Global climate models’ bias in surface temperature trends and variability

Richard Davy and Igor Esau

Nansen Environmental and Remote Sensing Center, Thormøhlensgt. 47, 5006 Bergen, Norway

E-mail: Richard.Davy@nersc.no

Received 24 September 2014, revised 20 October 2014
Accepted for publication 3 November 2014
Published 26 November 2014

Abstract

The Earth has warmed in the last century with the most rapid warming occurring near the surface in the Arctic. This Arctic amplification occurs partly because the extra heat is trapped in a thin layer of air near the surface due to the persistent stable-stratification found in this region. The amount of warming depends upon the extent of turbulent mixing in the atmosphere, which is described by the depth of the atmospheric boundary layer (ABL). Global climate models (GCMs) tend to over-estimate the depth of stably-stratified ABLs, and here we show that GCM biases in the ABL depth are strongly correlated with biases in the surface temperature variability. This highlights the need for a better description of the stably-stratified ABL in GCMs in order to constrain the current uncertainty in climate variability and projections of climate change in the surface layer.

Online supplementary data available from stacks.iop.org/ERL/9/114024/mmedia

Keywords: climate, climate change, global climate model, planetary boundary layer

The surface air temperature has become one of the most commonly-used metrics to assess our climate and climate change [1–3]. This is partly because it is such a readily and widely observed measure of the climate system (and so makes for a robust metric for climate models) but also because it is a very important parameter in the anthroposphere—the part of the environment that is inhabited and adapted by humans. A proper understanding of how the surface air temperature varies is essential to our understanding of Earth’s climate and how it responds to forcing. The magnitude of the surface air temperature response to forcing is determined by three components: the magnitude of the forcing, any feedback processes involved, and the effective heat capacity of the system [4]. The effective heat capacity of the atmosphere is defined by the region of turbulent mixing through which the heat is mixed i.e. it is defined by the depth of the ABL [5, 6]. So we can define the near-surface temperature response to forcing through an energy budget model of the form:

$$\frac{d\theta}{dt} = \frac{Q}{\rho c_p h},$$

where $Q$ (W m$^{-2}$) is the heat flux divergence within the ABL, $h$ (m) is the depth of the ABL, $\rho$ (kg m$^{-3}$) is the air density, $c_p$ (J kg$^{-1}$ K$^{-1}$) the heat capacity at constant pressure, and $\theta$ (K) is the potential temperature. Note that this is a reasonable approximation for well-mixed layers, where the potential temperature is constant with height, but may be more complicated in stable boundary layers where the potential temperature increases with height. An assessment of the response of stable boundary layers to increased radiative forcing using a single column model with a well-developed radiation scheme determined that one may expect an enhanced warming near the surface and a cooling aloft [7]. This has been ascribed to the increased turbulence within the ABL leading to better mixing, and so a weaker gradient in potential temperature with height. So a part of the signal of enhanced warming in stable boundary-layers may be due to the redistribution, rather than the accumulation, of heat within the ABL. While this highlights the complex relationship between radiative forcing and temperature response in stable boundary
layers, we can still expect that the magnitude of the SAT response to forcing will be modulated by the depth of the boundary-layer.

One important factor in the choice of a metric by which to assess the performance of a climate model is the availability and reliability of an equivalent observational dataset. This is a particular challenge when we wish to assess the performance of climate models with respect to the conditions of the ABL, since establishing the climatology of the ABL from observations has proved difficult. A comparison of the climatology of the ABL depth (as defined from the bulk Richardson number) between Radiosonde observations, reanalysis and global climate models (GCMs) over Europe and the continental US found that there were large uncertainties in the depth of the ABL: up to 50% for shallow, stable boundary layers, and around 20% for the deeper, convective boundary-layer [8–11]. There is also a general bias in the models towards deeper boundary-layers, as the models have difficulties representing stably-stratified conditions. Indeed, the ABL depth can even be lower than the first atmospheric model level of a GCM. With this bias towards deeper layers (and thus higher effective heat capacity) we would expect that the temperature response to forcing is under-estimated in models under conditions of stable-stratification, resulting in an under-estimation of the temperature trends and variability. Given the very different description of the ABL depth between GCMs in stably-stratified conditions, we may also expect the models to have the greatest differences in the temperature trends and variability under such conditions. We have tested these hypotheses by assessing the performance of GCMs for metrics that are directly affected by the ABL-response mechanism: the SAT mean, trends and variability, under different conditions of the boundary layer. These were used to assess model skill with respect to gridded observations of surface temperature and a reanalysis product, ERA-Interim [12]. Following an assessment of the controlling factors on the ABL depth, we use two proxies for the state of the boundary layer: for over-land we use the surface sensible heat flux and over ocean we use the vertical temperature gradient [13]. While the dependence of ABL height on sensible heat flux, wind speed and temperature gradient can be complicated in stably-stratified conditions [14], when we take a climatological-mean approach we can be confident that conditions with a mean negative sensible heat flux (cooling surface) will have a shallower ABL than conditions with a positive sensible heat flux. The ERA-Interim reanalysis assimilates observations from a world-wide Radiosonde network, and has relatively small root mean square (rms) forecast errors in the 850 hPa temperatures, giving us confidence in the climatological-mean vertical temperature gradient [12]. We established the climatology of the ABL in the GCM results using a bulk-Richardson approach, and compared the climatological mean ABL depth with the temperature variability across the models and reanalysis.

We used a statistical measure of model fidelity [15] to determine model departure from reanalysis for the historical simulations of the CMIP5 program, over the period 1979–2005. This is the full period of overlap between the CMIP5 historical runs and the ERA-Interim reanalysis dataset, which we use as a reference. We also assess the performance of these models as an ensemble by determining the inter-model mean and spread of these metrics. In this work we have focused on the Northern hemisphere because we use an approach which considers model error over a seasonal cycle. Nevertheless we obtain very similar results from a global analysis due to our approach which allows us to look at model performance with regard to the physics independently of the frequency of occurrence of a given state. So while the inclusion of the tropics in our analysis added more cases of convective conditions, it did not alter the result that the greatest model errors occur under stable-stratification.

The bias in the multi-model mean surface temperature variability, with respect to gridded observations, in the recent climatology is given in figure 1. The models consistently underestimate the SAT variability over the Siberian region—a region dominated by strongly-stable ABLs [5]—in every season but the boreal summer. Similarly, we see the models underestimate the variability over the North American continental interior during the winter months. This is consistent with our expectation that the GCMs have trouble representing the near-surface conditions under stably-stratified conditions [16], resulting in a consistent underestimation of surface temperature variability.

The typical geo-spatial pattern of the rms error in the mean, trend and variability of the surface air temperature is shown in figure 2. These errors represent the degree to which the models describe the climatological annual cycle of these fields at each location. Over ocean we see the greatest model error in mean SAT in the marginal ice zone from the Bering sea to the coast of Greenland. We expect this large departure from observations over the marginal ice zone to be related to differences in the sea–ice extent, since the presence or absence of sea–ice will strongly affect surface heat fluxes, and thus surface air temperature. There are similar hot-spots in the error in variability over the marginal ice zone, but we also see large errors over the continental interior in Asia and North America. These correspond with regions of large departure from observations in the SAT trends. These high-latitude continental interior regions are dominated by very shallow ABLs throughout the winter—and so a poor representation of the ABL in this region would be expected to lead to errors in the depicted annual cycle of SAT trends and variability. However, most locations have both stable and convective conditions, so for clarity we assess the models in parameter space, by their ability to match the observed climate variability under different states of the ABL.

The biases in the simulated SAT mean, trends and variability over land, with respect to ERA-Interim, are given in figure 3. They are plotted as a function of surface sensible heat flux. All model results approach a minimum bias in the mean SAT for weakly convective conditions which generally increases as we move towards more strongly convective conditions. But the largest biases are seen in the shallow, stably-stratified ABLs. There is a general warm bias in shallow ABLs with mean-model biases in excess of 10 K in very shallow layers. The ACCESS 1.3 model is the only model
which shows a strong bias in the mean SAT for weakly convective conditions: surface sensible heat flux in the range 20–70 W m$^{-2}$. It is worth noting that this model has included many new developments to the core ACCESS model which have not yet been fully tested for long-period climate runs [17].

All models show a consistent bias in the trend and variability in SAT over all convective conditions. If a model over-estimates the trend in weak convective conditions, it tends to over-estimate the trend in strong convective conditions, and vice versa. However, in stable conditions we see larger biases in the trends and variability, with a general negative bias in the models. This pattern is most apparent in the model biases in SAT variability: there are small, consistent biases in convective conditions which rapidly become increasingly negative as we move towards increasingly stable stratification. This is not surprising since a linear temperature trend picks out a single mode of variability, the nature of which can be very sensitive to the period of investigation, whereas temperature variance is an integrated measure of all modes of variability in the period under analysis, and so will give a clearer measure of any systematic biases that affect the temperature response.

This under-estimation of the magnitude of SAT trends and variability in stable conditions is consistent with our expectation: given that models are biased towards deeper ABLs under stable stratification, and that these shallow layers can strongly affect the magnitude of the SAT response to forcing, we expect a positive bias in ABL depth to result in a negative bias in SAT trends and variability.

The rms error of the SAT mean, trend and inter-annual variability reflects the pattern of model biases—large biases lead to large rms errors. These are given as a function of surface sensible heat flux for locations over land, and as a function of lower tropospheric temperature stability over ocean (figure 4). The models show the largest departure from the observed temperature trends and variability in shallow ABLs whether over land or ocean. The model error has a

Figure 1. Bias in the climatological-mean surface air temperature variability. The Multi-Model Mean of the CMIP5 historical models with respect to the HadCRUTv4 gridded observations for (A) spring (MAM), (B) summer (JJA), (C) autumn (SON) and (D) winter (DJF) over the period 1979–2010.
similar dependency on ABL depth in all models. The largest errors occur for shallow, stably-stratified layers, approach a minimum for weakly convective conditions, and then increase again towards strong convection. In stably-stratified boundary layers we can expect a strong increase in the model departure from observed temperature trends/variability since it is in these shallow boundary layers that the temperature response is most sensitive to the forcing, so a small difference in the depth of the ABL can lead to a large difference in the temperature trend/variability.

The agreement between the models as to the temperature trends and variability depends upon the state of the ABL. The inter-model mean and spread in the temperature trends and variability, as a function of the state of the ABL over land and ocean is given in figure 5. There is a large inter-model spread under stably-stratified conditions which approaches a minimum as we move to weakly convective conditions, before increasing again towards deep convection. We see a similar pattern in the inter-model spread for both the temperature trends and variability over land and ocean. The strongest trends are seen in stable conditions, where we also see the greatest spread between the models. The temperature variability can be divided into two regions with a sharp transition zone: in stable conditions there is a high inter-annual variability, and a large inter-model spread, whereas in convective conditions there is a much lower variability, and a smaller spread.

This is to be expected since the shallow ABLs that form in stable stratification amplify temperature changes relative to deep boundary-layers, under a given forcing. So in shallow layers we not only get the stronger temperature response to forcing—the greater trends and variability—but also the different climatology of the ABL between models leads to greater inter-model spread under these conditions. We also see an increasing inter-model spread as we move towards deep convective conditions. Climate models also have trouble representing the ABL depth under strongly convective conditions as many physical processes that occur in these conditions (e.g. self-organization of turbulent structures) are not accounted for in their parameterization schemes. However, due to the reciprocal relationship between ABL depth and temperature response, differences in the ABL depth are less important in determining the SAT trends and variability in deep convective layers, and so the inter-model spread is not as strong as in the shallow, stably-stratified conditions.

Figure 2. Geospatial pattern of typical model errors. The mean of the rms error in the (A) mean, (B) trend and (C) variability in the surface air temperature for the 36 member ensemble of CMIP5 model results.
We evaluate the ABL depth in the GCMs using a bulk-Richardson approach. The ABL height is defined to be the height at which the bulk Richardson number exceeds a critical value, 0.25, at which point it is assumed that turbulence ceases. This approach has its limitations in determining the vertical extent of the boundary-layer mixing [18], but in a study comparing many methodologies for deriving the ABL depth, the bulk-Richardson method was recommended for application to GCMs [13]. Overall there is a large variation in the global mean ABL depth in the models, ranging from approximately 400–800 m for low-lying, over-land locations. Generally speaking the bulk-Richardson method over-estimates the ABL depth over mountainous regions and does not capture the over-land variability that we see in the ERA-Interim reanalysis (Methods).

We found a strong positive relationship between the mean reciprocal ABL depth and the SAT variability in the models (figure 6), with most models having ABL depths greater than that derived from the reanalysis. This is consistent with our expectation that the more models over-estimate the mixing, the more this will dampen the response of the surface temperature to changes in the surface energy budget—leading to reduced surface temperature variability.

There has been a lot of development of the GCMs that have contributed to CMIP since the last phase of the project, CMIP3. However, there are still limitations in the models’ ability to describe the observed climate variability and trends, especially in cold conditions which tend to be dominated by stably-stratified conditions. Some of these limitations come from constraining the boundary conditions on the model, while others derive from the representation of physical processes in the models—the parameterization schemes.

Our assessment of the CMIP5 model biases and error, with respect to observations and reanalysis, has highlighted the relatively poor performance of these models in stably-stratified conditions. It is in these conditions that the models show the greatest departure from observations, and there is the greatest difference between the models in their representation of the surface conditions. This is partly due to the representation of the ABL in GCMs, which are biased towards over-estimating the amount of mixing and producing deeper-than-observed ABLs, which causes the models to
under-estimate the temperature trends and variability. The problems GCMs have representing the observed variations in the ABL are not surprising given that the parameterization schemes they use are tuned to neutral conditions. These schemes do a reasonably good job of reproducing the observed structure of the typical ABL of mid-latitudes [8], but miss some essential physics of the turbulence in the ABL under stable stratification and strong convective conditions [19]. This is especially important for shallow, stably-stratified layers which amplify the surface temperature response to forcing—for example in the Arctic. A recent decomposition of the processes that contribute to Arctic amplification [20–22] emphasized the importance of local temperature feedback processes [23], and thus the importance of an accurate

Figure 4. Model errors as a function of ABL state. The rms error in the (A) mean, (B) trend and (C) standard deviation in surface air temperature, as a function of the mean surface sensible heat flux over land, and in the (D) trend and (E) standard deviation in surface air temperature as a function of lower troposphere stability over ocean. The colored lines represent 36 of the model results from the CMIP5 simulations, and the multi-model mean and median are highlighted in solid and dashed black lines respectively. The thick red line is the mean of the model errors in each bin (i.e. the ‘typical’ model error).
representation of the ABL mixing in the Arctic, since this will determine the vertical profile of the warming.

Strongly stably-stratified conditions are rare, and so even large model error under these conditions may not significantly affect global-mean projections. However, given the importance of the ABL depth in determining the response of the surface temperature to enhanced forcing, this highlights the importance of improving the representation of stably-stratified ABLs in GCMs. A more accurate representation of shallow ABLs would improve the models ability to describe both the SAT mean and variability in the mean climatology, and the trend in surface temperature under an enhanced forcing. This is especially important in the Polar Regions which frequently have a strong surface inversion.

Acknowledgments

We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the US 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. We thank the ECMWF for providing ERA-Interim.

References

[1] Graverson R G et al 2008 Vertical structure of recent Arctic warming Nature 451 53–6
[2] Serreze M C, Barrett A P, Stroeve J C, Kindig D N and Holland M M 2009 The emergence of surface-based Arctic amplification Cryosphere 3 11–9
[3] Johannessen O et al 2004 Arctic climate change: observed and modeled temperature and sea–ice variability Tellus A 56 328–41
[4] Hasselmann K 1976 Stochastic climate models Part 1. Theory Tellus 28 473–85
[5] Esau I, Davy R and Outtin S 2012 Complementary explanation of temperature response in the lower atmosphere Environ. Res. Lett. 7 044026
[6] Esau I and Zilitinkevich S 2010 On the role of the planetary boundary layer depth in the climate system Adv. Sci. Res. 4 63–9
[7] McNider R T et al 2012 Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing J. Geophys. Res. 117 D14106
[8] Seidel D J et al 2012 Climatology of the planetary boundary layer over the continental United States and Europe J. Geophys. Res. 115 D16113
[9] Seidel D J, Ao C O and Li K 2010 Estimating climatological planetary boundary layer heights from Radiosonde observations: comparison of methods and uncertainty analysis J. Geophys. Res. 115 D16113
[10] Cuxart J et al 2012 Single-column model intercomparison for a stably stratified atmospheric boundary layer Bound.-Layer Meteorol. 118 273–303
[11] Svensson G and Holtslag A A M 2009 Analysis of model results for the turning of the wind and related momentum fluxes in the stable boundary layer Bound.-Layer Meteorol. 132 261–77
[12] Dee D P et al 2011 The ERA-interim reanalysis: configuration and performance of the data assimilation system Q. J. R. Meteorol. Soc. 137 553–97
[13] Medeiros B, Hall A and Stevens B 2005 What controls the mean depth of the ABL? J. Clim. 18 3157–72
[14] Basu S, Holtslag A A M, Van De Wiel B J H, Moene A F and Steeneveld G-J 2008 An inconvenient ‘truth’ about using sensible heat flux as a surface boundary condition in models under stably stratified regimes Acta Geophys. 56 88–99
[15] Gleckler P J, Taylor K E and Doutriaux C 2008 Performance metrics for climate models Geophys. Res. 113 D06104
[16] Holtslag A A M et al 2013 Stable atmospheric boundary layers and diurnal cycles: challenges for weather and climate models Bull. Am. Meteorol. Soc. 94 1691–706
[17] Davy R and Esau I 2014 Surface air temperature variability in global climate models Atmos. Sci. Lett. 15 13–20
[18] Von Engeln A and Teixeira J 2013 A planetary boundary layer height climatology derived from ECMWF re-analysis data J. Clim. 26 6575–90
[19] Pithan F and Mauritsen T 2014 Arctic amplification dominated by temperature feedbacks in contemporary climate models Nat. Geosci. 7 181–4
[20] Serreze M C and Francis J A 2006 The arctic amplification debate Clim. Change 76 241–64
[21] Holland M M and Bitz C M 2003 Polar amplification of climate change in coupled models Clim. Dyn. 21 221–32
[22] Taylor P C et al 2013 A decomposition of feedback contributions to polar warming amplification J. Clim. 26 7023–43
[23] Zilitinkevich S S, Elperin T, Kleeorin N, Rogachevskii I and Esau I 2012 A hierarchy of energy- and flux-budget (EFB) turbulence closure models for stably stratified geophysical flows Bound.-Layer Meteorol. 146 341–73