A Systematic Approach to Assessing the Sources and Global Impacts of Errors in Climate Models

S. D. SCHUBERT

Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, and Science Systems and Applications, Inc., Lanham, Maryland

Y. CHANG

Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, and Goddard Earth Sciences Technology and Research, Morgan State University, Baltimore, Maryland

H. WANG

Science Systems and Applications, Inc., Lanham, Maryland

R. D. KOSTER AND A. M. MOLOD

Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, Maryland

(Manuscript received 7 March 2019, in final form 18 July 2019)

ABSTRACT

We outline a framework for identifying the geographical sources of biases in climate models. By forcing the model with time-averaged short-term analysis increments [tendency bias corrections (TBCs)] over well-defined regions, we can quantify how the associated reduced tendency errors in these regions manifest themselves both locally and remotely through large-scale teleconnections. Companion experiments in which the model is fully corrected [constrained to remain close to the analysis at each time step, termed replay (RPL)] in the various regions provide an upper bound to the local and remote TBC impacts. An example is given based on MERRA-2 and the NASA/GMAO GEOS AGCM used to generate MERRA-2. The results highlight the ability of the approach to isolate the geographical sources of some of the long-standing boreal summer biases of the GEOS model, including a stunted North Pacific summer jet, a dry bias in the U.S. Great Plains, and a warm bias over most of the Northern Hemisphere land. In particular, we show that the TBC over a region that encompasses Tibet has by far the largest impact (compared with all other regions) on the NH summer jets and related variables, leading to significant improvements in the simulation of North American temperature and, to a lesser degree, precipitation. It is further shown that the results of the regional TBC experiments are for the most part linear in the summer hemisphere, allowing a robust interpretation of the results.

1. Introduction

Understanding and correcting errors in climate models has long relied on both intuition and trial and error. Efforts to diagnose the errors and provide some guidance to developers have been of some value, although such efforts, with few exceptions, have been more successful in identifying and documenting the errors in the model simulations than in identifying the model formulation deficiencies (hereafter referred to as model error) ultimately

responsible for them. Perhaps the most promising approach to identifying model error has been through the assessment of short-term forecast errors (typically obtained as part of a data assimilation system) that in principle allows the isolation of the different sources of model error since the forecasts are short enough that the errors do not have time to interact (e.g., [Klinker and Sardeshmukh 1992](#); [Schubert and Chang 1996](#)). Here, the underlying assumption is that by capturing the errors at the very early stages of growth (before nonlinearities develop), we should be able to associate those errors with specific deficiencies in the model's formulation of the relevant physical processes. A key limitation of

Corresponding author: Siegfried D. Schubert, siegfried.d.schubert@nasa.gov

this approach has been the quality/biases of the early reanalysis data, since the forecast errors are invariably estimated as differences with respect to the analyzed states, which themselves depend on short-term model forecasts. This approach has in recent years received renewed attention in a climate modeling context as a result of much improved long-term reanalysis data (e.g., Gelaro et al. 2017; C3S 2017; Saha et al. 2010; Kobayashi et al. 2015) together with the mounting evidence that many long-term climate biases already have some expression at very short time scales, typically associated with the model's parameterizations of the so-called "fast physics" (e.g., Phillips et al. 2004; Rodwell and Palmer 2007; Xie et al. 2012; Ma et al. 2015; Bhargava et al. 2018).

Recently, Chang et al. (2019) illustrated the effectiveness of the tendency bias correction (TBC) approach to correcting model error. Time-averaged 6-hourly analysis increments (first guess forecast minus analysis) obtained from MERRA-2 reanalysis data were added with opposite sign as additional forcing terms to the model equations during a free-running simulation of the GEOS AGCM (the model underlying MERRA-2). The application of TBC resulted in a substantial improvement to the AGCM's long-term climate, including marked improvements in the simulation of the boreal summer North Pacific jet and midlatitude weather transients, as well as improvements in the summer precipitation and surface air temperature over North America.

In the present study, we build on the results of Chang et al. (2019) by exploring in more detail the geographical sources of the model errors in the GEOS AGCM along with their global impacts—the ability, for example, of errors generated in specific regions by inadequate model formulations to manifest themselves, through global teleconnections, in remote areas. In essence, we perform a series of simulations in which TBCs are applied only in specific regions and quantify the degree to which improvements are seen across the globe. As we shall see, linearities underlying the system allow a systematic approach to identifying the geographical sources of model error, an approach that can be tailored to address specific biases (e.g., by telescoping the analysis regions) and that can be optimized for available computing resources. Furthermore, the general approach described herein is accessible to any modeling group; because it employs a "replay" methodology (hereafter referred to as RPL) developed in the GMAO (Orbe et al. 2017; Takacs et al. 2018; Chang et al. 2019) that constrains a model to remain close to an existing reanalysis (see section 2), modeling groups wanting to use the approach are not required to run a full data assimilation system to obtain the needed TBCs.

In addition to quantifying the impacts of applying TBCs in specific regions, we are also able to judge their efficacy by taking advantage of the fact that the regional RPL (where the model is forced to track the reanalysis in that region at each time step) provides an upper bound to what can be achieved by the TBC forcing in that region. Section 2 describes the overall methodology, the GEOS AGCM, and the specific experiments performed here. Section 3 presents the results. Discussion and the conclusions are provided in section 4.

2. Methodology and model experiments

a. Methodology

As described in Takacs et al. (2018), the RPL technique takes advantage of the incremental analysis update procedure (Bloom et al. 1996)¹ employed in the GEOS data assimilation system to force a model to track a pre-existing reanalysis (in our case MERRA-2; Gelaro et al. 2017). The equations governing RPL have the form

$$\frac{\partial x}{\partial t} = f(x) + \Delta x, \quad (1)$$

where $\Delta x = (\text{analysis} - \text{forecast})/6 \text{ h}$ is the instantaneous analysis increment (applied to the model's prognostic variables; see Table 1), and $f(x)$ consists of all the dynamics and physics terms of the model (basically the uncorrected model). The details of how the Δx values are computed during the course of RPL is provided in the appendix of Chang et al. (2019).

Following Chang et al. (2019), the governing equations for the TBC approach have the same form as (1), except that now the forcing term is a long-term mean of the increments. In particular,

$$\frac{\partial x}{\partial t} = f(x) + \overline{\Delta x}, \quad (2)$$

where $\overline{\Delta x}$ is the time average (denoted by the overbar) of the instantaneous Δx values computed over the years 1980–2016. The averaging is done independently for different times of the day and for different times of the year; in this way, the Δx values vary both diurnally and seasonally.

In the AGCM experiments described in section 2c, the Δx (in the case of RPL) and the $\overline{\Delta x}$ (in the case of

¹ It is worth noting that the incremental analysis update procedure (upon which the replay is based) differs from the traditional nudging approach in that the increments are introduced as a state-independent forcing term whereas in the case of nudging the entire model state is relaxed toward an analysis (Bloom et al. 1996).

TABLE 1. A summary of the AGCM experiments. Corrections are applied to the model's prognostic atmospheric variables consisting of temperature (T), specific humidity (q), horizontal winds (u , v), surface pressure (p_s), and ozone (O_3). See Fig. 1 for definitions of the regions.

Exp. No.	Exp. name	Description	Model
1	CNTL	37-yr control simulation for the period 1980–2016	Uncorrected AGCM
2	TBC/RPL_NP TBC/RPL_NM TBC/RPL_TR TBC/RPL_SM TBC/RPL_SP	37-yr simulations for the period 1980–2016 in which TBC or RPL is applied to selected zonal bands	AGCM with TBC or RPL
3	TBC/RPL_NM1 TBC/RPL_NM2 TBC/RPL_NM3 TBC/RPL_NM4 TBC/RPL_NM5 TBC/RPL_NM6	37-yr simulations for the period 1980–2016 in which TBC or RPL is applied to selected regions spanning the NM region	AGCM with TBC or RPL
4	TBC/RPL_TR1 TBC/RPL_TR2 TBC/RPL_TR3 TBC/RPL_TR4 TBC/RPL_TR5 TBC/RPL_TR6	37-yr simulations for the period 1980–2016 in which TBC or RPL is applied to selected regions spanning the TR region	AGCM with TBC or RPL
5	TBC_GLOBAL	37-yr simulations for the period 1980–2016 in which TBC is applied globally	AGCM with TBC
6	RPL_Tibet RPL_Tibet_T (T only) RPL_Tibet_uv (u , v only) RPL_Tibet_q (q only) RPL_Tibet_All (T , u , v , q , p_s , and O_3)	For each case: 31 runs initialized 30 April and integrated through 2 September of each year (1980–2010)	AGCM with RPL of selected variables for the Tibet region (70°–110°E; 17°–50°N)

TBC) are applied, in a series of simulations, to different regions of the globe (see section 2c). An obvious strategy would be to carry out the RPL experiments first and then average the resulting Δx in time to produce the $\bar{\Delta x}$ necessary to carry out the TBC experiments. For logistical reasons, we did not do that here. Instead [as in Chang et al. (2019)], we took advantage of the fact that the AGCM we used (section 2b) is the same as that underlying MERRA-2, though run at lower resolution. We could thus take the increments directly from the MERRA-2 archive (appropriately averaged to the reduced resolution of the AGCM) for the TBC experiments; the RPL experiments were in fact run later. As such, the Δx from the RPL experiments and those taken directly from MERRA-2 in producing the TBC experiments will differ to some degree, although we expect these differences to be unimportant to the conclusions of this study.

b. The GEOS-5 AGCM

Again, the GEOS AGCM used here is the same as that used to generate MERRA-2, although here the model is run at a coarser horizontal resolution (approximately 1°). In fact, the model is identical to that used in Chang et al. (2019). As described in Gelaro et al. (2017), this AGCM includes the finite-volume

dynamical core of Putman and Lin (2007), which uses a cubed sphere horizontal discretization and 72 hybrid-eta levels from the surface to 0.01 hPa. Recent upgrades to the physical parameterization schemes (in going from the original MERRA to MERRA-2) include increased re-evaporation of frozen precipitation and cloud condensate, changes to the background gravity wave drag, and an improved relationship between the ocean surface roughness and ocean surface stress. The model also includes a Tokioka-type trigger on deep convection as part of the relaxed Arakawa–Schubert (RAS; Moorthi and Suarez 1992) convective parameterization scheme (Bacmeister and Stephens 2011). A new glaciated land representation and seasonally varying sea ice albedo were implemented for MERRA-2, leading to improved air temperatures and reduced biases in the net energy flux over these surfaces (Cullather et al. 2014). The model includes the Catchment land surface model developed by Koster et al. (2000). Further details about this version of the GEOS AGCM can be found in Molod et al. (2015).

c. The experiments

The RPL and TBC experiments consist of two sets of GEOS AGCM simulations in which the correction terms (either $\bar{\Delta x}$ for TBC, or Δx for RPL; see section 2a)

FIG. 1. The regions over which the corrections were applied in the experiments with the global GEOS AGCM forced with observed SST. In each experiment, the correction term (either TBC: $\Delta\bar{x}$ or RPL: Δx) was limited to one of the 17 regions (outlined in red) specified above, and run for the period 1980–2016. A TBC run was also made in which the increments $\Delta\bar{x}$ were applied globally. A region centered on the Tibetan highlands (outlined in green) was the focus of additional runs in which the TBC was limited to specific variables. In addition, a CNTL run was made without any correction terms.

are applied in various regions (Fig. 1) that together span the globe. The choice of the regions, while somewhat arbitrary, was a compromise between the desire for the highest possible spatial resolution (i.e., the smallest possible regions) to better pinpoint the sources of model error and the need to limit the computational burden of running a large number of simulations. We in fact found it necessary to run the simulations for several decades to reduce the noise in the results. In addition, we found it helpful to further reduce the noise by applying a t -test filter before plotting the results; any differences (experiment minus control) not significant at the 5% level were set to zero.

The full set of AGCM experiments is given in Table 1. For all but the last set of experiments (RPL_Tibet),² the model was run for the period 1980–2016, forced with time-varying greenhouse gases (GHGs) as described in appendix A of Schubert et al. (2014) and with the observed SST and sea ice fraction used in MERRA-2.

² For these experiments we took an ensemble approach where, rather than running one long continuous simulation, the model was reinitialized on 30 April of each year and run only for 4 months to cover the boreal summer season (June–August) of each year.

As summarized in Gelaro et al. (2017), the MERRA-2 SSTs are based on a combination of different high-resolution daily NOAA OISST and OSTIA products, though prior to 1 January 1982, they are based on the Coupled Model Intercomparison Project (CMIP) mid-monthly 1° data.

The simulations include an uncorrected control (CNTL) run, a run in which the TBC is applied globally (as in Chang et al. 2019), and two sets of 17 runs in which either TBC or RPL is applied over selected regions. The 17 regions (shown in Fig. 1) consist of five zonal bands: the northern polar region (NP), the northern midlatitudes (NM), the tropical region (TR), the southern midlatitudes (SM), and the southern polar region (SP). The NM and TR latitudinal bands are each split into six subregions. Finally, some additional RPL experiments were performed that focus on the Tibet region (70°–110°E, 17°–50°N, outlined in green in Fig. 1); these experiments allow us to assess which of the variables (horizontal winds, T , or q) in that region seem to have the largest global impacts.

3. Results

We present here the results from the various experiments described in section 2c and summarized in Table 1.

Our focus on the boreal summer season is motivated in part by the desire to isolate the sources of some of the long-standing biases of the GEOS AGCM, including a stunted North Pacific summer jet (NPSJ), a dry bias over the U.S. Great Plains, and an overall warm bias over most of the summer continents—biases that were substantially corrected when applying the TBC globally (Chang et al. 2019). We thus focus here on the upper-tropospheric circulation, 2-m air temperature (T2m), and precipitation. The warm and dry biases over the Great Plains, in particular, appear to be common to many AGCMs, and therefore our results here should be of general interest (e.g., Lin et al. 2017). We leave to a future study a more comprehensive assessment of the model biases [but see Chang et al. (2019) for a discussion of the impact of TBC on clouds].

A note is warranted regarding our presentation approach and the organization of our results. Many of the figures in section 3a begin with a global map of the CTRL simulation's bias (relative to the analysis or observations) in a given variable. In fact, the negative of this bias is shown to allow a direct comparison with the changes (relative to CTRL) produced in the experiment simulations. If these latter changes agree in magnitude and pattern with the original biases, the change imposed in the experiment can be said to reduce or even eliminate the biases. Having said that, we want to be clear that the main focus of section 3a is to better understand and quantify how the model responds to the TBC in the different regions. We do that here by quantifying the extent to which the results obtained from applying TBC in a particular large region (e.g., over the entire global or in a particular zonal band) can be reconstructed as the sum of the results obtained by applying TBC to the various subregions encompassing the larger region. On the other hand, quantifying the extent to which applying TBC regionally actually corrects *model bias* is the subject of section 3b. Finally, in section 3c we take a more detailed look at the role of the Tibet region employing various replay experiments.

a. An assessment of regional TBC impacts

Figure 2 shows the 250-mb (1 mb = 1 hPa) *u*-wind results for the TBC experiments focused on the northern midlatitude region (NM), together with the results for the corresponding six subregions (NM_{*i*}, *i* = 1, 6). The results indicate that applying the TBC over the NM region acts to correct much of the long-term bias in that region (cf. top left and top center panels); this will be quantified in section 3b. The bottom six panels of Fig. 2 show the results of the TBC experiments for the six subregions of NM. The most striking aspect of these

latter results is the relatively large and hemispheric-wide response to the TBC in the NM₂ region: much of the Northern Hemisphere response to the TBC in the NM region is due to applying TBC over the NM₂ region alone.

The similarities of the various responses are quantified in terms of a normalized spatial inner product. This has the form

$$I = \frac{\mathbf{X} \cdot \mathbf{Y}}{|\mathbf{Y}|^2} = \frac{|\mathbf{X}|}{|\mathbf{Y}|} \cos\theta, \quad (3)$$

where $\mathbf{X}$ and $\mathbf{Y}$ are vectors with components corresponding to the values at the various (area-weighted) grid points making up a region. The vertical bars indicate the vector length or magnitude, and θ is the angle between the two vectors. The right-hand side of (3) makes it clear that (*I*) measures both the similarity of the patterns ($\cos\theta$), and the similarity in magnitudes (the ratio), with a perfect match ($\mathbf{X} = \mathbf{Y}$) corresponding to an amplitude ratio and $\cos\theta$ both equal to 1. We note, however, that the ratio of amplitudes is not constrained to be less than or equal to 1. In fact, as we shall see, double counting when adding the results from the various subregions can in some cases produce amplitude ratios (and values of *I*) substantially greater than 1, so that having both components of (*I*) can be especially useful in the interpretation of those results.

For the results shown in this section, $\mathbf{Y}$ is always the response obtained from applying the TBC over the larger of the two regions whose impacts are being compared. For example, for the left (first) bar in the top panel of Fig. 3, $\mathbf{X}$ is (TBC_NP – CNTL), and $\mathbf{Y}$ is (TBC_GLOBAL – CNTL) with the inner products computed over the domain NM. For the second bar $\mathbf{X}$ is (TBC_NM – CNTL), and $\mathbf{Y}$ is again (TBC_GLOBAL – CNTL) with the inner products computed over the domain NM, and so on. For the last bar, $\mathbf{X}$ is the sum of the responses in the five zonal bands (NP, NM, TR, SM, SP). Also, in the bottom panel of Fig. 3, for the left (first) bar, $\mathbf{X}$ is (TBC_NM₁ – CNTL), and $\mathbf{Y}$ is (TBC_NM – CNTL) with the inner products computed over the domain NM, and so on, for the other five subregions of NM. For the last bar, $\mathbf{X}$ is the sum of the responses in the six subregions. Results in the subsequent bar graph figures are computed analogously. In the following, we will describe the response $\mathbf{X}$ as accounting for a certain fraction (or percent) of the response $\mathbf{Y}$, based on the fact that inner product (3) is equal to the scalar projection of $\mathbf{X}$ onto $\mathbf{Y}$ ($|\mathbf{X}| \cos\theta$) divided by the magnitude of $\mathbf{Y}$.

Using this measure, Fig. 3 (top panel) shows that the correction obtained in the NM region by applying

FIG. 2. Results of the TBC experiments for the JJA 250-mb u wind (m s^{-1}). (top left) Minus the model bias $[-(\text{CNTL} - \text{MERRA-2})]$. (top center) The response to the TBC in the NM region ($\text{TBC_NM} - \text{CNTL}$). (top right) The sum of the individual results of the six NM regions (shown in the bottom six panels). Results are the averages over the years 1980–2016.

TBC globally is essentially captured by applying TBC in just the NM region (the normalized inner product for NM is approximately equal to 1): the TBC applied outside that region has negligible impacts on the mean 250-mb u wind in the NM region. This allows us to just focus on the NM region to further isolate the main sources of the impacts in that region. We can indeed isolate our focus even further; the bottom panel of Fig. 3 shows that the response to TBC in the NM_2 region accounts for more than 40% of the total response to NM TBC. In comparison, the NM_4 region (the northeast Pacific) accounts for about 20% of the total TBC impact in the NM region, with the other subregions showing less impact. Overall, these results highlight the

importance of errors over the Tibet (NM_2) region. We will take a more detailed look at the impact of the Tibet region in section 3c.

An important and somewhat unexpected result is that the sum of the responses to the TBC in the six NM subregions is approximately equal to the response to the TBC in the NM region as a whole. This linearity is evident from visually comparing the top middle and top right panels of Fig. 2, and it is quantified in the bottom panel of Fig. 3. In the latter, the bar labeled SUM is nearly equal to one; the sum of the responses would in fact exactly equal one under complete linearity and in the absence of sampling error. The fact that the sum slightly exceeds 1 is presumably a reflection of sampling

FIG. 3. Results for the TBC experiments for the JJA 250-mb u wind. The bar graphs are the normalized spatial inner products [see (3)] from the various experiments. (top) The contributions from the TBC in the five zonal bands (NP, NM, TR, SM, and SP) to the correction in the NM region achieved from applying the TBC globally (TBC_GLOBAL – CNTL). The last bar shows the sum of those five contributions. (bottom) The contributions from the six NM subregions to the correction in NM region achieved from TBC applied to the NM region (TBC_NM – CNTL). The last bar shows the sum of those six contributions. See the text for details of those calculations.

noise projecting on the results (recall that these results employ a 5% t -test filter on the responses).

We next turn, in Figs. 4 and 5, to the results for 2-m air temperature (T2m). The top left panel in Fig. 4 shows that the uncorrected model (CNTL) has a substantial warm bias (the figure shows the negative of the bias) over much of the mid- and high-latitude land areas, with the largest biases occurring over western North America and central Asia. The top center panel of Fig. 4 indicates that much of that bias can be corrected by applying the TBC in the NM region. In fact, the TBC in the NM region alone accounts for about 90% of the impact achieved by applying the TBC globally (TBC_GLOBAL – CNTL), with the rest due to the TBC in the NP region (top panel of Fig. 5). The bottom six panels of Fig. 4 indicate that those impacts come from a number of the subregions: this is quantified in the middle bar

graph in Fig. 5, which shows that regions NM₁, NM₂, NM₄, and NM₅ all play some role. Again, the results are to a large extent linear (the SUM term is nearly equal to 1). Focusing on the North American region (the NM₅ region, land only), we find that 65% of the impact achieved by applying TBC in the full NM region is due to remote forcing (sum of the impacts in NM₂ and NM₄), with only 29% being locally forced (within the NM₅ region); that is, remote TBC corrections provide more than twice the impact on T2m over North America as local corrections. It is interesting that the remote impacts appear to be primarily located over the northern Great Plains and Canada, while the local impacts tend to be more focused on the middle and southern Great Plains [consistent with Klein et al. (2006) regarding the location of the local impacts]. The implication of the above partition of the TBC impacts for correcting the T2m bias over North America will be discussed in the next section.

The results for precipitation are shown in Figs. 6 and 7. The top-left panel of Fig. 6 shows that the model has substantial long-term biases during JJA throughout the Northern Hemisphere (NH) and tropics. We focus here on the NH, and in particular in the NM region, where long-standing model biases include excessive precipitation in the region encompassing Tibet and insufficient precipitation in the U.S. Great Plains. In addition, dry biases over the North Pacific and North Atlantic appear to be associated with insufficient storm track activity in those regions (Chang et al. 2019).

As shown in Chang et al. (2019), all these biases are substantially corrected when applying the TBC globally. The top-center panel of Fig. 6 shows that applying TBC in the NM region alone captures most of the impact found by Chang et al. (2019) in that region. This is quantified in the top panel of Fig. 7; essentially all of the impact achieved from the global TBC is obtained by limiting the TBC to the NM region alone (the inner product is nearly 1). Furthermore, the NM₂ region accounts for more than 60% of that impact (middle panel of Fig. 7). We see from the bottom three panels of Fig. 6 that the TBC in the NM₂ region is critical for correcting some of the excessive precipitation biases over and near the Tibet region and that the TBC in three regions (NM₂, NM₄, and NM₅) appears to contribute to reducing the dry bias over the U.S. Great Plains.³ As was the case for T2m, the remote impacts on precipitation appear to be primarily located over the northern Great

³ In view of the small contributions from the NM₁, NM₃, and NM₆ regions we plot only the results for the NM₂, NM₄, and NM₅ regions in the bottom three panels of Fig. 6.

FIG. 4. As in Fig. 2, but for T2m (°C). The observations used to compute the bias in the top-left panel are the GHCN_CAMS gridded 2-m temperature land data (Fan and van den Dool 2008).

Plains and Canada, whereas the local impacts tend to be more focused on the middle and southern Great Plains [again consistent with Klein et al. (2006) regarding the location of the local impacts].

Focusing again on the North America region (NM₅, in this case including both land and ocean points), we note that TBC in the NM₂ region accounts for 46% of the impact achieved by applying TBC in the full NM region (bottom panel of Fig. 7), while regions NM₄ (18%) and NM₅ (28%; the local contribution) also have substantial impacts. Once again, the partition of the impacts in NM₅ from TBC in the full NM region is approximately 2/3 from remote sources and 1/3 from the local source. The results are for the most part linear, although they are noisier than the other fields examined

so far (as especially evident from the value (1.2) of SUM in the top panel of Fig. 7).

Turning to the tropics, we find that applying the TBC in TBC_TR substantially reduces the long-term JJA tropical precipitation biases (not shown), but has little impact on the NM region (top panel of Fig. 7). This is perhaps not unexpected in view of the northward expansion of the tropical easterlies during JJA, which tends to inhibit the response to tropical forcing from propagating into the summer hemisphere (e.g., Webster 1982). This is illustrated in Fig. 8, which shows the responses of the 250-mb eddy height field during JJA to the application of the TBC in various regions. Comparing the model bias in the top panel with the response to the NM TBC in the middle left panel, we see (similar

FIG. 5. Results for the TBC experiments for JJA T2m. The bar graphs are the normalized spatial inner products [see (3)] over land from the various experiments. (top) The contributions from the various zonal bands to the correction in the NM region achieved from applying the TBC globally (TBC_GLOBAL – CNTL). (middle) The contributions from the various NM subregions to the correction in NM region achieved from applying the TBC in the NM region (TBC_NM – CNTL). (bottom) The contributions from the various NM subregions to the correction achieved over North America (TBC_NM5) from TBC_NM. See text for details of the calculations.

to what we saw for the 250-mb u wind in Fig. 2) that the NM_TBC corrects much of the long-term bias there. Furthermore, the TBC in the NM₂ region again contributes significantly to that impact (middle-right panel of Fig. 8 and the top panel in Fig. 9). On the other hand, correcting the tropics (TBC_TR) has little impact on the NH eddy height (bottom-left panel of Fig. 8), although it does produce a response in the SH that tends to correct the wave bias in the middle and high latitudes in the Eastern Hemisphere (cf. the top-left and bottom-left panels of Fig. 8). Furthermore, we find that a substantial fraction of that SH correction comes from the TBC over the Indian Ocean region (TR₂); see the bottom-right panel of Fig. 8 and the bottom panel of Fig. 9.

It is important to note that the responses to the TBC_TR in the southern midlatitudes (SM) region are not linear, with the sum of the responses to the various

subregions of TR adding to more than 2 (bottom panel of Fig. 9). It is unclear whether this reflects the substantially greater variability (and therefore greater sampling noise) in the SH winter eddy height variability (compared to that in the NH summer) or whether there is something inherently more nonlinear in the responses to remote forcing during winter. This will be addressed in a follow-on study.

b. An assessment of the corrections to model bias

A key take-away from the above analysis is that most of the impacts on the model's JJA climate in the Northern Hemisphere (the NM region) obtained from the global TBC described in Chang et al. (2019) can be achieved by applying the TBC to the NM region alone—there is in particular very little contribution to the NM region from the TBC in the tropics. Furthermore, within the NM region, the Tibet region (NM₂) plays a dominant role. Indeed, teleconnections within the model allow the corrections applied in NM₂ (and, to a lesser extent, those applied in NM₄, the region just upstream of North America) to affect both T2M and precipitation in North America (NM₅)—together by a factor of 2 more than the corrections applied locally there.

This obviously has important implications for how to correct the long-term T2m bias over North America. It would not be enough to focus solely on the physics and dynamics over North America, since it appears that a substantial part of the bias originates upstream—improving the model's performance over the Tibetan (NM₂) and eastern Pacific (NM₄) regions should lead, through the model's teleconnections, to improvements over North America. One needs to keep in mind, however, that the above analysis of the TBC_NM impacts does not necessarily amount to a quantitative statement about the correction of the model bias. For that, we need to determine how much of the bias over North America is actually corrected by the TBC in the NM region. In other words, we need to determine the projection of the TBC impacts on the model bias itself.

In this section, we do just that, using the measure (3), where $\mathbf{Y}$ is now the negative of the model bias. We again focus on the u250mb winds in the midlatitudes and on the T2m and precipitation over the North American region (NM₅). Recall that the inner product (3) is equal to the scalar projection of $\mathbf{X}$ onto $\mathbf{Y}$ divided by the magnitude of $\mathbf{Y}$, so that we can in the following discuss the results in terms of the fraction (or percent) of the bias $\mathbf{Y}$ corrected by the impact $\mathbf{X}$.

The results are summarized in Table 2 where, in addition to the inner products, we also look separately at the $\cos\theta$ (pattern similarity) and amplitude contributions. Furthermore, in order to provide some measure

FIG. 6. Results of the TBC experiments for JJA precipitation (mm day^{-1}). (top left) Minus the model bias, where the observations used to estimate the bias are the GPCP v2.3 data (Adler et al. 2003). (top center) The response to the TBC in the NM region (TBC_NM – CNTL). (top right) The sum of the individual results of the 6 NM regions. The bottom three panels show the responses to TBC in regions NM₂, NM₄, and NM₅. Results are the averages over the years 1980–2016.

of the uncertainty in our results, we show in Table 2 two sets of results for each TBC experiment: 1) the numbers to the left are based on the biases computed with respect to MERRA-2 for u250mb, with respect to the gridded GHCN_CAMS data for T2m over land (Fan and van den Dool 2008), and with respect to GPCP V2.3 for precipitation (Adler et al. 2003), and 2) the values to the right are based on the biases computed with respect to an independent reanalysis, ERA5 (C3S 2017). As we shall see, perhaps not surprisingly, the results for the u250mb wind are quite stable (the results based on MERRA-2 and ERA5 differ very little), whereas the results for T2m and precipitation do depend to some extent on the verification dataset (although not so much as to modify our basic conclusions). It is of course very likely that both ERA5 and the observational products are themselves biased in their estimates of T2m and precipitation to some degree (although presumably the biases are different), with the observational products most likely having problems over mountainous regions where station observations are relatively scarce (e.g., Fan and van den Dool 2008; Adler et al. 2003). Early assessments of

ERA5 indicate that it provides improved (compared with ERA-Interim) forcing of land surface models (Albergel et al. 2018), and relatively small biases (compared with several other precipitation products) in summer precipitation over the northern Great Plains (Xu et al. 2019). In any event, providing two sets of bias correction estimates in Table 2 should illustrate the nature of the uncertainty we face in such an analysis.

We begin with a look at how TBC in the NM region and the various NM subregions contribute to reducing the u250mb bias in the entire NM region (the first three rows of the results in Table 2). We see that nearly 90% of the AGCM long-term JJA bias in u250mb in the NM region can be corrected by applying the TBC in just that (NM) region, with both the magnitude and pattern ($\cos\theta = 0.91$) matching very well. A substantial fraction of the bias correction is achieved by applying TBC in the NM₂ (Tibet) region, which alone corrects about 36% of the bias in the NM region (accounting for about half the magnitude and a pattern similarity of 0.65). Other important contributions come from regions NM₄ and NM₅. This is consistent with our

FIG. 7. As in Fig. 5, but for precipitation.

previous assessment of impacts with, for example, the NM₂ region accounting for more than 40% of the impacts on u250mb as shown in the bottom panel of Fig. 3. Overall, the results for u250mb are reasonably linear, although it appears that there is some tendency to overestimate the amplitude (SUM is about 120%).

Focusing on T2m over North America (the NM₅ region; the fourth, fifth, and sixth rows of the results in Table 2), we find that the TBC in the NM region corrects 58%–66% of the T2m bias, with the Tibet region (NM₂) on its own correcting about 17%–19% of the bias. Another important teleconnection is associated with the TBC in the region just upstream of North America (NM₄), which corrects 20%–22% of the T2m bias in the NM₅ region. The TBC applied locally (NM₅) corrects 17%–20% of the bias there. It is noteworthy that the pattern of the bias correction produced by the TBC in the full NM region (first two columns of the sixth row of Table 2) matches the actual bias pattern over NM₅ quite well ($\cos\theta$ greater than 0.9), while there is an underestimate of the amplitude (63%–70%). TBCs applied in the NM₂ and NM₄ regions also produce bias corrections that reflect the actual bias patterns quite well ($\cos\theta$ about 0.8), in fact better than the local correction does. Overall, the T2m bias over North America can be

partitioned as 17%–20% locally forced, 41%–46% remotely forced, and 34%–42% unaccounted for (i.e., not corrected by TBC in NM). In a relative sense (considering only the part of the bias actually corrected by TBC in NM), the partition is about 70% remote and 30% local, similar to the partition of the *impacts* reported on in the previous section (2/3 remote vs 1/3 local).

Focusing on the precipitation biases over North America (the seventh, eighth and ninth rows of the results in Table 2), we find that the TBC in the NM region corrects between 50% and 61% of the bias in NM₅. In assessing the relative impacts of the various NM sub-regions, we need to take into account the fact that there is, in this case, considerable nonlinearity in the results (the magnitude for the sum exceeds 140%, last two columns of the eighth row of the results in Table 2). As a first-order correction for this, when considering the relative contributions from the various subregions, we simply scale the inner products by the ratio of the total NM correction to that of the sum, which is approximately 0.78 for both sets of target observations. With that adjustment, we see that the TBC in the Tibet region (NM₂) corrects between 13% and 20% of the precipitation bias over North America, TBC in the region just upstream of North America (NM₄) corrects 6%–9% of this bias, and the local correction amounts to 21%–22%. Overall, for the precipitation bias over North America, remote and local TBC corrections account for 27%–39% and 21%–22%, respectively, of the bias reduction, with between 40% and 50% not corrected by TBC. Again, in a relative sense (considering only the part of the bias actually corrected), the partition ranges between 55% and 64% remote and between 36% and 45% local—a range of values that is not inconsistent with the partition of the *impacts* of TBC (2/3 remote vs 1/3 local; see section 3a).

c. RPL experiments

We next turn, in Figs. 10 and 11, to some of the RPL results, focusing on the 250-mb *u* wind in the NH middle latitudes. In view of the strong response to the TBC in the Tibetan region (NM₂; see Fig. 2), we are interested in determining whether the atmosphere is inherently more sensitive to perturbations in that particular region. It is worth recalling here the fundamental difference between the RPL and TBC experiments: with TBC, the climatological mean tendency error is removed from model state variables at every time step as an additional model “forcing” whereas with RPL the time-specific errors are removed, thereby forcing the evolution of the model’s weather in the designated region to agree with that in the analysis. That is, TBC corrects for a long-term mean error, whereas RPL effectively corrects the specific errors produced at each time step.

JJA 250mb Eddy Height (m)

FIG. 8. Results of the TBC experiments for 250-mb eddy height (m). (top left) Minus the model bias [$-(\text{CNTL} - \text{MERRA-2})$]. (middle left) The response to the TBC in the NM region ($\text{TBC_NM} - \text{CNTL}$). (middle right) The response to the TBC in the NM₂ region ($\text{TBC_NM}_2 - \text{CNTL}$). (bottom left) The responses to tropical zonal band ($\text{TBC_TR} - \text{CNTL}$), and (bottom right) for the tropical region ($\text{TBC_TR}_2 - \text{CNTL}$). Results are the averages over the years 1980–2016.

RPL thus imposes a much stronger correction to the states in the designated region.

The top two panels of Fig. 10 compare the TBC and RPL results for the NM region. (The TBC_NM result

was already shown in Fig. 2 but is repeated here for convenience.) The RPL results of course essentially reproduce MERRA-2 in the NM region and therefore provide an upper limit to what can be achieved with

FIG. 9. Results for the TBC experiments for JJA 250-mb eddy heights. The bar graphs are the normalized spatial inner products [see (3)] from the various experiments. (top) The contributions from the TBC in the six NM subregions to the correction in the NM region achieved by applying the TBC to the NM region ($\text{TBC_NM} - \text{CNTL}$). The last bar shows the sum of those six contributions. (bottom) The contributions from the six tropical (TR) subregions to the correction over the SM region compared to that achieved by applying the TBC to the full TR region ($\text{TBC_TR} - \text{CNTL}$). The last bar shows the sum of those six contributions. See the text for details of those calculations.

the TBC there. As shown in Table 2, the TBC_NM corrects nearly 90% of the model bias in that region. This emphasizes the fact that the NH jet biases are almost entirely the result of systematic errors in the AGCM forcing (the TBCs); there is apparently little room left for any contribution from nonsystematic (state-dependent) tendency errors. We will look more directly at the statistics of the increments (mean vs variance) later in this section.

The bottom six panels in Fig. 10 show the results of applying the RPL corrections to each of the six NM subregions. These results are analogous to the TBC results shown in the bottom six panels of Fig. 2. It is interesting that there is considerable overall similarity in the pattern of the response regardless of the subregion being replayed, with the similarity measure ($\cos\theta$ values in the top panel of Fig. 11) ranging

from 0.6 for NM_3 to 0.8 for NM_2 . Nevertheless, the RPL for the NM_2 region does appear to produce the strongest and most hemispheric-wide response. In fact, the top panel of Fig. 11 shows that the NM_2 region indeed produces the response with both the largest amplitude and with the greatest similarity in pattern to the model bias ($\text{RPL_NM} - \text{CNTL}$), although the contrast between its contribution and that of the others is not nearly as dramatic as we have seen for the TBC results in Fig. 2, where the TBC_NM_2 dominates the other responses (more on that below). (Note that the calculation of the $\cos\theta$ values in Fig. 11 does not include the patterns in the local region considered, since these are forced to track the analysis by design).

Because all of the RPL_NM_i regions produce relatively large and similar response patterns, the amplitude of the SUM (last set of three bars in the top panel of Fig. 11) is much larger than (more than 2.5 times) that of the RPL response to the full NM region (RPL_NM). Clearly the RPL results are not linear, and indeed there is no a priori reason to expect linearity, since for any region being forced to track the analysis one would expect that some of the neighboring regions would also track the analysis to a reasonable degree, simply by continuity, leading to potential double counting. In any event, the hemispheric nature of the responses suggests a controlling influence of the mean jets on the responses, although the mechanisms by which that occurs is unclear. One possibility is some type of resonance wave response, although the responses here have considerable north/south structure and wavenumbers that are much smaller than would be expected for typical NH summer jet speeds (wavenumbers 6–8; e.g., Kornhuber et al. 2017).

Returning to the comparison between the RPL and TBC results, we quantify (bottom panel of Fig. 11) the fraction of the RPL results that is accounted for by the TBC results in each subregion. Here [referring to Eq. (3)], $\mathbf{X}$ is ($\text{TBC_NM}_i - \text{CNTL}$) and $\mathbf{Y}$ is ($\text{RPL_NM}_i - \text{CNTL}$), where the summation is in each case carried out over the NM region, but excluding the considered subregion i . This again highlights the importance of the TBC in the Tibet region (NM_2) which produces more than 50% of the response achievable from RPL in that region. In fact, the relatively large response to TBC in the NM_2 region reflects both a strong similarity to the pattern of the RPL response ($\cos\theta$; red bars) and a relatively large amplitude compared to the responses in the other regions (blue bars). Note that TBCs in regions NM_4 and NM_5 produce response patterns similar to those in the corresponding RPL simulations, but with considerably weaker amplitudes.

TABLE 2. A summary of key results concerning bias correction from the regional TBC runs using the inner product decomposition (3). Here we show the inner product (I , as a percent) in bold font, the magnitude ratio (also in percent) in plain font, and $\cos(\theta)$ in italic font. For the left number in each pair of values, the biases are computed with respect to MERRA-2 for u250mb, gridded GHCN_CAMS for T2m over land (Fan and van den Dool 2008), and GPCP V2.3 for precipitation (Adler et al. 2003). The right numbers are the same, except all biases are computed with respect to ERA5 (C3S 2017). See text for details.

Target region (variable)	Bias corrected from TBC in NM region	From TBC in NM ₁	From TBC in NM ₂	From TBC in NM ₃	From TBC in NM ₄	From TBC in NM ₅	From TBC in NM ₆	Sum
NM (u250mb)	87.4, 88.4 96.0, 97.7 <i>0.91, 0.91</i>	9.6, 9.7 27.0, 27.5 <i>0.36, 0.35</i>	36.2, 36.8 55.4, 56.3 <i>0.65, 0.65</i>	9.5, 9.6 20.4, 20.7 <i>0.47, 0.46</i>	18.7, 19.0 29.3, 29.9 <i>0.64, 0.64</i>	13.3, 13.5 24.2, 24.6 <i>0.55, 0.55</i>	4.0, 4.0 18.6, 19.0 <i>0.21, 0.21</i>	91.5, 92.6 118.7, 120.8 <i>0.77, 0.77</i>
NM ₅ (T2m)	58.4, 65.9 63.1, 70.3 <i>0.92, 0.94</i>	1.6, 1.8 6.4, 7.2 <i>0.24, 0.25</i>	16.8, 19.0 20.9, 23.4 <i>0.80, 0.81</i>	3.0, 3.3 6.5, 7.3 <i>0.46, 0.45</i>	19.8, 22.2 23.9, 26.6 <i>0.83, 0.83</i>	17.2, 19.9 25.4, 28.7 <i>0.67, 0.69</i>	0.2, 0.0 5.0, 5.7 <i>0.03, 0.01</i>	58.5, 66.2 66.8, 74.7 <i>0.88, 0.89</i>
NM ₅ (Precip)	49.5, 60.5 91.3, 94.4 <i>0.54, 0.64</i>	3.2, 4.1 24.3, 25.1 <i>0.13, 0.16</i>	17.2, 25.5 59.6, 61.6 <i>0.29, 0.41</i>	4.4, 6.9 26.4, 27.3 <i>0.17, 0.25</i>	7.8, 11.5 37.0, 38.2 <i>0.21, 0.30</i>	27.8, 27.4 60.0, 62.0 <i>0.46, 0.44</i>	2.5, 2.3 26.3, 27.2 <i>0.10, 0.08</i>	62.9, 77.7 142.6, 147.4 <i>0.44, 0.53</i>

We now take a closer look at the Tibet region, using additional RPL experiments to parse out the causes of the hemispheric impacts. More specifically, we use the RPL approach to isolate those state variables that, if constrained to remain close to the reanalysis over the Tibet region at each time step, produce most of the remote impacts. Figure 12 shows the RPL results for the Tibet region where we now replay separately to the temperature (i.e., we force only the atmospheric temperature states to agree with the analysis over the Tibet region), the horizontal winds, and the moisture. We note that these model experiments (referred to as RPL_Tibet_T, RPL_Tibet_uv, and RPL_Tibet_q, respectively) differ somewhat from those described previously in that the Tibet region (70°–110°E, 17°–50°N) differs slightly from the NM₂ region, being more centered on the Tibetan highlands (see Fig. 1). Also, in order to reduce the computational burden, the runs were started each year on 30 April and only run through 2 September of that same year, with the JJA values averaged over the years 1980–2010. For convenience, the top panel of Fig. 12 shows the negative of the model bias to allow a more direct comparison with the RPL impacts. The middle left panel shows the results of replaying to all the variables in the Tibet region (RPL_Tibet_All). (This is analogous to RPL_NM₂ in Fig. 10; the strong similarity in these plots indicates a very low sensitivity of the results to the exact location and size of the Tibet region.) The remaining three panels in Fig. 12 show the separate impacts of RPL_Tibet_T (middle right), RPL_Tibet_uv (bottom left), and RPL_Tibet_q (bottom right). The remarkable similarity between the RPL_Tibet_T and RPL_Tibet_All responses indicates that the long-term bias is primarily driven by the temperature increments in that region.

Somewhat similar, although with weaker amplitude, results are obtained from only replaying to the moisture (RPL_Tibet_q). Replaying to the winds (RPL_Tibet_uv) produces a more locally confined response.

Further insight into what is driving the responses can be obtained from the impacts on precipitation (Fig. 13). Here also, the similarity between the RPL_Tibet_T and RPL_Tibet_All responses is remarkable, again pointing to the importance of the temperature increments over Tibet. Also noteworthy is that for both the RPL_Tibet_T and RPL_Tibet_q (less so for RPL_Tibet_uv) there is a substantial reduction to the excessive precipitation bias over the Himalayas, suggesting that the associated cooling anomalies may play some role in producing the hemispheric-wide responses in those experiments.

The above results indicate that we need to take a closer look at the spatial distribution and impacts of the temperature increments. The top-left panel of Fig. 14 shows the time-averaged temperature increments at 500 mb, highlighting that the largest (negative) increments do indeed occur over the Tibet region, with smaller but nevertheless pervasively negative values occurring over much of the Maritime Continent and surrounding regions, as well as over the western United States and Mexico. Positive increments occur over the storm track regions of the North Pacific and North Atlantic, presumably to correct for the too weak storm-related diabatic heating in those regions (Chang et al. 2019). The top-right panel of Fig. 14 again shows the mean increments, but now in units of standard deviation, making it clear that the mean values are a significant fraction of the daily variability of the increments in many regions, especially over the Tibet region, where they have a magnitude that is more than twice as large

FIG. 10. Results for the RPL experiments for JJA 250-mb u wind (m s^{-1}) with a focus on the NH (the NM region). (top center) The response to the RPL in the NM region (RPL_NM – CNTL). (top left) Shown for comparison are the results of the TBC in that region (TBC_NM – CNTL). The bottom six panels show the RPL results of the six NM subregions. Results are the averages over the years 1980–2016.

as the standard deviation. Given this, it is not surprising that the response to the TBC accounts for more than 50% of what is possible from RPL in that region (see bottom panel of Fig. 11).

To continue quantifying the impact of thermal forcing over the Tibet region, we next examine the response to an idealized cooling anomaly in that region utilizing a stationary wave model (SWM; Ting and Yu 1998).⁴ We

assume a vertical structure of the cooling anomaly that peaks at $\sigma = 0.5$ ($-6.0^\circ\text{C day}^{-1}$) and has a vertical integral of $-3.5^\circ\text{C day}^{-1}$. Our intent here is not to reproduce exactly the cooling associated with the temperature increments in the full AGCM, but rather to get a general sense of the global impacts of the thermal forcing in that region. In fact, the vertical structure of the temperature increments over Tibet (not shown) is such that the largest cooling (about $-3.5^\circ\text{C day}^{-1}$) occurs just above the surface, which, because of the high terrain, occurs between about 450 and 550 mb. In any event, the effective cooling anomaly that the AGCM feels in that region is likely some combination of the direct impact of the temperature increments and the cooling anomaly that results from the reduction in the excessive precipitation (see Fig. 13).

⁴The stationary wave model is the dry dynamical core of an AGCM. It is time-dependent and resolves stationary nonlinearity. It is based on the three-dimensional primitive equations in σ coordinates and has rhomboidal wavenumber-30 truncation in the horizontal and 14 unevenly spaced σ levels in the vertical (R30L14).

FIG. 11. Results for the RPL and TBC experiments for JJA 250-mb u wind. (top) The contributions from RPL in the six NM subregions to the bias correction achieved from RPL_NM. Here the last bar shows the contribution to the RPL_NM from the sum of the RPL results in those six subregions. The inner products for the subregions are computed only over those parts of the NM region falling outside each of the subregions. (bottom) Comparison of the contributions from the TBC in the six NM subregions and the NM region to the bias correction achieved from RPL in each of those regions. In the above, the black bars are the normalized spatial inner products, while the red (blue) bars are the contributions from $\cos\theta(|\mathbf{X}|/|\mathbf{Y}|)$. See the text and (3) for details of those calculations.

The bottom right panel of Fig. 14 shows that, despite the very idealized structure of the thermal forcing, the SWM's eddy streamfunction response bears a considerable resemblance to the AGCM's response to the TBC in region NM₂ (bottom left panel of Fig. 14). We note that the lower left panel shows essentially the same result as shown in the top right panel of Fig. 8 (JJA 250-mb eddy height response of the AGCM to the TBC in the NM₂ region), although converted here to streamfunction and interpolated to $\sigma = 0.257$ to be consistent with the SWM results. This comparison provides additional evidence that the excessive heating in that region is the main source of the AGCM's circulation biases (and related biases) that span the NH.

4. Discussion and conclusions

The results of this study show that regional TBC, as an extension of the global TBC examined in Chang et al. (2019), provides a powerful tool for identifying the geographical sources of a global climate model's biases. The approach requires a long term (multidecade) record of short-term analysis increments. These can

immediately be obtained as a by-product of using the model as part of a reanalysis effort; however, the use of the much simpler "replay" (RPL) technology developed by the GMAO, a strategy that constrains the model to track an existing reanalysis (either regionally or globally), can also provide the needed increments. This allows modeling groups to avoid the tremendous effort involved in running a full reanalysis and thus makes the strategy generally accessible. The long term means of the RPL increments are interpreted as TBCs that are introduced into the model equations (with opposite sign) as a non-state-dependent forcing term. The RPL results themselves provide an upper bound to the long-term mean climate corrections made possible by the TBCs.

Our dissection of the results of Chang et al. (2019) shows that almost all of the correction made to the GEOS AGCM's JJA climate in the Northern Hemisphere midlatitudes (the NM region) through TBC can be achieved by applying the TBC to the NM region alone. In the case of the u250mb zonal wind, we find that the TBC in the NM₂ (Tibet) region accounts for more than 40% of this correction. Furthermore, we find that hemispheric-wide teleconnections allow TBC in NM₂ (the Tibet region) and (to a lesser extent) in NM₄ (the region just upstream of North America) to affect significantly the T2M and precipitation over North America (the NM₅ region). In fact, two-thirds of the TBC impact on North American T2M and precipitation comes from remote sources.

While this result has obvious important implications for modelers trying to determine where geographically to focus their efforts, we must emphasize that it reflects only the extent to which model errors in one region affect (via teleconnections) the model climate in another region; it does not necessarily tell us how much of the original model bias is actually corrected by TBC in the different regions. For that, we must quantify the degree to which the regional TBC impacts actually project onto the model bias. We have done that here (section 3b) by comparing the regional TBC results against both observational data and an independently produced reanalysis dataset (ERA5). Based on these comparisons, we find that, within the NM region as a whole, nearly 90% of the AGCM's JJA bias in the u250mb can be corrected by applying the TBC in just that (NM) region, and that a substantial fraction of that correction (36%) is achieved by applying TBC in the NM₂ (Tibet) region alone. This analysis thus highlights the large projection of the TBC impacts, particularly the impacts of the TBC imposed in the Tibetan region, on the overall u250mb bias in northern midlatitudes. In the case of the precipitation and T2m biases over North America, the

FIG. 12. Results (JJA 250-mb u wind; m s^{-1}) of the RPL experiments for the Tibet region (indicated by the box). (top) Minus the model bias $[-(\text{CNTL} - \text{MERRA2})]$. (middle left) The response to replaying all variables in the Tibet region (RPL_Tibet_All - CNTL). (middle right) The response to replaying only to temperature (T) in the Tibet region (RPL_Tibet_T - CNTL). (bottom left) The response to replaying only to the winds (u, v) in the Tibet region (RPL_Tibet_uv - CNTL). (bottom right) The response to replaying only the specific humidity (q) in the Tibet region (RPL_Tibet_q - CNTL). Results are the JJA averages over the years 1980–2010.

TBC impacts project less strongly on the biases, accounting for between 58% and 66% of the bias for the T2m, and between 50% and 61% of the bias for precipitation. Nevertheless, in a relative sense, considering only the fraction of the bias over North America that is actually corrected by applying TBC in northern midlatitudes, we find that the sources of the biases are again roughly 2/3 *remote* and 1/3 *local*. It remains to be seen whether future improvements to reanalyses and the associated analysis increments will increase the TBC's

ability to correct the long-term summer biases over North America.

It is important to consider the above results concerning the precipitation and T2m bias corrections over North America in the context of the substantial body of work already done to diagnose the long-standing summertime dry and warm biases over the U.S. Great Plains that occur in many models (e.g., Lin et al. 2017; Morcrette et al. 2018; Van Weverberg et al. 2018; Ma et al. 2018; Zhang et al. 2018). As summarized in Steiner

FIG. 13. As in Fig. 12, but for precipitation (mm day^{-1}). In the top panel, the observations used to estimate the bias are the GPCP version v2.3 data (Adler et al. 2003).

(2018), the most recent consensus regarding the warm bias over the central United States is that it is tied to deficiencies in the simulation of deep convective clouds and the evaporative fraction at the surface. Furthermore, Lin et al. (2017) provide evidence from a diagnosis of a suite of models from phase 5 of the Coupled Model Intercomparison Project (CMIP5) that the dry and warm biases are intertwined, with the dry bias leading the warm bias in the Great Plains region, and with the precipitation deficit tied to the failure of models to capture strong rainfall events. Still other studies have focused on the diurnal cycle of precipitation, linking the inability of models to capture the

nocturnal maximum in precipitation over the Great Plains to their inability to simulate the eastward propagation of mesoscale convective systems from the Rocky Mountains, indicating the need for increased model resolution (e.g., Yamada et al. 2012). Furthermore, Klein et al. (2006) provide clear evidence in support of local errors playing an important role; they show that a substantial fraction of the overestimate of surface air temperature and underestimate of precipitation simulated by the GFDL climate model AM2 over the Southern Great Plains is already present in the first few days of forecasts with the model—a forecast period during which the large-scale circulation, and thus

FIG. 14. (top left) The time-mean (JJA 1980–2016) temperature analysis increments at 500 mb from MERRA-2 ($^{\circ}\text{C day}^{-1}$). (top right) As in the top left, but in units of standard deviation computed from the daily mean increments. (bottom left) The AGCM JJA eddy streamfunction response (interpolated to $\sigma = 0.257$) to the TBC in the NM₂ region (averaged over the years 1980–2016). (bottom right) The SWM eddy streamfunction response at $\sigma = 0.257$ to an idealized cooling anomaly (width is 50° lon $\times$ 16° lat) centered on 28°N , 95°E . The heavy purple contours show the vertically integrated cooling with a maximum amplitude of $3.5^{\circ}\text{C day}^{-1}$ (contour interval: $1^{\circ}\text{C day}^{-1}$). In the vertical, the cooling has a simple sinusoidal structure with the maximum amplitude at $\sigma = 0.5$, equal to $\sim 6^{\circ}\text{C day}^{-1}$. The basic state of the SWM is the JJA long-term mean (1980–2010) computed from MERRA-2. Units are $1 \times 10^6 \text{ m}^2 \text{ s}^{-1}$.

the contributions of remote sources, would presumably be well represented.

Our results do not undermine the above basic conclusions. In fact, our results provide little information on the actual model deficiencies that lead to the biases. Nevertheless, we believe that our approach provides important complementary information to modelers—information that helps distinguish between proximate and ultimate causes of the model biases. For example, to the extent that a part of the warm bias in North America is associated with erroneous circulation anomalies forced elsewhere (e.g., the Tibet region), that part of the warm bias would exist even if the local errors identified in the above-mentioned studies were corrected. Another issue involves the spatial scale considered. The fraction of the bias driven by remote errors for the GEOS AGCM over North America as a whole (the NM₅ region) is in the range of 41%–46% for T2m, and of 27%–39% for precipitation. Those numbers would almost certainly change (likely get smaller) if we were to focus even more closely on the U.S. Great Plains region, although

statistical sampling issues could be an impediment to going to such smaller regions (discussed further below). Given issues of scale and the fact that we do see some contribution of local error to the North American biases, our results are not inconsistent with studies that find a significant component of the bias over the U.S. Great Plains to be locally forced.

We believe our methodology is indeed flexible enough to allow one to zero in on the actual physical parameterization deficiencies that are at the root cause of the model biases. Limiting the TBCs to specific variables (e.g., the temperature, moisture, or wind increments) is one possible approach to further isolate the errors. Also, since the errors are identified early on before they achieve nonlinear growth, a more detailed statistical assessment might help associate the increments with specific deficiencies in model physical parameterizations [e.g., as in [Klinker and Sardeshmukh \(1992\)](#) and [Schubert and Chang \(1996\)](#)].

There are a number of issues that require further analysis. The choice of the regions examined is flexible, but of course the desire for making them as small

as possible must be weighed against the constraints imposed by the available computing resources. Our choice of rather broad zonal bands together with a more detailed subdivision of the northern midlatitude and tropical bands reflects our interest in examining the causes of the biases in the northern midlatitudes. We believe the size of the subregions examined (see Fig. 1) was a reasonable compromise: making them much smaller would almost certainly increase the statistical noise to an unacceptable degree (even with more than three decades of model runs), while making them much larger would presumably make it too blunt of a tool to be useful in isolating the geographical sources of the errors. Additional work is needed to better quantify the statistical sampling errors as a function of season, the sizes/locations of the regions, and the length of the model runs. The use of ensembles with perturbed initial conditions (rather than a single long model simulation) should also be considered when only a particular season is of interest.

There is the question of the applicability of the TBC approach to models that are not employed in the production of the reanalyses used to estimate the TBCs. As already mentioned, from a practical standpoint the implementation of TBC is certainly feasible by employing a replay approach. Scientifically, the general applicability of the TBC approach can only be fully assessed by repeating our analysis with a number of different models (and perhaps several different reanalyses). We are, however, confident that such efforts would be successful. Chang et al. (2019), for example, successfully applied the TBC approach (using MERRA-2 increments to compute the forcing terms) to a substantially updated version of the MERRA-2 AGCM coupled to an ocean model, a system with dramatically different atmospheric biases.

It would also be helpful to have a better understanding of the linearity we see in our results in the summer hemisphere. While there does appear to be some double counting, especially for precipitation impacts, some of that is likely caused by statistical sampling errors, although that needs further study. In any event, it is unclear whether the approximate linearity during summer reflects the dominance of the impact of the Tibetan region (with the other regions providing only smaller, approximately linear corrections to that) or whether it is a more general result representative of the responses to forcing within the relatively weak (compared with winter) summer basic state. In any event, the apparent lack of linearity for the TBC results for the winter hemisphere is of some concern. A general absence of linearity would limit the ability to make definitive statements through the approach as to which

regions are most important to the generation of model biases, although it might nevertheless still provide some broad insights as to what is important. The nonlinearity itself could even provide some clues as to how various errors interact to produce the model biases. This is clearly a subject that will require considerably more analysis.

Acknowledgments. This work was supported by the NASA MAP (NNG17HP01C and WBS 802678.02.17.01.33), and NOAA MAPP (NA14OAR4310221) programs. GHCN Gridded V2 and GPCP precipitation data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <https://www.esrl.noaa.gov/psd/>. We also thank three anonymous reviewers for their constructive comments that have helped to substantially improve the clarity of the manuscript.

REFERENCES

- Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- Albergel, C., E. Dutra, S. Munier, J.-C. Calvet, J. Munoz-Sabater, P. de Rosnay, and G. Balsamo, 2018: ERA-5 and ERA-Interim driven ISBA land surface model simulations: Which one performs better? *Hydrol. Earth Syst. Sci.*, **22**, 3515–3532, <https://doi.org/10.5194/hess-22-3515-2018>.
- Bacmeister, J. T., and G. L. Stephens, 2011: Spatial statistics of likely convective clouds in CloudSat data. *J. Geophys. Res.*, **116**, D04104, <https://doi.org/10.1029/2010JD014444>.
- Bhargava, K., E. Kalnay, J. A. Carton, and F. Yang, 2018: Estimation of systematic errors in the GFS using analysis increments. *J. Geophys. Res. Atmos.*, **123**, 1626–1637, <https://doi.org/10.1002/2017JD027423>.
- Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256–1271, [https://doi.org/10.1175/1520-0493\(1996\)124<1256:DAUIAU>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2).
- C3S, 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), accessed June 2019, <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- Chang, Y., S. D. Schubert, R. D. Koster, A. M. Molod, and H. Wang, 2019: Tendency bias correction in coupled and uncoupled global climate models with a focus on impacts over North America. *J. Climate*, **32**, 639–661, <https://doi.org/10.1175/JCLI-D-18-0598.1>.
- Cullather, R. I., S. M. J. Nowicki, B. Zhao, and M. J. Suárez, 2014: Evaluation of the surface representation of the Greenland Ice Sheet in a general circulation model. *J. Climate*, **27**, 4835–4856, <https://doi.org/10.1175/JCLI-D-13-00635.1>.
- Fan, Y., and H. van den Dool, 2008: A global monthly land surface air temperature analysis for 1948–present. *J. Geophys. Res.*, **113**, D01103, <https://doi.org/10.1029/2007JD008470>.
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2

- (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Klein, S. A., X. Jiang, J. Boyle, S. Malyshev, and S. Xie, 2006: Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophys. Res. Lett.*, **33**, L18805, <https://doi.org/10.1029/2006GL027567>.
- Klinker, E., and P. D. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608–627, [https://doi.org/10.1175/1520-0469\(1992\)049<0608:TDOMDI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<0608:TDOMDI>2.0.CO;2).
- Kobayashi, S., and Coauthors, 2015: The JRA-55 Reanalysis: General specification and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Kornhuber, K., V. Petoukhov, D. Karoly, S. Petri, S. Rahmstorf, and D. Coumou, 2017: Summertime planetary wave resonance in the Northern and Southern Hemispheres. *J. Climate*, **30**, 6133–6150, <https://doi.org/10.1175/JCLI-D-16-0703.1>.
- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based approach to modeling land surface processes in a general model: 1. Model structure. *J. Geophys. Res.*, **105**, 24 809–24 822, <https://doi.org/10.1029/2000JD900327>.
- Lin, Y., W. Dong, M. Zhang, Y. Xie, W. Xue, J. Huang, and Y. Luo, 2017: Causes of model dry and warm bias over central U.S. and impact on climate projections. *Nat. Commun.*, **8**, 881, <https://doi.org/10.1038/S41467-017-01040-2>.
- Ma, H.-Y., and Coauthors, 2015: An improved hindcast approach for evaluation and diagnosis of physical processes in global climate models. *J. Adv. Model. Earth Syst.*, **7**, 1810–1827, <https://doi.org/10.1002/2015MS000490>.
- , and Coauthors, 2018: CAUSES: On the role of surface energy budget errors to the warm surface air temperature error over the Central United States. *J. Geophys. Res.*, **123**, 2888–2909, <https://doi.org/10.1002/2017JD027194>.
- Molod, A. M., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geosci. Model Dev.*, **8**, 1339–1356, <https://doi.org/10.5194/gmd-8-1339-2015>.
- Moorthi, S., and M. J. Suarez, 1992: Relaxed Arakawa–Schubert: A parameterization of moist convection for general circulation models. *Mon. Wea. Rev.*, **120**, 978–1002, [https://doi.org/10.1175/1520-0493\(1992\)120<0978:RASAP0>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0978:RASAP0>2.0.CO;2).
- Morcrette, C. J., and Coauthors, 2018: Introduction to CAUSES: Description of weather and climate models and their near-surface temperature errors in 5 day hindcasts near the Southern Great Plains. *J. Geophys. Res.*, **123**, 2655–2683, <https://doi.org/10.1002/2017JD027199>.
- Orbe, C., L. D. Oman, S. E. Strahan, D. W. Waugh, S. Pawson, L. L. Takacs, and A. M. Molod, 2017: Large-scale atmospheric transport in GEOS replay simulations. *J. Adv. Model. Earth Syst.*, **9**, 2545–2560, <https://doi.org/10.1002/2017MS001053>.
- Phillips, T. J., and Coauthors, 2004: Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Amer. Meteor. Soc.*, **85**, 1903–1916, <https://doi.org/10.1175/BAMS-85-12-1903>.
- Putman, W., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>.
- Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Quart. J. Roy. Meteor. Soc.*, **133**, 129–146, <https://doi.org/10.1002/qj.23>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, <https://doi.org/10.1175/2010BAMS3001.1>.
- Schubert, S., and Y. Chang, 1996: An objective method for inferring sources of model error. *Mon. Wea. Rev.*, **124**, 325–340, [https://doi.org/10.1175/1520-0493\(1996\)124<0325:AOMFIS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<0325:AOMFIS>2.0.CO;2).
- , H. Wang, R. Koster, M. Suarez, and P. Groisman, 2014: Northern Eurasian heat waves and droughts. *J. Climate*, **27**, 3169–3207, <https://doi.org/10.1175/JCLI-D-13-00360.1>.
- Steiner, A., 2018: Diagnosing the warm bias in the central United States. *Eos Trans. Amer. Geophys. Union*, **99**, <https://doi.org/10.1029/2018EO095669>.
- Takacs, L. L., M. J. Suárez, and R. Todling, 2018: The stability of incremental analysis update. *Mon. Wea. Rev.*, **146**, 3259–3275, <https://doi.org/10.1175/MWR-D-18-0117.1>.
- Ting, M., and L. Yu, 1998: Steady response to tropical heating in wavy linear and nonlinear baroclinic models. *J. Atmos. Sci.*, **55**, 3565–3582, [https://doi.org/10.1175/1520-0469\(1998\)055<3565:SRTHI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<3565:SRTHI>2.0.CO;2).
- Van Weverberg, K., and Coauthors, 2018: CAUSES: Attribution of surface radiation biases in NWP and climate models near the U.S. Southern Great Plains. *J. Geophys. Res. Atmos.*, **123**, 3612–3644, <https://doi.org/10.1002/2017JD027188>.
- Webster, P. J., 1982: Seasonality of the local and remote atmospheric response to sea surface temperature anomalies. *J. Atmos. Sci.*, **39**, 41–52, [https://doi.org/10.1175/1520-0469\(1982\)039<0041:SITLAR>2.0.CO;2](https://doi.org/10.1175/1520-0469(1982)039<0041:SITLAR>2.0.CO;2).
- Xie, S., H. Ma, J. S. Boyle, S. A. Klein, and Y. Zhang, 2012: On the correspondence between short- and long-time-scale systematic errors in CAM4/CAM5 for the Year of Tropical Convection. *J. Climate*, **25**, 7937–7955, <https://doi.org/10.1175/JCLI-D-12-00134.1>.
- Xu, X., S. K. Frey, A. Boluwade, A. R. Erler, O. Khader, D. R. Lapen, and E. Sudicky, 2019: Evaluation of variability among different precipitation products in the Northern Great Plains. *J. Hydrol. Reg. Stud.*, **24**, 100608, <https://doi.org/10.1016/J.EJRH.2019.100608>.
- Yamada, T. J., M. Lee, M. Kanamitsu, and H. Kanamaru, 2012: Diurnal characteristics of rainfall over the contiguous United States and northern Mexico in the dynamically downscaled reanalysis dataset (US10). *J. Hydrometeor.*, **13**, 1142–1148, <https://doi.org/10.1175/JHM-D-11-0121.1>.
- Zhang, C., S. Xie, S. A. Klein, H.-y. Ma, S. Tang, K. Van Weverberg, C. J. Morcrette, and J. Petch, 2018: CAUSES: Diagnosis of the summertime warm bias in CMIP5 climate models at the ARM Southern Great Plains site. *J. Geophys. Res. Atmos.*, **123**, 2968–2992, <https://doi.org/10.1002/2017JD027200>.