

Köppen bioclimatic evaluation of CMIP historical climate simulations

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 Environ. Res. Lett. 10 064005

(<http://iopscience.iop.org/1748-9326/10/6/064005>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 210.77.64.110

This content was downloaded on 13/04/2017 at 02:03

Please note that [terms and conditions apply](#).

You may also be interested in:

[Implications of regional improvement in global climate models for agricultural impact research](#)

Julian Ramirez-Villegas, Andrew J Challinor, Philip K Thornton et al.

[The large influence of climate model bias on terrestrial carbon cycle simulations](#)

Anders Ahlström, Guy Schurgers and Benjamin Smith

[Effects of declining aerosols on projections of zonally averaged tropical precipitation](#)

L D Rotstayn, M A Collier and J-J Luo

[Projections of rapidly rising surface temperatures over Africa under low mitigation](#)

Francois Engelbrecht, Jimmy Adegoke, Mary-Jane Bopape et al.

[Sensitivity of regional climate to global temperature and forcing](#)

Claudia Tebaldi, Brian O'Neill and Jean-François Lamarque

[Preparing local climate change scenarios for the Netherlands using resampling of climate model output](#)

G Lenderink, B J J M van den Hurk, A M G Klein Tank et al.

[Potential change in forest types and stand heights in central Siberia in a warming climate](#)

N M Tchebakova, E I Parfenova, M A Korets et al.

[Global patterns of NDVI-indicated vegetation extremes and their sensitivity to climate extremes](#)

Guo Liu, Hongyan Liu and Yi Yin

[Observed and simulated temperature extremes during the recent warming hiatus](#)

Jana Sillmann, Markus G Donat, John C Fyfe et al.

Environmental Research Letters



LETTER

Köppen bioclimatic evaluation of CMIP historical climate simulations

OPEN ACCESS

RECEIVED

20 February 2015

REVISED

11 May 2015

ACCEPTED FOR PUBLICATION

13 May 2015

PUBLISHED

5 June 2015

Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



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, CA 94550, USA

E-mail: phillips14@llnl.gov**Keywords:** Köppen, bioclimate, CMIP, models, simulations, performance, metricsSupplementary material for this article is available [online](#)**Abstract**

Köppen bioclimatic classification relates generic vegetation types to characteristics of the interactive annual-cycles of continental temperature (T) and precipitation (P). In addition to predicting possible bioclimatic consequences of past or prospective climate change, a Köppen scheme can be used to pinpoint biases in model simulations of historical T and P . In this study a Köppen evaluation of Coupled Model Intercomparison Project (CMIP) simulations of historical climate is conducted for the period 1980–1999. Evaluation of an example CMIP5 model illustrates how errors in simulating Köppen vegetation types (relative to those derived from observational reference data) can be deconstructed and related to model-specific temperature and precipitation biases. Measures of CMIP model skill in simulating the reference Köppen vegetation types are also developed, allowing the bioclimatic performance of a CMIP5 simulation of T and P to be compared quantitatively with its CMIP3 antecedent. Although certain bioclimatic discrepancies persist across model generations, the CMIP5 models collectively display an improved rendering of historical T and P relative to their CMIP3 counterparts. In addition, the Köppen-based performance metrics are found to be quite insensitive to alternative choices of observational reference data or to differences in model horizontal resolution.

1. Introduction

Future deviations of temperature or precipitation from today's climatological normals may severely endanger multitudes of the Earth's living organisms (e.g. Williams *et al* 2007, Garcia *et al* 2014). Vegetation species are especially threatened, since they cannot easily migrate in response to changes in climate (Thomas *et al* 2004, Thuiller *et al* 2005, Jiang *et al* 2013, Xu *et al* 2013).

Wladimir Köppen (1900) was the first to systematically quantify perceived relationships between the climatological annual cycles of continental temperature and precipitation (hereafter, T and P) and associated generic vegetation types inhabiting different regions (e.g. tundra vegetation, evergreen or deciduous forests, grasslands, etc). Various modifications of Köppen's initial classification scheme were later introduced by Köppen and Geiger (1930), Trewartha

(1968), and Lamb (1972), for example. Other researchers have attempted to improve on Köppen's methodology, for instance by considering plant physiological factors in defining the operative climatic variables (e.g. Holdridge 1947, Thornthwaite 1948, Prentice 1990, Fedema 2005, Jolly *et al* 2006) or by applying statistical clustering techniques to more precisely define the boundaries of ecoregions whose climate characteristics evolve in time (Hargrove and Hoffman 1999, Hoffman *et al* 2005). In common with Köppen classification, these alternative schemes implicitly assume that the geographical distribution of dominant vegetation types is both in equilibrium with, and largely determined by, the climate state. Instead, probabilistic approaches such as that of Brovkin *et al* (1997) allow for the coexistence of several vegetation types in the same region that have different probabilities of occurrence which may vary continuously under changing climate conditions.

In spite of their shortcomings from an ecological perspective, Köppen and alternative deterministic schemes have proved very useful for identifying the first-order bioclimatic consequences of past climate change (Guetter and Kutzbach 1990, Wang and Overland 2004, Gerstengarbe and Werner 2009, Rubel and Kottek 2010, Chen and Chen 2013) or of prospective future climate change inferred from simulated projections of T and P under diverse greenhouse-gas emissions scenarios (Leemans *et al* 1996, de Castro *et al* 2007, Diaz and Eischeid 2007, Gao and Georgi 2008, Roderfeld *et al* 2008, Jylhä *et al* 2010, Feng *et al* 2012, 2014; Hanf *et al* 2012, Gallardo *et al* 2013, Mahlstein *et al* 2013, and Elguindi *et al* 2014). It is also noteworthy that Mahlstein *et al* (2013), Elguindi *et al* (2014), and Feng *et al* (2014) have conducted such studies using future-climate projections provided by the most recent generation of global coupled ocean-atmosphere climate model entries in phase 5 of the Coupled Model Intercomparison Project (CMIP5), described by Meehl and Bony (2011) and Taylor *et al* (2012).

For projections of future climate to be credible, however, models must demonstrate an ability to accurately simulate T and P in the historical climate record. Early applications of Köppen schemes for evaluating the performance of particular climate models were reported by Manabe and Holloway (1975) and Lohmann *et al* (1993). More recently, Gnanadesikan and Stouffer (2006) (hereafter, ‘G and S’) used a Köppen scheme to evaluate selected simulations of late-20th century continental climate rendered by multiple coupled ocean-atmosphere global climate model (AOGCM) entries in Phase 3 of the Coupled Model Intercomparison Project (CMIP3, see Meehl *et al* 2007). G and S pointed out that a model bias in T or P in one region may not have the same biological consequences as in another. They emphasized that, for climate models to provide useful bioclimatic predictions, they must correctly simulate specific thresholds of T and P that determine the natural regional boundaries of different vegetation types.

The present study aims to extend the work of G and S by providing an updated Köppen evaluation of CMIP5 simulations of the recent historical climate. (Elguindi *et al* 2014 also evaluate the CMIP5 historical climate simulations, but instead employ a revised Thornthwaite bioclimatic scheme.) We conduct our evaluation by comparing the Köppen mappings of vegetation types derived from each CMIP5 simulation with those obtained from observational reference values of T and P . To illustrate typical bioclimatic strengths and weaknesses of the models, we discuss detailed results for an example CMIP5 model simulation and that of its CMIP3 antecedent. In addition, we show how model errors in rendering Köppen vegetation types may be deconstructed to reveal the character of corresponding biases in regional T and P .

A secondary goal of our study is to compare the bioclimatic performance of the CMIP5 models with

that of their CMIP3 antecedents. This is an important exercise, since the collective improvement or deterioration of CMIP models impact periodic assessments of the Intergovernmental Panel on Climate Change (e.g. IPCC 2007, IPCC 2013), which rely heavily on the CMIP simulations of historical or future climate. To this end, we develop bioclimatic measures to quantify the performance of the CMIP3 and CMIP5 simulations of T and P relative to observational reference data. We also conduct a preliminary investigation of the effects of observational uncertainty, and of the impact of model horizontal resolution on these simulation performance measures.

Subsequent sections describe the methods and data employed (section 2) and the salient model-evaluation results (section 3). We offer concluding remarks in section 4, and discuss additional technical details in a supplementary material (SM) appendix available at stacks.iop.org/ERL/10/064005/mmedia.

2. Methodology and data

A Köppen classification scheme identifies generic vegetation types associated with regional climate zones defined by characteristics of the annual cycles of T and P (e.g. the value of maximum monthly temperature or the season of maximum precipitation). A Köppen vegetation type thus embodies the interplay of the amplitude and seasonal phase of the associated regional T and P annual cycles.

Because variants of Köppen classification define generic vegetation types somewhat differently, one should choose a scheme that is appropriate for evaluating the CMIP models. For example, the Köppen–Geiger classification differentiates some 30 vegetation types (Kottek *et al* 2006, Peel *et al* 2007)—probably too many for practical application on the coarse horizontal grid (resolving only several degrees latitude/longitude) that is typical of a CMIP model.

We instead adopt the scheme that G and S employed, which distinguishes 14 regional climates and associated vegetation types, but still sets a challenging standard for model evaluation. Choosing the G and S scheme also permits their initial evaluation of selected CMIP3 models to be extended consistently to the current-generation CMIP5 models. The criteria defining the 14 generic vegetation types and their associated regional climates are listed in table 1. Because these vegetation types are rather ambiguously described (e.g. ‘evergreen forest’, ‘evergreen broad-leaf forest’, etc), hereafter we will refer to the vegetation types by their corresponding Köppen regional climate designations (Dc, Cw, Cs, etc). Further details of the G and S Köppen scheme are discussed in the SM appendix.

By applying the defining criteria of table 1, the 14 vegetation types can be mapped from observations of the climatological annual cycles of regional T and P . Because fully global, satellite-based estimates of T and

Table 1. Köppen climate types (and associated vegetation-type description), with their corresponding continental temperature/precipitation (T/P) defining criteria, after Gnanadesikan and Stouffer (2006). (See further explanation in section S1 of the supplementary material).

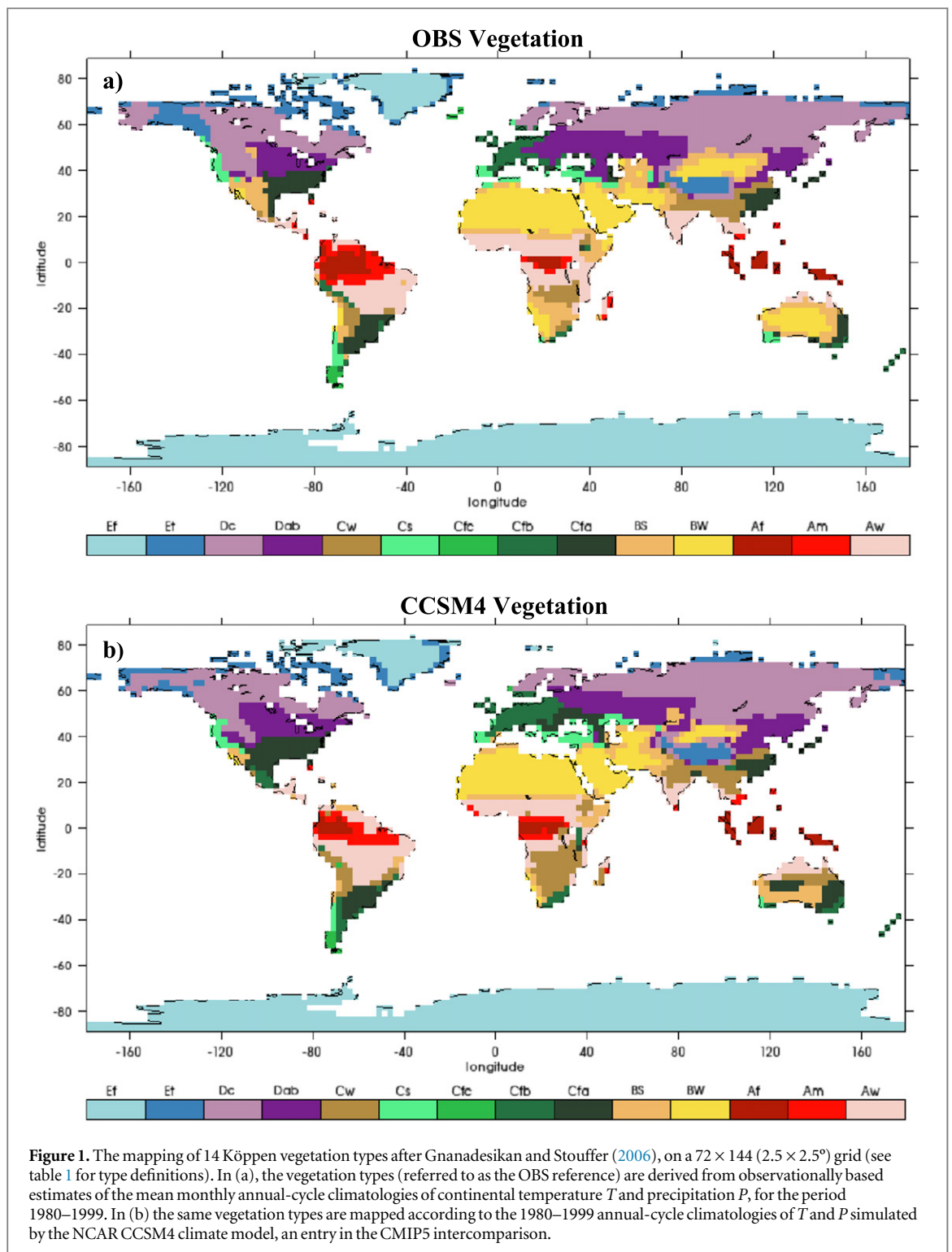
Köppen Regional Climate (Generic Vegetation Type)	T/P Criteria*
Ef: Polar Desert (scant vegetation)	$T_{\max} < 0$ Celsius (C)
Et: Tundra (dwarf trees, mosses)	$0\text{ C} < T_{\max} < 10\text{ C}$ and $T_{\min} < -3\text{ C}$
Dc: Cold Winters/Cool Summers (evergreen boreal forest)	$T_{\min} < -3\text{ C}$ and < 4 months warmer than 10 C , but not types BS or BW
Dab: Cold Winters/Warm Summers (deciduous forest)	$T_{\min} < -3\text{ C}$, $T_{\max} > 10\text{ C}$ and > 4 months warmer than 10 C , but not types BS or BW
Cw: Temperate, Wet Summers (evergreen forest)	$-3\text{ C} < T_{\min} < 18\text{ C}$ and $P_{\max} > 10P_{\min}$ with P_{\max} occurring in summer and P_{\min} in winter, but not types BS or BW
Cs: Temperate, Wet Winters (evergreen broad-leaf forest)	$-3\text{ C} < T_{\min} < -18\text{ C}$ and $P_{\max} > 3P_{\min}$, with P_{\max} occurring in winter and P_{\min} in summer, but not types BS or BW
Cfc: Temperate, Cool and Moist (needle-tree forest)	$-3\text{ C} < T_{\min} < 18$ and $T_{\max} < 22\text{ C}$, and with < 4 months warmer than 10 C , but not types BS , BW , Cs , or Cw
Cfb: Temperate, Warm and Moist (broad-leaf forest)	$-3\text{ C} < T_{\min} < 18$ and $T_{\max} < 22\text{ C}$, and with > 4 months warmer than 10 C , but not types BS , BW , Cs , or Cw
Cfa: Temperate, Hot and Moist (broad-leaf forest)	$-3\text{ C} < T_{\min} < 18\text{ C}$ and $T_{\max} > 22\text{ C}$, but not types BS , BW , Cs , or Cw
BS: Semiarid (bush or grassland)	$(T_{\text{avg}} + P_{\text{off}}) < P_{\text{year}} < 2(T_{\text{avg}} + P_{\text{off}})$
BW: Desert (wasteland, cactus/seasonal Vegetation)	$P_{\text{year}} < (T_{\text{avg}} + P_{\text{off}})$
Af: Tropical Wet (tropical evergreen rain forest)	$T_{\min} > 18\text{ C}$ and $P_{\min} > 6\text{ cm}$, but not types BS or BW
Am: Tropical Moist (tropical evergreen forest)	$T_{\min} > 18\text{ C}$ and $(250\text{ cm} - P_{\text{year}})/25 < P_{\min} < 6\text{ cm}$, but not types BS or BW
Aw: Tropical Dry (savanna/woodland)	$T_{\min} > 18\text{ C}$ and $P_{\min} < 6\text{ cm}$, $(250\text{ cm} - P_{\text{year}})/25$, but not types BS or BW

*Here $T_{\min, \max, \text{avg}}$ are, respectively, the minimum monthly, maximum monthly, and annual-average continental temperature T in degrees Celsius (C). $P_{\min, \max, \text{year}}$ are the minimum monthly, maximum monthly, and annually integrated continental precipitation P in centimeters (cm). The dimensionless precipitation seasonality index P_{off} is set to a value of 0 if $> 30\%$ of P_{year} falls in winter, to 7 if there is no distinctly wet season, and to 14 if $> 30\%$ of P_{year} falls in summer.

P were not available prior to 1979, and because the CMIP3 historical climate simulations did not extend past the year 2000 (Meehl *et al* 2007), we focused on the 20 year climatological period 1980–1999. Observationally based estimates of climatological monthly continental T for this 20 year period were obtained from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis surface air temperature field (Kalnay *et al* 1996), and climatological monthly

continental P amounts from the Global Precipitation Climatology Project (GPCP) data set (Adler *et al* 2003).

The vegetation types of table 1, derived from the chosen observational estimates of T and P (hereafter referred to as the ‘OBS’ vegetation reference) are mapped on a 72×144 ($2.5^\circ \times 2.5^\circ$) grid in figure 1(a). It is seen that the climate zones **A** to **E** in figure 1(a) align roughly according to latitude, but their respective classes/subclasses and associated vegetation types display



substantial longitudinal variation. For example, tundra (**Et**) and boreal forest (**Dc**) dominate northern Eurasia, but also the Tibetan plateau. Desert (**BW**) and semi-arid (**BS**) vegetation types occupy subtropical Africa and central Asia, but also intrude in ‘rain shadows’ to the east of the Rockies and Andes, and to the north of the Himalayas. Where subsiding air dominates the continental interiors of southern Africa and Australia, desert (**BW**) and semi-arid (**BS**) vegetation coexist with temperate types that occupy wetter

regions nearby. Tropical vegetation (**Af**, **Am**, **Aw**) populates southern and southeastern Asia, as well as Amazonia and equatorial Africa. Over North America, tundra (**Et**) and boreal forest (**Dc**) coexist with broad-leaf forests (**Dab**, **Cs**, **Cfa**), while semi-arid (**BS**) and desert (**BW**) vegetation dominates the southwest US and Mexico. A patchwork of temperate forests (**Cs**, **Cw**, **Cfa**, **Cfb**, **Cfc**) also occupy portions of South America, Africa, Australia, Europe, and China—notwithstanding the absence of forests resulting from

historical land-clearing practices in many of these locations.

The OBS vegetation mapping of figure 1(a) serves as a reference standard against which the G and S vegetation types derived from CMIP3/5 historical climate simulations of T and P can be evaluated. Our study considers 1980–1999 historical climate simulations of T and P from 27 CMIP5 models and their 18 CMIP3 antecedents (for the same modeling groups). In many cases, a CMIP model produced multiple realizations of historical climate that were usually distinguished by a different specification of the initial conditions of the model's coupled ocean-atmosphere state. In such cases, we arbitrarily chose only the first in a series of these realizations (denoted as 'run 1' of a CMIP3 model's 20th Century ('20c3m') experiment, or 'r1i1p1' of a CMIP5 model's 'historical' experiment—see web address http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax_v0-25_clean.pdf for notational details on the CMIP5 realizations). All chosen simulations were mapped to the same 72×144 grid as that of the OBS reference. Detailed information on technical features of individual CMIP3 models may be obtained at web address www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php, and for individual CMIP5 models at <http://es-doc.org>. The CMIP3/5 model names, their native horizontal grid resolutions, and the associated modeling groups are listed in table 2.

3. Köppen evaluation results

3.1. Vegetation mappings

Köppen vegetation maps such as those shown in figure 1(a) were derived for each of the CMIP3/5 model historical simulations. As an illustrative example, we focus on the simulation of the widely used Community Climate System Model Version 4 (CCSM4), a CMIP5 entry. The CCSM4 vegetation mapping (figure 1(b)) displays strengths and weaknesses that are typical of many other CMIP models.

In general, the CCSM4 vegetation types replicate salient features of the OBS mapping (figure 1(a)), but with differences in position and areal extent. **Ef** and **Et** types are generally well-simulated, with the exception of Alaska. Marked discrepancies include the constriction of desert (**BW**) and semi-arid (**BS**) vegetation areas in the lee of the Rockies, Andes, and Himalayas, as well as over Mexico and in the southern African and Australian interiors, where they are displaced in many regions by temperate vegetation types (**Cw**, **Cfa**). In central and eastern Europe also, temperate broadleaf forests (**Cfb**) erroneously encroach on deciduous cold-winter forests (**Dab**). Some discrepancies in simulated vegetation types occur elsewhere, such as Amazonia and equatorial Africa, where the divisions between different types of tropical forests (**Af**, **Am**, **Aw**) are not well reproduced.

3.2. Deconstruction of simulation biases

Because Köppen vegetation maps are built up from characteristics of the annual-cycle climatologies T and P (e.g. fields of $T_{\max/\min}$, T_{avg} , $P_{\max/\min}$, P_{year} , P_{off} —see table 1), we can readily deconstruct the specific biases in modeled T and/or P that are responsible for distinctive regional vegetation errors. Here, the CCSM4 simulation of the vegetation map in figure 1(b) again provides an illustrative example. The more egregious CCSM4 vegetation-type errors (relative to the OBS standard in figure 1(a)) tend to occur in the drier Köppen climatic zones such as leeward of topography or over southern Africa and the Australian interior, where the annual accumulation of precipitation P_{year} is excessive for these regions (not shown). Other CCSM4 dry-zone errors over central Europe and Mexico instead result from an interplay of the biases in T and P .

For example, we deconstruct the CCSM4 T and P biases over Mexico in figure 2. The modeled vegetation types (figure 2(a)) are quite different from those seen in the OBS reference (figure 2(b)). In CCSM4, there is a dearth of semi-arid (**BS**) vegetation that appears in the OBS mapping over central Mexico; instead, temperate vegetation types (**Cfa** and **Cfb**, denoted by shades of green) occupy these model grid cells. In panels 2(c) and (d), the CCSM4–OBS differences in annual-average temperature (ΔT_{avg}) and in annually integrated precipitation (ΔP_{year}) are shown. In some locations, the CCSM4 simulation of P_{year} exceeds that of the OBS reference by ~ 90 cm, while the modeled T_{avg} falls below that of the OBS by ~ 4 C. By examining monthly T and P , it is found that both the CCSM4 T deficits and P excesses over Mexico are present throughout the year, but are most extreme in May–July (not shown).

In addition, there are sizeable temporal phase errors in the modeled precipitation, as displayed by a field of the seasonality index P_{off} (see table 1 and SM section 1) for the CCSM4 (figure 2(e)) that contrasts with the observed P_{off} (figure 2(f)), notably over central Mexico. Here, instead of a pronounced summer peak in the observed P ($P_{\text{off}} = 14$) coinciding with T_{\max} , CCSM4 precipitation is distributed more evenly over the entire year ($P_{\text{off}} = 7$). Hence, the model's maximum and annual-averages temperatures T_{\max} and T_{avg} are both depressed relative to observations (figure 2(c)). In consequence, the simulated Köppen estimate of potential evaporation $E_p = (T_{\text{avg}} + P_{\text{off}})$ over Mexico (see SM section S1) is substantially smaller than that for the OBS, and the defining criteria for semi-arid vegetation **BS** ($E_p < P_{\text{year}} < 2 E_p$, from table 1) is not satisfied (since simulated $P_{\text{year}} > 2 E_p$). Instead, temperate broadleaf forest types **Cfa** and **Cfb** (figure 2(a)) erroneously displace much of the Mexican semi-arid (**BS**) vegetation (figure 2(b)) in locations where the model's depressed annual T_{\max} satisfies the corresponding temperate criteria (see table 1).

Table 2. Selected participating CMIP modeling groups (and home countries) listed with associated climate models (and their native horizontal grids, expressed as the number of latitudes \times longitudes). Globally aggregated performance scores **VH** and **VA** (optimal values = 100%) also are listed for each model's simulation of historical climate evaluated relative to the OBS reference (see text for further details). **VH** and **VA** scores are shaded green where CMIP5 model scores improve on those from 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)	32	10	BCC-CSM1-1 (64 x 128)	60	87
Canadian Centre for Climate Modelling & Analysis (Canada)	CGCM3.1 (48 x 96)	59	83	CanCM4 (64 x 128)	63	82
	CGCM3.1-T63 (64 x 128)	62	76	CanESM2 (64 x 128)	61	82
National Center for Atmospheric Research (USA)	CCSM3.0 (128 x 256)	62	73	CCSM4 (192 x 288)	68	84
National Science Foundation-Department of Energy-National Center for Atmospheric Research Earth System Model Contributors (USA)	CCSM3.0 (128 x 256)	62	73	CESM1-BGC (192 x 288)	70	83
				CESM1-CAM5-1-FV2 (192 x 288)	63	81
				CESM1-FASTCHEM (192 x 288)	67	84
				CESM1-WACCM (96 x 144)	59	81
Centre National de Recherches Meteorologiques/ Centre Europeen de Recherche et Formation avancee en Calcul Scientifique (France)	CNRM-CM3 (64 x 128)	60	75	CNRM-CM5 (128 x 256)	61	82
Commonwealth Scientific and Industrial Research Organisation with Queensland Climate Change Centre of Excellence (Australia)	CSIRO-Mk-3.0 (96 x 192)	60	81	CSIRO-Mk-3-6 (96 x 192)	56	76

3.3. Bioclimatic performance metrics

For objective comparison of CMIP model simulations, it is necessary to develop quantitative measures of bioclimatic performance. For example, we define a vegetation 'hits' metric $\mathbf{h}(v_i)$ as an area-weighted measure of the percentage of one-to-one matches of each model-derived vegetation type with the corresponding OBS type v_i (where $i = 1$ to 14) in each grid cell. Here $\mathbf{h}(v_i)$ is weighted by the grid-cell areas which (given the convergence of the meridians toward the pole) decrease as the cosine of the latitude.

In effect, the metric \mathbf{h} constitutes the diagonal of a hits matrix $\mathbf{H}(v_i, v_j)$ that relates occurrences of model vegetation type v_j in grid cells where the observed (OBS) vegetation type is v_i . In figure 3, hits matrices are shown for the CCSM4 historical climate simulation and that of its CMIP3 antecedent, the CCSM3 model. The color-coded model-OBS hits percentages $\mathbf{h}(v_i)$ for each OBS vegetation type v_i are arrayed along

the matrix diagonals. The off-diagonal patches represent the model 'misses' in vegetation type, where their vertical distances from the diagonal broadly indicate the degree of mismatch with the OBS vegetation types. Overall, the CCSM4 simulation displays fewer misses in vegetation type than the CCSM3.

The hits metric \mathbf{h} is a stringent measure of a model's bioclimatic performance, since slight but consistent errors in the spatial locations of simulated vegetation types can have a sizeable negative impact. A more 'forgiving' performance metric instead compares the percentage of total land area $\mathbf{a}(v_i)$ occupied by each simulated vegetation type v_i with that of the corresponding OBS reference v_i , regardless of whether there is a one-to-one match of vegetation types in each grid box.

Both these metrics are plotted for the CCSM3 and CCSM4 simulations in figure 4. For the two simulations, the hits percentages tend to be higher for

Table 2. (Continued.)

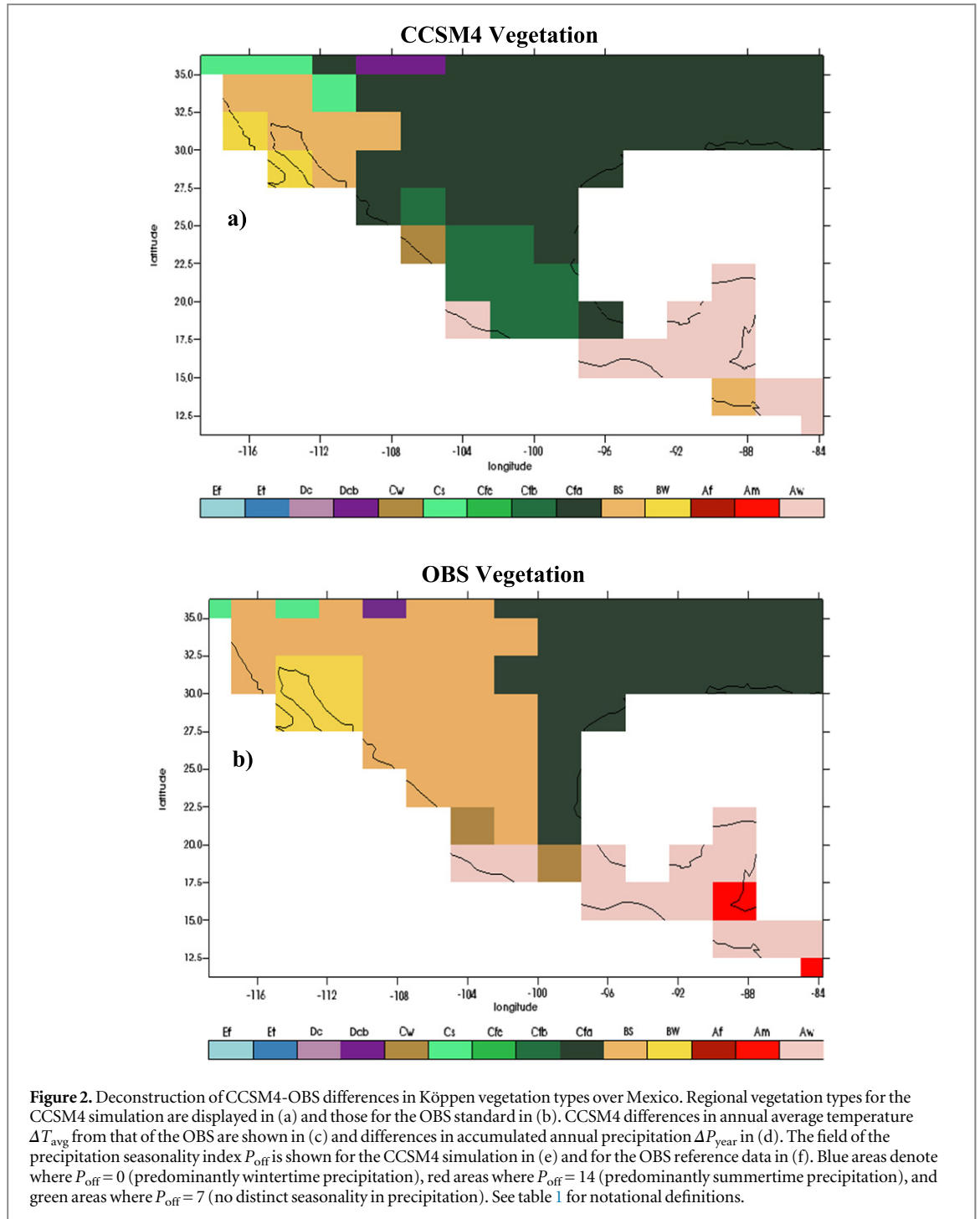
Modeling Group (Country)	CMIP3 Model (Horizontal Grid)	VH	VA	CMIP5 Model (Horizontal Grid)	VH	VA
NOAA Geophysical Fluid Dynamics Laboratory (USA)	GFDL-CM2.0 (90 x 144)	56	75	GFDL-CM3 (90 x 144)	66	81
				GFDL-ESM2G (90 x 144)	61	77
	GFDL-CM2.1 (90 x 144)	58	79	GFDL-ESM2M (90 x 144)	59	81
NASA Goddard Institute for Space Studies (USA)	GISS-EH (46 x 72)	49	87	GISS-E2-H (90 x 144)	56	78
	GISS-ER (46 x 72)	55	76	GISS-E2-R (90 x 144)	58	80
Met Office Hadley Centre (UK) with Instituto Nacional de Pesquisas Espaciais (Brazil)	HadCM3 (73 x 96)	64	89	HadGEM2-CC (145 x 192)	67	87
	HadGEM1 (145 x 192)	66	91	HadGEM2-ES (145 x 192)	69	86
Institute for Numerical Mathematics (Russia)	INM-CM3.0 (45 x 72)	54	76	INM-CM4 (120 x 180)	61	68
Institut Pierre-Simon Laplace	IPSL-CM4 (72 x 96)	52	68	IPSL-CM5A-LR (96 x 96)	59	74
				IPSL-CM5A-MR (143 x 144)	61	81
Atmosphere and Ocean Research Institute (Univ. of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (Japan)	MIROC3.2-MEDRES (64 x 128)	62	85	MIROC4h (320 x 640)	68	88
				MIROC5 (128 x 256)	60	84
				MIROC-ESM (64 x 128)	60	83
				MIROC-ESM-CHEM (64 x 128)	61	83
Max Planck Institute for Meteorology (Germany)	MPI-ECHAM5 (96 x 192)	66	82	MPI-ESM1-LR (96 x 192)	67	85
Meteorological Research Institute (Japan)	MRI-CGCM2.3.2a (64 x 128)	65	81	MRI-CGCM3 (160 x 320)	61	80
Norwegian Climate Centre (Norway)	BCCR-BCM2.0 (64 x 128)	58	76	NorESM1-M (96 x 144)	65	80

vegetation in the Polar, Boreal, and Temperate (**E**, **D**, and **C**) zones than in the Arid (**B**) or Tropical (**A**) zones (figure 4(a)). Both CCSM3 and CCSM4 display particular weaknesses in matching the locations of the semiarid (**BS**) and the moist tropical (**Am**) vegetation types, with CCSM3 performing somewhat better for **BS**, and CCSM4 for **Am**. For the majority of vegetation types, however, the hits percentages are higher for CCSM4 than for CCSM3.

The aggregate percentage of total land area $a(v_i)$ occupied by different OBS vegetation types v_i (black line in figure 4(b)) ranges widely, with Evergreen Boreal Forest (type **Dc**) occupying the largest percentage area (~17%), and Temperate Needle-Tree Forest (**Cfc**) the smallest (~1%). Semiarid (**BS**), desert (**BW**),

and tropical dry (**Aw**) vegetation types also cover comparatively large (~11–16%) percentage areas. Both the CCSM3 and CCSM4 simulations display good agreement with the OBS tropical (**Af**, **Am**, **Aw**) vegetation areas. The close matching of the simulated moist tropical (**Am**) vegetation area with that of the OBS appears to result from compensating errors, with too little of this type being simulated over Amazonia, and too much over equatorial Africa. This fortuitous matching of total **Am** areas stands in contrast to a relatively low hits percentage (figure 4(a)), evidenced most clearly by the erroneous spatial displacement of the **Am** type over Amazonia (figure 1(b)).

Both CCSM models also under-predict desert vegetation areas (**BW**), as noted in section 2. In



addition, CCSM4 under-predicts the area of the semi-arid (BS) vegetation, in agreement with a markedly low hit percentage for this type (figure 4(a)). However, CCSM4 matches Polar, Tundra, and Boreal Forest (Ef, Et, Dc, and Dcb) vegetation areas much better than CCSM3. Both models reproduce temperate vegetation areas comparatively well, but with CCSM4 tending to over-predict Cw, and Cfa types (figure 4(b)). The CCSM4 simulation better reproduces the majority of OBS vegetation areas, however.

In order to assess the overall bioclimatic performance of each CMIP model, measures that are aggregated across all 14 vegetation types v_i are also

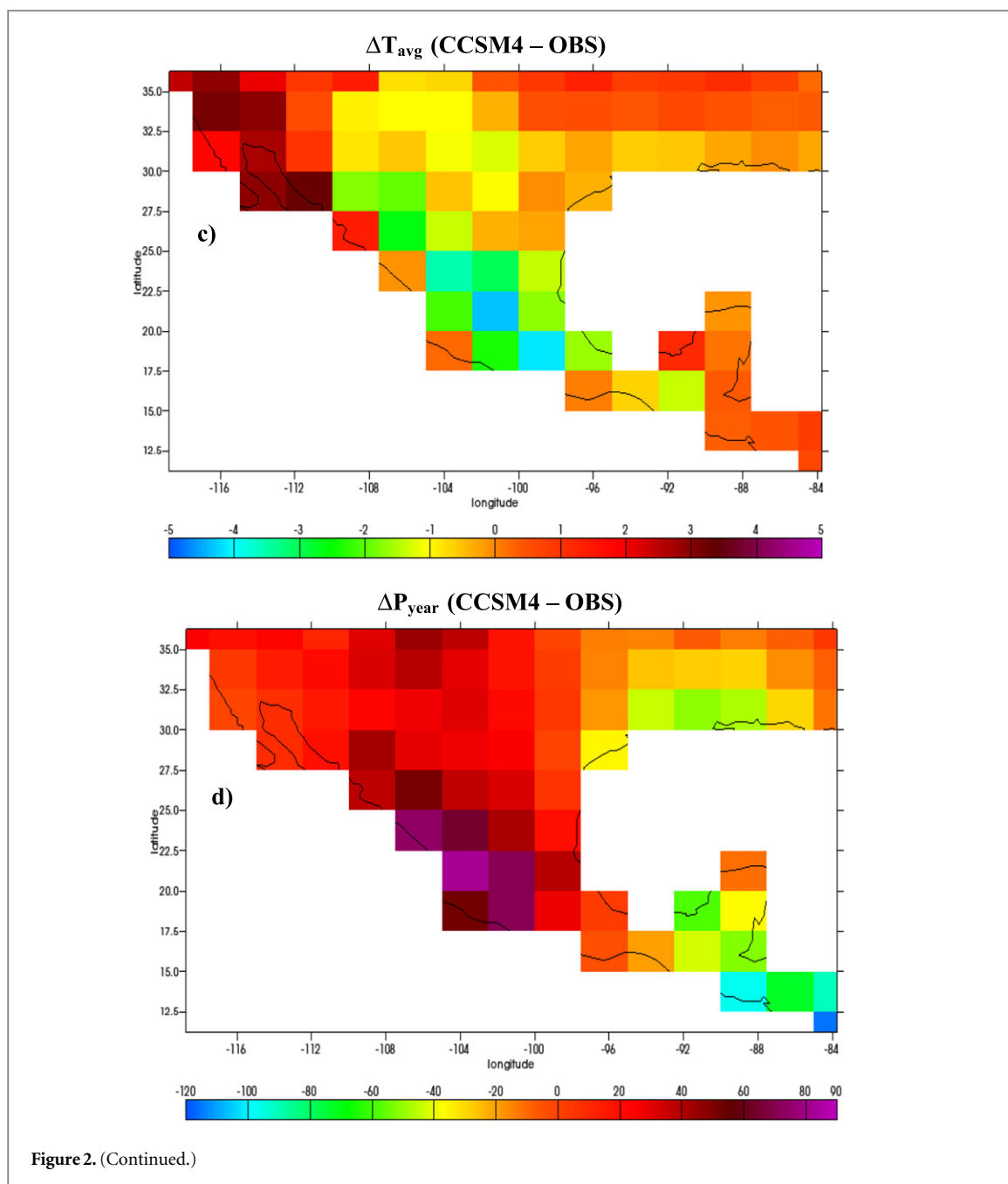
necessary. For example, aggregate performance scores VH and VA can be derived from the vegetation-specific indices $h(v_i)$ and $a(v_i)$:

$$\text{VH} = \sum_i h(v_i) / \sum_i i = \sum_i h(v_i) / 14, \quad (1)$$

$$\text{VA} = 100 - \sum_i |\Delta_i|, \quad (2)$$

where Δ_i signifies the model-OBS difference in $a(v_i)$ percentage.

Note that both VH and VA have optimum values of 100%.



The VH score measures an aggregate agreement in both vegetation type and location. In similar evaluative studies such as that of Elguindi *et al* (2014) this agreement is instead expressed by a *kappa* statistic (Cohen 1960, Monserud 1990, Prentice *et al* 1992). Although VH and *kappa* both attempt to measure overall agreement between two fields of vegetation types, we would point out several basic differences in these measures. Our study's objective is to measure the area-weighted, gridbox-scale agreement of *each modeled vegetation type* with that of the OBS reference vegetation field, expressed in the area-weighted hits metric values $h(k)$. Thus, we implicitly treat the OBS reference vegetation field as a 'truth' target that a simulation of vegetation types matches according to its $h(k)$ values. The VH aggregate score then is

built up by averaging a simulation's $h(k)$ over all vegetation types. The *kappa* statistic, as implemented by Elguindi *et al* (2014), does *not* implicitly treat the vegetation reference as a truth target. The *kappa* statistic also is strictly an aggregate measure of agreement, and is *not* built up from any vegetation-specific measure, nor are vegetation types area-weighted. Finally, unlike the VH score, the *kappa* statistic attempts to account for chance occurrences of agreement in the vegetation fields, and so *kappa* is considered a very conservative measure.

Table 2 lists the VH and VA metrical values (rounded to the nearest percent) for the CMIP3 and CMIP5 simulations. For example, the overall performance of the CCSM4 model (VH = 68, VA = 84) is substantially better than that of CCSM3 (VH = 62 and VA = 73).

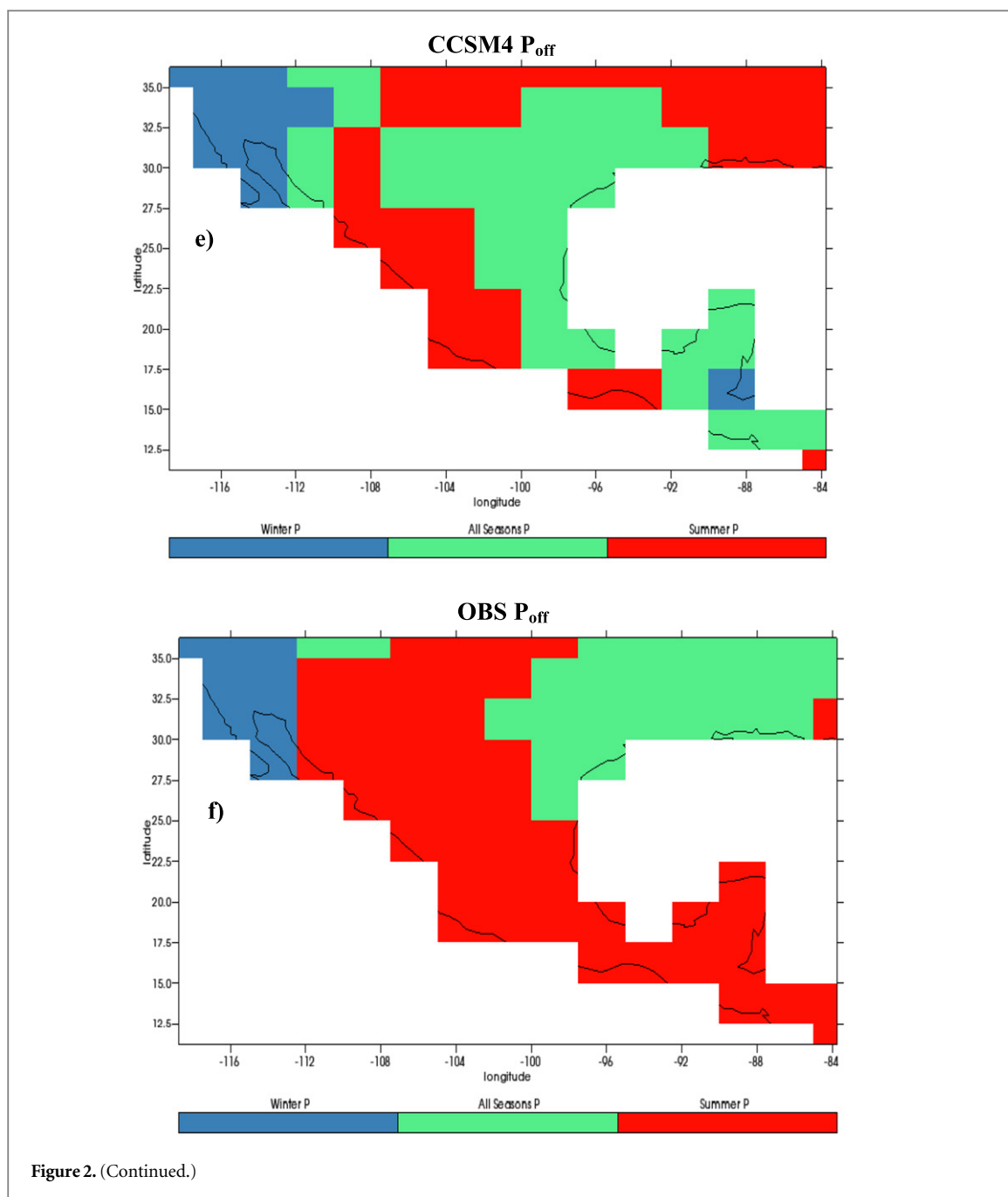
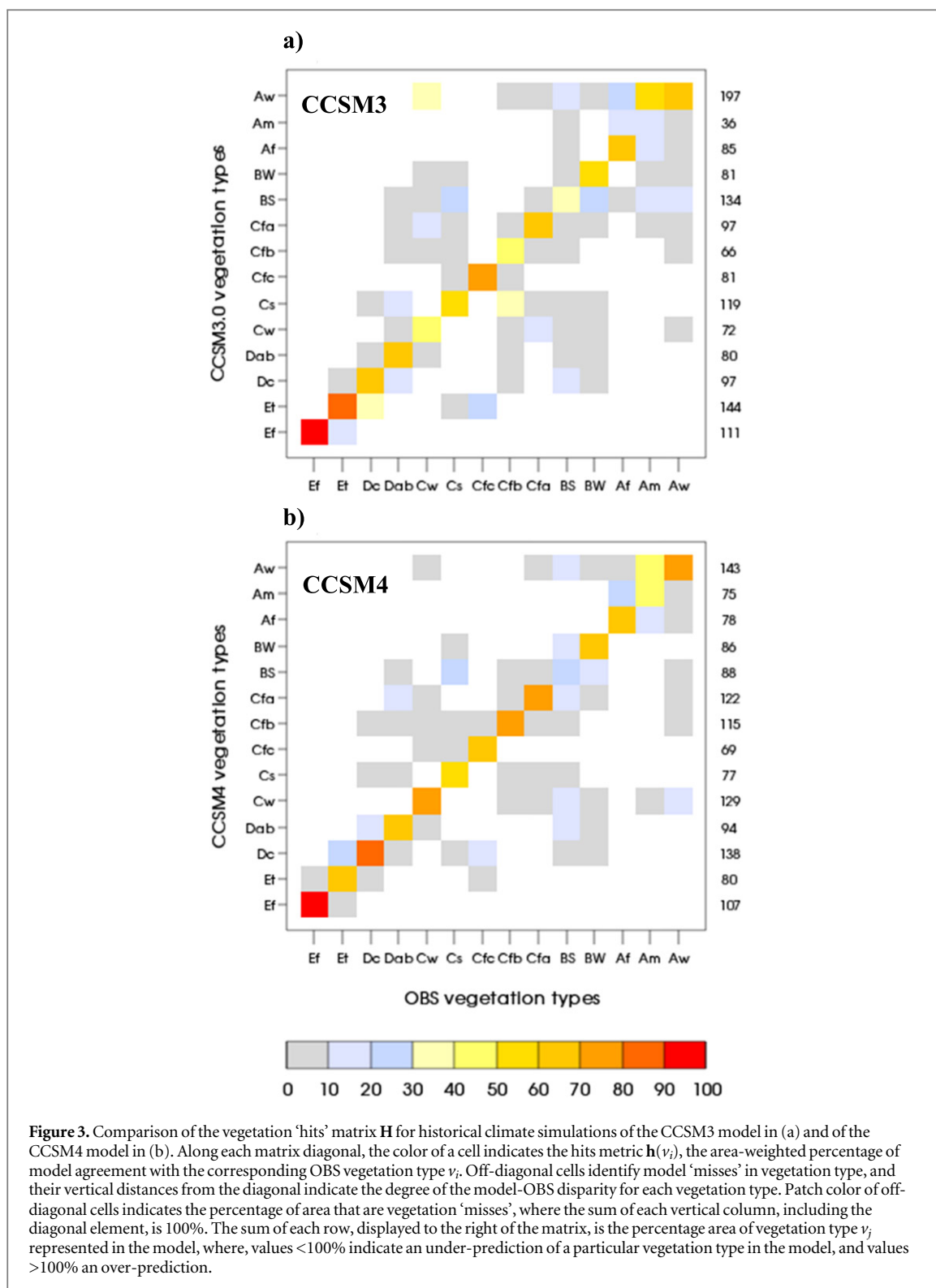


Figure 2. (Continued.)

Adjacent to the CCSM4 results in table 2, performance scores are listed for different versions of the successor Community Earth System Model (CESM1) which include more complex representations of atmospheric radiation and cloud-aerosol interactions than CCSM4 (see www.cesm.ucar.edu/models/cesm1.0/notable_improvements.html and linked pages). In addition, emissions of biogenic aerosols and their deposition on ice, snow, and vegetation are treated in the CESM1-FASTCHEM, while the CESM1-WACCM simulates the stratosphere at finer vertical resolution (although at only half the horizontal resolution of the other CESM1 model versions—see table 2). CESM1-BGC also predicts variations in the global carbon cycle, including flux exchanges between ocean, land, and atmosphere, as well as related variations in

vegetation types and areas. Note, however, that in the CMIP5 historical climate experiments, all such carbon-cycle prognostics are ‘switched off’. Nevertheless, for such an Earth System Model (ESM), the Köppen evaluation subjects the simulation of T and P to careful scrutiny, as a prerequisite for a realistic simulation of the land’s carbon cycle and vegetation cover.

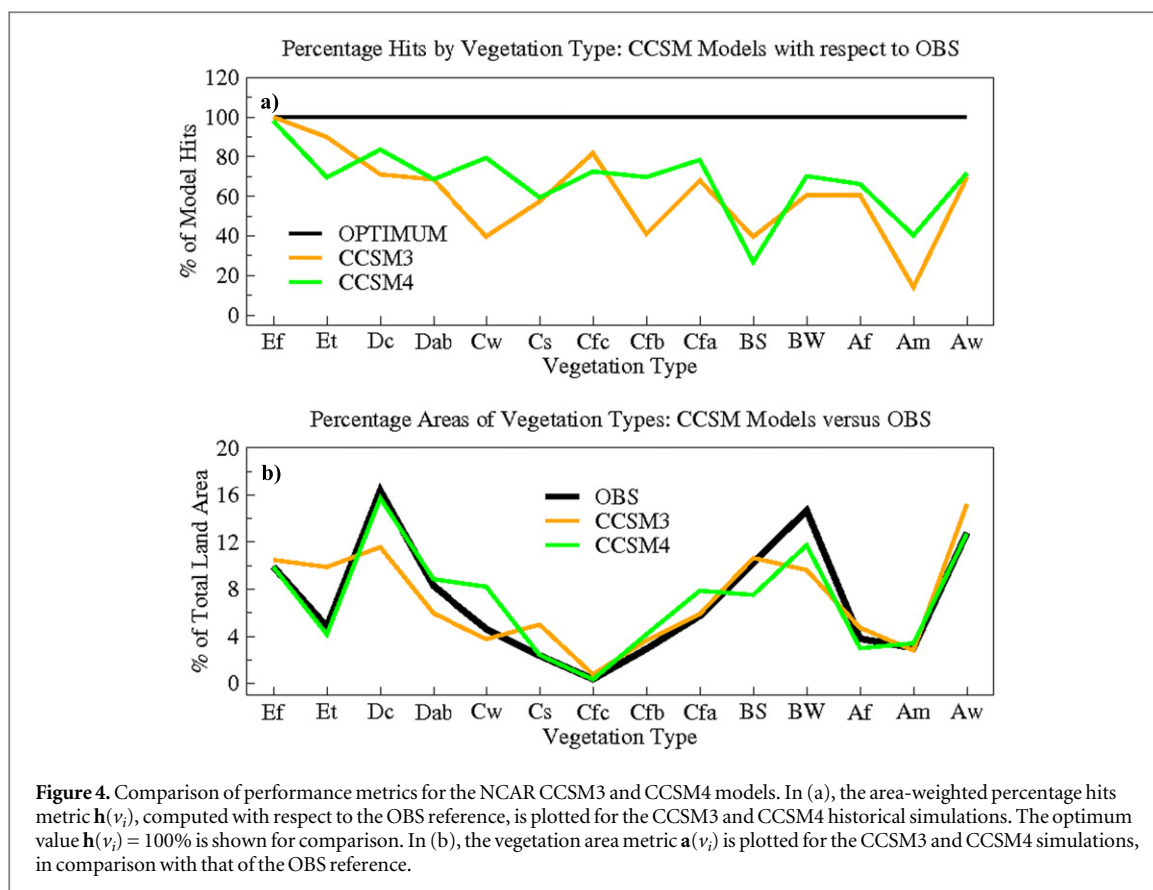
The simulations of all the CESM1 historical simulations display biases in vegetation types (not shown) that are qualitatively similar to those of the CCSM4 (figure 1(b)). The performance scores **VH** and **VA** for the CESM1-BGC and CESM1-FASTCHEM—despite the greater complexity of their physical-process representations—also are similar to those of the CCSM4 (table 2). (The **VH** and **VA** scores for the CESM1-CAM5-1-FV2 are somewhat lower than CCSM4s, and



are lower still for CESM1-WACCM; but the WACCM version, having only half the horizontal resolution of its companion models, probably suffers an inherent disadvantage in such a comparison—see also section S3.1 of the supplementary material on the effects of model horizontal resolution.) Performance scores for other ESMs that are analogous to CESM-BGC in table 2 (e.g. CanESM2, GFDL-ESM2G, GFDL-ESM2M, HadGEM2-ES, INM-CM4, MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM1-LR, and

NorESM1-M) are also close to those of the less complex CMIP5 counterpart(s) from the same modeling group.

Additional salient model performance results in table 2 are summarized as follows. For the CMIP3 simulations, the **VH** score ranges between values of 32–66, and for CMIP5, between 56 and 70. In most cases, the **VH** scores for a particular CMIP5 model simulation roughly equal, or in many instances exceed, those for its CMIP3 antecedent (see green-



shaded cells in table 2). Notable examples of such improved model performance include the simulations by BCC-CSM1-1, CNRM-CM5, GFDL-CM3, GISS-E2-H, GISS-E2-R, IPSL-CM5A-LR, IPSL-CM5A-Mr, MIROC4h, and NorESM1-M.

A collective improvement in the CMIP5 models relative to their CMIP3 counterparts is also evidenced in multi-model averages of the respective hits index $h(v_i)$, shown in figure 5. However, the CMIP5 improvement mostly occurs in the vegetation types of the Boreal (D) and Temperate (C) climate zones. The collective VH scores of CMIP5 are close to those of CMIP3 in the Polar (E) and Arid (B) zone, and are only slightly higher for some Tropical (Am and Aw) vegetation classes. The persistent difficulties in replicating the dry vegetation (BS and BW) types may be related to an erroneous 'drizzle effect' exhibited by many climate models, in which simulated precipitation events occur more frequently than observed (e.g. Stephens *et al* 2010). From inspection of Köppen vegetation maps of individual CMIP models, the collective failure to reproduce the moist tropical vegetation type (Am) results mainly from erroneous simulation of precipitation amounts or patterns over Amazonia.

From table 2, it is seen that a model's VA score almost always surpasses its VH value. This is as expected, since VH measures the overall ability of a model to reproduce vegetation types that correctly match the OBS reference at each grid cell, while VA measures its ability to simulate only the aggregate percentage areas

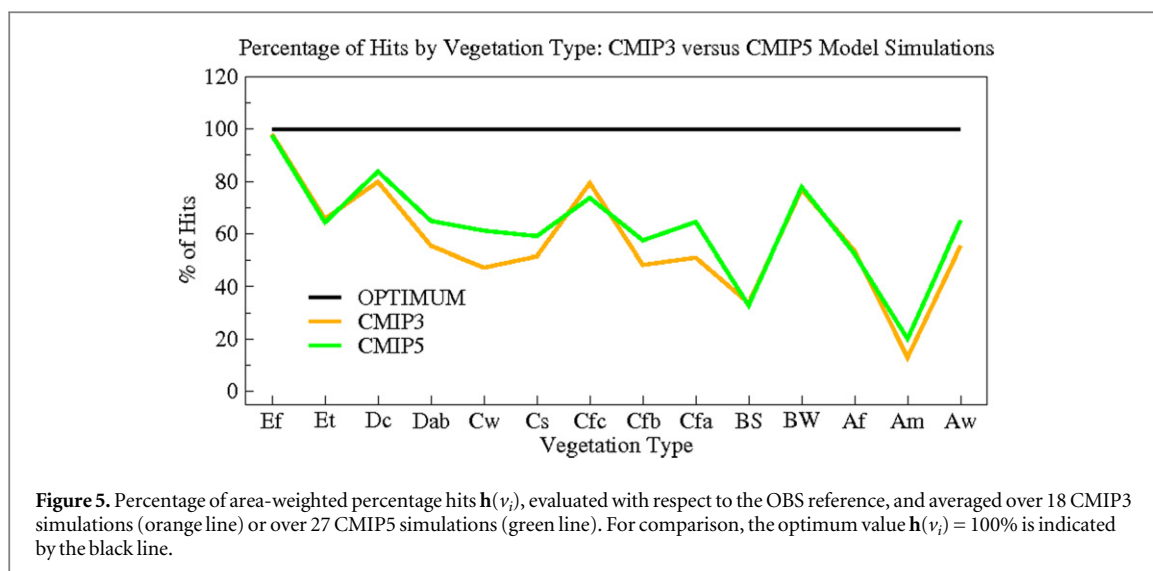
of vegetation types. In many CMIP5 simulations there are also improvements in the VA scores relative to their CMIP3 antecedents, but with a number of exceptions to this pattern (note gray-shaded cells in the rightmost column of table 2). These outcomes imply that VA is a less consistent measure of overall model performance than VH.

3.4. Sensitivity to observational uncertainty and model resolution

It is possible that our bioclimatic performance metrics are sensitive to observational uncertainties (e.g. depending on the choice of observed reference data) or to the grid resolution of the Köppen vegetation mappings. However, further investigation (see SM sections S2 and S3) implies that the bioclimatic performance measures are quite insensitive to both these factors.

4. Concluding remarks

Our study demonstrates the efficacy of a Köppen-based bioclimatic evaluation of the collection of CMIP model simulations of the annual-cycle climatologies T and P , as critical determinants of habitability for living organisms. In particular, a Köppen scheme pinpoints where simulated T and P deviate from values associated with observed regional bioclimatic zones. Because the Köppen vegetation types are derived from specific characteristics of the annual cycles of T and P , it is straightforward to deconstruct the particular



simulation biases that produce distinctive regional vegetation errors. This approach also lends itself to developing metrics and related graphical depictions of model strengths and weaknesses in simulating T and P , and thus for tracking changes in model performance across development cycles.

Our bioclimatic performance evaluation is moderately encouraging, in the sense that most CMIP5 models display improvements over their CMIP3 antecedents, especially in representing T and P in high- and mid-latitudes (i.e. Köppen climate zones E, D and C). It is also reassuring that the performance of the ESMs compares well with their less complex model counterparts. Moreover, these improvements appear robust to different choices of observational reference data and grid resolution.

On the other hand, the CMIP5 simulations agree with the observational reference vegetation types in, at most, about 70 percent of the grid cells (see VH values in table 2 and S1), and obvious deficiencies remain in simulating the Arid (B) and Tropical (A) zones. This harsh appraisal should perhaps be softened somewhat, however, in view of the generally more accurate CMIP5 simulations of aggregate areas of Köppen vegetation types.

These outcomes highlight the stringency of a Köppen-based regional evaluation of global climate simulations that are mostly implemented at rather coarse horizontal resolution. Our study suggests, however, that such regional discrepancies should *not* be attributed mainly to resolution deficiencies. Rather, the lackluster bioclimatic performance of today's climate models reflects a collective inability to predict essential physical characteristics of T and especially of P , whose realistic simulation requires accurate representations of frontal dynamics, convection, and topographic uplift (e.g. Qian *et al* 2009, Catto *et al* 2013, Hirota and Takayabu 2013). In addition, representations of the complex aerosol-cloud and biogeochemical interactions relevant to cloud microphysics and precipitation

formation have only recently been introduced in some of today's climate models (e.g. Gettelman *et al* 2015).

With the advent of the ESMs, climate-biosphere interactions stand at the forefront of current modeling efforts. Hence, there is an acute need to more precisely simulate details of T and P on regional scales, as an essential prerequisite for the prediction of the global carbon cycle, vegetation cover, and related processes. We thus anticipate that Köppen-based evaluations of climate simulations will continue to prove their worth.

Acknowledgments

This work was funded by the U.S. Department of Energy Office of Science and was performed at the Lawrence Livermore National Laboratory under Contract DE-AC52-07Na27344. We also acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the modeling groups (listed in table 2 of this paper) for producing and making available their simulation outputs. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison (PCMDI) provides coordinating support and leads the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Finally, we gratefully acknowledge the generous computational assistance provided by Charles Doutriaux.

References

- Adler R *et al* 2003 The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present) *J. Hydrometeorol.* **4** 1147–67
- Brovkin V, Ganopolski A and Svirezhev Y 1997 A continuous climate-vegetation classification for use in climate-biosphere studies *Ecol. Model.* **101** 251–61
- Catto J L, Jakob C and Nicholls N 2013 A global evaluation of fronts and precipitation in the ACCESS model *Aust. Meteorol. Oceanogr. Soc. J.* **63** 191–203

- Chen D and Chen H W 2013 Using Köppen classification to quantify climate variation and change: an example for 1901–2010 *Environ. Dev.* **6** 69–79
- Cohen J 1960 A coefficient of agreement for nominal scales *Educ. Psychol. Meas.* **20** 37–46
- de Castro M, Gallardo C, Jylhä K and Tuomenvirta H 2007 The use of climate-type classification for assessing climate change effects in Europe from an ensemble of nine regional climate models *Clim. Change* **81** 329–41
- Diaz H F and Eischeid J K 2007 Disappearing ‘alpine tundra’ Köppen climatic type in the western United States *Geophys. Res. Lett.* **34** L18707
- Elguindi N, Grundstein A, Bernardes S, Turuncoglu U and Feddema J 2014 Assessment of CMIP5 global model simulations and climate change projections for the 21st century using a modified Thornthwaite climate classification *Clim. Change* **122** 523–38
- Feddema J J 2005 A revised Thornthwaite-type global classification *Phys. Geogr.* **26** 442–66
- Feng S, Hu Q, Huang W, Ho C-H, Li R and Tang Z 2014 Projected climate regime shift under future global warming from multi-model, multi-scenario CMIP5 simulations *Glob. Plan. Change* **112** 41–52
- Feng S, Ho C-H, Hu Q, Oglesby R J, Jeong S-J and Kim B-M 2012 Evaluating observed and projected future climate changes for the Arctic using the Köppen–Trewartha climate classification *Clim. Dyn.* **38** 1359–73
- Gao X and Giorgi F 2008 Increased aridity in the Mediterranean region under greenhouse gas forcing estimated from high resolution simulations with a regional climate model *Glob. Planet. Change* **62** 195–209
- Gallardo C, Gil V, Hagel E, Tejada C and de Castro M 2013 Assessment of climate change in Europe from an ensemble of regional climate models by the use of Köppen–Trewartha classification *Int. J. Climatol.* **33** 2157–66
- Garcia R, Cabeza M, Rahbek C and Araujo M B 2014 Multiple dimensions of climate change and their implications for biodiversity *Science* **344** 1247579
- Gerstengarbe F W and Werner P C 2009 A short update on Köppen climate shifts in Europe between 1901 and 2003 *Clim. Change* **92** 99–107
- Gottelman A, Morrison H, Santos S, Bogenschütz P and Caldwell P M 2015 Advanced two-moment bulk microphysics for global models: II. Global model solutions and aerosol–cloud interactions. *J. Clim.* **28** 1288–1307
- Guetter P and Kutzbach J E 1990 A modified Köppen classification applied to model simulations of glacial and interglacial climates *Clim. Change* **16** 193–215
- Gnanadesikan A and Stouffer R J 2006 Diagnosing atmosphere–ocean general circulation model errors relevant to the terrestrial biosphere using the Köppen climate classification *Geophys. Res. Lett.* **33** L22701
- Hanf F, Körper J, Spanghel T and Cubasch U 2012 Shifts of climate zones in multi-model climate change experiments using the Köppen climate classification *Meteorol. Z.* **21** 111–23
- Hargrove W W and Hoffman F M 1999 Using multivariate clustering to characterize ecoregion borders *Comput. Sci. Eng.* **1** 18–25
- Hirota N and Takayabu Y N 2013 Reproducibility of precipitation distribution over the tropical oceans in CMIP5 multi-climate models compared to CMIP3 *Clim. Dyn.* **41** 2909–20
- Hoffmann F M, Hargrove W W, Erickson D J and Oglesby R J 2005 Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models *Earth Interact.* **9** 1–27
- Holdridge L R 1947 Determination of world formations from simple climatic data *Science* **105** 127–53
- IPCC 2007 *Climate Change 2007: The Physical Science Basis* ed S Solomon *et al* (Cambridge, UK: Cambridge University Press) p 996
- IPCC 2013 *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed T F Stocker *et al* (Cambridge, UK: Cambridge University Press) p 1535
- Jiang X, Rauscher S A, Ringler T D, Lawrence D M, Williams A P, Allen C D, Steiner A L, Cai D M and McDowell N G 2013 Projected future changes in vegetation in Western North America in the twenty-first century *J. Clim.* **26** 3671–87
- Jolly W M, Nemani R and Running S W 2006 A generalized bioclimatic index to predict foliar phenology *Glob. Change Biol.* **11** 619–32
- Jylhä K, Tuomenvirta H, Ruosteenoja K, Niemi-Hugaerts H, Keisu K and Karhu J A 2010 Observed and projected future shifts of climatic zones in Europe, and their use to visualize climate change information *Weather Clim. Soc.* **2** 148–67
- Kalnay E *et al* 1996 The NCEP/NCAR 40-year reanalysis project *Bull. Am. Meteorol. Soc.* **77** 437–71
- Köppen W 1900 Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt *Geogr. Z.* **6** 657–79
- Köppen W and Geiger R 1930 *Handbuch der Klimatologie* (Berlin: Gebrüder Bornträger)
- Kottek M J, Grieser C, Beck B, Rudolf and Rubel F 2006 World map of the Köppen–Geiger climate classification updated *Meteorol. Z.* **15** 259–63
- Lamb H H 1972 Climate: present, past and future *Fundamental and Climate Now* vol 1 (New York: Methuen) p 613
- Leemans R, Cramer W and van Minnen J G 1996 Prediction of global biome distribution using bioclimatic equilibrium models. *Effects of Global Change on Coniferous Forests and Grassland* ed J M Melillo and A Breymeyer (New York: Wiley) pp 413–50
- Lohmann U, Sausen R, Bengtsson L, Cubasch U, Perlwitz J and Roeckner E 1993 The Köppen climate classification as a diagnostic tool for general circulation models *Clim. Res.* **3** 177–93
- Mahlstein I, Daniel J S and Solomon S 2013 Pace of shifts in climate regions increases with global temperature *Nat. Clim. Change* **3** 739–43
- Manabe S and Holloway J L 1975 The seasonal variation of the hydrologic cycle as simulated by a global model of the atmosphere *J. Geophys. Res.* **80** 1617–49
- Meehl G A, Covey C, Delworth T L, Latif M, McAvaney B, Mitchell J F B, Stouffer R J and Taylor K E 2007 The WCRP CMIP3 multimodel dataset: a new era in climate change research *Bull. Amer. Meteor. Soc.* **88** 1383–94
- Meehl G A and Bony S 2011 Introduction to CMIP5 CLIVAR *Exchanges* **16** 4–5
- Monserud R A 1990 *Methods for Comparing Global Vegetation Maps* (Laxenberg, Austria: IIASA) WP-90-40
- Peel M C, Finlayson B L and McMahon T A 2007 Updated world map of the Köppen–Geiger climate classification *Hydrol. Earth Syst. Sci.* **11** 1633–44
- Prentice K C 1990 Bioclimatic distribution of vegetation for general circulation model studies *J. Geophys. Res.* **95** 830
- Prentice I C, Cramer W, Harrison S P, Leemans R, Monserud R A and Solomon A M 1992 A global biome model based on plant physiology and dominance, soil properties and climate *J. Biogeogr.* **19** 117–34
- Qian Y, Ghan S J and Leung L R 2009 Downscaling hydroclimatic changes over the Western US based on CAM subgrid scheme and WRF regional climate simulations *Int. J. Climatol.* **30** 675–93
- Roderfeld H E, Blyth R, Dankers G, Huse D, Slagstad I, Ellingsen A, Wolf and Lange M A 2008 Potential impact of climate change on ecosystems in the Barents Sea Region *Clim. Change* **87** 283–303
- Rubel F and Kottek M 2010 Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen–Geiger climate classification *Meteorol. Z.* **19** 135–41
- Stephens G L, Ecuver T L, Forbes R, Gettleman A, Golaz J-C, Bodas-Salcedo A, Suzuki K, Gabriel P and Haynes J 2010 Dreary state of precipitation in global models *J. Geophys. Res.* **115** D24211

- Taylor K E, Stouffer R J and Meehl G A 2012 An overview of CMIP5 and the experiment design *Bull. Am. Meteorol. Soc.* **93** 485–98
- Thomas C D *et al* 2004 Extinction risk from climate change *Nature* **427** 145–8
- Thornthwaite C W 1948 An approach to the rational classification of climate *Geogr. Rev.* **38** 55–94
- Thuiller W, Lavorel S, Araújo M B, Sykes M T and Prentice I C 2005 Climate change threats to plant diversity in Europe *Proc. Natl Acad. Sci.* **120** 8245–50
- Trewartha G T 1968 *An Introduction to Climate* 5th edn (New York: McGraw-Hill) p 437
- Wang M and Overland J E 2004 Detecting arctic climate change using Köppen climate classification *Clim. Change* **67** 43–62
- Williams J W, Jackson S T and Kutzbach J E 2007 Projected distributions of novel and disappearing climates by 2100 AD *Proc. Natl Acad. Sci.* **104** 5738–42
- Xu L *et al* 2013 Temperature and vegetation seasonality diminishment over northern lands *Nat. Clim. Change* **3** 581–6