Supplementary Material for
Köppen bioclimatic evaluation of CMIP historical climate simulations

Thomas J. Phillips and Céline J.W. Bonfils
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
Lawrence Livermore National Laboratory, Mailcode L-103
7000 East Avenue, Livermore, California 94550, USA
Email: phillips14@llnl.gov

This supplementary material (SM) includes further explanation of the adopted Köppen scheme in Section S1. In addition, the impact of observational uncertainty on the Köppen evaluation of CMIP models (as inferred from a different choice of observational reference data) is discussed in Section S2. Finally, a preliminary investigation of the impact of model horizontal resolution on the Köppen evaluation is described in Section S3.

S1. Details of the Köppen scheme

We adopted the Köppen scheme of Gnanadesikan and Stouffer (2006), referred to as ‘G & S’. Like other varieties of Köppen classification, G & S account for five major climate “zones”, designated as A (Tropical, coolest month warmer than 18 C), B (Arid, insufficient rainfall to balance potential evaporation), C (Temperate, coolest month between -3 C and 18 C), D (Boreal, temperature of warmest month > 10 C), and E (Polar, temperature of warmest month < 10 C, and of the coldest month < -3 C).

The zones are further divided into several regional “classes” which specify whether there is a summer dry season (class s), a winter dry season (class w), a monsoon climate (e.g. dry winter and wet summer—class m), or no dry season (class f). Within the Tropical A zone, for instance, three regional classes are distinguished: Af (Tropical Wet), Am (Tropical Moist), and Aw (Tropical Dry), which are associated with the generic vegetation types: tropical evergreen rain forest, evergreen forest, and savanna/woodland, respectively.
In addition, regional “subtypes” also are distinguished, depending on whether the summer is “hot” (warmest monthly average temperature greater than 22 C, subtype a), “warm” (more than 4 months with average temperatures greater than 10 C, subtype b), or “cool” (fewer than 4 months with average temperature exceeding 10 C, subtype c). The Arid Zone B instead is divided into steppe S and desert W subtypes according to annual precipitation amount, and the Polar Zone E also comprises two subtypes: tundra t or polar desert f, depending on whether the maximum monthly temperature exceeds or falls below 0 C. Finally, G & S group the traditional Köppen types Dwa, Dwb, Dfa, and Dfb into a single type, designated as Dab, that is associated with deciduous forest, and they group types Dwc and Dfc into a new Dc type that is associated with needle-tree forest.

Table 1 (in the main text) summarizes the particular criteria (based on characteristics of the annual-cycle climatologies of continental T and P) for deriving the 14 Köppen regional climates and associated generic vegetation types. These criteria include the minimum or maximum of monthly average temperature $T_{min}$ or $T_{max}$ and of monthly average precipitation $P_{min}$ or $P_{max}$; the annual-average temperature $T_{avg}$; and the total accumulated yearly precipitation $P_{year}$. (Here, all temperatures are expressed in degrees Celsius, and all precipitation amounts in centimeters of water.)

Other defining criteria for the regional climates/vegetation types in the main text’s Table 1 require more explanation. For instance, the annual average potential evaporation $E_{pot}$ (the amount of water evaporated from a fully wet surface) is approximated (after Köppen 1936) as a linear function of the annual-average temperature: $E_{pot} = T_{avg} + P_{off}$. Here $P_{off}$ is a dimensionless seasonality index that is set to a value of 0 if more than 30% of $P_{year}$ falls in winter, to a value of 14 if more than 30% of $P_{year}$ falls in summer, and to a value of 7 when there is no distinctly wet season. (The seasonal designations are hemisphere-specific, e.g. where ‘summer’ includes the months of June, July, and August in the Northern Hemisphere, and December, January, and February in the Southern Hemisphere.) The magnitude of $P_{year}$ relative to potential evaporation $E_{pot} = (T_{avg} + P_{off})$ differentiates the semiarid BS from the BW desert.
type. In BW, $P_{\text{year}}$ is not sufficient to balance potential evaporation: $P_{\text{year}} < (T_{\text{avg}} + P_{\text{off}})$, while in semiarid BS zones, $P_{\text{year}}$ exceeds $E_p$, but is less than $2E_p$: $(T_{\text{avg}} + P_{\text{off}}) < P_{\text{year}} < 2(T_{\text{avg}} + P_{\text{off}})$.

The Tropical (A) climate zone instead is defined by a threshold for minimum monthly average temperature: $T_{\text{min}} > 18 \, \text{C}$. The value of the minimum monthly average precipitation $P_{\text{min}}$ then is used to differentiate the Tropical wet (Af), moist (Am), and dry (Aw) climate classes: $P_{\text{min}} > 6 \, \text{cm}$, $(250 \, \text{cm} - P_{\text{year}}) < P_{\text{min}} < 6 \, \text{cm}$, and $P_{\text{min}} < 6 \, \text{cm}$ and $< (250 \, \text{cm} - P_{\text{year}})$, respectively (cf. Köppen and Geiger 1930).

**S2. Impacts of observational uncertainty**

Our choice of reference data (NCEP-NCAR Reanalysis T and GPCP P) is somewhat arbitrary, and so it is conceivable that substantively different model performance results might result from using alternative reference fields of T and P. To investigate this potential sensitivity, we repeated the Köppen-based evaluation of the CMIP3 and CMIP5 model historical simulations using alternative Climatic Research Unit (CRU) observationally based estimates of T and P (cf. Harris et al. 2014). Because CRU data are not available over Antarctica, we excluded this area from this alternative Köppen evaluation of the CMIP model simulations.

The map of Köppen vegetation types derived from the CRU reference data, and projected on the same 72 x 144 grid as before, is shown in Figure S1. In addition, the percentage of hits $h(v_i)$ and percentage areas $a(v_i)$ for vegetation types $v_i$ of the CCSM3 and CCSM4 historical simulations—all evaluated with respect to the CRU reference vegetation (denoted CRUOBS)—are plotted in Figure S2. Both figures indicate that the chief differences between the CRUOBS vegetation mapping and that of the OBS are in the semi-arid (BS) and desert (BW) zones, which are more spatially extensive in CRUOBS.

Because the collective CMIP3/5 model simulation of these dry zones is already seen to be deficient (Figure 5), it is not surprising that the aggregate VH and VA scores of model performance with respect to CRUOBS (Table S1) generally fall somewhat below those based on the OBS reference (Table 2). Nevertheless, the performance of an individual CMIP model, relative to that of other models in Table S1, is quite similar to what is implied by Table 2. In particular, the generally improved performance of the
CMIP5 simulations over those of CMIP3 noted previously (Table 2, Figure 5) also is preserved in the CRUOBS results (Table S1, Figure S3). It thus seems reasonable to assume that the Köppen-based model performance metrics are fairly insensitive to different choices of observational reference data.

Figure S1: As in Figure 1 of the main text, the mapping of 14 Köppen vegetation types on a 72 x 144 (2.5 x 2.5 degree) grid are shown (see Table 1 in the main text for type definitions). Here instead, the vegetation types (referred to as the CRUOBS reference) are derived from the Climatic Research Unit (CRU) observationally based estimates of the mean monthly 1980-1999 annual-cycle climatologies of continental temperature T and precipitation P (cf. Harris et al. 2014). Note the absence of vegetation type(s) over Antarctica, owing to missing CRU data for this region.
Figure S2: As in Figure 4 of the main text, except that the CCSM3 and CCSM4 performance metrics $h(v_i)$ in a) and $a(v_i)$ in b) are compared with respect to the CRUOBS reference derived from the CRU observational estimates of continental T and P. Note that the values of $h$ and $a$ for the Polar Desert climate Ef are computed with the exclusion of Antarctica, where CRU reference data are missing.
Figure S3: As in Figure 5 of the main text, except that the area-weighted percentage hits $h_{v_i}$, are evaluated with respect to the CRU OBS reference, derived from the CRU observational estimates of continental T and P.
Table S1. As in Table 2 of the main text, selected participating CMIP modeling groups (and home countries) are listed with associated climate models (and their native horizontal grids, expressed as the number of latitudes × longitudes). Globally aggregated performance scores $V_H$ and $V_A$ (optimal values = 100 %) also are listed for each model’s historical climate simulation, but these are evaluated with respect to vegetation types (denoted CRUOBS) derived from the CRU T and P observational data (cf. Harris et al. 2014). Here, model vegetation over Antarctica is excluded from consideration, consistent with its absence from the CRUOBS reference vegetation. $V_H$ and $V_A$ scores are shaded green where CMIP5 model scores improve relative to those of the CMIP3 antecedent model(s).

| Modeling Group (Country)                                                                 | CMIP3 Model (Horizontal Grid) | VH  | VA  | CMIP5 Model (Horizontal Grid) | VH  | VA  |
|---------------------------------------------------------------------------------------|-------------------------------|-----|-----|-------------------------------|-----|-----|
| Beijing Climate Center , China Meteorological Administration (China)                  | BCC-CM1 (64 x 128)            | 31  | 12  | BCC-CSM1-1 (64 x 128)         | 56  | 81  |
|                                                                                       | CGCM3.1 (48 x 96)             | 54  | 81  | CanCM4 (64 x 128)             | 60  | 81  |
|                                                                                       | CGCM3.1-T63 (64 x 128)        | 55  | 77  | CanESM2 (64 x 128)            | 59  | 82  |
| National Center for Atmospheric Research (USA)                                         | CCSM3.0 (128 x 256)           | 56  | 68  | CCSM4 (192 x 288)             | 64  | 72  |
| National Science Foundation-Department of Energy-National Center for Atmospheric Research Earth System Model Contributors (USA) | CCSM3.0 (128 x 256)           | 56  | 68  | CESM1-BGC (192 x 288)         | 66  | 72  |
|                                                                                       | CESM1-CAM5-1-FV2 (192 x 288)  | 60  | 68  | CESM1-FASTCHEM (192 x 288)    | 64  | 73  |
|                                                                                       | CESM1-WACCM (96 x 144)        | 57  | 70  |                               |     |     |
| Centre National de Recherches Meteorologiques/Centre Europeen de Recherche et Formation avancee en Calcul Scientifique (France) | CNRM-CM3 (64 x 128)           | 56  | 69  | CNRM-CM5 (128 x 256)          | 58  | 74  |
| Commonwealth Scientific and Industrial Research Organisation with Queensland Climate Change Centre of Excellence (Australia) | CSIRO-Mk-3.0 (96 x 192)       | 55  | 82  | CSIRO-Mk-3-6 (96 x 192)       | 53  | 74  |
| Modeling Group (Country)                                                                 | CMIP3 Model (Horizontal Grid) | VH | VA | CMIP5 Model (Horizontal Grid) | VH | VA |
|----------------------------------------------------------------------------------------|-------------------------------|----|----|-------------------------------|----|----|
| NOAA Geophysical Fluid Dynamics Laboratory (USA)                                        | GFDL-CM2.0 (90 x 144)         | 51 | 68 | GFDL-CM3 (90 x 144)           | 63 | 73 |
|                                                                                       | GFDL-CM2.1 (90 x 144)         | 54 | 72 | GFDL-ESM2G (90 x 144)         | 58 | 70 |
|                                                                                       |                               |    |    | GFDL-ESM2M (90 x 144)         | 56 | 72 |
| NASA Goddard Institute for Space Studies (USA)                                         | GISS-EH (46 x 72)             | 46 | 84 | GISS-E2-H (90 x 144)          | 59 | 78 |
|                                                                                       | GISS-ER (46 x 72)             | 52 | 75 | GISS-E2-R (90 x 144)          | 58 | 80 |
| Met Office Hadley Centre (UK) with Instituto Nacional de Pesquisas Espacials (Brazil) | HadCM3 (73 x 96)              | 59 | 87 | HadGEM2-CC (145 x 192)        | 65 | 81 |
|                                                                                       | HadGEM1 (145 x 192)           | 60 | 83 | HadGEM2-ES (145 x 192)        | 67 | 82 |
| Institute for Numerical Mathematics (Russia)                                          | INM-CM3.0 (45 x 72)           | 50 | 74 | INM-CM4 (120 x 180)           | 56 | 64 |
| Institut Pierre-Simon Laplace                                                         | IPSL-CM4 (72 x 96)            | 48 | 69 | IPSL-CM5A-LR (96 x 96)        | 56 | 73 |
|                                                                                       |                               |    |    | IPSL-CM5A-MR (143 x 144)      | 60 | 78 |
| Atmosphere and Ocean Research Institute (Univ. of Tokyo), National Institute for     | MIROC3.2-MEDRES (64 x 128)    | 56 | 75 | MIROC4h (320 x 640)           | 66 | 78 |
| Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (Japan)|                               |    |    | MIROC5 (128 x 256)            | 61 | 72 |
|                                                                                       |                               |    |    | MIROC-ESM (64 x 128)          | 57 | 71 |
|                                                                                       |                               |    |    | MIROC-ESM-CHEM (64 x 128)     | 57 | 70 |
| Max Planck Institute for Meteorology (Germany)                                        | MPI-ECHAM5 (96 x 192)         | 62 | 82 | MPI-ESM1-LR (96 x 192)        | 64 | 84 |
| Meteorological Research Institute (Japan)                                             | MRI-CGC2M3.2a (64 x 128)      | 58 | 83 | MRI-CGC3M3 (160 x 320)        | 60 | 79 |
| Norwegian Climate Centre (Norway)                                                    | BCCR-BCM2.0 (64 x 128)        | 52 | 69 | NorESM1-M (96 x 144)          | 61 | 73 |
S3. Impacts of horizontal resolution

Possible impacts of diverse model horizontal resolutions on the Köppen evaluation of CMIP model bioclimatic performance also should be addressed. Owing to recent computational advances, many CMIP5 models were run at higher horizontal resolution than their CMIP3 counterparts. We first consider whether the improved collective performance of CMIP5 models relative to CMIP3 simulations (see Figure 5 in the main text) is mainly a result of higher resolution.

S3.1 Impacts on collective model performance

In investigating this aspect of horizontal resolution, some insights can be gleaned from considering cases where a CMIP5 model retained the same resolution as its CMIP3 predecessor. Examples in this category include certain BCC, CCCMA, CSIRO, GFDL, Hadley Centre, and MPI CMIP3 and CMIP5 model versions. If VH is employed as the more consistent measure of overall simulation performance, the CMIP5 versions in these cases almost all display better performance than their CMIP3 counterparts (see Tables 2 and S2). The implication is that the superior performance of these CMIP5 models results from their better representations of physical processes that impact T and P.

In contrast, a recent evaluation of CMIP3 versus CMIP5 model simulations of historical continental climate (focusing on temperature, precipitation, and sea-level pressure) by Watterson et al. (2014) attributed most of the increased skill of the CMIP5 simulations relative to CMIP3 to increases in CMIP5 horizontal resolution. Our study’s seeming differences with Watterson et al. may be attributable to the use of bivariate measures of model skill. That is, the Köppen-based VH and VA scores are aggregate indicators of how realistically the combined annual cycles of T and P are simulated, whereas Watterson et al.’s results are based on correlation and RMS measures of univariate skill with respect to several reference data sets.

S3.2 Impacts on individual model performance

We previously noted (in Section 3.3 of the main text) that the lower aggregate VH and VA scores of the CESM1-WACCM model relative to other CESM1 versions might be partly attributable to its substantially coarser horizontal resolution; but in this case a resolution effect cannot be cleanly separated
from the effects of model physics differences. There is one example, however, where the impact of increasing resolution, *absent changes in model physics*, can be inferred by comparing the VH score of the low-resolution CMIP5 model version IPSL-CM5A-LR with that of its medium-resolution counterpart, IPSL-CM5A–MR. Here a 140 % increase in horizontal resolution (from a 96 x 96 → 143 x 144 grid) produces only modest improvements in overall model performance (VH = 56→60 in Table S1, and 59 → 61 in Table 2). While such selected examples are not definitive, it seems reasonable to attribute most of this CMIP3→CMIP5 improvement to better model physical-process representation, rather than to higher resolution.

Another aspect of the potential impact of horizontal resolution--that of the remapping of model data to a common grid--also warrants attention. In projecting a relatively high-resolution simulation of T and P onto a coarser 72 x 142 common grid, it is possible that the resulting Köppen vegetation types might be degraded relative to the types derived from fields of T and P simulated on the native grid of the model.

We conducted a preliminary investigation of this possibility by selecting simulations of 5 CMIP5 models with relatively high resolution (see leftmost column of Table S2). Because the original CRU data are available on an ultrahigh resolution (0.5 x 0.5-degree) grid, it is feasible to project these reference data to the native grid of an individual model, and then conduct its Köppen evaluation on this model-specific grid.

In Table S2, the VH and VA scores (rounded to the nearest whole number) resulting from these native-grid evaluations are contrasted with the scores obtained on the coarse 72 x 144 grid (obtained from Table S1). The scores computed on each model’s native grid are similar, but mostly a bit lower than those determined on the coarse grid. Possibly these score differences may be attributed, in part, to the different sampling sizes (number of grid boxes) involved in computing the respective VH and VA scores. While these results must be regarded as preliminary, they suggest that the Köppen evaluation scores are rather insensitive to the mapping of the CMIP models to a common 72 x 144 grid.
Table S2. Impacts of horizontal resolution on aggregate performance scores VH and VA for 5 CMIP5 model simulations of Köppen vegetation types, evaluated with respect to those derived from the CRUOBS reference vegetation. The metrical results listed in column 2 correspond to model performance evaluations where both the CMIP5 simulations and the CRUOBS are interpolated to a common 72 x 144 grid (from Table S1). The VH and VA scores listed in column 3 are obtained from Köppen performance evaluations where the CRUOBS-derived reference vegetation types are mapped on the native horizontal grid for each model (grid resolution shown in parentheses in the leftmost column).

| CMIP5 Model (native grid) | VH, VA (72x144 grid) | VH, VA (native grid) |
|---------------------------|-----------------------|----------------------|
| CCSM4 (192 x 278)         | 64, 72                | 64, 68               |
| MRI-CGCM3 (160 x 320)     | 60, 79                | 62, 77               |
| HadGEM2-ES (145 x 192)    | 67, 82                | 66, 76               |
| CNRM-CM5 (128 x 256)      | 58, 74                | 54, 74               |
| MIROC4h (320 x 640)       | 66, 78                | 65, 76               |

References

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

Harris, I., P.D. Jones, T.J. Osborn, and D.H. Lister 2014: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. Int. J. Climatol. 34 623-642

Köppen, W. 1936 Das geographische System der Klimate. Handbuch der Klimatologie I-C, Gebrüder Bontraeger, Berlin

Watterson, I.G., J. Bathols, and C. Heady 2014 What influences the skill of climate models over the continents Bull. Amer. Meteor. Soc. 95 689-700