and GOGA2 simulations for Saskatoon, Toronto and Siderovsk that is, the variance across 10 × 36 years. Shaded range depicts the 2.5th–97.5th percentile range of the variance for each day of the year determined from 1000 samples generated by bootstrapping with replacement an equivalent number of years to that in the 10 GOGA1/2 members and recalculating the variance. (d)–(f) Probability distributions (PDFs) of sub-seasonal daily average T_{2m} for Saskatoon, Toronto and Siderovsk. (g)–(i) PDFs of daily average *T*850 for Saskatoon, Toronto and Siderovsk. Red = GOGA1, Blue = GOGA2. Black = observation based datasets (note that only ERA5 is used for *T*850). In (d)–(i) a Gaussian Kernel density estimate is used to depict the PDFs and the red and blue shading shows the minimum to maximum range across the 10 GOGA1 and GOGA2 ensembles, respectively.

variance over central Russia. The GOGA simulations with prescribed SSTs and sea ice exhibit the same differences between CESM1 and CESM2 as the coupled simulations (Figures 1g–1i) indicating that changes to the representation of the ocean or sea ice are not playing a role.

A more detailed comparison of the GOGA variability at three locations (Saskatoon, Toronto and Siderovsk, locations depicted by black circles in Figure 1i) is shown in Figure 2. We will continue to use these three locations throughout our analysis to build up an understanding of the cause behind this change in T_{2m} variability. Figures 2a–2c demonstrate that this large reduction in T_{2m} variance between CESM1 and CESM2 is primarily a wintertime feature, although it lasts longer in Siderovsk (Figure 2c) than in the Canadian locations (Figures 2a and 2b). While the reduction in variance is substantial throughout the whole winter season, it is largest in the early winter at each location, such that the peak wintertime variance is shifted later by a month or so in CESM2 compared to CESM1. The summertime changes in variance are much more muted and exhibit a different spatial structure to those shown in Figures 1f and 1i (not shown) and we do not consider them further here.

Probability distributions (PDFs) of T_{2m} are shown for our three locations in Figures 2d–2f on a logarithmic scale to emphasize the tails of the distribution. At each location, the T_{2m} distributions are broader at both ends of the distribution in CESM1 compared to CESM2 that is, in CESM1, cold extremes were colder and warm extremes were warmer. Furthermore, the shaded ranges in Figures 2d–2f which depict the range across the 10 GOGA members, demonstrate that the change that has occurred between CESM1 and CESM2 is highly significant. These PDFs also provide a more detailed view of what was already observed in Figure 1. That is, at the Canadian locations (Figures 2d and 2e) the CESM2 PDF of T_{2m} is much more comparable to observed since CESM1 had too much T_{2m} variability. At the Russian location (Figure 2f), it is less clear which of CESM1 or CESM2 is closer to observed as it depends somewhat on which observation-based product is being considered. However, we can conclude that CESM1 had too much variability at this location and CESM2 now likely has too little.

The large change in the PDFs of T_{2m} between CESM1 and CESM2 (Figures 2d–2f) is not accompanied by a change in the PDF of temperature in the lower troposphere at 850 hPa (T850, Figures 2g–2i) and T850 in both CESM1 and CESM2 compares well with that in ERA5 at these three locations. So, we need to understand why this change in temperature variability, which is localized to near the surface, has occurred between CESM1 and CESM2.

3.1. The Relative Roles of CAM and CLM

Having eliminated a role for the ocean or sea ice components in contributing to the T_{2m} variability change, we are left with either CAM, CLM or some combination of the two as potential drivers of this change, so we now use the sensitivity experiments described in Section 2.1.3 to isolate their relative contributions. Figure 3b demonstrates that the difference in T_{2m} variance seen between GOGA2 and GOGA1 (Figure 3a) can be reproduced within the CESM2 code-base by comparing CAM6_CLM5 with the simulation in which both the atmosphere and land have been reverted back to their CESM1 versions (CAM5_CLM4), as expected.

The role of CAM in this variance reduction can be isolated by taking the difference between CAM6_CLM5 and CAM5_CLM5 (Figure 3c) and the role of CLM can be isolated by taking the difference between CAM6_CLM5 and CAM6_CLM4 (Figure 3d). Summing up the CAM and CLM contributions, reveals that they approximately add up to the total change seen between CESM1 and CESM2 (compare Figure 3e with Figure 3a or 3b or consider Figure 3h which shows the percentage of the overall variance change explained by the sum of the CAM and CLM contributions).

Figures 3c and 3d demonstrate that there is a role for both the CAM5 to CAM6 transition and the CLM4 to CLM5 transition in the reduction in T_{2m} variance between CESM1 and CESM2. CAM is particularly important over Alaska and North West Canada and plays some role over Western Russia (Figures 3c and 3f). However, aside from over Alaska, the CLM4 to CLM5 transition is the dominant contributor and has resulted in a hemisphere-wide reduction in variance over high-latitude NH land (Figures 3d and 3g).

Our primary focus from now on will be on understanding the reasons behind the CLM influence on T_{2m} variability, given its dominant influence. However, CAM does play a role, particularly over Alaska and in supplemental Figure 3 we show some analyses of simulations in which individual parameterizations within CAM6 are reverted back to those of CAM5, to further understand the origins of this change. This reveals that there is not one single parameterization that is responsible. However, the combination of changes to the deep convection scheme, orographic form drag and shallow convection scheme go a long way to producing the variance reduction seen over Alaska. The introduction of the new orographic form drag scheme has contributed to reduced variance over Alaska because it has resulted in a reduction in lower tropospheric meridional wind variance in that region (see Figure 7g of Simpson et al. (2020)). The reason behind the influence of the other two schemes is less apparent. Overall, it should be concluded that there is not one single reason behind the changes in T_{2m} variance in CAM6 compared to CAM5.

3.2. The Influence of Snow Density on T_{2m} Variability

Having demonstrated an important role for the CLM4 to CLM5 transition in the reduction in T_{2m} variance in CESM2, we now investigate which particular CLM change is responsible. Comparing the change in variance in the transition from CLM4 to CLM5 (reproduced now in Figure 4a) with the difference between CAM6_CLM5

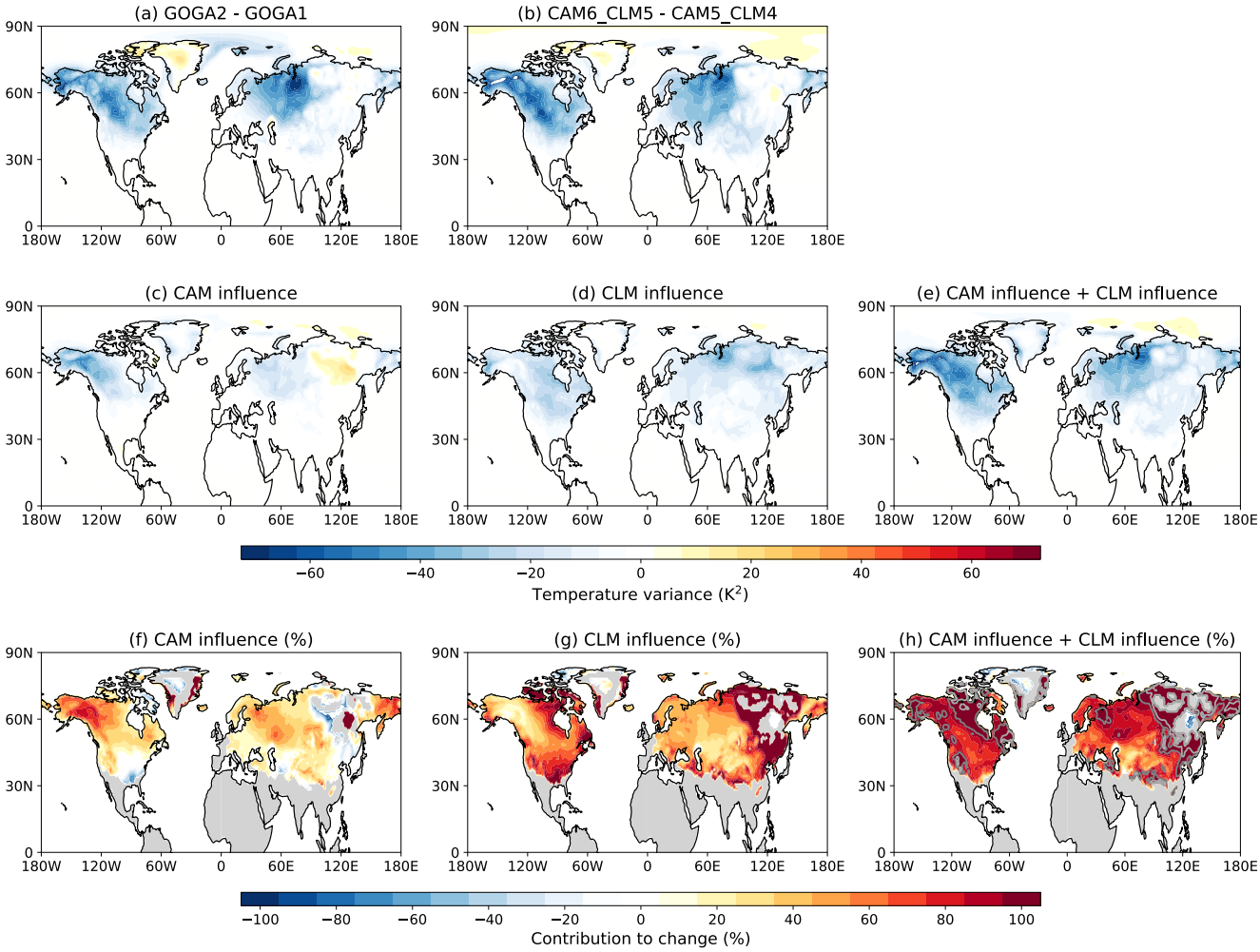

Figure 3. (a)–(e) Difference in sub-seasonal daily average T_{2m} variance between (a) GOGA2 and GOGA1, (b) CAM6_CLM5 and CAM5_CLM4, (c) CAM6_CLM5 and CAM5_CLM5 (the CAM influence), (d) CAM6_CLM5 and CAM6_CLM4 (the CLM influence), (e) the sum of panels (c) and (d). (f)–(h) show the percent difference between CAM6_CLM5 and CAM5_CLM4 explained by (f) Community Atmosphere Model alone that is, (CAM6_CLM5 − CAM5_CLM5) and (g) Community land model alone that is, (CAM6_CLM5 − CAM6_CLM4) and (h) the sum of panels (f) and (g). Percent differences are only shown where the magnitude of the total change in variance was greater than 5 K^2. Regions where the magnitude of the change in variance is less than 5 K^2 are shaded gray and in panel (h) the dark gray/light gray contours depict the 110% and 130% values.

and SNWDENS (where CLM5 is used but the snow density and densification settings listed in Table 2 are reverted back to those in CLM4, Figure 4b) reveals that much of the reduction in T_{2m} variance when using CLM5 has resulted from the change in these snow density and densification settings. This is also clear when considering the T_{2m} PDFs at our three locations (Figures 4c–4d). At each location, the T_{2m} PDF of SNWDENS (green) is very similar to that of CAM6_CLM4 (red) and exhibits a much broader distribution than CAM6_CLM5 (red). Of these three locations, the only aspect that is not fully explained by the snow density settings is the changes to the warm tail at Toronto (Figure 4d) but, overall, the snow density and densification settings clearly dominate in the reduction in T_{2m} variance that has been found in transitioning from CLM4 to CLM5.

The column average density of snow can be calculated by the sum of the snow ice and liquid water contents (CLM variables SNOWICE and SNOWLIQ) divided by the snow depth (CLM variable SNOWDP). Over NH land regions (except Greenland), this can be seen to have increased in CAM6_CLM5 compared to SNWDENS (Figures 5d–5f) as may be expected, given that the majority of the changes described in Section 1 would act to increase snow density. The DJF averaged snow fraction (Figures 5a–5c) is very similar in SNWDENS and CAM6_CLM5 and it is clear that the regions that show a reduction in T_{2m} variance in the transition from CLM4 to CLM5 have a DJF averaged snow fraction of close to one that is, are snow covered for much of the winter season.

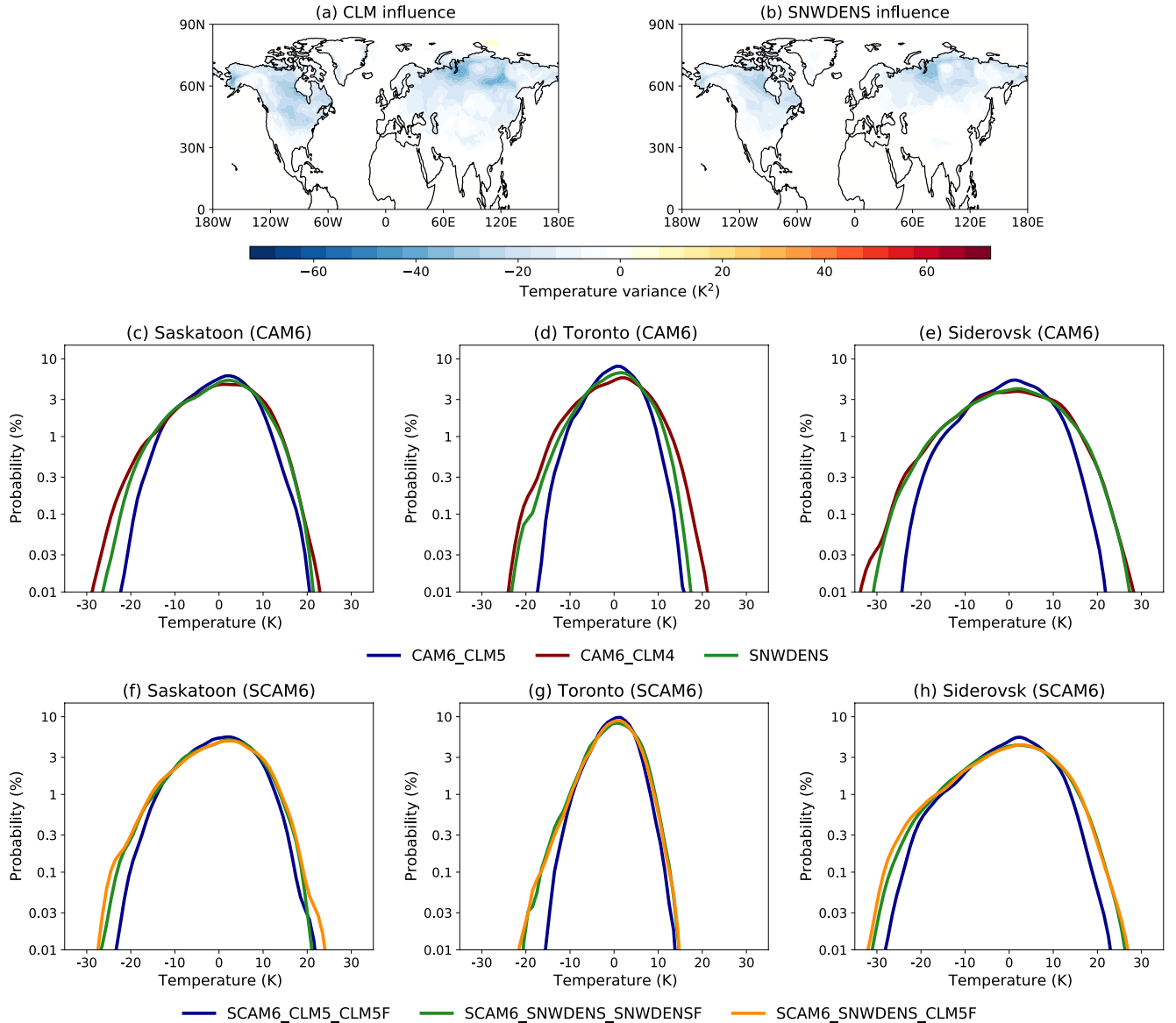

Figure 4. (a) The difference in December-January-February (DJF) sub-seasonal T_{2m} variability between CAM6_CLM5 and CAM6_CLM4 that is, the influence of changes in Community land model. (b) The difference in DJF sub-seasonal T_{2m} variability between CAM6_CLM5 and SNWDENS that is, the influence of snow density and densification changes between CLM4 and CLM5. (c)–(e) DJF sub-seasonal T_{2m} Probability distributions at Saskatoon, Toronto and Siderovsk for CAM6 simulations with (blue) CLM5 (red) CLM4 and (green) CLM5 but with the old snow density and densification settings. (f)–(h) DJF sub-seasonal T_{2m} PDFs at Saskatoon, Toronto and Siderovsk for single column model experiments with (blue) CLM5 and large scale forcings from CAM6_CLM5, (green) CLM5 with CLM4 snow density and densification settings and large scale forcings from SNWDENS and (orange) CLM5 with CLM4 snow density and densification settings and large scale forcing from CAM6_CLM5.

The seasonality of snow cover at our three locations can also explain their differing seasonalities of T_{2m} variance change. Siderovsk is snow covered for much longer than the Canadian locations and sees substantial increases in snow density with the new snow settings from October through to April (Figure 5i) corresponding to the time in which T_{2m} variance has reduced (Figure 2c). In Saskatoon and Toronto, reductions in T_{2m} variance are seen between November and March (Figures 2a and 2b), corresponding to the time when snow is consistently present and there is an increase in snow density with the new snow settings (Figures 5g and 5h). In addition, the increase in snow density in CAM6_CLM5 compared to SNWDENS is not uniform throughout the winter. It is largest in the early winter, which can help to explain why the reduction in T_{2m} variance is largest then, leading to a shift in the peak wintertime T_{2m} variance to later in the season (compare Figures 5g–5i with Figures 2a–2c) Furthermore, spatially, the regions that see the largest reduction in T_{2m} variance in Figure 4b (eastern Canada, particularly

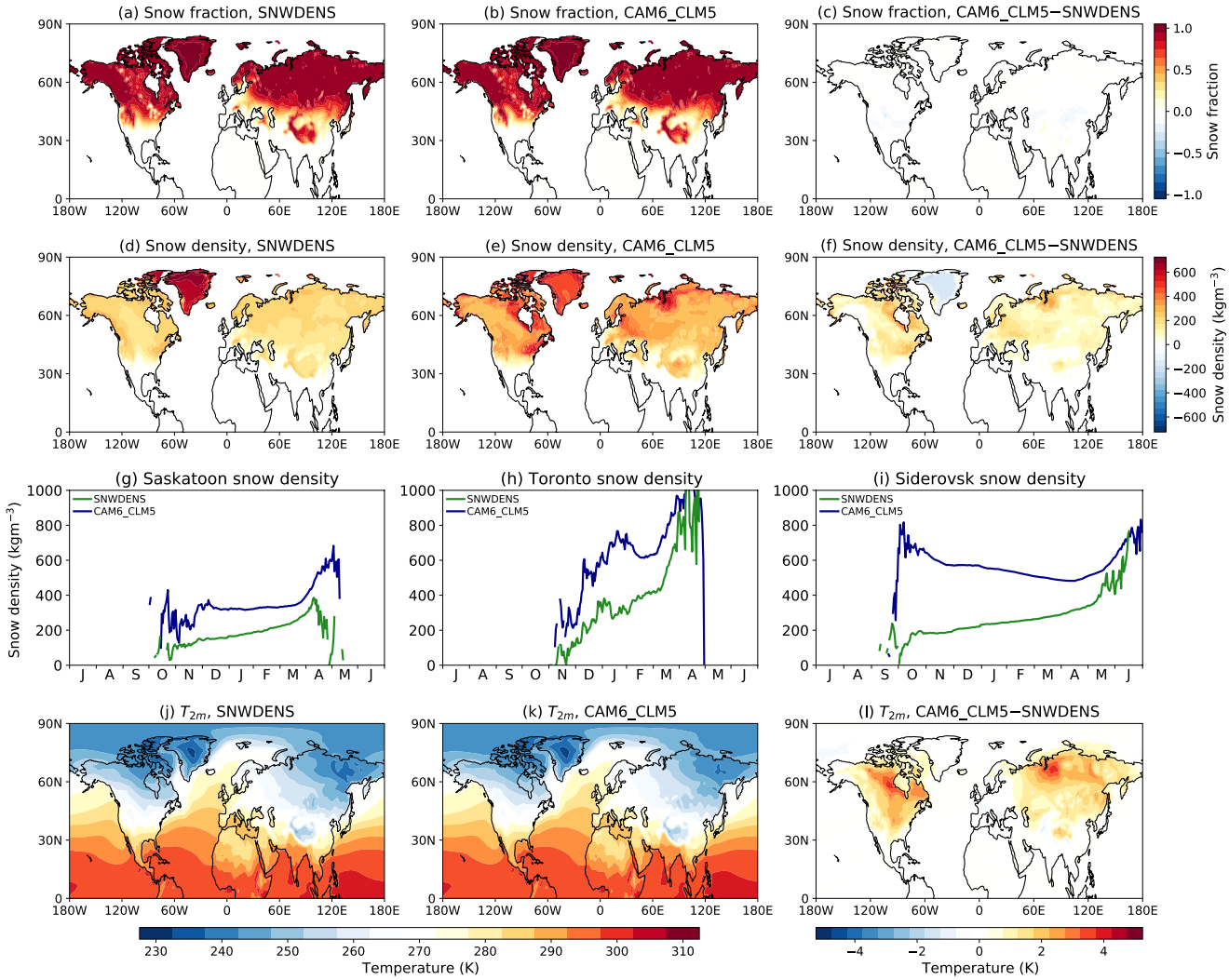

Figure 5. (a)–(c) December-January-February (DJF) averaged snow fraction for SNWDENS, CAM6_CLM5 and CAM6_CLM5 − SNWDENS difference. (d)–(f) DJF average snow density for SNWDENS, CAM6_CLM5 and the CAM6_CLM5 − SNWDENS difference. (g)–(i) seasonal cycle of snow density at Saskatoon, Toronto and Siderovsk for (green) SNWDENS and (blue) CAM6_CLM5 (j)–(l) DJF climatology of T_{2m} for SNWDENS, CAM6_CLM5 and CAM6_CLM5 − SNWDENS.

around Hudson's Bay and Northern Russia in the region to the south of the Kara sea) also correspond to where the snow density difference between CAM6_CLM5 and SNWDENS is the largest (Figure 5f).

3.2.1. Local Versus Non-Local Influences

The snow density settings not only change the T_{2m} variability, but also change the climatological average T_{2m}, with the majority of regions that see a snow density increase (Figure 5f) also showing an increase in the mean T_{2m} (Figure 5l). This raises the following question: does the snow density alter the T_{2m} variance primarily by changing the behavior of the column physics or are there non-local influences through altered mean temperature gradients and/or circulation variability and the associated changes to temperature advection?

To answer this question, we use the SCAM simulations summarized in Section 2.1.4 which allow us to isolate the relative roles of the column physics versus the non-local influences of changes in temperature advection. Comparison of the blue and green PDFs in Figures 4f–4h with those in Figures 4c–4e demonstrate that SCAM can reasonably well reproduce the increase in T_{2m} variability seen with the older snow density settings when it is given both the change in snow density and the large scale forcing from the SNWDENS simulations that is, both local and non-local influences. The correspondence between the SCAM PDFs and the full 3D simulation PDFs is not perfect and there are a number of possible reasons for this: the different vertical advection scheme used in

SCAM and the relaxation of the aerosol species and temperature described in Section 2.1.4. Despite these imperfections, it is clear that SCAM can reproduce the leading order effect of the snow density changes at these three locations. We can, therefore, use it to parse out the relative influence of the local column physics versus non-local effects in producing the change in T_{2m} variability.

The orange PDFs in Figures 4f–4h are for SCAM simulations where the large scale dynamics tendencies are from the CAM6_CLM5 simulation (i.e., the temperature tendencies from the large scale dynamics are derived from a simulation in which the new snow density settings were used) but locally, within SCAM, the snow density settings are reverted back to those of CLM4. In other words, these SCAM simulations only have the snow density influence on the column physics (the local influence) reverted back to that of CLM4. This completely reproduces the change in the T_{2m} PDF that is found when reverting both the non-local and local influences back to those of CLM4 (green PDF) so it can be concluded that the important effects of the snow density changes are arising through influences on the column physics as opposed to the non-local effects of altered temperature advection.

4. Understanding the Snow Density Influence on T_{2m} Variability

Here, we explain how the increased snow density in CESM2 acts to reduce the T_{2m} variability by consideration of the surface energy balance from two perspectives: (a) composites conditioned on T_{2m} using the full 3D version of CESM and (b) lagged regressions onto $T850$ using SCAM. The surface energy balance can be written as

$$-G\downarrow \approx F\uparrow = -SW_{net}\downarrow + LW_{net}\uparrow + SH\uparrow + LH\uparrow. \quad (1)$$

here $G\downarrow$ is the heat flux across the atmosphere-land interface that is, the heat flux into the snow when snow-covered or the ground otherwise, referred to as ground heat flux hereafter. So minus 1 times this quantity is equal to the net energy flux into the atmosphere ($F\uparrow$) which, to leading order, is equal to the sum of four surface energy fluxes: minus 1 times the net downward shortwave radiative flux ($SW_{net}\downarrow$); the net upward longwave radiative flux $LW_{net}\uparrow$; the upward sensible heat flux $SH\uparrow$ and the upward latent heat flux $LH\uparrow$. (Here, we use the surface fluxes from the atmosphere component which include additional correction terms to account for the melting of snow or freezing of rain as it hits the ground, depending on the surface temperature. So the left and right of Equation 1 are not exactly equal but, as will be shown, this residual is negligible.)

4.1. T_{2m} Conditioned Surface Energy Balance Composites

We begin by considering composites of the sub-seasonal anomalies in the surface energy balance at our three locations, conditioned on T_{2m} (Figure 6). To produce these composites, DJF days are binned into 10 bins according to their T_{2m} values. The edges of these bins are determined from the CAM6_CLM5 distribution, which is the narrower distribution and they correspond to the 10 10-percentile ranges from that distribution, except for the coldest and warmest bins. For the coldest and warmest bins, instead of using the 0th-tenth percentile range and the 90th–100th percentile range, we use the first-10th and 90th-99th, to avoid large differences in composite mean T_{2m} between CAM6_CLM5 and SNWDENS that can occur when sampling their very different distribution tails. The result is 10 composites of days ranging from cold to warm, which, by construction, have similar T_{2m} values for CAM6_CLM5 and SNWDENS (Figures 6a, 6d, 6g). Note that all days from CAM6_CLM5 that lie between the 1st and 99th percentiles are incorporated into these composites, while many more days from SNWDENS that are colder than the first and warmer than the 99th percentiles of the CAM6_CLM5 distribution are neglected. The purpose is to assess, how does the surface energy balance differ between CAM6_CLM5 and SNWDENS on days when they have similar T_{2m}?

The composites of $G\downarrow$ and $F\uparrow$ shown in Figure 6 (middle and right columns) indicate that, on cold days, there is a net upward energy flux from ground to atmosphere ($G\downarrow$ is negative, $F\uparrow$ is positive and $-G\downarrow \approx F\uparrow$) and the opposite is true on warm days. In other words, if these upward energy energy flux anomalies from ground to atmosphere are in the form of heat or radiation that will be absorbed in the lower atmosphere, they will dampen atmospheric temperature anomalies. Furthermore, there is a clear and systematic difference in the magnitude of these ground to atmosphere energy flux anomalies between CAM6_CLM5 and SNWDENS (compare Figure 6 middle and right columns). With the new snow density and densification settings in CAM6_CLM5 (Figure 6 right column) compared to the old snow settings in SNWDENS (Figure 6 middle column) there is a much larger upward energy flux anomaly from ground to atmosphere on cold days and a much larger downward energy flux

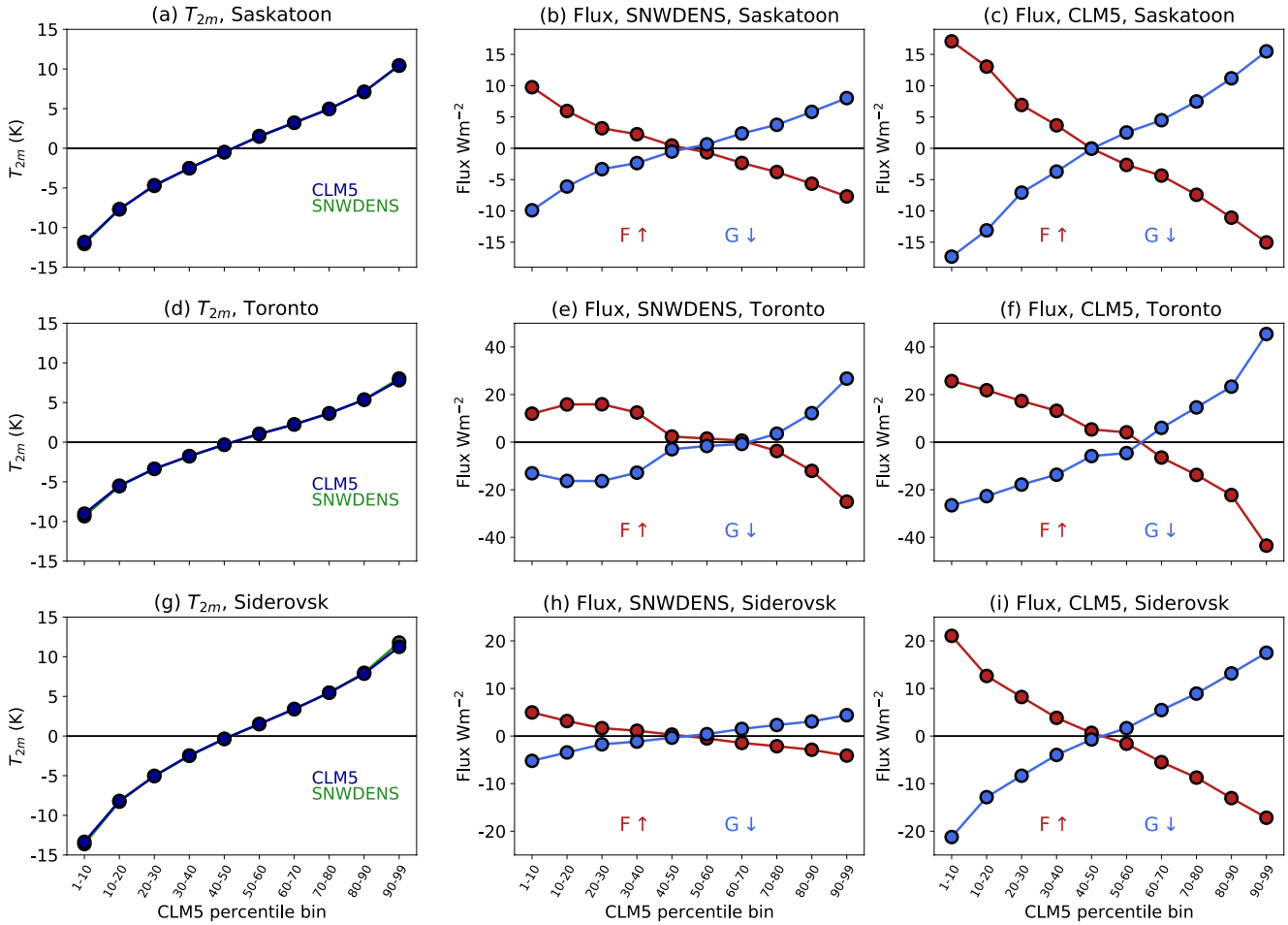

Figure 6. Composites of sub-seasonal daily average fields conditioned on T_{2m} with composite bins based on the CLM5 T_{2m} distribution (x axis). (a) T_{2m} composites for Saskatoon for (blue) CAM6_CLM5 and (green) SNWDENS. (b) Composites of the net upward heat flux $F\uparrow$ and the ground heat flux $G\downarrow$ for Saskatoon for the SNWDENS simulation. (c) Composites of the net upward heat flux $F\uparrow$ and the ground heat flux $G\downarrow$ for Saskatoon for the CAM6_CLM5 simulation. (d)–(f) are as (a)–(c) but for Toronto. (g)–(i) are as (a)–(c) but for Siderovsk. Note. that CLM5 and SNWDENS are overlapping in panels (a), (d) and (g) by construction since this is showing composites of T_{2m} conditioned on T_{2m}.

anomaly from atmosphere to ground on warm days (see also the difference in $G\downarrow$ and $F\uparrow$ between CAM6_CLM5 and SNWDENS in the top row of Figure 7). In other words, the dampening effect of the energy flux from ground to atmosphere is stronger with the new snow settings than with the old snow settings, consistent with the reduced temperature variance in CAM6_CLM5 compared to SNWDENS.

Why would there be a bigger anomalous upward energy flux from ground to atmosphere when it is cold (and vice-versa when it is warm) with the new snow settings? Let us first consider the vertical structure of temperature throughout the atmosphere and the snow column for our coldest and warmest composite bins, shown in Figure 8. CLM5 can have up to 12 snow layers, but the maximum number at any of our three locations is 5. The number of snow layers can change at every timestep so, given the challenges of outputting timestep resolution fields globally, in Figure 8 the T_{2m} composite analysis has been performed using the SCAM simulations in which we have output timestep level fields for these three locations. The composites are based on daily average T_{2m} and the averages within the snow layers performed at timesteps when the snow layers exist. When it is anomalously cold/warm at the surface, the cold/warm anomalies within the snow become smaller with depth. In other words, a temperature gradient is induced across the snow layer. Such a temperature gradient would be accompanied by a heat flux across the snow layer ($F_{sno}\uparrow$, positive upward) by thermal conduction given by

$$F_{sno}\uparrow = \lambda \frac{\partial T}{\partial z} \tag{2}$$

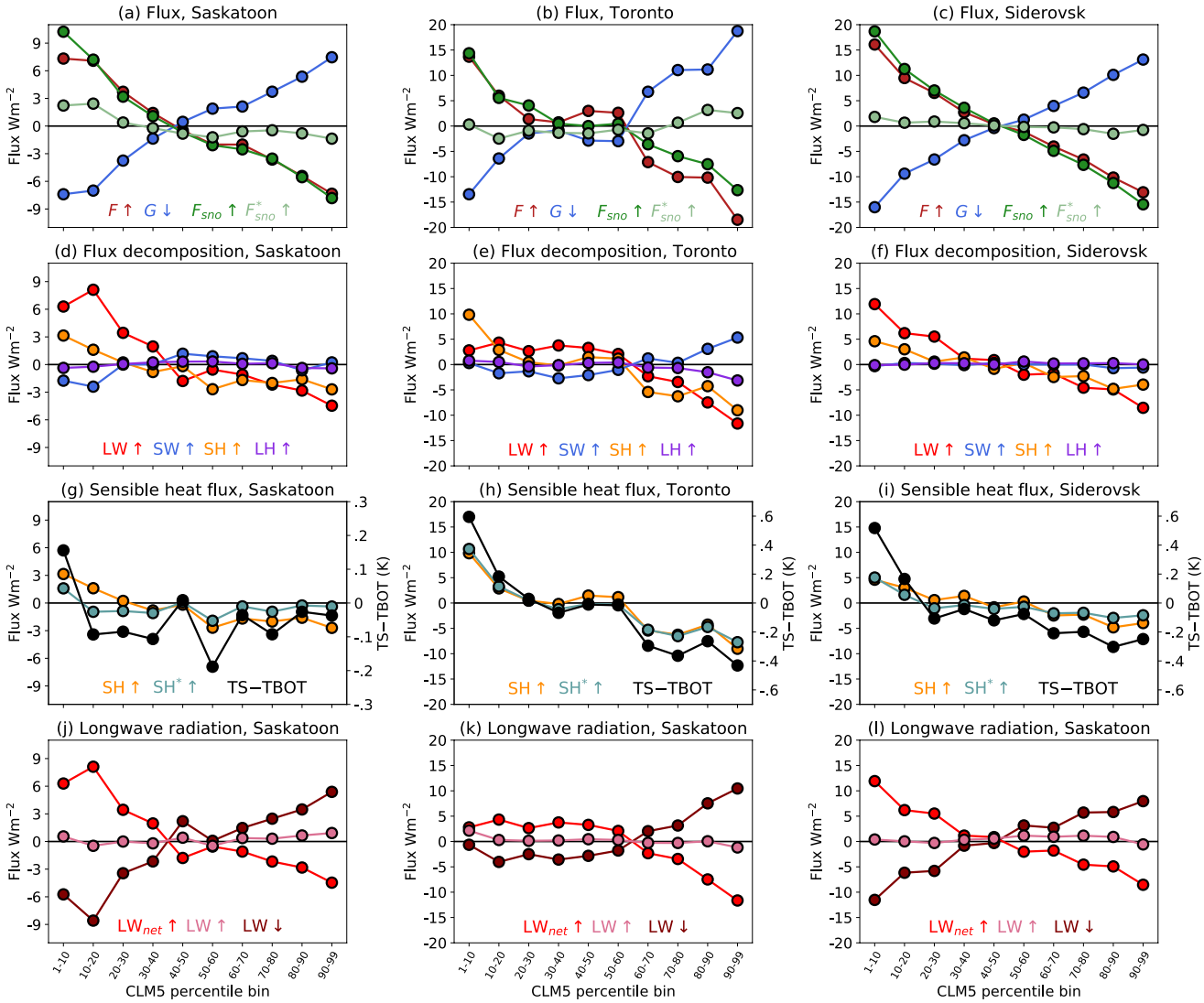

Figure 7. Composites of daily average fields conditioned on daily average T_{2m} with composite bins based on the CLM5 T_{2m} distribution (x-axis). All panels show the difference between CAM6_CLM5 and SNWDENS for (left to right) Saskatoon, Toronto, Siderovsk. (a)–(c) Net upward heat flux F and ground heat flux G together with the diagnosed bulk flux across the snow ($F_{sno}\uparrow$, Equation 4) as well as the diagnosed bulk flux across the snow assuming a constant density equal to that of the CLM5 average ($F^*_{sno}\uparrow$). (d)–(e) Net upward heat flux decomposition into net longwave (LW), net shortwave, sensible heat flux (SH) and latent heat flux (LH). (g)–(i) The SH and the component of the sensible heat flux that can be explained by K (TS − TBOT) where K is derived from CLM5 (SH*) along with the difference between TS and TBOT (right axis) (j)–(l) Net long wave radiation LW_{net} and the contribution from the upward and downward components.

where $\partial T/\partial z$ is the vertical temperature gradient across the snow layer at depth, z, with z defined as positive downward and λ is the thermal conductivity of the snow (Lawrence et al., 2018, chapter 6). The formulation for the thermal conductivity in CLM is

$$\lambda = \lambda_{air} + \left(7.75 \times 10^{-5}\rho_{sno} + 1.105 \times 10^{-6}\rho^2_{sno}\right)(\lambda_{ice} - \lambda_{air}) \quad (3)$$

which follows Jordan (1991) where coefficients are based on an empirical fit to laboratory-based data from Yen (1962). The thermal conductivity of ice (λ_{ice}) is 2.29Wm^{-1}K^{-1} and the thermal conductivity of air (λ_{air}) is 0.023Wm^{-2}K^{-1} and ρ_{sno} is the density of snow. So, the higher the density of the snow, the higher the conductivity.

The PDFs of snow density in Figures 8d–8f indicate that, in CLM5, the snow density is increased in each of the snow layers compared to the old snow settings. While, in reality, the conductance calculation is performed numerically at each of the individual snow layers, to diagnose the influence of the snow density changes we can attempt to simplify matters by instead approximating the snow column as a constant flux layer characterized by

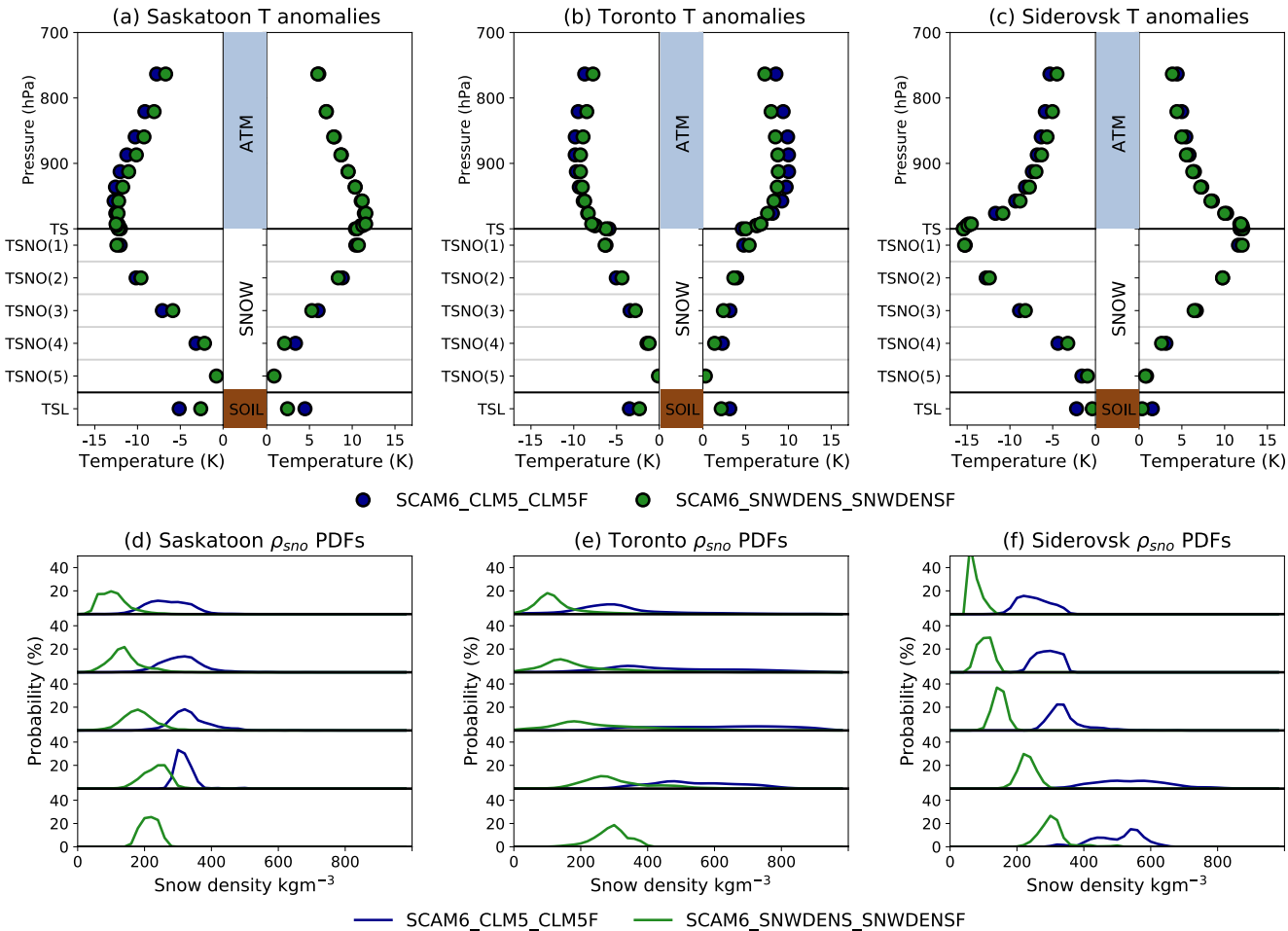

Figure 8. (top) Composites of temperature anomalies at various levels for (left) the coldest days (first-tenth percentile of the CLM5 T_{2m} distribution) and (right) the warmest days (90th–99th percentile of the CLM5 T_{2m} distribution). From top to bottom of each panel, the points depict the atmospheric temperature anomalies on model levels (assuming a surface pressure of 1000 hPa), the surface temperature *TS*, the temperature within the snow layers and the soil temperature. These composites are based on the single column model run with CLM5 as the land component and forcings from the CAM6_CLM5 simulations (SCAM6_CLM5_CLM5F, blue) and run with the old snow settings in the land components and forcings taken from the SNWDENS simulations (SCAM6_SNWDENS_SNWDENSF, green). Note. that the *y*-axis is not to scale and also since the number of snow layers varies at every timestep, different days contribute to the composites at different levels (bottom) Probability distributions of snow density in each layer at the timestep level, expressed as probability per 20 kgm^{-3} density bin, for (green) SCAM6_SNWDENS_SNWDENSF and (blue) SCAM6_CLM5_CLM5F. From top to bottom shows the first to fifth snow layer and the fifth snow layer is never occupied with CLM5 at Saskatoon and Toronto.

the column averaged density $(\overline{\rho}_{sno})$ and, therefore, column averaged conductance $\overline{\lambda}$. Returning to the full CAM simulations, as opposed to SCAM, we diagnose an approximate bulk heat flux across the snow layer according to

$$F_{sno}\uparrow = \overline{\lambda}\frac{TSL - TSNO(1)}{\Delta z} \tag{4}$$

where $TSNO(1)$ is the temperature of the top snow layer (CLM variable SNOTTOPL), TSL is the temperature at the top of the soil column (CLM variable TSL) and Δz is the snow depth (CLM variable SNOWDP). As before, the column averaged snow density is given by (SNOWICE + SNOWLIQ)/SNOWDP. $F_{sno}\uparrow$ is only calculated on days when the snow depth is greater than 5 cm and to deseasonalize this quantity which can have very sharp jumps at the seasonal edges, prior to fitting the first 4 harmonics to the seasonal cycle, F_{sno} is tapered to zero over a 30 day period after the end of March and before the beginning of November.

The difference in the T_{2m} conditioned composite of $F_{sno}\uparrow$ between CAM6_CLM5 and SNWDENS is shown by the dark green lines in Figures 7a–7c and this corresponds very well to the difference in the upward energy flux from ground to atmosphere as a function of T_{2m} between CAM6_CLM5 and SNWDENS. We can further

demonstrate that this change in $F_{sno}\uparrow$ has primarily resulted from the altered snow density by replacing $\overline{\lambda}$ in Equation 4 with the time averaged conductance from CAM6_CLM5 in each case, thereby diagnosing the $F_{sno}\uparrow$ for each day of each simulation that would occur if the snow density were constant and equal to that of the CLM5 average ($F^{*}_{sno}\uparrow$, light green in Figures 7a–7c). As expected, this does not correspond to the difference in $F_{net}\uparrow$ between CAM6_CLM5 and SNWDENS, indicating that it is, indeed, the change in the snow density that is primarily responsible for the alterations to the heat flux across the snow.

Panels d–l in Figure 7 illustrate how the terms in the atmospheric surface energy balance are altered in association with the difference in heat flux across the snow. The terms that exhibit a systematic difference between CAM6_CLM5 and SNWDENS are the net longwave radiation ($LW_{net}\uparrow$) and the sensible heat flux ($SH\uparrow$). On cold/warm days, both of these upward fluxes show a relative increase/decrease in CAM6_CLM5 compared to SNWDENS (Figures 7d–7f).

Consider $SH\uparrow$ first. As described in Neale et al. (2012) (Section 4.11.1), this is given by

$$SH\uparrow = \rho_1 c_p \frac{\theta_s - \theta_1}{r_{ah}} \tag{5}$$

where ρ_1 is the density at the lowest model level, c_p is the specific heat capacity of dry air at constant pressure, r_{ah} is the aerodynamic resistance calculated using Monin-Obhukov theory, θ_s is the potential temperature at the surface and θ_1 is the potential temperature at the lowest model level. To assess the reason for the change in $SH\uparrow$, we simplify Equation (5) and diagnose the following flux which assumes that $SH\uparrow$ is linearly related to the temperature difference between the surface and the lowest model level and neglects variations in the other parameters of (5) and in the pressure difference between the surface and the lowest atmospheric model level:

$$SH^{*}\uparrow \sim K(TS - TBOT) \tag{6}$$

where TS is the surface temperature, $TBOT$ is the temperature at the mid-point of the lowest atmospheric model level (∼50m above the ground) and K is empirically derived, via linear regression, from the daily values of $SH\uparrow$, TS and $TBOT$. The difference in $SH^{*}\uparrow$ between CAM6_CLM5 and SNWDENS when using a single value of K, that was derived from CAM6_CLM5, is shown in teal in Figures 7g–7i. This matches rather well the actual change in $SH\uparrow$, particularly for Toronto and Siderovsk. The match is less perfect in Saskatoon, suggesting the approximations that go into (6) are less valid there. But overall, this indicates that the dominant reason for the altered $SH\uparrow$ is the change in the temperature difference between the surface and the lowest atmospheric model level ($TS - TBOT$, black in Figures 7g–7i) for a given T_{2m} anomaly. With the new snow settings, at a given cold value of T_{2m} the surface is relatively less cold than the atmosphere above, and associated with this is an enhanced $SH\uparrow$. The surface is relatively less cold than the atmosphere because of this enhanced heat flux across the snow in the presence of enhanced density and conductance. The opposite is true on warm days.

Considering now $LW_{net}\uparrow$, when it is cold/warm, there is a relative increase/decrease in $LW_{net}\uparrow$ in CAM6_CLM5 compared to SNWDENS, and in this T_{2m} conditioned composite view, this is primarily associated with a difference in the downward LW radiation (Figure 7 bottom). The upward longwave radiation will depend on the temperature of the surface which isn't very different between CAM6_CLM5 and SNWDENS by construction here since we have conditioned our composites on T_{2m}. The downward LW radiation will depend on the temperature in the atmospheric column above and clouds. It is the change in the clear sky fluxes that dominate in the $LW_{net}\uparrow$ anomalies (not shown) so impacts of changes in clouds are a secondary effect. This behavior of $LW_{net}\uparrow$ can be understood given the differences in the atmospheric temperature profile (see Figures 8a–8c for the SCAM profiles; the CAM profiles look similar and are not shown). With the new snow settings, to reach a given cold/warm T_{2m} value, the atmosphere in the column above has to be colder/warmer. This leads to CAM6_CLM5 exhibiting a relative reduction/increase in downward longwave radiation on cold/warm days compared to SNWDENS.

This change in the dependence of surface energy balance terms on T_{2m} can be generalized to other locations in the NH by calculating the linear regression of the flux onto T_{2m} and taking the difference in the regression coefficient between CAM6_CLM5 and SNWDENS. This is shown for $F\uparrow$, $SH\uparrow$ and $LW_{net}\uparrow$ in Figure 9. Across the snow covered regions of the NH it can be seen that in CAM6_CLM5 compared to SNWDENS, the slopes of $F\uparrow$, $SH\uparrow$ and $LW_{net}\uparrow$ against T_{2m} are more negative, consistent with what was found in Saskatoon, Toronto and Siderovsk.

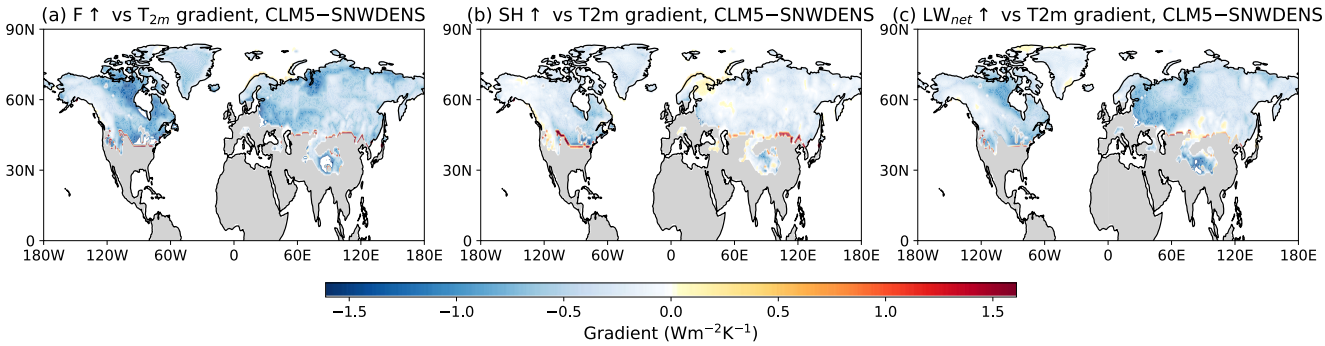

Figure 9. The change between the new snow settings and the old snow settings in the slope of the linear regression of a given flux against T_{2m} for (a) net upward energy flux from ground to atmosphere ($F\uparrow$), (b) sensible heat flux ($SH\uparrow$) and (c) net longwave radiation ($LW_{net}\uparrow$). The gradient is only shown for land points where the December-January-February averaged snow fraction is greater than 0.5 and other land regions are shaded in gray.

4.2. SCAM Regressions

To confirm the above arguments we also consider how the snow settings affect the evolution of the surface energy balance in the days leading up to and following a temperature anomaly. T_{2m} is highly correlated with the 850 hPa temperature ($T850$) and, as shown in Figures 2g–2i, the $T850$ variability is unaffected by the snow settings. So, we consider lagged regressions of fields onto $T850$ and for this we use the SCAM simulations forced with the large scale circulation from the CLM5 simulations (i.e., SCAM6_CLM5_CLM5F and SCAM6_SNWDENS_CLM5F). The SCAM simulations provide a cleaner picture because they are each effectively given the same weather with the only difference being the snow settings, but similar results are found for the CAM6_CLM5 and SNWDENS simulations, although they are much noisier (not shown).

Figure 10 shows lagged regressions onto $T850$ for a variety of fields, multiplied by minus one so that we will consider how these fields evolve as $T850$ becomes cold. By construction, the $T850$ anomalies are -1 for each case at day 0 and the snow settings do not alter how $T850$ evolves into and out of that cold state, as expected (Figures 10a–10c). However, accompanying that $T850$ anomaly at lag zero, is a colder T_{2m} anomaly with the old snow settings (green) than with the new snow settings (blue; Figures 10d–10f). As T_{2m} starts to get colder, the new CLM5 snow settings result in a larger net upward flux from ground to atmosphere (red in Figures 10g–10i) which is very well explained by the diagnostic bulk heat flux across the snow layer (Equation 4, green in Figures 10g–10i). At negative lags, the sensible heat flux dominated the enhanced upward flux from ground to atmosphere with the new snow settings (Figures 10j–10l, orange) and this can be reasonably well explained by the change in the $TS - TBOT$ difference using Equation 6 with K derived from the SCAM6_CLM5_CLM5F simulation (Figures 10m–10o, turquoise). So, as the cold anomalies in the lower atmosphere increase, the surface does not become as cold with the new snow settings because of the enhanced upward heat flux across the snow, TS is relatively warmer than $TBOT$ and the upward sensible heat flux increases, which will act to reduce the cold temperature anomalies in the atmospheric layers above, although Figures 2g–2i suggests that by the 850 hPa level, this has a minimal effect on temperature variance. The difference in the net upward longwave radiation lags the sensible heat flux by about a day (Figures 10j–10l, red) and it then dominates in the altered surface energy balance at positive lags. In this lagged regression view, since the T_{2m} is not constrained to be the same in each case like it was in the T_{2m} conditioned composites above, a difference in the upward LW radiation dominates. The surface has not become as cold with the new settings and so, compared to the old settings, there is a relative upward longwave flux during times when $T850$ becomes cold.

To summarize, the T_{2m} conditioned composites of the surface energy balance in the full version of the model and the lagged regression of surface enery balance terms onto $T850$ in the single column model, paint a similar picture. From the perspective of what happens when it gets cold (with the opposite being true for when it gets warm), a cold anomaly in the lower atmosphere and at the surface that has not yet penetrated down to affect the ground temperature below, will induce a larger anomalous upward heat flux across the snow layer with the new snow settings, since the snow density and conductance are higher. This enhanced heat flux across the snow will prevent the surface (at the top of the snow) from becoming as cold as it would have with the CLM4 snow settings and is balanced by both a relative upward sensible heat flux and upward longwave radiative flux from ground

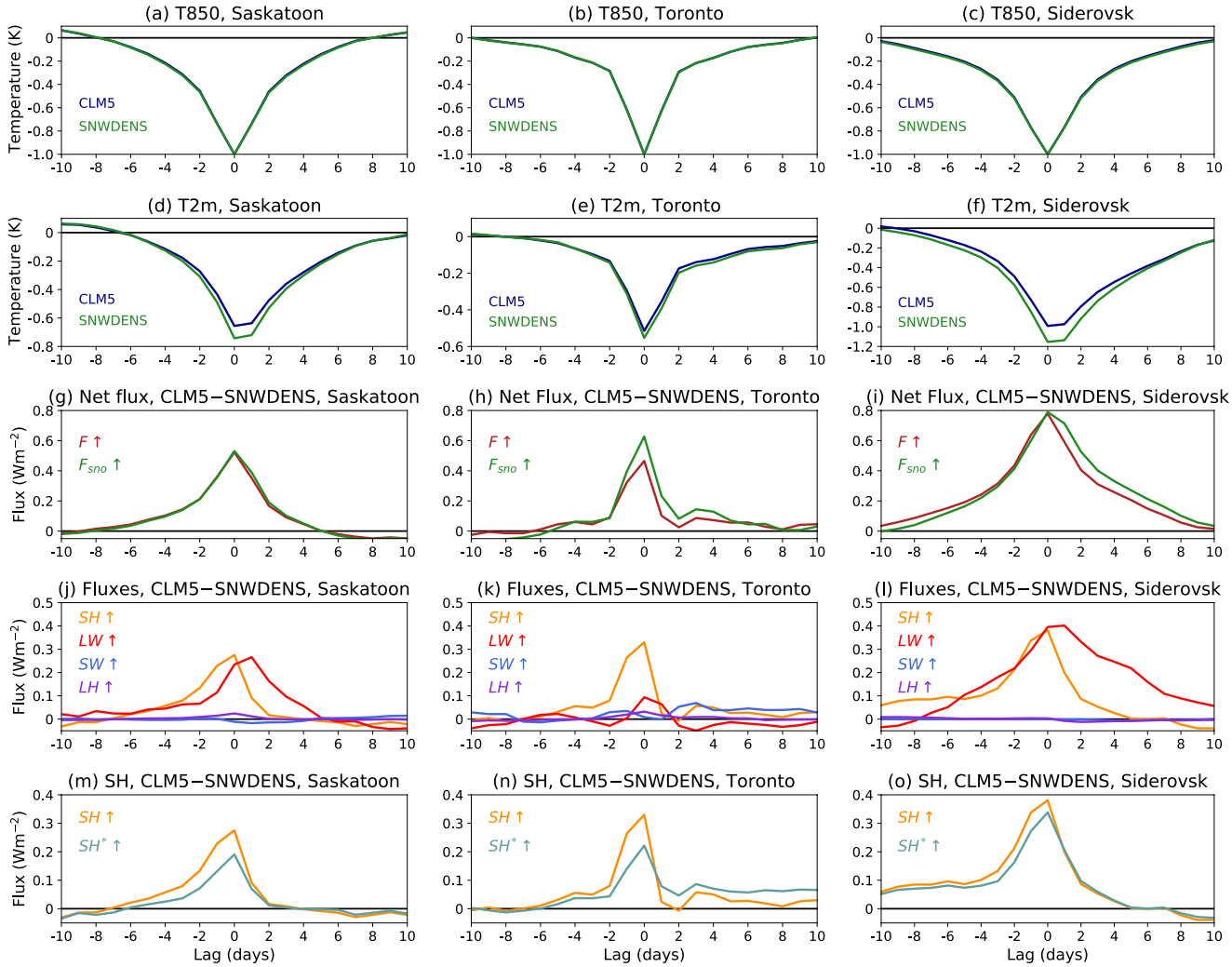

Figure 10. Lagged regressions onto 850 hPa temperature using Single Column Atmospheric Model (SCAM) with forcing taken from CLM5, multiplied by minus 1 to represent anomalies that accompany cold temperatures at 850 hPa. (a)–(c) 850 hPa temperature for SCAM run with (blue) CLM5 and with (green) CLM5 with snow density and densification settings reversed (SNWDENS). (d)–(f) T_{2m} for (blue) CLM5 and (green) SNWDENS. (g)–(h) (red) the net upward heat flux from ground to atmosphere (green) F_{sno}, which is an estimate of the upward flux of heat across the snow layer given by the conductance times the temperature gradient between the upper most layer of the snow and the soil (j)–(l) the various components of the surface energy balance (yellow) sensible heat flux (red) longwave radiation (blue) shortwave radiation (purple) latent heat flux (m)–(o) the sensible heat flux (orange) and the sensible heat flux estimated using Equation 6 where K is derived from CAM6_CLM5 using linear regression.

to atmosphere, which will dampen the atmospheric temperature anomalies aloft. The increased heat flux across the snow layer, therefore acts to dampen temperature anomalies at the surface and lower atmosphere, leading to reduced surface air temperature variance.

5. Implications for Future Projections

Future projections of the mean change in NH land surface temperatures under anthropogenic forcing are subject to considerable uncertainty due to internal variability (e.g., Deser et al., 2012) and this uncertainty has been quantified specifically for LENS1 by a number of studies (e.g., Kay et al., 2014; Thompson et al., 2015). Of course, the magnitude of this uncertainty as simulated by a model will depend on that model's representation of internal variability and since we have seen that the characteristics of daily T_{2m} variability are different between CESM1 and CESM2, this motivates an assessment of differences in the uncertainty in forced climate projections due to internal variability between LENS1 and LENS2. For this, we consider the ensemble spread in Future − Past differences in DJF averaged T_{2m} for LENS1 and LENS2, using the period 1979–2014 for the "Past". Since CESM1

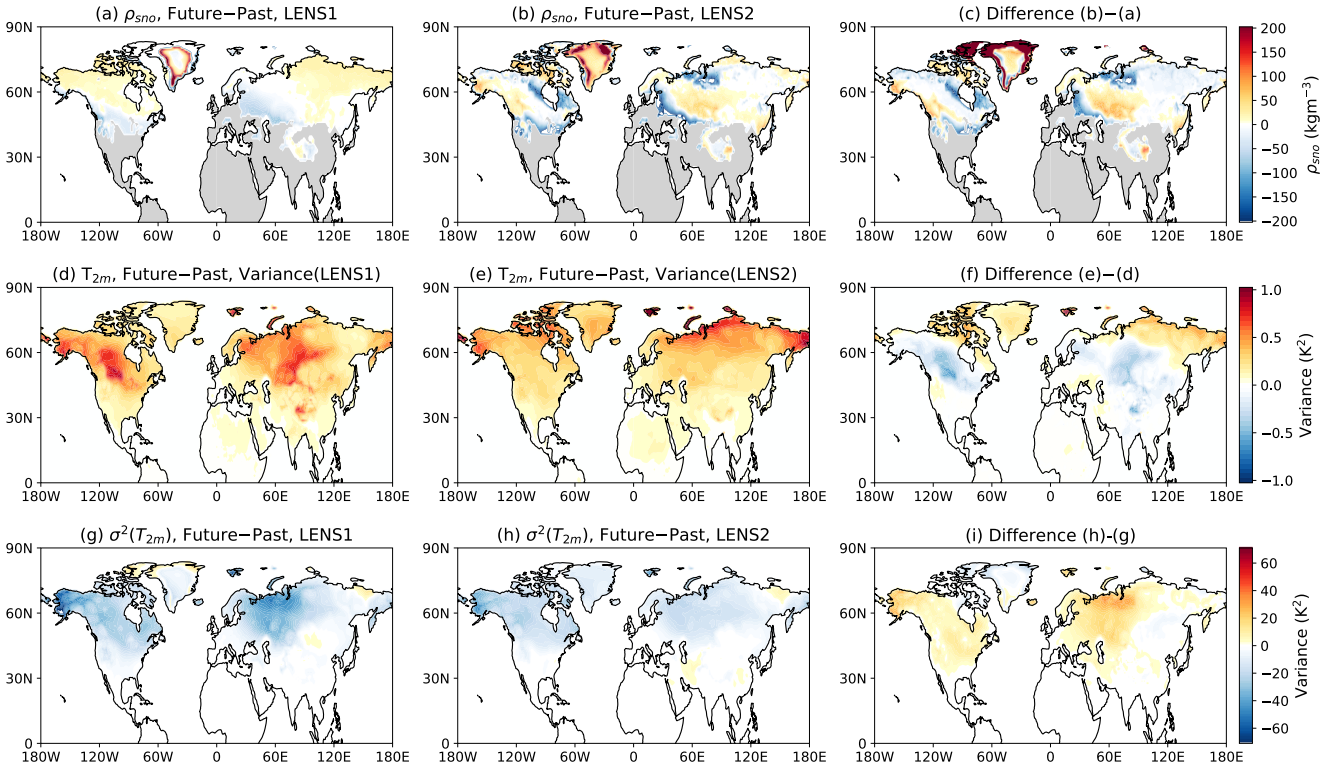

Figure 11. A comparison of the Community Earth System Model Version 1 and the Community Earth System Model Version 2 large ensembles for December-January-February. (a)–(c) Future − Past difference in snow density for LENS1, LENS2 and the difference between LENS2 and LENS1. (d)–(f) The variance, across ensemble members, in the Future − Past difference of T_{2m} for LENS1, LENS2 and the difference between LENS2 and LENS1. (g)–(i) The Future − Past difference in the daily variance of T_{2m} for LENS1, LENS2 and the difference between them.

and CESM2 have different climate sensitivities and the large ensembles are run under different forcing scenarios, rather than using the same time period for the "Future" in each case, we examine time periods when LENS1 and LENS2 have similar changes in global mean surface temperature. This is achieved by finding the 30 year period in LENS1 (which is under the higher forcing scenario) when the global mean surface temperature increase compared to 1979–2014 is the same magnitude as that for years 2070–2099 of LENS2. This period turned out to be 2060–2089. So, for LENS1, the future is 2060–2089 and for LENS2 it is 2070–2099.

First, the future projected change in snow density is rather different between CESM1 and CESM2 (Figures 11a–11c). In particular, in the regions in which CESM2 has a large increase in present day snow density compared to CESM1 (around Hudson Bay and south of the Kara sea, Figure 5f), the future projections of snow density exhibit a decline in LENS2 (Figure 11b), while they showed an increase or near zero change in LENS1 (Figure 11a). We speculate that the reason for this is the inversion of the dependence of density on temperature included in CLM5 - at very cold temperatures, the snow is now denser, so as the temperature warms, we may expect the density in very cold regions to decrease (see Section 1 and Figure 1c of van Kampenhout et al. (2017)). For the mean projected T_{2m} change, the ensemble spread in LENS1, as measured by the across-member variance in Future − Past differences, was very large in Alaska, Central Canada and South-Central Russia, in particular (Figure 11d). In LENS2, the variance in the Future − Past difference across ensemble members has reduced considerably in these regions (Figures 11e and 11f). Many things are different between LENS1 and LENS2 which precludes a definitive attribution of changes in ensemble spread to the factors that have given rise to the altered daily T_{2m} variability. But, it seems likely that in these regions, the reduced internal variability in daily T_{2m} in CESM2 is translating into a reduced uncertainty in future projected climate change. Lower day-to-day variability in T_{2m} means that there will be a lesser role for the internal variability "noise" in the climatological averages that are being examined here, leading to lower spread among the ensemble members (e.g., Thompson et al., 2015). It is clear that other things are happening around the sea ice edge, particularly in Northern Russia, such that reduced internal variability is not translating into reduced uncertainty in future climate projections. This may be

because other aspects related to the representation of sea ice have changed between CESM1 and CESM2 (e.g., DeRepentigny et al., 2020).

Future projections are also characterized by a reduction in daily T_{2m} variability in the NH wintertime (Figures 11g–11h), related to the fact that the Arctic exhibits relatively amplified warming compared to the mid-latitudes, which reduces the temperature variability associated with meridional advection (Holmes et al., 2016; Schneider et al., 2015; Screen, 2014). A reduction in daily T_{2m} variability is found in the future projections in both LENS1 and LENS2, but this reduction in variability is larger in LENS1 (Figures 11g–11i). Because the T_{2m} variability associated with given atmospheric circulation anomalies is larger in CESM1 than it is in CESM2, there is more variability to lose as the meridional temperature gradient weakens and the variability in thermal advection is reduced.

6. Discussion

Throughout this analysis we have considered the influence of all of the snow settings listed in Table 2 together. While we have not decomposed the relative contributions of these parameters further with full CAM simulations, additional SCAM simulations in which we only revert the settings of fresh snow density back to those of CLM4 reveal that this reproduces the difference in temperature variability from CLM5 that is found when all snow settings are reverted (Figure S4 in Supporting Information S1).

We have diagnosed that T_{2m} variability has improved in CESM2 through comparison with observations, but diagnosing whether T_{2m} variability has improved for the right reasons is more challenging. First, snow density was generally considered to be too low in CESM1 and so the increase in snow density in CESM2 is likely an improvement. van Kampenhout et al. (2017) demonstrated reduced biases in Greenland snow density and Lawrence et al. (2019) argued that the denser snow over Alaska was more in-line with observation based estimates. However, demonstrating more generally whether the snow density is now correct is a challenge given the lack of global snow density measurements. Second, we can take some comfort in the fact that CESM2 now has an improved representation of T_{2m} variability while apparently also correctly simulating the variability in temperature in the free troposphere (as indicated by the $T850$ PDFs in Figures 2g–2i). This gives us some confidence that the changes in land-atmosphere coupling through the snow density are not compensating for a bias in free tropospheric processes.

A third line of reasoning that would give us confidence that T_{2m} variability is now represented correctly would be if the behavior of the surface fluxes in association with T_{2m} variability were now improved relative to observations. We attempt such an analysis in Figure 12 but rather than leading to definitive conclusions, this illustrates some of the challenges when comparing with observation-based surface fluxes. Figures 12a–12c show the slope of the linear regression line of $SH\uparrow$ against T_{2m} that is, as was shown in Figure 9b for CAM6_CLM5 − SNWDENS. Recall that we had found that in CAM6_CLM5 compared to SNWDENS this gradient is more negative that is, when it gets cold, there is an enhanced SH ↑ and vice-versa. If this were an improvement, we would hope to see a slope of $SH\uparrow$ against T_{2m} that was too positive for SNWDENS and is now improved in CAM6_CLM5. This is broadly what is seen in Figures 12b and 12c, which show the bias in this regression slope relative to ERA5 for CAM6_CLM5 (b) and SNWDENS (c). The bias relative to FLUXNET2015 is overlayed in these panels in the dots. Figure 12c shows a regression slope of $SH\uparrow$ in SNWDENS that was too positive and in CAM6_CLM5 this bias has been reduced (Figure 12b), although a positive bias in the slope still remains over much of Russia and now there are some portions of Canada that exhibit a negative bias.

In Figures 12d–12f, the result is less compelling. Here we show the slope of the regression of net upward radiation (LW ↑ − SW ↓) against T_{2m}. These patterns are dominated by $LW\uparrow$ so it is closely related to the $LW\uparrow$ regression slope that was shown in Figure 9c but we use net radiation here since it was available for more FLUXNET2015 stations. Our analysis in Figure 9c had demonstrated that the LW ↑ regression slope had become more negative in CLM5 compared to SNWDENS, so we would hope to see here a positive bias in the slope for SNWDENS (Figure 12f) and this bias now being alleviated in CAM6_CLM5 (Figure 12e). Over Canada, there is some indication of this: the slope was too positive before, now there is an inconsistent sign of the bias over Canada and some indications that while things might have improved over much of the region, it may have gone too far in the opposite direction around Hudson's Bay. Over Russia, however, the bias in the regression slope has

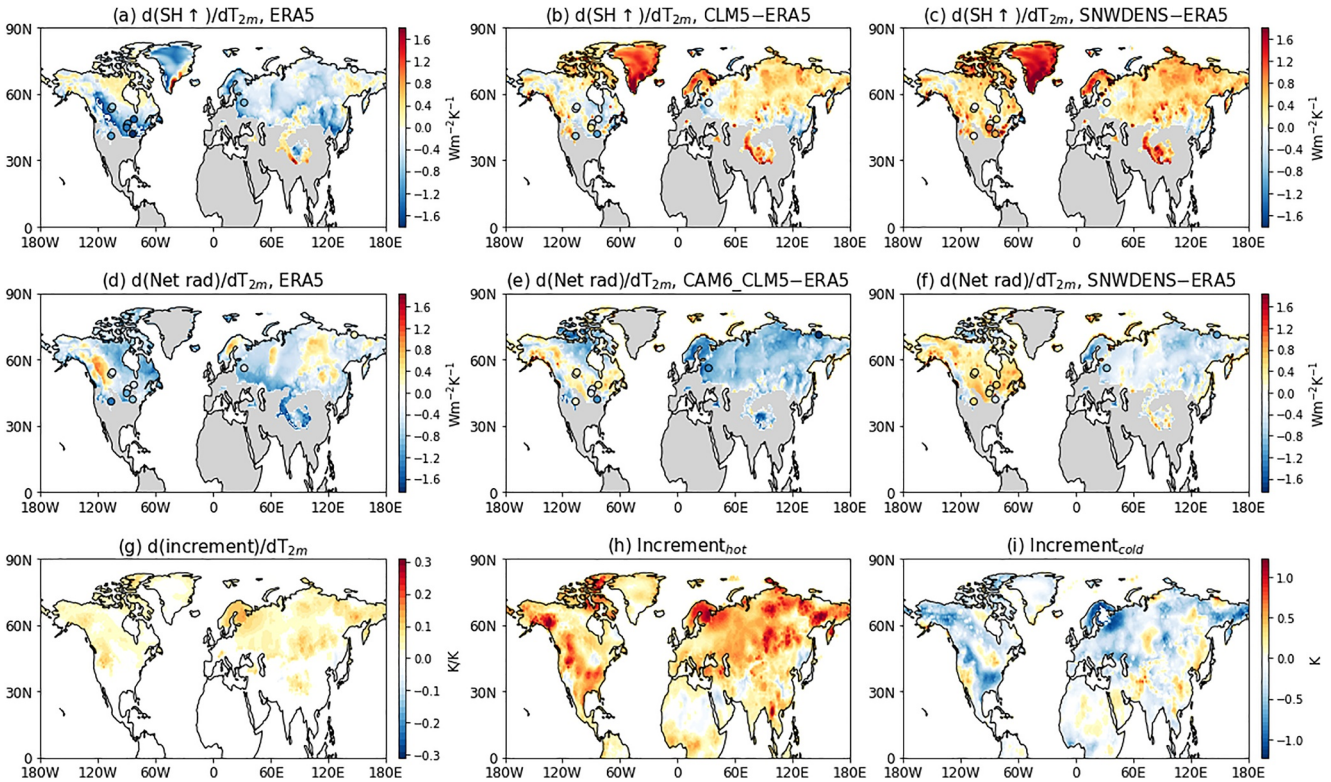

Figure 12. (a)–(c) Regression slope of sensible heat flux (SH ↑) onto T_{2m} over Northern Hemisphere land regions for (a) ERA5 reanalysis, (b) The difference between CAM6_CLM5 and ERA5 reanalysis and (c) the difference between SNWDENS and ERA5 reanalysis. The circular points depict the equivalent regression slopes for the FLUXNET stations in (a) and the difference relative to the fluxnet stations in (b) and (c). Regions where the December-January-February averaged snow fraction in CAM6_CLM5 is less than 0.5 are masked out in gray. (d)–(f) are as (a)–(c) but for net radiation (LW ↑ − SW ↓). (g) Shows the regression slope of the ERA5 analysis increments onto T_{2m}. (h) Shows the ERA5 analysis increment on days that fall into the 90th-99th percentile range of the CAM6_CLM5 distribution and (h) shows the ERA5 analysis increment on days that fall into the first to tenth percentile range of the CAM6_CLM5 distribution that is, equivalent to the warmest and coldest bins used in Figures 6–8.

worsened. SNWDENS started off with a negative bias in the slope over Russia (Figure 9f) and it has now become more negative that is, more biased (Figure 12e).

The comparison with ERA5 radiative fluxes, therefore, is not very compelling. But, of course, the ERA5 radiative fluxes are model-based and could suffer from issues in the underlying forecast model. Indeed, we find a rather systematic role for analysis increments in constraining T_{2m} in ERA5 (Figures 12g-12i). The regression slope of the analysis increments against T_{2m} is positive across the NH (Figure 12g). This means that on warm days, the analysis increments are acting to make it warmer (Figure 12h) and on cold days the analysis increments are acting to make it colder (Figure 12i). Or in other words, on warm days, the underlying forecast model is drifting to colder temperatures than it should and on cold days, it is drifting to warmer temperatures than it should and these anomalies can be of the order of ∼1K over the 12 hr forecast window for days at the tail end of the T_{2m} distribution (Figures 12h and 12i). It is, therefore, unclear whether the ERA5 surface fluxes can really be regarded as the truth. There is some support for increased bias in the behavior of net radiation in CAM6_CLM5 from the two Russian FLUXNET2015 stations (Figure 12e, dots), but with only two locations, confidence is lacking.

Overall, this analysis is somewhat inconclusive. We have to be cautious when considering the ERA5 surface fluxes to be the truth given the important role for analysis increments in shaping the ERA5 T_{2m} distribution. There is some indication that the change in behavior of SH ↑ as a function of T_{2m} with the new snow settings is an improvement (Figures 12a–12c) but there is also some indication that, over Russia, the change in behavior of the radiative fluxes with the new snow settings is actually a degradation.

Finally, Figure 13 illustrates whether other models in the CMIP6 archive overestimate wintertime T_{2m} variability in a manner similar to CESM1. Sub-seasonal T_{2m} variability during DJF is shown for 20 models. Here, it can be

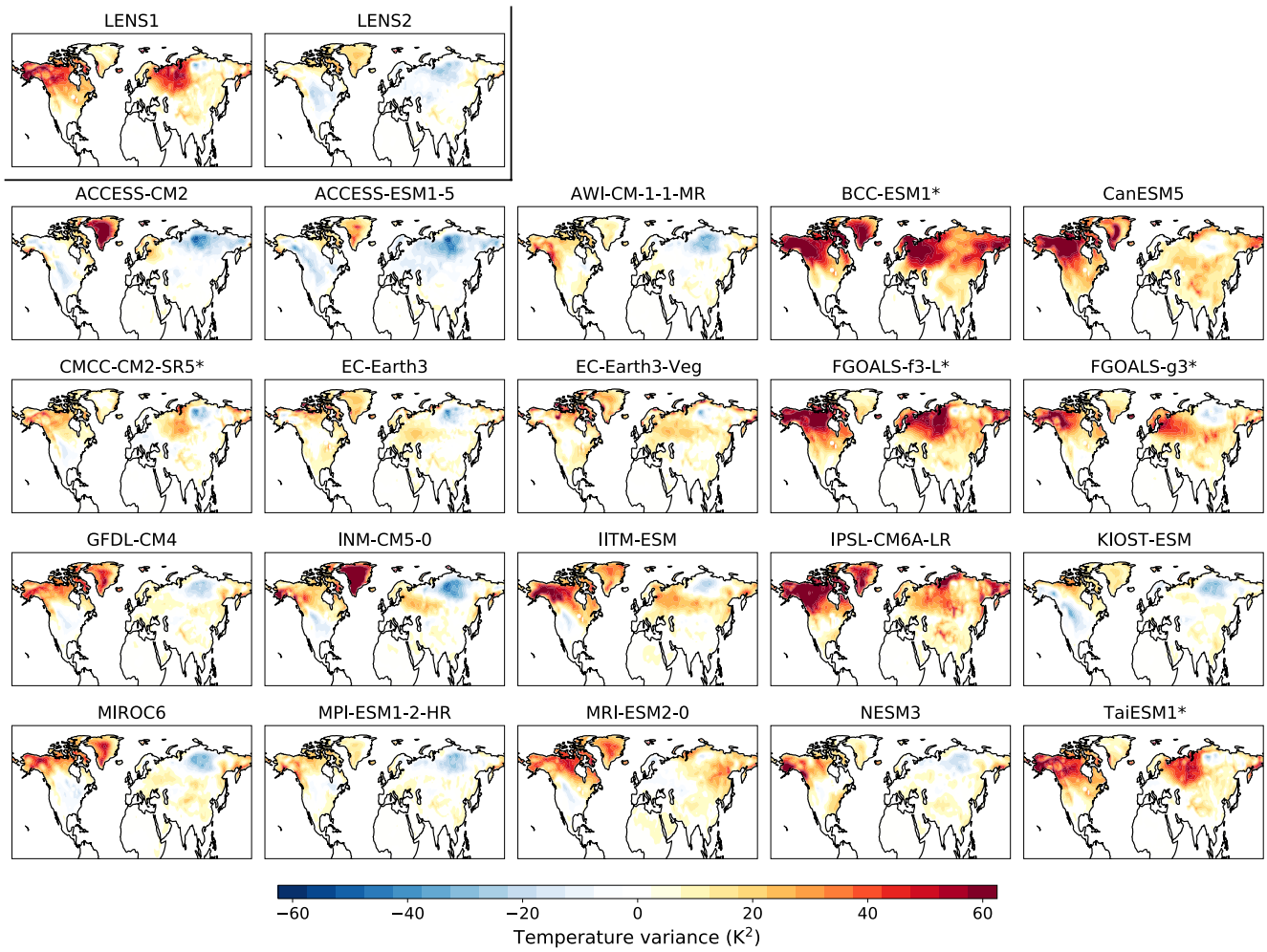

Figure 13. The bias in sub-seasonal December-January-February T_{2m} variance relative to Berkeley Earth Surface Temperature daily product for 1979–2014 for (top) LENS1 and LENS2 and (remaining panels) 20 models from CMIP6 that had daily average T_{2m} data available (The models marked with a* are using a land model which is related to CLM4 or CLM4.5).

seen that an overestimate of T_{2m} variability is, by no means, ubiquitous across the models in the CMIP6 archive. There are, however, several models that do exhibit daily T_{2m} variability that is much too high in a similar manner to CESM1: BCC-ESM1; CanESM5; FGOALS-f3-L; FGOALS-g3; IITM-ESM; IPSL-CM6A-LR, TaiESM1. Some of these models are directly related to CLM4 and for the others there are a variety of possible reasons for this bias and further investigation into this is beyond the scope of this study, but our experience based on CESM suggests that investigation by these modeling groups into their representation of snow density may prove useful.

7. Conclusions

In earlier versions of CESM, there was a substantial overestimate of wintertime daily T_{2m} variance over snow covered regions of the NH and here we have demonstrated that this bias has largely been alleviated in CESM2. The primary cause of this improvement is a change in the representation of fresh snow density that was implemented in the transition from CLM4 to CLM5. Surface energy balance arguments have been used to infer the mechanisms whereby this snow density change affects T_{2m} variability. Increased snow density in CLM5 leads to enhanced conductance of the snow layer. As a result, when the surface gets cold, an enhanced upward heat flux is induced across the snow which will dampen the surface temperature variability and the variability in near surface air temperatures above. Similarly, when it gets warm, an enhanced downward heat flux will be induced across the snow, taking heat away from the surface and, again, dampening surface and near surface temperature variability.

The result is an overall reduction in T_{2m} variability in CESM2 that is more aligned with observed. An attempted comparison with observed surface fluxes to assess whether CESM2 is getting T_{2m} variability correct for the correct reasons has proven somewhat inconclusive given the lack of direct observations of surface fluxes and a more in-depth analysis of reanalysis and station-based surface fluxes as well as snow density itself is warranted. However, it is at least reassuring that CESM2 now captures both the free tropospheric temperature variability and the near surface temperature variability with fidelity suggesting there is not some compensation between errors in the atmospheric circulation influences and the land-atmosphere coupling influence.

This change in T_{2m} variability has implications for future projections and we have shown that accompanying the reduced T_{2m} variability is a reduced ensemble spread in future projections of mean T_{2m} change in the CESM2 large ensemble compared to the CESM1 large ensemble over NH high latitudes. In addition, CESM1 which had greater NH wintertime variability in T_{2m} in its historical simulation, exhibited a larger decline in T_{2m} variability under climate change than CESM2 now does. Overall, a fortuitous result of changes in snow density that were implemented with other motivations in mind has now made CESM2 a more accurate tool for studies of NH wintertime temperature variability and change.

Data Availability Statement

The Community Earth System Model Version 1 (CESM1) and Community Earth System Model Version 2 (CESM2) large ensemble simulations as well as the Global Ocean Global Atmosphere (GOGA) simulations are freely available (see https://www.cesm.ucar.edu/projects/community-projects/LENS/ for the CESM1 large ensemble, https://www.cesm.ucar.edu/projects/community-projects/LENS2/ for the CESM2 large ensemble an https://www.cesm.ucar.edu/working_groups/CVC/ for the GOGA simulations. The Berkeley Earth Surface Temperature daily data are available from http://berkeleyearth.org/data/, ERA5 is available for download from the Copernicus Climate Change Services at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 and the Integrated Surface Database station data can be downloaded from https://www.ncei.noaa.gov/data/global-summary-of-the-day/archive/ and the FLUXNET2015 data can be downloaded from https://fluxnet.org/data/fluxnet2015-dataset/. The data required to reproduce the figures in this manuscript can be found at https://doi.org/10.5065/80jh-za14 and the analysis codes used to produce this data and to produce the figures can be found at https://github.com/islasimpson/snowpaper_2022. Other data from these simulations can be made available from Isla Simpson on request.

Acknowledgments
This work is supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under the Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne supercomputer (https://doi.org/10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory at NCAR. We also acknowledge the Community Earth System Model Version 1 Large Ensemble Community Project, the Community Earth System Model Version 2 Large Ensemble Community Project and supercomputing resources provided by the IBS Center for Climate Physics in South Korea (https://doi.org/10.5194/esd-2021-50) and the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for Coupled Model Intercomparison Project (CMIP), and we thank the climate modelling groups (listed on Figure 13) for producing and making available their model output. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. KAM acknowledges support from NSF AGS award 1939988.

References

Anderson, E. A. (1976). *A point energy and mass balance model of a snow cover* (Tech. Rep. No. NOAA tech. Rep. NWS 19). NOAA National Weather Service.

Bogenschutz, P. A., Gettlman, A., Hannay, C., Larson, V. E., Neale, R. B., Craig, C., & Chen, C.-C. (2018). The path to CAM6: Coupled simulations with CAM5.4 and CAM5.5. *Geoscientific Model Development*, *11*(1), 235–255. https://doi.org/10.5194/gmd-11-235-2018

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2019). The community Earth system model 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001916. https://doi.org/10.1029/2019MS001916

DeRepentigny, P., Jahn, A., Holland, M. M., & Smith, A. (2020). Arctic sea ice in two configurations of the CESM2 during the 20th and 21st centuries. *Journal of Geophysical Research: Oceans*, *125*(9), e2020JC016133. https://doi.org/10.1029/2020jc016133

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3–4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Diro, G. T., Sushama, L., & Huziy, O. (2018). Snow-atmosphere coupling and its impact on temperature variability and extremes over North America. *Climate Dynamics*, *50*(7–8), 2993–3007. https://doi.org/10.1007/s00382-017-3788-5

Durre, I., Wallace, J. M., & Lettenmaier, D. P. (2000). Dependence of extreme daily maximum temperatures on antecedent soil moisture in the contiguous United States during summer. *Journal of Climate*, *13*(14), 2641–2651. https://doi.org/10.1175/1520-0442(2000)013<2641:doedmt>2.0.co;2

Dutra, E., Schär, C., Viterbo, P., & Miranda, P. M. (2011). Land-atmosphere coupling associated with snow cover. *Geophysical Research Letters*, *38*(15), L15707. https://doi.org/10.1029/2011GL048435

Fischer, E. M., Lawrence, D. M., & Sanderson, B. M. (2011). Quantifying uncertainties in projections of extremes - A perturbed land surface parameter experiment. *Climate Dynamics*, *37*(7–8), 1381–1398. https://doi.org/10.1007/s00382-010-0915-y

Fischer, E. M., Seneviratne, S. I., Lüthi, D., & Schär, C. (2007). Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters*, *34*(6), L06707. https://doi.org/10.1029/2006gl029068

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., et al. (2019). The whole atmosphere community climate model version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, *124*(23), 12380–12403. https://doi.org/10.1029/2019JD030943

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Holmes, C. R., Woollings, T., Hawkins, E., & De Vries, H. (2016). Robust future changes in temperature variabilty under greenhouse gas forcing and the relationship with thermal advection. *Journal of Climate*, *29*(6), 2221–2236. https://doi.org/10.1175/jcli-d-14-00735.1

Huang, B., Banzon, V. F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T. C., et al. (2014). Extended reconstructed Sea Surface temperature version 4 (ERSST.v4): Part 1. Upgrades and intercomparisons. *Journal of Climate*, *28*, 911–930.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended reconstructed Sea Surface temperature, version 5 (ERSSTv5): Upgrades, validations and intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. https://doi.org/10.1175/jcli-d-16-0836.1

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., & Rosinski, J. (2008). A new Sea Surface temperature and sea ice boundary dataset for the community atmosphere model. *Notes and Correspondence*, *21*(19), 5145–5153. https://doi.org/10.1175/2008jcli2292.1

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The community Earth system model: A framework for collaborative Research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/bams-d-12-00121.1

Jordan, R. (1991). *A one-dimensional temperature model for a snow cover* (tech. Rep. No. 91-16). U.S. Army Corps of Engineers, Cold Regions Research and Engineering Laboratory.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2014). The community Earth system model (CESM) large ensemble Project. *Bulletin of the American Meteorological Society*, *96*, 1333–1349.

Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., Bonan, G., et al. (2019). The Community Land Model, version 5: Description of new features, benchmarking and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018ms001583

Lawrence, D., Fischer, R., Koven, C., Oleson, K., Swenson, S., Vertenstein, M., et al. (2018). In *Technical description of version 5.0 of the community land model (CLM)* (Tech. Rep). (pp. 1–337). NCAR National Center for Atmospheric Research. Retrieved from https://www.cesm.ucar.edu/models/cesm2/land/CLM50_Tech_Note.pdf

Linz, M., Chen, G., Zhang, B., & Zhang, P. (2019). A framework for understanding how dynamics shape temperature distributions. *Geophysical Research Letters*, *47*, e2019GL085684. https://doi.org/10.1029/2019GL085684

Neale, R., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., et al. (2012). *Description of the NCAR community atmosphere model (CAM 5.0)* (tech. Rep. Nos. TN-486+STR). NCAR National Center for Atmospheric Research.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Nature Scientific Data*, *7*, 225. https://doi.org/10.1038/s41597-020-0534-3

Rodgers, K., Lee, S. S., Rosenbloom, N., Timmerman, A., Danabasoglu, G., Deser, C., et al. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*. https://doi.org/10.5194/esd-12-1393-2021

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., et al. (2013). Berkeley Earth temperature averaging process. *Geoinformatics Geostatistics: An overview*, *1*(2), 1-13. https://doi.org/10.4172/gigs/1000103

Sato, T., Kosugi, K., Mochizuki, S., & Nemoto, M. (2008). Wind speed dependences of fracture and accumulation of snowflakes on snow surface. *Cold Regions Science and Technology*, *51*(2–3), 229–239. https://doi.org/10.1016/j.coldregions.2007.05.004

Schneider, T., Bischoff, T., & Plotka, H. (2015). Physics of changes in synoptic midlatitude temperature variability. *Journal of Climate*, *28*(6), 2312–2331. https://doi.org/10.1175/jcli-d-14-00632.1

Screen, J. A. (2014). Arctic amplification decreases temperature variance in northern mid- to high- latitude. *Nature Climate Change*, *4*(7), 577–582. https://doi.org/10.1038/nclimate2268

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Simpson, I. R., Bacmeister, J., Neale, R. B., Hannay, C., Gettelman, A., Garcia, R. R., et al. (2020). An evaluation of the large-scale atmospheric circulation and its variability in CESM2 and other CMIP models. *Journal of Geophysical Research: Atmospheres*, *125*(13), e2020JD032835. https://doi.org/10.1029/2020jd032835

Smith, A., Lott, N., & Vose, R. (2011). The integrated surface Database: Recent developments and partnerships. *Bulletin of the American Meteorological Society*, *92*(6), 704–708. https://doi.org/10.1175/2011bams3015.1

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

van Kampenhout, L., Lenarts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M., Slater, A. G., & van den Broeke, M. R. (2017). Improving the representation of polar snow and firn in the community Earth system model. *Journal of Advances in Modeling Earth Systems*, *9*(7), 2583–2600. https://doi.org/10.1002/2017ms000988

Vargas Zeppetello, L. R., Tètreault-Pinard, E., Battisti, D. S., & Baker, M. B. (2020). Identifying the sources of continental summertime temperature variance using a diagnostics model of land-atmosphere interactions. *Journal of Climate*, *33*(9), 3547–3564. https://doi.org/10.1175/jcli-d-19-0276.1

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., et al. (2012). The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geoscientific Model Development*, *5*(3), 773–791. https://doi.org/10.5194/gmd-5-773-2012

Xu, L., & Dirmeyer, P. (2011). Snow-atmosphere coupling strength in a global atmospheric model. *Geophysical Research Letters*, *28*(13), L13401. https://doi.org/10.1029/2011GL048049

Yen, Y.-C. (1962). Effective thermal conductivity of ventilated snow. *Journal of Geophysical Research*, *67*(3), 1091–1098. https://doi.org/10.1029/jz067i003p01091