GEOPHYSICAL RESEARCH LETTERS, VOL. 33, L22701, doi:10.1029/2006GL028098, 2006

Diagnosing atmosphere-ocean general circulation model errors relevant to the terrestrial biosphere using the Köppen climate classification

Anand Gnanadesikan[1] and Ronald J. Stouffer[1]

Received 6 September 2006; revised 12 October 2006; accepted 18 October 2006; published 16 November 2006.

[1] Coupled atmosphere-ocean-land-sea ice climate models (AOGCMs) are often tuned using physical variables like temperature and precipitation with the goal of minimizing properties such as the root-mean-square error. As the community moves towards modeling the earth system, it is important to note that not all biases have equivalent impacts on biology. Bioclimatic classification systems provide means of filtering model errors so as to bring out those impacts that may be particularly important for the terrestrial biosphere. We examine one such diagnostic, the classic system of Köppen, and show that it can provide an "early warning" of which model biases are likely to produce serious biases in the land biosphere. Moreover, it provides a rough evaluation criterion for the performance of dynamic vegetation models. State-of-the art AOGCMs fail to capture the correct Köppen zone in about 20–30% of the land area excluding Antarctica, and misassign a similar fraction to the wrong subzone.
Citation: Gnanadesikan, A., and R. J. Stouffer (2006), Diagnosing atmosphere-ocean general circulation model errors relevant to the terrestrial biosphere using the Köppen climate classification, *Geophys. Res. Lett.*, *33*, L22701, doi:10.1029/2006GL028098.

1. Introduction

[2] The process of developing climate models involves a vast array of choices. Numerical schemes, parameterizations of subgrid-scale processes, parameter values within these parameterizations- a host of issues face any model development team. The recent process of model development for the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) at the Geophysical Fluid Dynamics Lab (GFDL) involved both qualitative metrics (e.g. maintenance of a reasonable thermohaline overturning and El Nino) and quantitative metrics (e.g. minimization of the maximum error and RMS error in sea surface temperature, minimization of the near-surface air temperature and precipitation errors). An implicit assumption for many such metrics, however, is that biases in temperature and precipitation are weighted equally regardless of location. For example, a 0.3 m/yr underprediction in precipitation in the Amazon rainforest (where the mean precipitation is high) is weighted the same as a 0.3 m/yr underprediction in the middle of North America (which could flip a region from grassland to desert).

[3] From the point of view of physical climate, this is a very defensible approach. It protects modelers from being accused of tuning the "skill score" of their particular model so as to maximize the good points of that model. However, as we will demonstrate in this note, not all model errors are equivalent in how they impact biology. There is evidence that the terrestrial biosphere is affected by key thresholds which must be properly simulated in order for models to be able to simulate the present-day biosphere and to be credible in simulating its future evolution. This demands that developers of coupled models consider not only the raw physical fields such as temperature and precipitation, but additional diagnostics which are sensitive to the impact of these fields on biospherically relevant processes.

[4] An illustration of how models are evaluated from a physical point of view is shown in Figure 1, which shows surface temperatures and precipitation over land in a simulation using the CM2.1 global coupled climate model [*Delworth et al.*, 2006] recently developed at GFDL. This model consists of ocean, sea ice, and atmosphere models, with a land model based on that of *Milly and Shmakin* [2002] that does not include vegetative feedback. Examination of the mean and RMS precipitation errors (Figure 1) shows that there are large errors over the tropical regions, particularly the Amazon, and smaller errors over the temperate midlatitudes. Large temperature errors are also seen over the Sahara. At the edges of mountainous areas (particularly the Tibetan plateau), smoothing of topography within the model leads to significant errors in temperature. The percentage precipitation errors are largest over the Sahara, where precipitation rates are low, and in mountainous regions.

[5] Different errors, however, will not have equivalent impacts when considered from the point of view of individual organisms. An example of this that is familiar to every gardener is the ability of plants to survive cold winters. A 2C cold bias may well mean that a particular plant is unable to survive a winter, while a 2C warm bias may have little effect. Similarly, vegetation may be very sensitive to whether or not the ground dries out during the summer. This paper examines these effects using a classic bioclimatic model, proposed originally by the climatologist W. Köppen. We show that the distribution of Köppen climate types within models provides a way of isolating errors that can be important for land biology, thus providing guidance for developers of physical climate models.

2. Koppen Climate Index

[6] During the first part of the 20th century, the climatologist W. Köppen developed a scheme for predicting

[1]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey, USA.

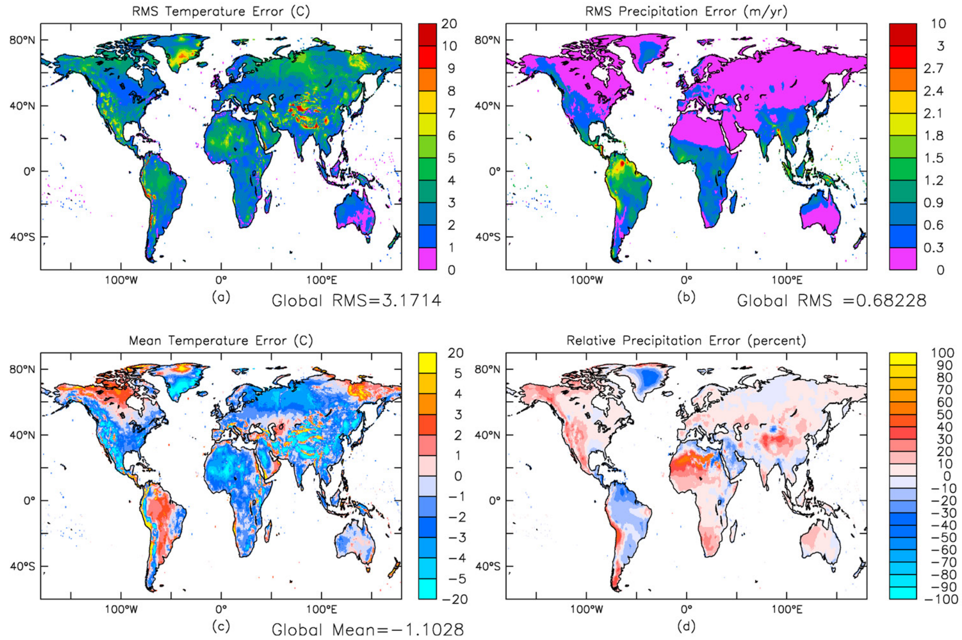

Figure 1. Temperature and precipitation errors computed relative to the UEA CRU05 [*New et al.*, 1999] dataset in CM2.1. Model output is mapped to dataset before differences are computed. (a) RMS temperature error in C. (b) RMS precipitation error in m/yr. (c) Annual mean temperature error in C. (d) Relative precipitation error in %.

biome distributions that took some of these thresholds into account. Such bioclimatic classification schemes have advantages over complex dynamic vegetation models in that they are simple, empirical, and easy to apply to readily available output from different models. A particular attraction of the Köppen scheme is that it is sensitive to the details of the seasonal cycle and takes account of the interactions between temperature and precipitation. A Köppen-like scheme was used by *Henderson-Sellers* [1990] to evaluate the feasibility of estimating the parameters in "big-leaf" biosphere models from climatic variables. *Kleidon et al.* [2000] used the scheme to identify regions that may be particularly sensitive to changes in the land parameterization.

[7] We have adapted the version published by *Lamb* [1972, Chapter 11]. This defines 5 climate types: A, tropical (coolest month warmer than 18C); B, arid (insufficient rain to balance potential evaporation- criteria described in Table 1); C, temperate (coolest month between −3C and 18C); D, boreal forest and snow (warmest month warmer than 10C but coldest month less than −3C); E, cold snow climates (warmest month less than 10C and coldest month less than −3C). Each zone may be subdivided into classes, depending on whether there is a summer dry season, a winter dry season, a monsoon climate, or no dry season (s,w, m or f) and whether the summers are hot (warmest month warmer than 22C, subtype a), warm (more than 4 months warmer than 10C, subtype b) or cool (fewer than 4 months warmer than 10C, subtype c). Arid climates are subdivided into steppe (S) and desert (W) depending whether there is sufficient precipitation over the course of the year. Polar regions are subdivided into tundra (T) and polar desert (F, the division determined by whether the maximum monthly temperature reaches 0C). Since we are interested in the impact of climate on vegetation, we chose to combine the Dwa, Dwb, Dfa and Dfb types (which are said to go with deciduous forest) into a type which we designate Dab, and the Dwc and Dfc types (which are said to go with needle-tree forest) into a type which we designate as Dc. The full list of subtypes that we will use, as well as the criteria determining the different classes is given in Table 1.

[8] The real-world distribution of the Köppen climates as evaluated using the University of East Anglia Climate Research Unit half-degree monthly climatology (UEA CRU05 [*New et al.*, 1999]) of precipitation and temperature is shown in Figure 2a. The climatology covers the time period from 1961–1990. Examination of the map shows that the climatology captures many known geographic features, such as the distribution of deserts, semiarid regions and rain forests. While it is easy to criticize the Köppen classification as being too rigid in its boundaries [*Prentice*, 1990], and too simplistic in its classification types [*Sanderson*, 1999], it can be thought of as a nonlinear filter on the temperature and precipitation that produces a first-order picture of biome distributions. Our goal in this note is to use such distributions to highlight potential biases in the physical climate of modern climate models. We do not necessarily claim that the biases seen will necessarily lead to changes in biomes, only that they have the potential to do so. We are currently developing a more detailed discussion of the utility of the Köppen scheme compared with more ecologically-based schemes (A. Gnanadesikan et al., Evaluating coupled climate models in the context of nonlinear terrestrial ecosystem responses, manuscript in preparation, 2006).

[9] The CM2.1 model does a reasonable job at simulating the overall distribution of Köppen types (Figure 2b). However, a closer examination shows that there are many differences between the simulation and the observational inferred distribution. Particular regions of disagreement are the midlatitude semiarid zones in both hemispheres. While

Table 1. Köppen Vegetation Types Used in This Paper With Nominal Vegetation Types[a]

Type	Name (Nominal Vegetation Types)	Criteria
1. Af	Tropical wet (Tropical evergreen rain forest)	$T_{min} > 18C$, not BS or BW, $P_{min} > 6$
2. Am	Tropical moist (Tropical evergreen rain forest)	$T_{min} > 18C$, not BS or BW, $6 > P_{min} > (250 - P_{year})/25$
3. Aw	Tropical Dry (Savanna/Woodland)	$T_{min} > 18C$, not BS or BW, $6, (250 - P_{year})/25 > P_{min}$
4. BS	Semiarid (Bush to grassland)	$2(T_{ave} + P_{off}) > P_{year} > (T_{ave} + P_{off})$ $P_{off} = 0$, > 30% of rain in winter $P_{off} = 7$, no wet season $P_{off} = 14$, > 30% of rain in summer
5. BW	Desert (waste to cactus/seasonal vegetation)	$(T_{ave} + P_{off}) > P_{year}$
6. Cs	Temperate winter wet (Evergreen broad-leaf forest)	$18C > T_{min} > -3C$, not BS or BW, $P_{max} > 3P_{min}$, winter max., summer min.
7. Cfa	Hot temperate moist (Broad-leaf forest)	$18C > T_{min} > -3C$, not BS or BW, Not Cs or Cw, $T_{max} > 22C$
8. Cfb	Warm temperate moist (Broad-leaf forest)	$18C > T_{min} > -3C$, not BS or BW, Not Cs or Cw, $T_{max} < 22C$. 4+ months warmer than 10C.
9. Cfc	Cool temperate moist (Needle-tree forest)	$18C > T_{min} > -3C$, not BS or BW, Not Cs or Cw, $T_{max} < 22C$. Less than 4 months warmer than 10C.
10. Cw	Temperate summer wet (Evergreen forest)	$18C > T_{min} > -3C$, not BS or BW, $P_{max} > 10P_{min}$, summer max., winter min.
11. Dab:	Cold winters/warm summers (deciduous forest)	$T_{max} > 10C$, $-3C > T_{min}$, not BS or BW 4+ months warmer than 10C.
12. Dc:	Cold winters/cool summers (evergreen forest)	$-3C > T_{min}$, not BS or BW 4+ months warmer than 10C.
13. Et:	Tundra (tundra, dwarf trees, mosses)	$10C > T_{max} > 0C$, $T_{min} < -3C$
14. Ef	Polar desert (permanent ice or rock, little plant life)	$T_{max} < 0C$

[a]As discussed by *Prentice* [1990], actual vegetation may differ. $T_{min,max,ave}$ are the minimum monthly, maximum monthly, and annual-average temperature in C. $P_{min,max,year}$ are the minimum monthly, maximum monthly, and annually-integrated precipitation in cm.

the model does, for example, reproduce the Namib and Atacama deserts, it does not capture the Argentinian Pampas, the Kalahari, the Central Asian steppes or the rain shadow of the Rockies. In fact, (Figure 2c) 45.2 Mkm^2 of the land surface is classified in the wrong type, or 30.8% of the land area excluding Antarctica. Moreover, 38.0 Mkm^2 (25.9%) is in the wrong subtype (Figure 2d), with a particularly significant region being the Amazonian rain forest. In total over half of the land outside Antarctica is classified in the wrong Koppen type or subtype (using the list in Table 1). Some of the regions (such as the Amazon) are subject to large precipitation and temperature errors.

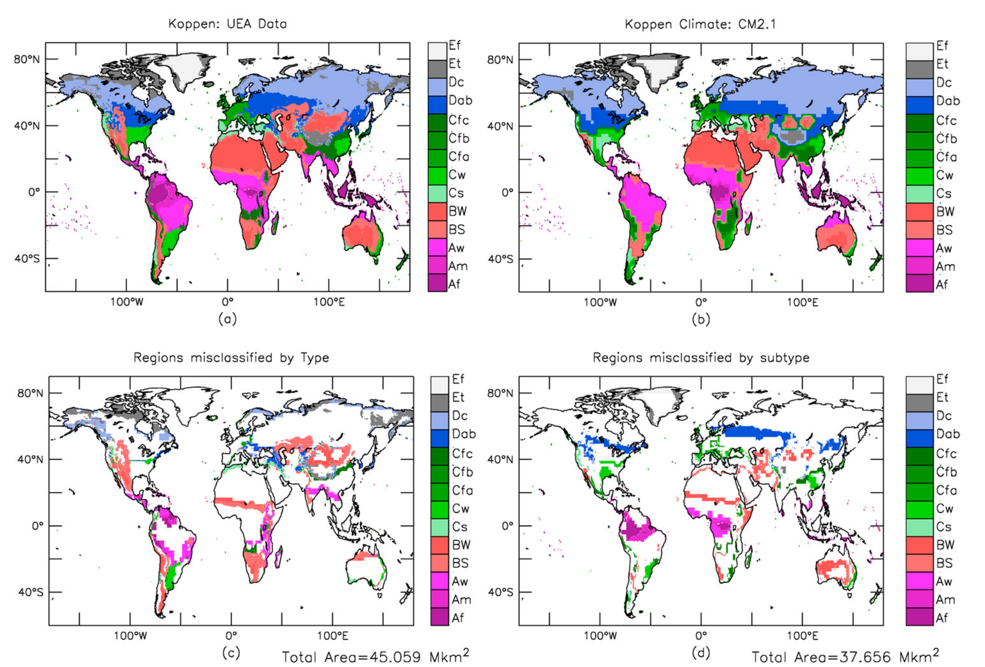

Figure 2. Biases in the climate model revealed using the Köppen scheme. (a) Köppen types from data. (b) Köppen types in CM2.1. (c) Regions (data) misidentified in CM2.1 by type. (d) Regions (data) misidentified in CM2.1 by subtype.

Table 2. Errors in Koppen Type and Subtype Compared With RMS Temperature and Precipitation Errors[a]

Model	Area in Wrong Type, Mkm^2	Area in Wrong Subtype, Mkm^2	RMS Temp. Error, °C	RMS Precip. Error, m/yr
HadGEM	36.1	36.0	3.71	0.69
MPI	41.9	37.7	2.58	0.62
GFDL CM2.1	45.1	37.7	3.17	0.68
CCSM 3.0	44.9	39.9	2.95	0.70
CSIRO	47.6	39.9	3.43	0.70
MIROC-med	51.8	40.4	3.27	0.66
IPSL	50.9	42.3	3.70	0.80
HadCM3	51.8	44.2	2.91	0.60
CCCMA T63	54.2	50.1	3.68	0.68
PCM	58.7	45.3	4.14	0.87

[a]Compared with Koppen types and subtypes computed from the UEA CRU dataset.

However, in other regions (such as steppes of central Asia) the errors are less obvious.

[10] It is interesting to compare the distributions of errors across 10 models submitted to the IPCC AR4 (Table 2). While it is obvious that a perfect simulation of temperature and precipitation would produce a perfect simulation of Koppen climate types (assuming perfect observations of precipitation and temperature), larger errors do not necessarily correspond to worse simulation of climate types. This can be clearly seen by comparing the two models from the Hadley Centre. The newer HadGEM model [*Johns et al.*, 2006] has the best simulation of the distribution of Köppen climates of any of the 10 models considered here, despite ranking 9th in RMS temperature error and 6th in RMS precipitation error. By contrast the HadCM3 model [*Gordon et al.*, 2000] ranks 1 and 2nd in temperature and precipitation errors respectively while ranking 8th with respect to the Köppen type simulation. It is worth noting that the significant improvement in, for example, the distribution of the Am type (which accounts for much of the improvement in the distribution of subtypes) is not obvious from an examination of the figures of *Johns et al.* [2006].

[11] By running the Köppen diagnostic in different models, it is possible to make some statements about the source of the important biases. Figures 3a and 3b show the errors in Köppen type and subtype in a slab model [e.g., *Manabe and Stouffer*, 1979], run with the same atmosphere and land model as in the CM2.1 coupled model, but with a slab ocean whose mean temperatures are computed using flux adjustments that significantly reduce errors in temperature [*Findell et al.*, 2006]. Significant improvements over the AOGCM are seen in Southern Africa and in the southern part of the rain shadow of the Rockies, but the failure to produce enough precipitation in the Amazon basin and to capture the semiarid regions in the Caucasus remains. Interestingly, when the model is run with the dynamic land model (LM3) of *Shevliakova et al.* [2006] the errors do not change substantially (Figures 3c and 3d). This shows that our land model does not cause the climate to shift to a new state, implying no strong feedbacks between the land and the large-scale atmospheric circulation that impact our simulation. It also suggests that certain errors in the land model simulation (such as the putting forest in semiarid zones throughout Midwest North America) can be attributed to biases in the physical simulation.

3. Summary

[12] One of the great challenges in modeling the biosphere is that global biogeochemical cycles may be sensitive to aspects of the physical climate which may not be at

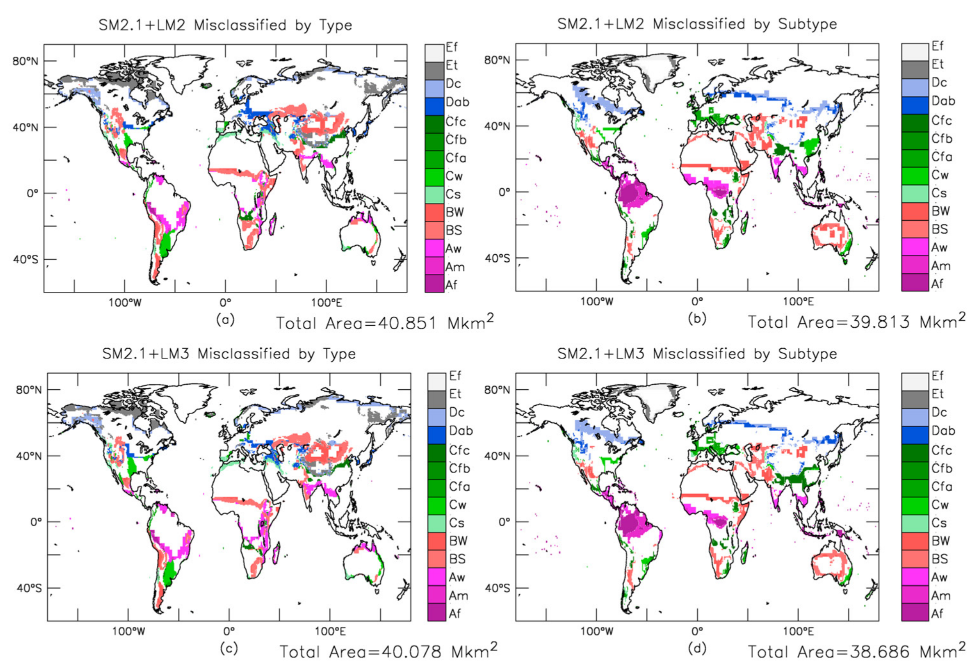

Figure 3. Biases in the Koppen diagnostic in two slab models run with the same atmosphere as the GFDL coupled climate model. (a and b) Models with the same land model as in CM2.1, the so-called LM2 model which is based on the LaD model of *Milly and Shmakin* [2002] in which vegetation is fixed. (c and d) An early version of the LM3 model of *Shevliakova et al.* [2006] in which vegetation is allowed to evolve.

the top of the list for physical climate modelers to evaluate when building a model or which may have very subtle signatures. This is particularly true when looking at systems which exhibit nonlinear responses (of which the threshold-type ecosystem responses we examine here are a special case). Other examples include deep watermass formation due to open-ocean convection, which responds asymmetrically to heating and cooling [*Stouffer and Manabe*, 2003] and disease patterns, which can depend on the spectrum of interannual variability as well as the mean state of climate [*Koelle et al.*, 2005]. Nonlinear responses can result in relatively small errors in regions whose climate is near a threshold having a bigger impact than larger errors in regions whose climate is far from a threshold. In our models, for example, the errors associated with the failure to reproduce the semiarid zones in Central Asia are smaller than those found in the coastal North China Plain- which is nonetheless relatively well simulated from a bioclimatic point of view.

[13] While we have argued looking at threshold-type behavior can make a significant difference when diagnosing the realism of coupled models, this paper does not purport to be a complete study of all the issues surrounding such responses. We wish to conclude by highlighting a number of critical issues.

[14] 1. Which thresholds are actually important? Some thresholds, like whether the ground freezes or dries out can clearly be linked to biogeochemical responses. But what about the boundary between the temperate zone and tropical zones (determined by whether the minimum monthly temperature dips below 18C)?

[15] 2. How sharp are thresholds in reality? It is not clear that the boundary between ecosystems should be anywhere near as sharp as in a Köppen-type scheme. If a physical model is close to one side of an overly sharp threshold, the result will be to make the model far too sensitive to changes in climate that push it across that threshold.

[16] 3. To what extent are thresholds based on mean monthly behavior actually proxies for other behavior? For many organisms, it is not the mean monthly temperature which is important, but whether or not temperatures drop below the frost damage point for buds or leaves on individual organisms for a single day during that month [*Loehle*, 1998].

[17] Further exploration of these issues, particularly in the context of land models, is clearly warranted. The Köppen scheme can be thought of as a simple, empirical, but not particularly sophisticated land model. It should prove interesting to revisit these issues using some more recently developed schemes. *Feddema* [2005] presents a revision of the *Thornthwaite* [1948] classification, long favored by physical climatologists as being more mechanistically based than the Köppen scheme. *Hofmann et al.* [2005] suggest using a multivariate spatial clustering scheme that picks out separate regimes based on the raw climate variables-essentially allowing the data to choose where thresholds occur. *Jolly et al.* [2006] propose a bioclimatic classification scheme which evaluates whether foliar phenology is limited by minimum daily temperature, photoperiod, or maximum vapor pressure deficit- using satellite products and in-situ data to evaluate key thresholds that determine whether or not foliage is evergreen or deciduous. It will be important for Earth System Modelers to identify whether their own land models are governed by similar thresholds and to make a good case for the realism of such thresholds. Only by doing this will the community be able to evaluate whether a simulation is truly skillful, or whether errors in the physics have been buried by unrealistic assumptions (and incorrect thresholds) in biospheric models. In the meantime, we recommend that bioclimatic schemes such as the Köppen climate classification be considered a standard part of future efforts to develop and evaluate AOGCMs.

[18] **Acknowledgments.** We wish to thank Elena Shevliakova, Sergey Malyshev, and Cyril Crevoisier for useful discussions and two anonymous reviewers for useful comments. AG wishes to thank his students from the Community Middle School Science Olympiad program for reintroducing him to the Köppen scheme.

References

Delworth, T. L., et al. (2006), GFDL's CM2 global coupled climate models-part 1: Formulation and simulation characteristics, *J. Clim.*, *19*, 643–674.

Feddema, J. J. (2005), A revised Thornthwaite-type global classification, *Phys. Geogr.*, *26*, 442–466.

Findell, K. L., T. Knutson, and P. C. D. Milly (2006), Weak simulated extratropical responses to complete tropical deforestation, *J. Clim.*, *19*, 2835–2850.

Gordon, C., C. Cooper, C. A. Senior, H. T. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dyn.*, *16*, 147–168.

Henderson-Sellers, A. C. (1990), Predicting generalized ecosystem groups with the NCAR CCM: First steps toward an interactive biosphere, *J. Clim.*, *3*, 917–940.

Hofmann, F. M., W. W. Hargrove, D. J. Erickson, and R. J. Oglesby (2005), Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models, *Earth Interactions*, *9*(10), 1–27, doi:10.1175/EI110.1.

Johns, T. C., et al. (2006), The new Hadley Centre climate model (HadGEM1):Evaluation of coupled simulations, *J. Clim.*, *19*, 1327–1353.

Jolly, W. M., R. Nemani, and S. W. Running (2006), A generalized, bioclimatic index to predict foliar phenology, *Global Change Biol.*, *11*, 619–632.

Koelle, K., X. Rodo, M. Pascual, M. Yunus, and G. Mostafa (2005), Refractory periods and climate forcing in cholera dynamics, *Nature*, *436*, 696–700.

Kleidon, A., K. Fraedrich, and M. Heimann (2000), A green planet versus a desert world: Estimating the maximum effect of vegetation on the land surface climate, *Clim. Change*, *44*(4), 471–493.

Lamb, H. H. (1972), *Climate: Present, Past and Future*, vol. 1, *Fundamentals and Climate Now*, 613 pp., Methuen, New York.

Loehle, C. (1998), Height growth rate tradeoffs determine northern and southern range limits for trees, *J. Biogeogr.*, *25*, 735–742.

Manabe, S., and R. J. Stouffer (1979), A CO_2-climate sensitivity study with a mathematical model of the global climate, *Nature*, *282*, 491–493.

Milly, P. C. D., and A. B. Shmakin (2002), Global modeling of land, water, and energy balances, *J. Hydrometeorol.*, *3*, 283–299.

New, M. G., M. Hulme, and P. D. Jones (1999), Representing twentieth-century space-time climate variability. part I: Development of a 1961–90 mean monthly terrestrial climatology, *J. Clim.*, *12*, 829–856.

Prentice, K. C. (1990), Bioclimatic distribution of vegetation for general circulation model studies, *J. Geophys. Res.*, *95*(D8), 11,811–11,830.

Sanderson, M. (1999), The classification of climates from Pythagoras to Koppen, *Bull. Am. Meteorol. Soc.*, *80*, 669–673.

Shevliakova, E., R. J. Stouffer, M. Spelman, S. Malyshev, and S. Pacala (2006), Feedbacks Between Terrestrial Biosphere and Climate in the GFDL Dynamic Land/Slab Ocean Climate Model, *Eos Trans. AGU*, *87*(52), Fall Meet. Suppl., Abstract A54A-06.

Stouffer, R. J., and S. Manabe (2003), Equilibrium response of thermohaline circulation to large changes in atmospheric CO_2, *Clim. Dyn.*, *20*, 759–773.

Thornthwaite, C. W. (1948), An approach to the rational classification of climate, *Geogr. Rev.*, *38*, 55–94.

A. Gnanadesikan and R. J. Stouffer, NOAA Geophysical Fluid Dynamics Laboratory, Biospheric Processes Group, P.O. Box 308, Princeton, NJ 08542, USA. (anand.gnanadesikan@noaa.gov)