



Evaluation of CMIP5 present climate simulations using the Köppen-Trewartha climate classification

Michal Belda^{1,*}, Eva Holtanová¹, Tomáš Halenka¹, Jaroslava Kalvová¹,
Zdeněk Hlávka²

¹Charles University in Prague, Dept. of Atmospheric Physics, 182 00 Prague, Czech Republic

²Charles University in Prague, Dept. of Probability and Mathematical Statistics, 186 75 Prague 8, Czech Republic

ABSTRACT: We used the Köppen-Trewartha classification on the CMIP5 family of global climate model (GCM) simulations and global Climatic Research Unit (CRU) data for comparison. This evaluation provides preliminary insight on GCM performance and errors. For the overall model intercomparison and evaluation, we used 2 simple statistical characteristics: normalized error, which assesses the total relative difference of the area classified by the individual model with respect to the area resulting from CRU data, and overlap, calculating relative area of matching grid boxes in model results and CRU data. With the additional analysis of the classification on world maps, we show that there are some common features in the model results. Many models have problems capturing the rainforest climate type *Ar*, mainly in Amazonia. The desert climate type *BW* is underestimated by as many as half of the models, with Australia being a typical example of a region where the *BW* is not well represented. The boreal climate type *E* is overestimated by many models, mostly spreading over to the areas of observed tundra type *Ft*. All applied metrics indicate that with the current generation of GCMs, there is no clear tendency for models to improve the representation of climate types with higher spatial resolution.

KEY WORDS: Köppen-Trewartha · Köppen · Climate classification · CMIP5 · Global climate model · CRU TS 3.22 dataset · Overlap measure of similarity · Hierarchical cluster analysis

Resale or republication not permitted without written consent of the publisher

1. INTRODUCTION

Recently, a new coordinated set of experiments with global climate models (GCMs) was performed in the framework of the fifth phase of the Coupled Model Intercomparison Project (CMIP5, see Taylor et al. 2012 for project description). The new generation of CMIP5 GCMs differs from the previous (AR4 GCMs, Meehl et al. 2007) in many aspects. The horizontal resolution is finer in most cases, and more physical processes are included in the models. Moreover, in some GCMs, biogeochemical processes are considered in detail and therefore the simulation of a full carbon cycle is possible. Such GCMs are usually called earth system models (ESMs). Examples of

ESMs include MIROC-ESM-CHEM, MPI-ESM-LR, and GFDL-ESM2G (for more details see Section 2 and Table 1). Further progress in comparison to AR4 GCMs is that changes in land use are taken into account in historical simulations of the 20th century (Taylor et al. 2012).

Before future climate changes are estimated based on the outputs of GCMs, it is always necessary to evaluate the model performance in simulating observed climate in a reference period. For such tasks, climate classifications can provide an interesting complement to the commonly used evaluation methods. Climate classifications depict a fairly broad range of climate features in 1 simple characteristic. Most often, the annual and monthly mean values of

*Corresponding author: michal.belda@mff.cuni.cz

air temperature and precipitation and their annual cycle are summarized in the form of climate types and subtypes. For analysis of GCM outputs, the classifications that correspond to vegetation distribution are usually incorporated. One of the most frequently used classifications of this kind was developed by Köppen (1923, 1931, 1936) and modified by various authors (e.g. Geiger 1954, Trewartha & Horn 1980; see Belda et al. 2014 for comparison).

Probably the first paper dealing with the application of a Köppen climate classification (KCC) for the validation of GCM outputs was by Manabe & Holloway (1975), who evaluated a simulation of the GFDL atmospheric general circulation model with a horizontal resolution of about 265 km and 11 vertical levels. Frequently cited work incorporating the Köppen classification is that of Lohmann et al. (1993), who diagnosed the outputs of ECHAM3 (Roeckner et al. 1992) at resolutions T21 and T42. Kalvová et al. (2003) applied the Köppen classification to the simulations of 4 state of the art GCMs, viz. HadCM2 (Mitchell & Johns 1997), ECHAM4 (Roeckner et al. 1996), CSIRO-Mk2b (Hirst et al. 2000), and CGCM1 (Boer et al. 2000). Peel et al. (2007) created digital global maps of Köppen-Geiger climate (KGC) types (Geiger 1954) based on a large amount of long-term monthly precipitation and temperature station data. They also discussed the geographical distribution of KGC types over individual continents as well as the influence of observed temperature trends on results of the classification.

More recently, Rubel & Kottek (2010) created a series of digital world maps of Köppen climate types for the period 1901–2100. These maps are based on Climatic Research Unit (CRU) TS2.1 (Mitchell & Jones 2005), GPCC Version 4 data (<http://gpcc.dwd.de/>) and GCM outputs from the TYN SC 2.0 dataset (Mitchell et al. 2004).

Another broadly used climate classification is the modification of the Köppen system proposed by Trewartha (Trewartha 1968, Trewartha & Horn 1980). They modified original Köppen thermal and dryness criteria to make the climate types better correspond with the observed boundaries of natural landscapes and to eliminate certain vagueness of the KCC definitions. This classification, sometimes referred to as Köppen-Trewartha or K-T classification (denoted as KTC in the following text), provides a more detailed description of climate type distributions than the original Köppen scheme, modifying the thresholds between types and adding a further level of division of types *E* and *F* (see Belda et al. 2014 for more details and a comparison of KCC and KTC). The KTC

scheme was used by e.g. Feng et al. (2012) to analyze observed and projected climate changes and their impact on vegetation for the area north of 50° N over the period 1900–2099. They focused on the Arctic region, and in addition to the observed data, they incorporated the simulations of 16 AR4 GCMs under SRES scenarios B1, A1B, and A2. Another example of the application of KTC on the outputs of GCMs is the study by Guetter & Kutzbach (1990), who investigated the last interglacial and glacial climates (126 000 and 18 000 yr before present) using simulations of an atmospheric general circulation model.

In studies using rule-based classifications, such as KCC or KTC, it is sometimes not obvious which rules were applied first. For example, according to Kottek et al. (2006), the KTC rules determining polar climate class must be applied first, then those for arid climate, and finally those for tropical, temperate, and cold classes. In KTC, the first step selects the polar areas *F*, and dry climate class *B* is evaluated subsequently, i.e. the *B* type cuts across all climate groups except for polar climate *F* (Trewartha & Horn 1980).

A different approach to climate classification is represented by methods based solely on statistical techniques, most often cluster analysis (e.g. Fovell & Fovell 1993, Stern et al. 2000). A disadvantage of such procedures is that the physical interpretation of the results might be difficult. Alternatively, Cannon (2012) used a rule-based clustering algorithm, multivariate regression tree (MRT). He applied the MRT method on long-term monthly mean temperature and precipitation amounts and concluded that the MRT performed significantly better than the Köppen-Geiger classification, which may be suboptimal for applications that are sensitive to spatial variations in precipitation.

For GCM validation, however, the rule-based classification schemes are still preferred, mainly because they are simple, transparent, and easy to interpret, and resulting climate types are related to vegetation distribution, which is important for estimates of climate change impacts.

Most recently, analyses of future climate by means of KTC from the latest generation GCMs were performed by e.g. Mahlstein et al. (2013) and Feng et al. (2014) on a subset of CMIP5 models. However, to our knowledge, no comprehensive study of model validation for the present climate has been conducted using climate classification. Our interest is the evaluation of the new generation GCMs with regard to geographical differences in model skill and inter-comparison of the models. The motivation for such in-depth analysis is manifold. First, one can derive

from it to which extent the latest generation of GCMs improves climate simulations. Second, it lays the groundwork for further assessment of future climate projections by the ensemble of CMIP5 models. Third, and not least important for our application, is the geographical distribution of model errors, which can provide preliminary information to the regional climate modeling communities as to which models may be better suited for driving the regional models in individual regions.

Here we used the KTC to evaluate the performance of CMIP5 GCMs in simulating the climate of the reference period 1961–1990. The KTC scheme used is described only briefly in this text; for a full description see Belda et al. (2014). Additional maps and other graphical materials can be found on a supplementary website at <http://kfa.mff.cuni.cz/projects/trewartha>.

2. DATA

The CRU TS 3.22 dataset (Harris et al. 2014) was used as reference data representing the observed state of the climate in the reference period 1961–1990. We used monthly mean surface air temperature and precipitation to derive the Köppen-Trewartha climate types (see Section 3 for details). Our analysis covers global land areas (excluding Antarctica) for which the CRU TS 3.22 is available at a high spatial resolution of $0.5^\circ \times 0.5^\circ$.

A list of CMIP5 GCM simulations involved in the present study and some basic information about them are provided in Table 1. The data are freely available for non-commercial purposes at <http://cmip-pcmdi.llnl.gov/cmip5/>. We used the outputs from the experiment denoted as 'historical'. These model runs were forced by observed atmospheric composition changes and time-evolving land cover during the period from the mid-19th century to the near present (Taylor et al. 2012). For ESMs with a carbon cycle, the carbon dioxide concentrations are prescribed in this experiment. Where more ensemble members were available, we chose the ensemble member 'r1p1i1'.

In some cases, runs from several model versions are available and therefore it is possible to analyze the uncertainty coming from 1 or more aspects in model formulation when these runs are compared. For example, GFDL-ESM2G and GFDL-ESM2M have the same atmosphere component but different formulation of the ocean components (www.gfdl.noaa.gov/cm2m-and-cm2g). Similarly, GISS-E2-H

and GISS-E2-R differ only in the model of the ocean (<http://data.giss.nasa.gov/modelE/ar5/>). IPSL-CM5A and IPSL-CM5B use different physical parameterizations in the atmospheric models. 'LR' and 'MR' in the abbreviations of these model runs relate to the spatial resolution of the atmospheric component (<http://icmc.ipsl.fr/index.php/icmc-models-2/icmc-ipsl-cm5>). MIROC-ESM-CHEM is the same model as MIROC-ESM, except with an additional interactive model of tropospheric and stratospheric chemistry (Watanabe et al. 2011).

3. METHODS

For validation of CMIP5 GCMs, we applied the Köppen-Trewartha climate classification (for further details about KTC see Trewartha & Horn 1980 and Belda et al. 2014) in the present study. The KTC system has 6 main climate groups. Five of them (denoted *A*, *C*, *D*, *E*, and *F*) are basic thermal zones, and the sixth group (*B*) is the dry climatic zone that cuts across the other climate types, except for the polar climate *F*. The main climate types are, similarly to the Köppen classification scheme, determined according to long-term annual and monthly means of surface air temperature and precipitation amounts. In Belda et al. (2014) and in the present study, we used Patton's formula (Patton 1962, Trewartha & Horn 1980) for dryness criteria, expressed as:

$$R = 2.3T - 0.64Pw + 41 \quad (1)$$

where R denotes the mean annual precipitation threshold in cm, T is the mean annual air temperature in $^\circ\text{C}$, and Pw is the percentage of annual precipitation concentrated in winter (October to March in the northern hemisphere and April to September in the southern Hemisphere). If the mean annual precipitation amount P_{mean} is lower than R and higher than $0.5R$, type *BS* is defined in such a grid box. Where the mean annual precipitation P_{mean} is lower than $0.5R$, type *BW* is defined. A brief description of climate types and subtypes is provided in Table 2.

The evaluation of the GCM performances in simulating the Köppen-Trewartha climatic types is complicated due to differences in model horizontal resolution. The KTC types were first calculated for both CRU TS 3.22 and CMIP5 GCMs in their original grids, i.e. for each grid point, and based on this information we calculated land areas falling into each climatic subtype (expressed in terms of relative areas, percentage of total global land area except Antarctica). Furthermore, the maps of KTC climatic types for all

Table 1. CMIP5 global climate models analyzed in this paper

CMIP5 model	Resolution	Modeling center
ACCESS1.3	1.88° × 1.24°	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia
BCC-CSM1.1	2.8° × 2.8°	Beijing Climate Center, China Meteorological Administration
BCC-CSM1.1m	1.13° × 1.13°	
CanESM2	2.8° × 2.8°	Canadian Centre for Climate Modelling and Analysis
CCSM4	1.25° × 0.94°	National Center for Atmospheric Research
CESM1-BGC	1.25° × 0.94°	Community Earth System Model Contributors
CESM1-CAM5	1.25° × 0.94°	
CESM1-CAM5.1-FV2	2.50° × 1.88°	
CESM1-FASTCHEM	1.25° × 0.94°	
CESM1-WACCM	2.50° × 1.88°	
CMCC-CESM	3.75° × 3.75°	Centro Euro-Mediterraneo sui Cambiamenti Climatici
CMCC-CM	0.75° × 0.75°	
CNRM-CM5	1.4° × 1.4°	Centre National de Recherches Meteorologiques; Centre Européen de Recherche et Formation Avancées en Calcul Scientifique
CNRM-CM5.2		
CSIRO-Mk3.6.0	1.9° × 1.9°	Commonwealth Scientific and Industrial Research Organisation; Queensland Climate Change Centre of Excellence
EC-EARTH	1.1° × 1.1°	EC-EARTH consortium
FGOALS-g2	2.81° × 3.00°	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University
GFDL-CM2p1		Geophysical Fluid Dynamics Laboratory
GFDL-CM3	2.5° × 2°	
GFDL-ESM2G	2.5° × 2°	
GFDL-ESM2M	2.5° × 2°	
GISS-E2-H	2.5° × 2°	NASA Goddard Institute for Space Studies
GISS-E2-H-CC	2.5° × 2°	
GISS-E2-R	2.5° × 2°	
GISS-E2-R-CC	2.5° × 2°	
HadCM3	3.75° × 2.5°	Met Office Hadley Centre
HadGEM2-AO		
HadGEM2-CC	1.875° × 1.25°	
HadGEM2-ES	1.875° × 1.25°	
INM-CM4	2° × 1.5°	Institute for Numerical Mathematics
IPSL-CM5A-LR	3.75° × 1.9°	Institut Pierre-Simon Laplace
IPSL-CM5A-MR	2.5° × 1.3°	
IPSL-CM5B-LR	3.75° × 1.9°	
MIROC5	1.4° × 1.4°	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology
MIROC-ESM	2.8° × 2.8°	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies
MIROC-ESM-CHEM	2.8° × 2.8°	
MPI-ESM-LR	1.9° × 1.9°	Max Planck Institute for Meteorology
MPI-ESM-MR	1.9° × 1.9°	
MPI-ESM-P	1.9° × 1.9°	
MRI-CGCM3	1.125° × 1.125°	Meteorological Research Institute
MRI-ESM1	1.13° × 1.13°	
NorESM1-M	2.5° × 1.9°	Norwegian Climate Centre
NorESM1-ME	2.5° × 1.9°	

Table 2. Definition of Köppen-Trewartha classification climate types according to Trewartha & Horn (1980), with a dryness threshold defined by Patton (1962). T_{mo} : long-term monthly mean air temperature; T_{cold} (T_{warm}): monthly mean air temperature of the coldest (warmest) month; P_{mean} : mean annual precipitation (cm); P_{dry} : monthly precipitation of the driest summer month; R : Patton's precipitation threshold, defined as $R = 2.3T - 0.64P_w + 41$, where T is mean annual temperature ($^{\circ}\text{C}$), and P_w is the percentage of annual precipitation occurring in winter

Type/ subtype	Criteria Rainfall/temperature regime
A	$T_{cold} > 18^{\circ}\text{C}$; $P_{mean} > R$
<i>Ar</i>	10 to 12 mo wet; 0 to 2 mo dry
<i>Aw</i>	Winter (low-sun period) dry; >2 mo dry
<i>As</i>	Summer (high-sun period) dry; rare in type A climates
B	$P_{mean} < R$
<i>BS</i>	$R/2 < P_{mean} < R$
<i>BW</i>	$P_{mean} < R/2$
C	$T_{cold} < 18^{\circ}\text{C}$; 8 to 12 mo with $T_{mo} > 10^{\circ}\text{C}$
<i>Cs</i>	Summer dry; at least 3 times as much rain in winter half year as in summer half-year; $P_{dry} < 3$ cm; total annual precipitation < 89 cm
<i>Cw</i>	Winter dry; at least 10 times as much rain in summer half-year as in winter half-year
<i>Cf</i>	No dry season; difference between driest and wettest month less than required for <i>Cs</i> and <i>Cw</i> ; $P_{dry} > 3$ cm
D	4 to 7 mo with $T_{mo} > 10^{\circ}\text{C}$
<i>Do</i>	$T_{cold} > 0^{\circ}\text{C}$
<i>Dc</i>	$T_{cold} < 0^{\circ}\text{C}$
E	1 to 3 mo with $T_{mo} > 10^{\circ}\text{C}$
F	All months with $T_{mo} < 10^{\circ}\text{C}$
<i>Ft</i>	$T_{warm} > 0^{\circ}\text{C}$
<i>Fi</i>	$T_{warm} < 0^{\circ}\text{C}$

CMIP5 models listed in Table 1 were created and can be found at <http://kfa.mff.cuni.cz/projects/trewartha>.

For the overall model intercomparison and evaluation, we used a simple statistical characteristic normalized error (NE), defined as:

$$\text{NE} = \sum_{i=1}^K \frac{|m_i - o_i|}{o} \quad (2)$$

where K is the number of climatic types (13 for KTC incorporated in this study), m_i is the continental area for type i in model m , o_i is the corresponding area according to CRU, and o is the global continental area (according to CRU and the respective models).

The NE was calculated based on both original and re-gridded model data (interpolated to the 0.5° regular grid used in CRU). Although the actual NE values are obviously not equal, the model rank based on NE does not change with re-gridding. In the following analysis, we used NE calculated from the original model data (i.e. not interpolated), to avoid introducing inconsistencies from an arbitrary choice of interpolation method.

To compare spatial representation of the model data, a very simple similarity measure, also known as the overlap, was used. For each model, we calculated the total area of grid points where the model-simulated climate type was the same as the CRU one. The models can be sorted in terms of this matched area, or rather, relatively with respect to the overall continent area providing an overlap characteristic (Table 3). Unlike the NE measure, this method requires interpolating model data to a common grid (in our case a CRU grid) before calculating climate types. The interpolation into the CRU grid was performed using a simple bilinear interpolation method. There is, however, an obvious drawback to such an overlap measure. Clearly, it only takes matching grid points into account without accounting for mismatches, i.e. events when the climate model gives different climate types than CRU (for example, GCM simulates climate type *BS* while CRU gives type *BW*). We did not attempt to correct this problem, since there are many possible ways to assign weights for individual type mismatches. Analyzing these similarity measures in more detail (see e.g. Boriah et al. 2008) is beyond the scope of this paper; therefore, we used no penalization of mismatches during further assessment.

Furthermore, the similarity between any 2 CMIP5 models expressed as overlap can be used for hierarchical cluster analysis. A matrix of pairwise distances is obtained by subtracting the similarity from 100%, i.e. the distance between 2 models is defined as a percentage of the total Earth surface (excluding Antarctica) where the climate types do not agree. The resulting matrix of similarities is visualized by hierarchical cluster analysis. The statistical analysis was performed in the statistical computing environment R (R Core Team 2012). The Ward algorithm (R, library MASS, function `hclust()`) was used to create a dendrogram (see Fig. 2).

All mentioned metrics were calculated for individual models as well as for the ensemble mean, calculated by applying the KTC classification to the ensemble average of temperatures and precipitation. The multi-model mean values are denoted as CMIP5 ENS in the included tables and figures.

Table 3. Model to observation statistics calculated for 1961–1990. CMIP5 ENS: ensemble mean; NE: normalized error of Köppen-Trewartha classification climate type areas for CMIP5 GCMs vs. CRU TS 3.22; NE rank: ranking of the models based on normalized error; Overlap: percentage of continental area, excluding Antarctica, where the model and CRU climate types agree; Overlap rank: ranking of the models based on overlap

Model	NE	NE rank	Overlap (%)	Overlap rank
CMIP5 ENS	0.189	19	70.07	1
ACCESS1.3	0.142	5	66.46	12
BCC-CSM1.1-m	0.111	1	59.21	37
BCC-CSM1.1	0.156	8	62.78	25
CanESM2	0.184	16	65.01	17
CCSM4	0.281	37	63.28	22
CESM1-BGC	0.286	40	63.34	21
CESM1-CAM5	0.253	30	65.29	16
CESM1-CAM5.1-FV2	0.322	43	61.83	29
CESM1-FASTCHEM	0.289	41	63.16	23
CESM1-WACCM	0.298	42	59.18	38
CMCC-CESM	0.174	13	63.41	20
CMCC-CM	0.157	10	66.18	13
CNRM-CM5	0.208	22	67.60	7
CNRM-CM5.2	0.196	20	67.05	10
CSIRO-Mk3.6.0	0.226	23	59.60	35
EC-EARTH	0.236	26	66.13	14
FGOALS-g2	0.257	31	56.50	44
GFDL-CM2p1	0.228	24	59.73	34
GFDL-CM3	0.238	27	63.84	18
GFDL-ESM2G	0.274	34	57.46	41
GFDL-ESM2M	0.235	25	59.41	36
GISS-E2-H	0.158	11	61.51	30
GISS-E2-H-CC	0.156	9	61.26	31
GISS-E2-R	0.153	7	62.36	26
GISS-E2-R-CC	0.144	6	62.18	27
HadCM3	0.172	12	60.82	33
HadGEM2-AO	0.184	17	67.72	6
HadGEM2-CC	0.207	21	67.35	8
HadGEM2-ES	0.177	14	68.94	3
INM-CM4	0.325	44	61.95	28
IPSL-CM5A-LR	0.277	36	63.83	19
IPSL-CM5A-MR	0.251	29	65.73	15
IPSL-CM5B-LR	0.250	28	61.14	32
MIROC-ESM	0.283	39	56.59	43
MIROC-ESM-CHEM	0.271	33	56.83	42
MIROC5	0.281	38	62.98	24
MPI-ESM-LR	0.137	4	67.82	5
MPI-ESM-MR	0.115	2	68.26	4
MPI-ESM-P	0.127	3	68.97	2
MRI-CGCM3	0.187	18	66.82	11
MRI-ESM1	0.180	15	67.20	9
NorESM1-M	0.260	32	58.12	39
NorESM1-ME	0.275	35	57.47	40

4. KTC TYPES FROM CMIP5 GCMS FOR THE PERIOD 1961–1990

As a first look at the results, we show the world maps of classification calculated from CRU TS 3.22 and CMIP5 simulation results (Fig. S1 in the Supplement, www.int-res.com/articles/suppl/c064p201_supp.pdf). The percentages of continental areas (excluding Antarctica) covered by KTC climate types according to CRU TS 3.22 and CMIP5 GCMs are summarized in Fig. 1. In these figures, we present the results based on data in original model grids. The results for re-gridded data do not show significant differences and are presented at the accompanying website (<http://kfa.mff.cuni.cz/projects/trewartha>).

The most frequent KTC type in CRU TS 3.22 is the desert climate *BW*, which covers 19% of the investigated continental area. Twenty out of 43 studied GCMs also give *BW* as the most frequent KTC type. On the other hand, 11 (7) GCMs produce type *E* (*Aw*) as the most abundant. Both versions of NorESM1 give type *Cf* with the highest coverage, and CESM1-FASTCHEM, CESM1-BGC, and HadGEM2-AO produce type *Dc* as the most frequent type.

In addition to *BW*, there are 4 KTC types with more than 10% continental coverage in the observed data: *Aw*, *BS*, *Dc*, and *E*. A similar result is seen for almost all GCMs and the ensemble average. In a few cases, type *Aw* is underestimated. According to 12 GCMs (e.g. most of the CESM1 simulations), type *BS* covers less than 10% of the continental area. Based on the CRU dataset, type *Cf* covers 8.1% of continental area, but approximately half of the GCMs give more than 10% continental coverage for this type. This overestimation results mainly from too large a simulated *Cf* area in southern Africa, South and Central America, northern India, and Australia (Fig. S1).

We will now comment on the skill of the GCMs at simulating the geographical distribution of the most widespread climatic types. About half of the GCMs underestimate the area covered by *BW*. However, as evident in Fig. S1, it does not appear possible to depict one problematic region that is common to all models. For example, all versions of NorESM1, CESM1, FGOALS-g2, and MIROC do not simulate *BW* over Australia, whereas in CRU TS 3.22, most of this continent is covered by this type. On the other hand, INMCM4 represents the distribution of KTC types over Australia correctly, but underestimates *BW* over all other continents. The smallest coverage of *BW* is simulated by FGOALS-g2, which in the southern part of Africa, over Australia, and in central Asia does not simulate any desert at all. The best

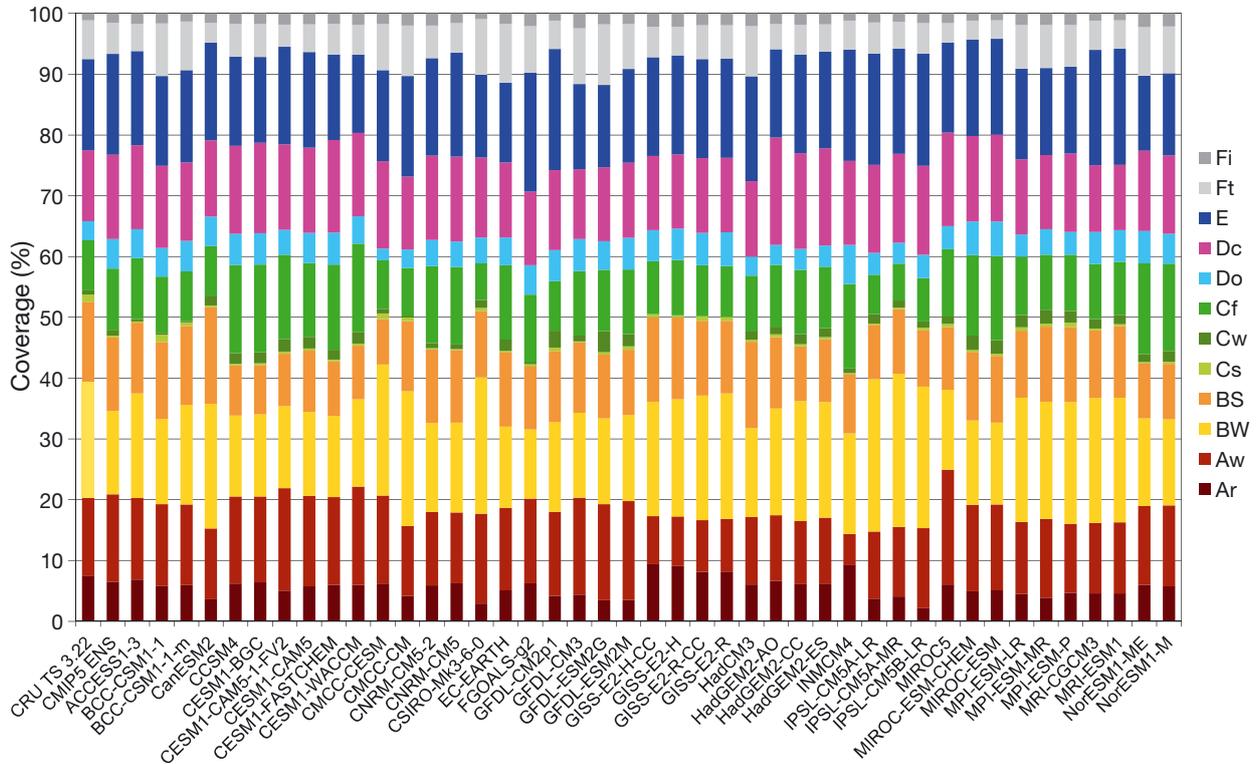


Fig. 1. Percentage of continental area (excluding Antarctica) covered by Köppen-Trewartha classification climate types (see Table 2 for descriptions of climate types) according to CRU TS3.22 and CMIP5 global climate models for the period 1961–1990

results in terms of both total coverage and actual distribution of type *BW* were achieved by MPI-ESM-LR.

Most of the models (39 out of 43 GCMs) underestimate the percentage of global continents covered by *BS*. The exceptions are CanESM2, HadCM3, and GISS-E2-H, GISS-E2-H-CC, which overestimate *BS* mainly over Australia and South America (HadCM3 only over Australia). The area of *BS* is most underestimated by CMCC-CESM, mainly because of a small coverage of this type over North America and Asia. The percentage of continents covered by *BS* is best simulated by GISS-E2-H, and both simulations of BCC-CSM1, but the distribution of this type in these models is not quite correct. They overestimate its area over Australia, while over North America this type is barely present.

Type *Aw* is best simulated by MPI-ESM-MR, IPSL-CM5B-LR, and BCC-CSM1.1m, even though the first 2 overestimate the area covered by *Aw* in South America and underestimate it over Africa and southern Asia. Regarding other GCMs, there is no general tendency to overestimate or underestimate the percentage of continental area belonging to *Aw*. MIROC5 overestimates this type quite strongly (by 6 percentage points, i.e. 6% of the continental area), mostly over Africa, where it pushes the borders of this type

farther from the equator in both directions, and over South America, where it replaces type *Ar* with *Aw*. On the other hand, the model INMCM4 only has 5% of continental areas covered by *Aw*, in comparison to 12.9% in CRU TS 3.22. In equatorial areas over Africa and South America, this model gives *Ar* and *Cf* over areas belonging to *Aw* in the observed data.

Type *Dc* is overestimated by the CMIP5 GCMs. The only exceptions are both MRI simulations and GFDL-CM3, which underestimate the *Dc* coverage, but by less than 1 percentage point. In CRU TS 3.22, *Dc* is found mainly over Europe, central Asia and North America, and these areas are depicted quite well by most of the models. The strongest overestimation of more than 3 percentage points is seen in the HadGEM2 simulations and MIROC5. These models simulate *Dc* over areas that belong to *BS* or *E* in the observed data.

The GCMs tend to overestimate the area covered by type *E* (26 GCMs overestimate type *E* area, 17 underestimate it, and the absolute bias is smaller in the case of underestimation); this is most evident in the simulations of GFDL-CM2p1, FGOALS-g2, and both simulations of MRI. These models give type *E* where, according to observed data, *Ft* should be placed mainly over northern Canada, Alaska (USA),

and north-eastern Asia. FGOALS-g2 and MRI even give *E* over western parts of the USA, where *Dc* or *BS* are present according to the observed data. Interestingly, the model CNRM-CM5, which also overestimates the area of *E* by about 2 percentage points, simulates *E* even on the shores of Greenland and in several grid points over the Arctic islands. Most of these problems are not present in the CNRM-CM5.2 simulation. Following Fig. 1, the closest models to observations in terms of the percentage of continents covered by type *E* are CMCC-CESM and MPI-ESM-LR. The geographical distribution is depicted quite well over Europe and Asia in these 2 models (Fig. S1). However, they fail in distinguishing *E* and *Ft* types over Alaska and northern Canada.

Regarding other climate types (besides the above analyzed 5 most abundant), it is worth denoting that most of the GCMs are not capable of characterizing type *Ar*, mainly over South America. Some of them simulate dry climate types *BS* and *BW* over areas belonging to *Ar* in CRU TS 3.22 over this region (e.g. CanESM2, CSIRO-Mk3.6.0, all MPI simulations). The percentage of global continental area covered by *Ar* is underestimated by most GCMs, sometimes by >50%. The exceptions are the GISS simulations and INMCM4, which overestimate the coverage of *Ar* by up to 1.8 percentage points.

When we assess the overall GCM performance according to the NE (Eq. 2, values shown in Table 3), the models closest to observations are BCC-CSM1.1m, ACCESS1.3, and 3 MPI-ESM.

When we use a simple overlap similarity measure, the rank of the models is different, (last column in Table 3). This metric, when subtracted from 100%, can also be used as a dissimilarity measure for hierarchical cluster analysis. The dendrogram (Fig. 2) suggests that the models on the right-hand side are more similar to CRU than models on the left-hand side. CRU seems to be most similar (closest) to CMCC-CM (within 1 cluster). In the next step, we may add the cluster consisting of INMCM4, MPI-ESM-LR, and MRI-CGCM3 and then clusters of IPSL experiments and CNRM GCMs. In most cases, GCMs originating from 1 modeling center have close locations within the dendrogram, regardless of differences in physical parameterizations and spatial resolution, e.g. models MPI, IPSL, and HadGEM2. One such group of models consists of MRI-CGCM3 and MRI-ESM1, where MRI-ESM1 was basically created from MRI-CGCM3 by adding chemical and biogeochemical modules, and therefore the dynamical and thermodynamical processes are entirely the same in both models (Adachi et al. 2013).

Regarding the performance of the multi-model mean (CMIP5 ENS), we found the following main results. Unlike CRU TS 3.22, CMIP5 ENS gives type *E* as the most abundant climate, and continental coverage of *Cf* is larger than 10%. It underestimates relatively strongly the area of *BW* on practically all continents (by more than 5 percentage points overall). On the other hand, CMIP5 ENS overestimates the area of *Dc*, mainly over North America and the Middle East. The overestimation of type *E* areas by CMIP5 ENS is relatively large, and is mainly due to pushing the northern border of type *E* too far over northern Canada and eastern Asia. For other KTC types, the CMIP5 differences from CRU are mostly smaller than 1.5 percentage points. According to the NE, CMIP5 ENS rank is approximately in the middle of all studied GCMs. On the other hand, based on the overlap characteristic, CMIP5 ENS is in the best agreement with CRU.

5. DISCUSSION AND CONCLUSIONS

We evaluated the skill of CMIP5 GCMs (Taylor et al. 2012) based on the ability of GCMs to represent the climatic types according to the Köppen-Trewartha climate classification (Trewartha & Horn 1980, Belda et al. 2014). The distribution of KTC types was analyzed for the 30 yr reference period 1961–1990. In connection to the choice of averaging period, it is necessary to discuss the influence of natural variability. Since an in-depth analysis of the sources of uncertainty is beyond the scope of this study, we reference our previous study (Belda et al. 2014), where the evolution in 30 yr moving averages was analyzed on CRU data as the first step. Observed changes of KTC types during the 20th century inferred from these values are mostly within a few tenths of percentage points, except for *BS*, which is within ± 0.6 percentage points (i.e. changes of 0.6% of the continental area).

However, Deser et al. (2012) showed that within a large model ensemble, the natural variability contributes considerably to the uncertainty of future climate on the multi-decadal scale (in their case, the period 2006–2060). In our CMIP5 study, we examined the influence of internal climate variability on simulated KTC types by analysis of perturbed initial conditions within a 10-member ensemble of CSIRO-Mk3.6.0 simulations. Areas covered by individual KTC types differ by <0.7% of total continental area between individual ensemble members (not shown). Therefore, in our opinion, the uncertainty connected to the choice of a specific reference period is much

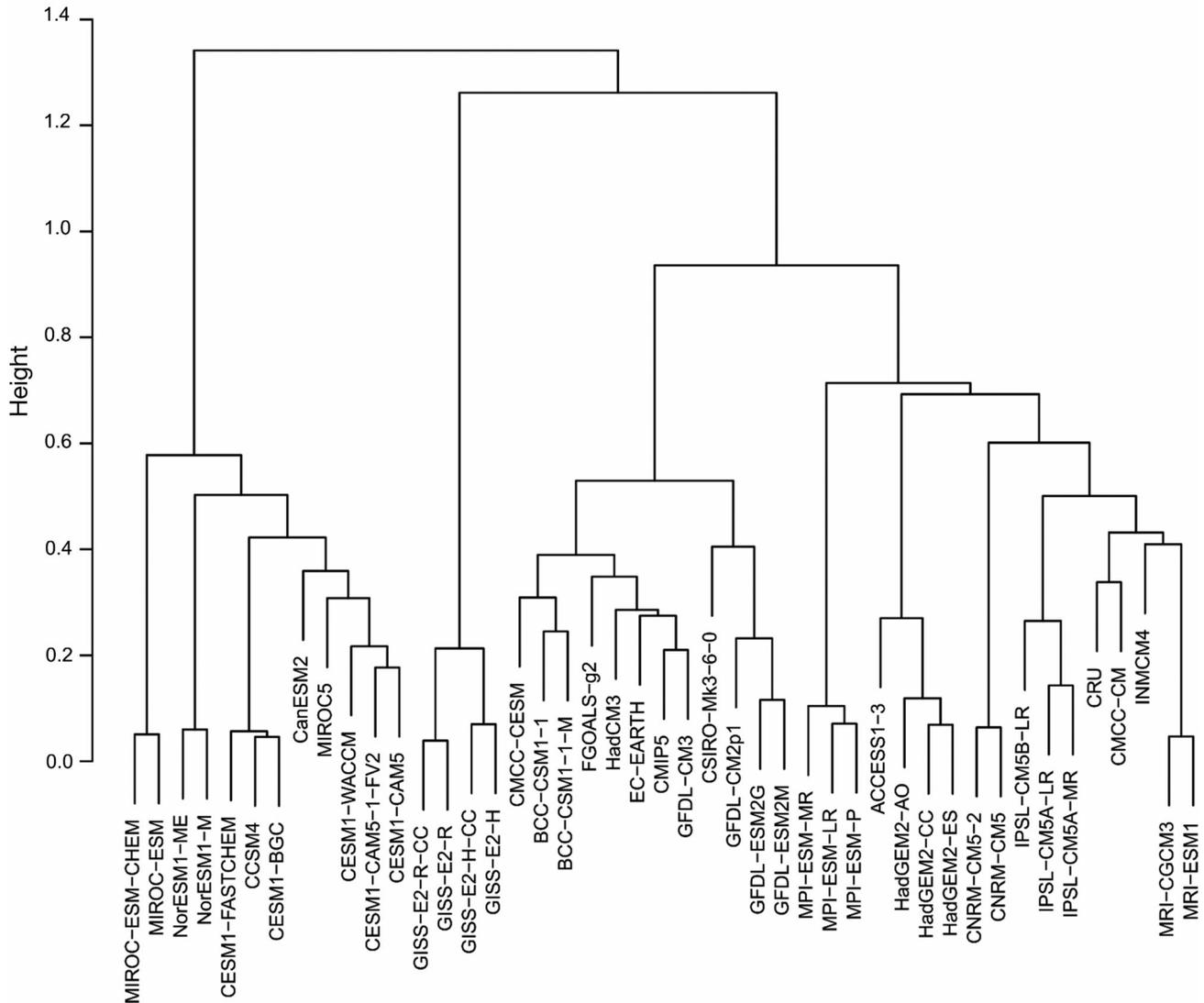


Fig. 2. Cluster dendrogram of model correspondence within the ensemble and with the CRU dataset; the Ward algorithm was used based on overlap measures

lower than model errors; also, the spread of individual models is larger than between different realizations of the selected model CSIRO-Mk3.6.0.

We used several different perspectives of model performance evaluation, ranging from the percentage of continental area (excluding Antarctica) covered by KTC climate types and a simple visual assessment of simulated geographical distribution of climatic types through relative error and overlap, to visualizing the matrix of similarities (the similarity between any 2 climate models is measured by a percentage of overlap of identical climate types) by hierarchical cluster analysis.

The GCMs' performance was assessed according to the percentage of continental areas (excluding Antarctica) covered by KTC climate types. We con-

centrated mainly on the 5 most abundant climate types *BW*, *BS*, *Aw*, *Dc*, and *E*. We can see some general tendencies to overestimate or underestimate areas belonging to individual climate types that seem to be common to most of the GCMs. About half of GCMs underestimate the area covered by *BW* and *Aw*, *BS* is underestimated by most of the models, and *E* and *Dc* are overestimated. However, problematic geographical regions differ between models, e.g. it is not possible to depict a region where most of the models unrealistically simulate the *Aw* type.

In the case of type *BW*, both total coverage and actual distribution are best simulated by MPI-ESM-LR. In terms of area covered by *BS*, models GISS-E2-H and both BCC-CSM1 simulations are closest to the observed state, but the geographical distribution of

this type in these models is not quite realistic. Type *Aw* is best simulated by MPI-ESM-MR, IPSL-CM5B-LR, and BCC-CSM1.1m. Climate type *Dc* occurs mainly over Europe, central Asia, and North America, and most of the models depict these areas quite well. Regarding the total land area covered by type *E*, the closest models to observations are CMCC-CESM and MPI-ESM-LR, but, similarly to *BS*, the geographical distribution of type *E* is not depicted correctly in these 2 models.

Most of the GCMs cannot correctly characterize the equatorial climate over South America, with some of them even giving types *BS* and *BW* over areas belonging to *Ar* in CRU TS 3.22 (e.g. CanESM2, CSIRO-Mk3.6.0, and all MPI simulations). As a whole, the area of tropical rainforest climate *Ar* is underestimated by most CMIP5 GCMs.

The above summarized evaluation of GCMs in terms of the percentage of continental areas (excluding Antarctica) covered by KTC climate types was based on the GCMs' outputs in their original spatial resolution. Calculations with model outputs interpolated into the CRU TS 3.22 grid gave almost identical results. Therefore, we consider the error caused by interpolation negligible.

A GCM can give correct values of global land area fraction belonging to a specific KTC climate type even though the geographical distribution is not well captured by the model. Furthermore, the same GCM can skillfully simulate global land area fractions for some of the climate types but can fail in the case of others.

The evaluation of the models by a specific characteristic depends on what the characteristic depicts, and the results obtained using different metrics can significantly differ. This fact has already been shown in many studies, e.g. Gleckler et al. (2008), Pierce et al. (2009). The agreement of simulated and observed KTC types also depends on the choice of criterion. For example, the model BCC-CSM1.1m is the best model when considering NE of all climate types (Table 3). However, the most widespread type *BW* is not simulated very consistently by this model (Fig. 1); for example, *BW* is absent over Australia according to BCC-CSM1.1m.

A different perspective to the model evaluation is provided by the overlap measure of similarity (Table 3), which is based on the relative area of grid points in which the climate type given by the GCM matches CRU TS 3.22. The best models according to this assessment are MPI-ESM-P, HadGEM2-ES, and MPI-ESM-MR. The cluster dendrogram based on overlap similarity measures showed that the models most similar to CRU TS 3.22 are CMCC-CM,

INMCM4, MPI-ESM-LR, and MRI-CGCM3. The next cluster is formed by IPSL experiments and then by CNRM-CM5 simulations.

The horizontal resolution of GCMs analyzed in our study varies from $0.75^\circ \times 0.75^\circ$ to $3.75^\circ \times 3.75^\circ$. Our results do not indicate any general tendency that GCMs with finer horizontal resolution give better representation of KTC types. For example, we can see in Fig. 2 that in the cluster nearest to CRU TS 3.22 are GCMs with relatively high spatial resolution; however, they were developed in different modeling centers, and it is not clear whether their success is caused by their resolution. On the other hand, model CSIRO-Mk3.6.0 with a relatively high resolution of $1.9^\circ \times 1.9^\circ$ is quite far from that cluster. Both IPSL-CM5A-LR and IPSL-CM5A-MR, which share the same physical parameterizations but differ in their spatial resolution, belong to 1 cluster, indicating a lower effect of resolution on simulated KTC types. The described ambiguous effect of spatial resolution is not particularly surprising because model performance is affected not only by coarse or fine resolution, but also by the parameterizations used, numerical schemes, etc. (e.g. Duffy et al. 2003). However, a clearer benefit of finer resolution might occur if we add another level of subtypes into the classification. This may prove useful especially when applying the classification to regional models rather than GCMs in regions with a much more diverse structure of surface characteristics.

The GCMs coming from the same modeling center are often grouped in the same cluster. A similar result was already described by Knutti et al. (2013). Even though they used a different methodology, their conclusions are quite similar to our results. For example, Knutti et al. (2013) also found that the simulations of CESM1 are close to CCSM4, although most of the major parameterizations were changed going from CCSM4 to CESM1.

According to our results, the multi-model ensemble mean did not outperform all individual GCMs, except for the overlap characteristic. Various previous studies evaluating GCM performances according to different characteristics and metrics indicated superiority of the multi-model mean (e.g. Gleckler et al. 2008, Pierce et al. 2009, Miao et al. 2014). A possible explanation is that unlike usual validation metrics, the climate classification schemes combine various aspects of both air temperature and precipitation fields and therefore no simple cancellation of errors can be expected.

Further analysis of the reasons why the CMIP5 GCMs have problems with simulation of the ob-

served distribution of some KTC types in some regions for the reference period would require a detailed analysis of simulated air temperature and precipitation fields. However, such an investigation is beyond the scope of this paper. We concentrated on the evaluation of CMIP5 GCMs using KTC climate types to show how a climate classification can provide a compact analysis tool integrating important temperature and precipitation characteristics.

Acknowledgements. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model outputs. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provided coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The CRU TS 3.22 dataset was provided by the Climatic Research Unit, University of East Anglia. This study was supported by project UNCE 204020/2012 funded by Charles University in Prague and by research plan no. MSM0021620860 funded by the Ministry of Education, Youth and Sports of the Czech Republic. In addition, the work is part of the activity under the Program of Charles University PRVOUK No. 02 'Environmental Research.'

LITERATURE CITED

- Adachi Y, Yukimoto S, Deushi M, Obata A and others (2013) Basic performance of a new earth system model of the Meteorological Research Institute (MRI-ESM1). *Papers Meteorol Geophys* 64:1–19
- Belda M, Holtanová E, Halenka T, Kalvová J (2014) Climate classification revisited: from Köppen to Trewartha. *Clim Res* 59:1–13
- Boer GJ, Flato G, Reader MC, Ramsden D (2000) A transient climate change simulation with greenhouse gas and aerosol forcing: experimental design and comparison with the instrumental record for the twentieth century. *Clim Dyn* 16:405–427
- Boriah S, Chandola V, Kumar V (2008) Similarity measures for categorical data: a comparative evaluation. *Proceedings of the 8th SIAM International Conference on Data Mining*, p 243–254
- Cannon AJ (2012) Köppen versus the computer: comparing Köppen-Geiger and multivariate regression tree climate classifications in terms of climate homogeneity. *Hydrol Earth Syst Sci* 16:217–229
- Deser C, Knutti R, Solomon S, Phillips AS (2012) Communication of the role of natural variability in future North American climate. *Nat Clim Change* 2:775–779
- Duffy PB, Govindasamy B, Iorio JP, Milovich J and others (2003) High-resolution simulations of global climate. 1. Present climate. *Clim Dyn* 21:371–390
- Feng S, Ho CH, Hu Q, Oglesby RJ, Jeong SJ, Kim BM (2012) Evaluating observed and projected future climate changes for the Arctic using the Köppen–Trewartha climate classification. *Clim Dyn* 38:1359–1373
- Feng S, Hu Q, Chen F, Ho CH, Li R, Tang Z (2014) Projected climate shift under future global warming from multi-model, multi-scenario, CMIP5 simulations. *Global Planet Change* 112:41–52
- Fovell RG, Fovell MYC (1993) Climate zones of the conterminous United States defined using cluster-analysis. *J Clim* 6:2103–2135
- Geiger R (1954) *Landolt-Börnstein—Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, alte Serie Vol 3*. Springer, Berlin
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res* 113:D06104, doi:10.1029/2007JD008972
- Guetter PJ, Kutzbach JE (1990) A modified Koeppen classification applied to model simulation of glacial and interglacial climates. *Clim Change* 16:193–215
- Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. *Int J Climatol* 34:623–642
- Hirst AC, O'Farrell SP, Siobhan P, Gordon HB (2000) Comparison of a coupled ocean-atmosphere model with and without oceanic eddy-induced advection. 1. Ocean spinup and control integrations. *J Clim* 13:139–163
- Kalvová J, Halenka T, Bezpalcová K, Nemešová I (2003) Köppen climate types in observed and simulated climates. *Stud Geophys Geod* 47:185–202
- Knutti R, Masson D, Gettelman A (2013) Climate model genealogy: generation CMIP5 and how we got there. *Geophys Res Lett* 40:1194–1199
- Köppen W (1923) *Die Klimate der Erde: Grundriss der Klimakunde*. Walter de Gruyter & Co., Berlin
- Köppen W (1931) *Grundriss der Klimakunde*. Walter de Gruyter & Co., Berlin
- Köppen W (1936) *Das geographische System der Klimate*. In: Köppen W, Geiger R (eds) *Handbuch der Klimatologie*. Gebrüder Borntraeger, Berlin, p C1–C44
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. *Meteorol Z* 15:259–263
- Lohmann U, Sausen R, Bengtsson L, Cubasch U, Perlwitz J, Roeckner E (1993) The Köppen climate classification as a diagnostic tool for general circulation models. *Clim Res* 3:177–193
- Mahlstein I, Daniel JS, Solomon S (2013) Pace of shifts in climate regions increases with global temperature. *Nat Clim Change* 3:739–743
- Manabe S, Holloway JL Jr (1975) The seasonal variation of the hydrological cycle as simulated by a global model of the atmosphere. *J Geophys Res* 80:1617–1649
- Meehl GA, Covey C, Delworth T, Latif M and others (2007) The WCRP CMIP3 multi-model dataset: a new era in climate change research. *Bull Am Meteorol Soc* 88:1383–1394
- Miao C, Duan Q, Sun Q, Huang Y and others (2014) Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. *Environ Res Lett* 9:055007, doi:10.1088/1748-9326/9/5/055007
- Mitchell JFB, Johns TC (1997) On modification of global warming by sulphate aerosols. *J Clim* 10:245–267
- Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climatol* 25:693–712
- Mitchell TD, Carter TR, Jones PD, Hulme M, New M (2004) A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed records (1901–2000) and 16 scenarios (2001–2100). Working

- paper 55. Tyndall Centre of Climate Change Research, Norwich
- Patton CP (1962) A note on the classification of dry climate in the Köppen system. *Calif Geogr* 3:105–112
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci* 11:1633–1644
- Pierce DW, Barnett TP, Santer BD, Gleckler PJ (2009) Selecting global climate models for regional climate change studies. *Proc Natl Acad Sci USA* 106:8441–8446
- R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.R-project.org/
- Roeckner E, Arpe K, Bengtsson L, Brinkop S and others (1992) Simulation of the present-day climate with ECHAM model: impact of model physics and resolution. Rep 93. Max-Planck-Institut für Meteorologie, Hamburg
- Roeckner E, Arpe K, Bengtsson L, Christoph M and others (1996) The atmospheric general circulation model ECHAM-4: model description and simulation of present-day climate. Rep 218. Max-Planck-Institut für Meteorologie, Hamburg
- Rubel F, Kottek M (2010) Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorol Z* 19:135–141
- Stern H, De Hoedt G, Ernst J (2000) Objective classification of Australian climates. *Aust Meteorol Mag* 49:87–96
- Taylor K, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498
- Trewartha GT (1968) *An introduction to climate*. McGraw-Hill, New York, NY
- Trewartha GT, Horn LH (1980) *Introduction to climate*, 5th edn. McGraw Hill, New York, NY
- Watanabe S, Hajima T, Sudo K, Nagashima T and others (2011) MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments. *Geosci Model Dev* 4:845–872

Editorial responsibility: Tim Sparks, Cambridge, UK

*Submitted: October 6, 2014; Accepted: May 27, 2015
Proofs received from author(s): August 11, 2015*