

High Resolution Köppen-Geiger Classifications of Paleoclimate Simulations

Christian Willmes, Daniel Becker, Sebastian Brocks, Christoph Hütt and Georg Bareth

Institute of Geography, University of Cologne

Abstract

The development and application of an algorithm to compute Köppen-Geiger climate classifications from the Coupled Model Intercomparison Project (CMIP) and Paleo Model Intercomparison Project (PMIP) climate model simulation data is described in this study. The classification algorithm was applied to data from the PMIP III paleoclimate experiments for the Last Glacial Maximum, 21k years before present (yBP), Mid-Holocene (6k yBP) and the Pre-Industrial (0k yBP, control run) time slices. To infer detailed classification maps, the simulation datasets were interpolated to a higher resolution. The classification method presented is based on the application of Open Source Software, and the implementation is described with attention to detail. The source code and the exact input data sets as well as the resulting data sets are provided to enable the application of the presented approach.

1 Introduction

Climate classifications are helpful tools to better identify and understand certain patterns in climate data. At first sight, Köppen-Geiger climate classification maps make it easy to identify regions of similar climate conditions. The Climate Model Intercomparison Project (CMIP) and Paleo Model Intercomparison Project (PMIP) models supply 50 output variables in the atmosphere realm alone. For researchers looking at climate patterns and not for the details of Earth System Models (ESM), this number of variables may be hard to grasp. Looking at a classification map combining several input parameters instead obviously saves time and effort compared to looking at each variable by itself. The work presented here was carried out in the framework of a large interdisciplinary research project, the Collaborative Research Centre 806 (<http://www.sfb806.de>), concerning “Culture-Environment Interaction and Human Mobility in the Late Quaternary” (Richter et al. 2012), where the presented climate classifications of paleoclimate model simulations are of interest for research and further applications.

To infer higher resolution maps from the given 192×96 ($1.875^\circ \times 1.875^\circ$) Earth system model (Paleo) of the Max Planck Institute for Meteorology (MPI-ESM-P) climate model grids, we increased the resolution by applying an interpolation to a $3,600 \times 1,800$ ($0.1^\circ \times 0.1^\circ$) grid. The grids cover the earth's surface from 90° N to 90° S and from 180° W to 180° E. The implementation of the classification was carried out using the Open Source GIS software packages of GRASS GIS (Neteler et al. 2012; GRASS Development Team 2014), QGIS (QGIS Development Team 2014) and GDAL (Warmerdam 2008). Because the main aim of this study is to provide

Address for correspondence: Christian Willmes, Institute of Geography, University of Cologne, Germany. E-mail: c.willmes@uni-koeln.de

Acknowledgements: We want to thank the anonymous reviewers of the FOSS4G academic track and from the TGIS review process for their valuable feedback that clearly improved the article. This research was conducted within the Collaborative Research Centre 806 (<http://www.sfb806.de>) and funded by the German Research Foundation (DFG).

an easily reproducible method for deriving Köppen-Geiger climate classifications from CMIP/PMIP climate model simulation outputs using accessible open source tools, the computations and single GIS processing steps are described in detail, to facilitate the reapplication of the presented method discussed. Thus, the aim of this work is not to compare the results of different climate models, for which the presented method would be suitable as well, but to present a reproducible method, application, or tool, to derive the Köppen-Geiger classifications as basis or input for further applications and research. Because there is clearly an increasing momentum towards *open science* (Hey and Payne 2015) in which the authors of this study want to participate, close attention is paid to the reproducibility of the presented method and its application. To support this, in addition to using open source software for the computations, all resulting and source datasets are provided by the authors.

This study proceeds by first looking at related research of applying climate classifications to climate model simulations in Section 2. As described and applied in some of these studies, the here presented classification method can also be used to cross-validate and evaluate climate models. In the following sections of this article, we describe the data and methods applied in this study in Section 3. The CMIP/PMIP models and experiments are introduced in Section 3.1, the input variables for the classifications are described in Section 3.1.2 and the exact data sources for the classification maps presented in this article are given in Section 3.1.3. This is followed by a description of the process of creating the land masks needed for time slices with different sea levels in Section 3.2 and the interpolation procedure applied to the input data in Section 3.3. In Section 3.4, we describe the updated Köppen-Geiger classification scheme after Peel et al. (2007) and Kottek et al. (2006), which we implemented in this contribution. Section 4 concerns the actual practical implementation details of how the classifications are computed. The resulting classifications for the three CMIP/PMIP experiments (LGM, Mid-Holocene and Pre-Industrial) are shown in Section 5, followed by a discussion, in Section 6, of the results and the implementation approach we applied. Finally, some concluding remarks and possible applications for the presented methods are given.

2 Related Work

In this section, we give an overview of related recent studies concerning Köppen-Geiger climate classifications and their applications for climate model testing and evaluation.

The methodological basis of this work is mainly built on the work of Kottek et al. (2006) and Peel et al. (2007) for the definition of the classification and implementation of the Köppen-Geiger algorithms. These two contributions created an updated Köppen-Geiger classification scheme on the basis of worldwide observed climate data.

Lohmann et al. (1992) already developed Köppen-Geiger classifications for comparing climate models more than 20 years ago. They compared coupled (ECHAM1-LSG) and uncoupled (ECHAM2/ECHAM3) simulations between each other and against real observed climate data on the basis of Köppen-Geiger classifications. Overall, they concluded that climate classifications are a suitable tool for the evaluation of climate models against each other and in particular against observed environmental data.

Kalvova et al (2003) assessed the impact of global warming using Köppen-Geiger classifications, as simulated by the general circulation models ECHAM4 and HadCM2. This contribution is interesting because of its interpolation approach, which actually is an extrapolation approach of fitting data to a regular grid coarser than the grid in which the input data was available, to integrate the different resolutions of the studied climate model simulations. It was

concluded that using climate classifications for large scale comparison of long-term observed and simulated data is a useful tool.

Kleidon et al. (2000) successfully used the Köppen-Geiger classification to compare model scenarios, to illustrate the differences in climatic character and what it would imply for the vegetation type.

Gnanadesikan and Stouffer (2006) used the Köppen-Geiger classification to identify the effect of model errors on biosphere coupling, and show that it can successfully be used as an indicator for the performance of vegetation models within atmosphere-ocean-land-sea models (AOGCMs).

These few studies show that Köppen-Geiger climate classifications are a valid and useful tool for recognizing climate and vegetation patterns in simulated and observed temperature and precipitation data. Thus, a simply adaptable tool for creating Köppen-Geiger classifications from CMIP and PMIP model simulation data, as presented in this study, is a useful contribution.

3 Data and Methods

3.1 PMIP III Data

As input data for the climate classifications, we acquired data from the Paleoclimate Modelling Intercomparison Project version 3 (PMIP III; <https://pmip3.lscce.ipsl.fr/>). The PMIP initiative emerged from two parallel endeavors (Braconnot et al. 2011). The first endeavor was the Cooperative Holocene Mapping Project (COHMAP-MEMBERS 1988), which successfully facilitated methods to combine model simulations and syntheses of paleoenvironmental in-situ data. The second endeavor was the establishment of the Atmospheric Modelling Intercomparison Project (AMIP; <http://www-pcmdi.llnl.gov/projects/amip/index.php>), which was the first of several model intercomparison projects of which the Coupled Model Intercomparison Project (CMIP, <http://cmip.llnl.gov/>) and PMIP are parts (Braconnot et al. 2011, 2012).

In its first phase (PMIP I; <http://pmip.lscce.ipsl.fr/>), the project focused on atmosphere models only, since the second phase (PMIP II; <http://pmip2.lscce.ipsl.fr/>), the focus was extended to the aspects of ocean-atmosphere and atmosphere-vegetation models (Braconnot et al. 2012).

In the third phase (PMIP III), the models of the CMIP5 (<http://cmip.llnl.gov/cmip5/>) paleoclimate simulations for the Last Millennium (the last 1,000 years), Mid-Holocene (6k yBP) and Last Glacial Maximum (LGM, 21k yBP) were evaluated. Additionally, non-CMIP models for early Holocene and the 8.2k yBP event, as well as the last Interglacial (130k yBP to 115k yBP) and a model on Mid-Pliocene were set up and planned within PMIP III.

3.1.1 Paleoclimate model

For this study, the focus is on the three experiments shown below, because at the time of writing, these were the available paleoclimate experiments from PMIP III:

1. 0ka yBP, *Pre-Industrial*;
2. 6ka yBP, *Mid-Holocene*; and
3. 21ka yBP, *Last Glacial Maximum (LGM)*.

The method presented here can be adapted to the two other time slices as soon as they are available. Certain boundary model forcing variables are defined by the CMIP/PMIP community for each of the different time slice experiments. These parameters can be obtained in detail from the above mentioned PMIP III and CMIP 5 websites as well as from Taylor et al. (2012), where the general model setup is explained.

The PMIP III Pre-Industrial control (piControl) run has the same parameters as the CMIP5 piControl experiments also given in Taylor et al. (2012). The experiment serves as the baseline for the analysis of historical and future scenario runs with prescribed greenhouse gas concentrations and orbital parameters (Taylor et al. 2012). For the Mid-Holocene (6ka yBP) model, the orbital and greenhouse gas concentration parameters were adjusted as boundary conditions fitting the modeled time slice, which are also given on the above mentioned websites and in Taylor et al. (2012). The LGM (21ka yBP) time slice experiment, in addition to adjusted orbital parameters and greenhouse gas concentrations, and different land masses including glaciation extents, has to fit the LGM sea level of 120 m below the level of today (Clark and Mix 2002), and to fit the different environmental conditions due to the glaciations.

3.1.2 Climate variables

The Köppen-Geiger classification is based on two climate parameters, temperature and precipitation. These parameters are traditionally obtained at climate stations, which are widely distributed around the world (Peel et al. 2007), mostly in degrees Celsius (°C) and in millimeters (mm). Thus, Köppen (1900) developed his classification more than 100 years ago from those parameters which were accessible to him.

The (paleo-)climate model data used as input for the computations presented here contain sufficient variables which can be used to replace the original measured inputs. For the temperature parameter, the *Near-Surface Air Temperature* (tas) variable is used, which represents the temperature at 2 m height and is given in degrees Kelvin (°K). For the precipitation parameter, the *Precipitation* (pr) variable was chosen, representing the precipitation at surface and including both liquid and solid phases. The variable is defined as the precipitation flux in $\text{kg m}^{-2} \text{s}^{-1}$, that equates to mm/second and can replace the precipitation measurements of climate stations in mm/day by simplifying the fraction.

3.1.3 Data sources

Several different model results from different institutions are available from the PMIP III output database. In the context of this study, it was decided to obtain the MPI-ESM-P (Max-Planck-Institut – Earth System Model – Paleo Version) model (Giorgetta et al. 2013) simulations. From this model, the experiments and variables listed in Table 1 were used.

The files were obtained from the Earth System Grid Federation (ESGF) German Climate Computing Centre (DKRZ) node (<http://esgf-data.dkrz.de/esgf-web-fe/>). It was decided to use the experiments of the r1i1p1 realm, which are consistent with PMIP III Mid-Holocene and LGM conditions, respectively, and do not include dynamical calculation of the vegetation distribution including feedback on climate, as the r1i1p2 experiments do (Giorgetta et al. 2013). This decision is arbitrary and can be substituted by other experiments and models from the CMIP and PMIP model output databases, to derive Köppen-Geiger classifications as presented here. More detailed instruction for accessing the data, as well as copies of the here applied source datasets are accessible from Willmes et al. (2014d). The aim of this study is not to evaluate or compare the results of the different models, but to present a method to derive climate classifications from these model simulation datasets.

3.2 Landmasks

To create the classification maps, as explained in Section 4.6, landmass polygons are necessary to clip the vectorized classification data to the world's land surfaces. Since the coast lines

Table 1 Overview of CMIP 5/PMIP 3 data sources used as input for the classifications. In total six datasets, for each time slice the precipitation (pr) and temperature (tas) variable datasets, were obtained from the DKRZ ESGF node

Experiment	Var	Resolution	Model	Realm	Version
piControl	pr tas	96 x 192	MPI-ESM-P Max-Planck-Institut	r1i1p1	20120602
MidHolocene	pr tas		Earth System Model Paleo		
LGM	pr tas		(Raddatz et al. 2007)		

change according to the sea level, different landmass polygons are necessary for different sea levels corresponding to the different time slices. Hence, landmass polygons according to the LGM, Mid-Holocene and Pre-Industrial time slices are applied in this study. See Table 2 for an overview of the datasets and its sea levels.

For the calculation of the LGM landmask, a sea level of –120 m was anticipated according to Clark and Mix (2002). Based on global bathymetric and elevation data obtained from GEBCO (Helffrich 1996; General Bathymetric Chart of the Oceans, GEBCO 2010), all areas that have a height between 10,000 and –120 m were reclassified as land. This newly created land mask had to be further extended by areas that were covered with the Ice Sheet of the LGM. This glaciation extent geodata was acquired from the supplemental online material of Ehlers et al. (2011).

Shennan and Horton (2002) indicated that the sea level during the Mid-Holocene was the same as of today. For this reason, today’s extent of the land surface was applied to create the landmasks of Pre-Industrial and Mid-Holocene. Natural Earth Data (Kelso and Patterson 2010) land polygons including major islands in small scale 1:10M resolution were applied, which is an appropriate level of detail for this global data set.

3.3 Interpolation

It is assumed that the interpolation of the input data before their classification increases the information provided by the original input datasets in the context of the presented application. The resolution of the input data is quite low, and delivering classifications of higher resolutions is very valuable for the applications that use these data as an input. For precipitation and

Table 2 Sea levels for LGM, Mid-Holocene and Pre-Industrial time slices according to state-of-the-art peer-reviewed research

Time slice	Sea level difference	Source of sea level	Dataset
LGM	–120 m	(Clark and Mix 2002)	(Willmes et al. 2014a)
Mid-Holocene	0 m	(Shennan and Horton 2002)	(Kelso and Patterson 2010)
Pre-Industrial	0 m	(Kelso and Patterson 2010)	(Kelso and Patterson 2010)

temperature data it is possible to assume continuity and smoothness up to the first order derivative. This assumption will be discussed in more detail in Section 6. To derive higher resolution climate classifications from the 192×96 ($1.875^\circ \times 1.875^\circ$) input grids, bilinear interpolations to a $3,600 \times 1,800$ ($0.1^\circ \times 0.1^\circ$) grid were applied.

The interpolations were applied to the 24 monthly mean raster files of the input variables (*tas* and *pr*), derived at original resolution from the input NetCDF files, before any computations for the classifications were undertaken. It is crucial for the interpolation of the data to a higher resolution, to assume that the input data is a continuous surface with a steady gradient of the values between the grid cells. The chosen interpolation method should compute the values inbetween the input grid values on a continuous slope. Bilinear interpolation fits these demands well, keeping the original values of the input data and computing the interpolated values on a simple but realistic surface between those input values. Bilinear interpolation is a widely applied standard interpolation method for resampling regular grids (Neteler and Mitasova 2004) such as the regularly gridded model output data given by CMIP/PIMIP in this application. Bicubic interpolation was also tested, no significant differences were observed between the results of bilinear and bicubic resampling. But the bicubic interpolation is more complex, has more support and thus smoothes the data more. To avoid unnecessary complexity, it was decided to apply the simpler bilinear resampling instead of bicubic.

Nearest neighbor resampling does not preserve the continuity of the data, and thus did not increase the information of the interpolated or resampled grid, since the result is equivalent to increase the resolution without using the hypothesis of a continuous and smooth signal, and thus was withdrawn.

3.4 Köppen-Geiger Climate Classifications

The Köppen-Geiger classification is one of the most widely used climate classification systems (Peel et al. 2007; Kottek et al. 2006). It was developed by the Russian-German climatologist and botanist Wladimir Köppen in the last decade of the 19th century (Köppen 1900, 1918; Köppen and Geiger 1936). Later, the German meteorologist Rudolph Geiger developed the classification system further (Geiger and Pohl 1954). Since then, the system is referred to as the *Köppen-Geiger climate classification*.

The Köppen-Geiger climate classification is based on vegetation-referenced temperature and precipitation thresholds, which differentiate in their combinations. According to the here applied updated classification scheme of Peel et al. (2007), six climate zones or main groups (first letter) are sub-divided by 12 climate types (second and third letter) based on the combination of temperature and precipitation patterns, resulting in 30 distinct climates. See Table 3 for the semantics of the letters in the classification codes. The climate zones are defined by temperature, except for the arid B climates, which are defined by precipitation. The climate types are mainly differentiated by annual precipitation and its yearly distribution, while the climate subtypes are differentiated by its seasonal variability.

A detailed description of the classification scheme applied in this contribution is given in Section 3.4.1.

In this contribution, the classification scheme developed by Peel et al. (2007) was applied, which follows the scheme of Köppen and Geiger (1936) with the exception of the boundary between the C and D climates. This boundary was adopted by Peel et al. (2007) from Russell (1931), who used $>0^\circ\text{C}$, instead of $>-3^\circ\text{C}$ as threshold for the coldest month's temperature of the year.

Table 3 Semantics of the indicating letters within the Köppen-Geiger classification codes

Climate zones	Climate types (precipitation)	Sub-types (temperature)
<ul style="list-style-type: none"> • A: tropics • B: arid • C: warm temperate • D: snow • ET: polar tundra • EF: polar frost 	<ul style="list-style-type: none"> • W: desert • S: steppe • f: fully humid • s: summer dry • w: winter dry • m: monsoonal 	<ul style="list-style-type: none"> • h: hot arid • k: cold arid • a: hot summer • b: warm summer • c: cool summer • d: extr. continental

3.4.1 Updated Köppen-Geiger classification Scheme

Based on the Kottek et al. 2006 publication of an updated algorithm for computing the Köppen-Geiger classifications, Peel et al. (2007) published an even further updated classification scheme, which is given in Table 4. The classifications of this contribution are derived using these definitions.

To be able to compute classes based on the differentiation between summer and winter, summer is defined as the warmer period of October to March or April to September. This is computed by simply adding the monthly means of the seasons and then assigning the value 1 for all cells with $ONDJFM \leq AMJJAS$, see Listing 4.

As given in Table 4, the Köppen-Geiger classification defines 30 possible climate types. These climate types are divided into three tropical (Af, Am, Aw), four arid (BWh, BWk, BSh, Bsk), nine temperate (Csa, Csb, Csc, Cfa, Cfb, Cfc, Cwa, Cwb, Cwc), 12 cold (Dsa, Dsb, Dsc, Dsd, Dfa, Dfb, Dfc, Dfd, Dwa, Dwb, Dwc, Dwd) and two polar (ET and EF) types.

This is different from the Kottek et al. (2006) implementation, where four tropical types are defined with the addition of an As type for equatorial savannah with dry summer, and Aw for equatorial savannah with dry winter. In this contribution, the definition of Peel et al. (2007), where only a single savannah type (Aw) is defined without distinction between summer and winter, was used.

Due to some rare cases where the criteria for Cs and Cw or for Ds and Dw climates can both be met, an additional computation as in Peel et al. (2007, p. 1637) was made to distinguish between the two respective climate types in those cases (see Listing 1).

Listing 1: GRASS Map Algebra for differentiation of Cs/Cw and Ds/Dw doubles.

```

Dsreal = if(Ds == 1 && Dw == 1, if(pr_s_mm < pr_w_mm, 1, 0), Ds)
Dwreal = if(Ds == 1 && Dw == 1, if(pr_s_mm >= pr_w_mm, 1, 0), Dw)
Dw = Dwreal
Ds = Dsreal
    
```

4 Implementation of the Climate Classifications

The classification algorithm was implemented using GRASS GIS, GDAL and QGIS tools. The data import, the interpolation and the computation of the classification were implemented using GRASS GIS, while for scripting the algorithm, the pyGRASS (Zambelli et al. 2013)

Table 4 The Köppen-Geiger classification criteria after Peel et al. (2007) based on Kottek et al. (2006)

1st	2nd	3rd	Description	Criteria
A			Tropical	$T_{cold} \geq 18$
	f		-Rainforest	$P_{dry} \geq 60$
	m		-Monsoon	Not(Af) AND $P_{dry} \geq 100 - P_{ann}/25$
	w		-Savannah	Not(Af) AND $P_{dry} < 100 - P_{ann}/25$
B			Arid	$P_{ann} < 10 \times P_{threshold}$
	W		-Desert	$P_{ann} < 5 \times P_{threshold}$
	S		-Steppe	$P_{ann} \geq 5 \times P_{threshold}$
		h	-Hot	$T_{ann} \geq 18$
		k	-Cold	$T_{ann} < 18$
C			Temperate	$T_{hot} > 10$ AND $0 < T_{cold} < 18$
	s		-Dry Summer	$P_{sdry} < 40$ AND $P_{sdry} < P_{wwet}$
	w		-Dry Winter	$P_{wdry} < P_{swet}/10$
	f		-Without dry season	Not(Cs) OR Not(Cw)
		a	-Hot Summer	$T_{hot} \geq 22$
		b	-Warm Summer	Not(a) AND $T_{mon10} \geq 4$
		c	-Cold Summer	Not(a or b) AND $1 \leq T_{mon10} < 4$
D			Cold	$T_{hot} > 10$ AND $T_{cold} \leq 0$
	s		-Dry Summer	$P_{sdry} < 40$ AND $P_{sdry} < P_{wwet}/3$
	w		-Dry Winter	$P_{wdry} < P_{swet}/10$
	f		-Without dry season	Not(Ds) AND Not(Dw)
		a	-Hot Summer	$T_{hot} \geq 22$
		b	-Warm Summer	Not(a) AND $T_{mon10} \geq 4$
		c	-Cold Summer	Not(a OR b OR d)
		d	-Very cold Winter	Not(a OR b) AND $T_{cold} < -38$
E			Polar	$T_{hot} < 10$
	T		-Tundra	$T_{hot} > 0$
	F		-Frost	$T_{hot} \leq 0$

T_{cold} = coldest month temperature, T_{hot} = hottest month temperature, P_{dry} = driest month precipitation, P_{ann} = mean annual precipitation, T_{ann} = mean annual temperature, P_{sdry} = precipitation of the driest month in summer, P_{swet} = precipitation of wettest month in summer, P_{wwet} = precipitation of wettest month in winter, P_{wdry} = precipitation of driest month in winter, T_{mon10} = count of months per year where temperature is above 10°C.

python bindings for GRASS GIS were used. The cartography was conducted in QGIS manually.

The computation of the classifications from the input data to the Köppen-Geiger maps follows this algorithm (see Figure 1):

1. Data import \rightarrow monthly mean *tas* and *pr*
2. Reprojection of input data to equal area projection (see Section 4.2) \rightarrow equal area monthly means
3. Interpolation \rightarrow Interpolated *tas* and *pr*

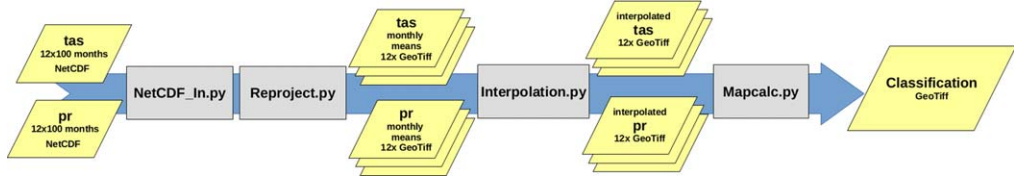


Figure 1 Data and pyGRASS scripts flowchart

4. Computation of the classifications → classifications
5. Vectorization → classification area polygons
6. Cartography → classification maps

The pyGRASS scripts created for this implementation are published online and can be acquired from Willmes et al. (2014d).

4.1 Data Import

The classifications are computed from the monthly means of the model outputs. Because the input data is provided in model runs over several years, for Mid-Holocene and LGM in 100 and for Pre-Industrial in 150 years, the monthly means for *tas* and *pr* of each month were computed.

The computation is implemented in the pyGRASS Python script and given in Willmes et al. (2014d). The script first defines rasters for the 12 months, looping through the NetCDF input raster layers adding the monthly values to the temporary overall month value raster, which is then divided by the count of years to compute the monthly arithmetic mean.

4.2 Reprojection

To preserve the spatial properties of the simulated model output, the interpolation must be computed in an area preserving projection. The Lambert based Cylindrical Equal Area Projection (CEA) was chosen to meet this demands. Thus, the NetCDF imported data was reprojected into a new GRASS GIS mapset using the *r.reproj* command, see Listing 2.

Listing 2: Projecting the geographic coordiante monthly mean layers to cylindrical equal area (cea).

```
r.proj input=layer location='KoeppenGeiger' mapset='import' -overwrite
```

The GRASS GIS Location *KoeppenGeiger* containing the *import* mapset is originally defined in Lat/Lon WGS84 Geographic Coordinates (EPSG:4326). The command for Listing 2 is triggered from a Location and Mapset in the Lambert Cylindrical Equal Area Projection but preserving the same resolution, 192×96 cells for the $360^\circ \times 180^\circ$ earth surface covering input data. This is applied in a loop through all monthly mean temperature and precipitation data layers of the referenced GRASS GIS Mapset in the script *reproject.py* pyGRASS script, available from (Willmes et al. 2014c).

4.3 Interpolation

The interpolation, as described in Section 3.3, was carried out using the GRASS *r.resamp.interp* tool (see Listing 3) and implemented in the *Interpolation.py* pyGRASS script available from Willmes et al. (2014d).

Listing 3: Resampling of the monthly mean pr and tas rasters.

```
r.resamp.interp input=tas output=tasout method=bilinear -overwrite
```

The script first sets the GRASS region to the desired interpolated extent and resolution, in this case to the extent of 90, 180, -90, -180 and a resolution of $1,800 \times 3,600$, and then loops through the 24 monthly mean input variable rasters applying the GRASS interpolation command.

4.4 Computation of the Classification

The classification was implemented in the *mapcalc.py* pyGRASS script given in Willmes et al. (2014d), applying 108 consecutive map algebra calculations to implement the classification criteria given in Table 4 on the 24 interpolated input variable rasters, using the GRASS *r.mapcalc* functionality.

To implement the criteria shown in Table 4, several combined variables from the temperature *tas* and precipitation *pr* input datasets need to be derived. At first, the following variables are computed:

- Mean Annual Precipitation P_{ann}
- Driest Month P_{dry}
- Wettest Month P_{wet}
- Mean Annual Temperature T_{ann}
- Coldest Month T_{cold}
- Hottest Month T_{hot}

To be able to determine seasonality-dependent variables, the differentiation between summer and winter is needed. The differentiation was computed as shown in Listing 4.

Listing 4: Distinction between summer and winter (GRASS *r.mapcalc* syntax)

```
ONDJFM = tas10 + tas11 + tas12 + tas01 + tas02 + tas03
AMJJAS = tas04 + tas05 + tas06 + tas07 + tas08 + tas09
nsummer = if(ONDJFM <=AMJJAS, 1, 0)
```

Using this differentiation, the following seasonality-dependent variables were computed:

- Driest month in summer P_{sdry}
- Driest month in winter P_{wdry}
- Wettest month in summer P_{swet}
- Wettest month in winter P_{wwet}

For some classes, the value for the number of months with a mean temperature above 10°C is required. It is computed by counting the months above 10°C in variable T_{mon10} .

To classify the arid *B* climates, a precipitation threshold rule is applied. The threshold varies according to the following rules: If more than 70% of P_{ann} occurs in winter, $P_{threshold}=2\times P_{ann}$, if more than 70% of P_{ann} occurs in summer, $P_{threshold}=2\times T_{ann}+28$, otherwise the threshold will be $P_{threshold}=2\times T_{ann}+14$. The according mapcalc sequence is given in Listing 5.

Listing 5: Mapcalc formulation of precipitation thresholds for calculating *B* climates

```
#Calculate if 70% of precipitation occurs in Winter
pr_winter = if(nsummer,(pr10 + pr11 + pr12 + pr01 + pr02 + pr03),\
  (pr04 + pr05 + pr06 + pr07 + pr08 + pr09))
pr_tr1 = if(pr_winter >= (pr_ann_sum * 0.7),(tas_ann_c * 2), 0)

#Calculate summer threshold
pr_summer = if(nsummer,(pr04 + pr05 + pr06 + pr07 + pr08 + pr09),\
  (pr10 + pr11 + pr12 + pr01 + pr02 + pr03))
pr_tr2 = if(pr_summer > (pr_ann_sum * 0.7),(tas_ann_c * 2) + 28, 0)

#if not tr1 or tr2
pr_tr3 = if((pr_tr1 == 0) && (pr_tr2 == 0),(tas_ann_c * 2) + 14, 0)

#combine
pr_threshold = pr_tr1 + pr_tr2 + pr_tr3
```

Using the combined variables defined above, the classification criteria given in Table 4 can be formulated in mapcalc syntax and computed. The resulting classification is then given in a raster file, with integer values representing the 30 defined Köppen-Geiger classes as values of the raster cells.

4.5 Computation of the Landmasks

To derive the landmask of the LGM from the elevation model and the glaciation extents, as explained in Section 3.2, GRASS GIS (GRASS Development Team 2014) and GDAL (Warmerdam 2008; GDAL Development Team 2013) were applied as described in the following steps.

A new location and mapset were defined in GRASS GIS using WGS84 as the coordinate reference system. The extent and resolution of the mapset was adjusted to match the elevation and bathymetry data, obtained from the General Bathymetric Chart of the Oceans (GEBCO) (2010) using the `g.region` functionality. It was then imported into GRASS using `r.in.gdal`. To reclassify the imported raster into land (0) and sea (1), the `r.reclass` function was used (see Listing 6).

Listing 6: New Grass Mapset “gebco” with wgs84 as reference system

```
g.region n=90 s=-90 e=180 w=-180 rows=21600 cols=43200
r.in.gdal input=gebco_08.nc output=gebco_08 -k -l -o
r.reclass input=gebco_08@gebco output=lgm rules=lgm.txt title=lgm
```

For future studies that possibly will include time slices with different sea levels, the relevant values can be easily adapted in the parameter defining textfile (*lgm.txt*, see Listing 7) to

compute different landmasks from the General Bathymetric Chart of the Oceans (GEBCO) (2010) dataset.

Listing 7: Content of `lgm.txt`

```
-20000 thru -120 = 1
-119 thru 10000 = 0
```

The reclassified raster file was transferred to the vector domain using the `r.to.vect` tool with the `simplify` option (`-s`). The `dissolve` tool simplifies the distinct polygons to two multipolygons. One multipolygon contains the area of the land, the other one contains the sea. As this study only uses the land area, the multipolygon that has value 0 (land) was exported (see Listing 8) and used as the input for the next section.

Listing 8: Dissolve and extract

```
r.to.vect -s input=lgm output=lgm_vektor feature=area
v.dissolve input=lgm_vektor@gebco output=lgm_vektor_dissolve
v.extract input=lgm_vektor_dissolve@gebco
output=lgm_vektor_dissolve_land where= value = 0
```

To include areas that are below -120 m of today's bathymetry, but were covered by the last glacial maximum's ice sheet, the glaciation extents of the LGM from Ehlers et al. (2011) were combined with the previously computed -120 m landmass polygons (see Listing 9).

Listing 9: Read the `lgm` ice sheet data and merge it with the `lgm` landmask

```
v.in.ogr dsn=lgm.shp output=lgm_icesheet
v.overlay ainput=lgm_vektor_dissolve_land@gebco
binput=lgm_icesheet@gebco output=lgm_and_icesheet
v.dissolve input=lgm_and_icesheet@gebco
output=lgm_and_icesheet_dissolve column=dissolve
```

As a result we have a landmask combining the -120 m bathymetry coastline derived from the General Bathymetric Chart of the Oceans (GEBCO) (2010) dataset and the LGM glaciation extents dataset provided by Ehlers et al. (2011).

4.6 Cartography

The cartography of the paleoclimate classifications was created using the print composer of QGIS (Version 2.2). In this section, the map creation process is described in detail.

Because of better visualization capabilities for vector geometries containing categorized values compared with raster files containing that same information, such as Köppen-Geiger classes, the rasters containing the classifications were vectorized before the data was visualized in QGIS. The vectorization of the classification rasters was carried out using the `gdal_polygonize` tool. The command applied for this task is given in Listing 10.

Listing 10: GDAL Polygonize

```
gdal_polygonize Classification.tif -f "ESRI Shapefile"
  ClassPoly.shp KG-Classification KGClass
```

The resulting vector dataset in Esri Shapefile format contains polygons surrounding adjacent pixels of the same Köppen-Geiger class. It would also have been possible to use the GRASS GIS *r.to.vect* tool for this task; there was no particular objective reason why *gdal_polygonize* was chosen over *r.to.vect* in this case.

In the next step, the landmasks as described in Section 3.2 and derived in Section 4.5 were applied. The landmasks were used to clip the classification data to limit the data shown in the maps to just the landmass. The clipping process was accomplished with the GRASS GIS tool *v.overlay* using the “and” operator.

This was followed by choosing and applying an appropriate map projection. In order to present the results in visually appealing maps, the Robinson projection (EPSG:54030) was used as the maps’ coordinate reference system. The authors think, that the resulting maps look more natural than for example plate carrée mapped WGS84-data, the default in most GIS due to the lower distortion of the polar areas.

The colors applied to the Köppen-Geiger climate classes were derived from Kottek et al. (2006). This was accomplished in the QGIS layer properties, where the correct labels and the corresponding colors of the climate classification were assigned to the values of the vector layer.

The 10° graticules, producing the grid shown on the maps, were generated using the *v.mkgrid* tool of GRASS GIS.

Listing 11: GRASS v.mkgrid

```
v.mkgrid map=grid_10degree grid=18,36
```

The sea water body background was generated with the *v.in.region* tool of GRASS GIS and visualized in QGIS afterwards.

5 Results

In this section the resulting classification maps (see Figure 2) are presented. The corresponding geospatial datasets of the classifications are published online in vector (Esri Shapefile) and raster (GeoTiff) format and can be accessed from:

- Pre-Industrial (Willmes et al. 2014c);
- Mid-Holocene (Willmes et al. 2014b); and
- Last Glacial Maximum (Willmes et al. 2014a).

These datasets are published on the CRC806-Database web portal (Willmes et al. 2014e), as file downloads and OGC conform web service in WMS, WFS and WCS format and accessible from the CRC806-Database SDI (<http://crc806db.uni-koeln.de/maps/>). Additionally, the source code of the Köppen-Geiger classification implementation in the form of pyGRASS scripts, including documentation in the form of a step-by-step tutorial for executing the scripts in the GRASS GIS (GRASS Development Team 2014) environment, are also published online on the

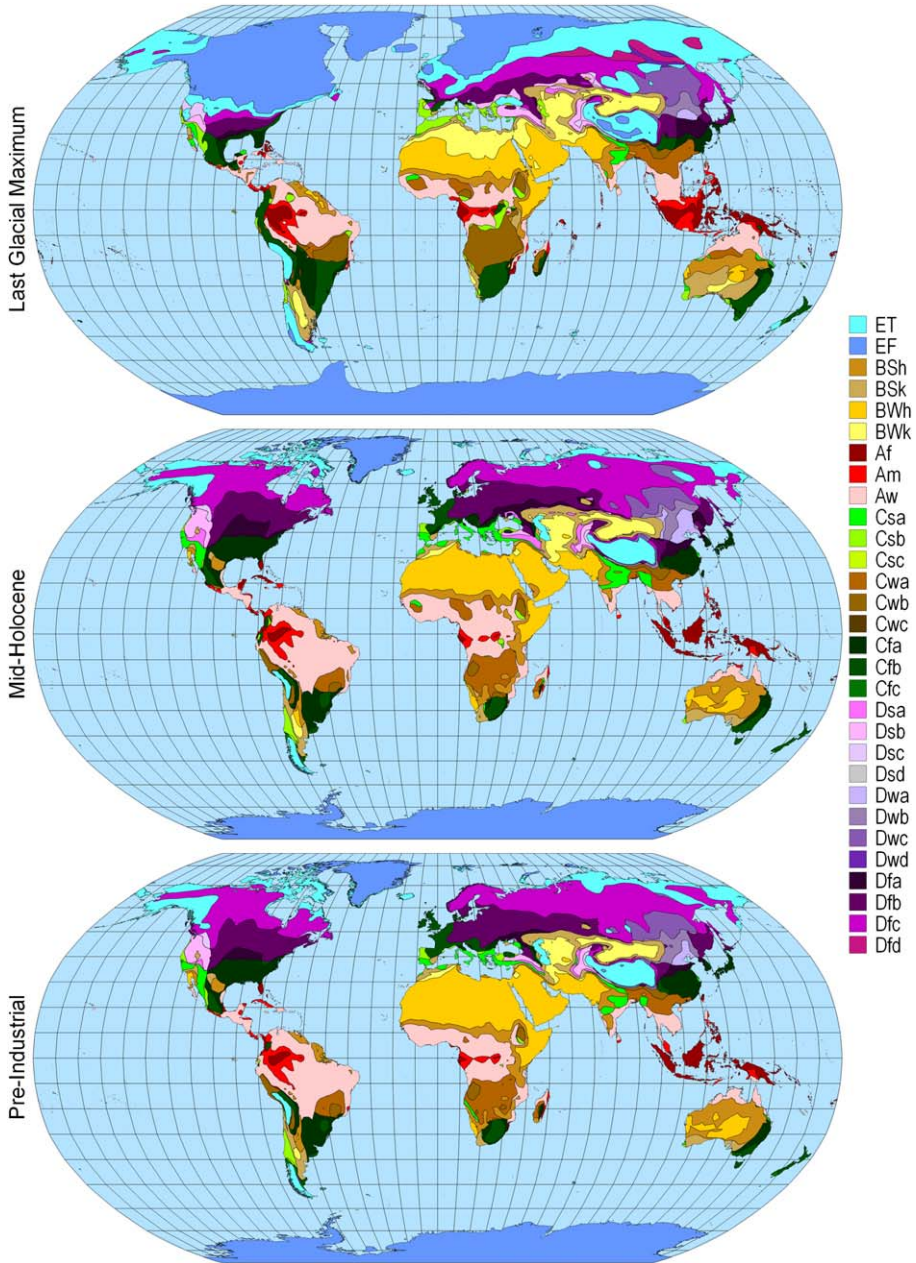


Figure 2 The Köppen-Geiger classification maps for LGM, Mid-Holocene and Pre-Industrial time slices. Data: Willmes et al. (2014a, b, c)

CRC806-Database portal (Willmes et al. 2014d). For optimal reuse and thus the reproducibility of the presented application, the scripts are made available under the creative commons CC-BY license.

Figure 2 shows the maps visualizing the simulated climate conditions for the three time slices (LGM, Mid-Holocene and Pre-Industrial) in a clear and summarized way. Additionally, due to the fact that the Köppen-Geiger classification was developed on the basis of vegetational and environmental patterns, these vegetational and environmental patterns can now be inferred from the climate model simulations. Due to the application of the landmasks and glaciation extents for the different sea level conditions of the LGM, the environment of that time is shown in an informative comprehensive way. This allows, for example, researchers concerning archaeological or paleoenvironmental questions to assess this additional information for their research. Obviously, it needs to be kept in mind that these maps are based on simulated model experiments and do not show the whole truth, but at least a fair enough approximation to gain additional insight about, in this case, the LGM and Mid-Holocene paleoenvironment.

6 Discussion and Conclusions

The methods and the algorithm to compute the classifications provided by this contribution can be adapted to derive the Köppen-Geiger classifications for any model simulation from the CMIP and PMIP databases. By using only Open Source software tools for the presented approach, the reproducibility for anyone interested in applying this classification tool is ensured. Furthermore, the presented method can be adapted to any further CMIP/PMIP models, it is not restricted to the MPI-ESM-P models. CMIP/PMIP variables are standardized, the parameters that may differ in the simulation output files are the resolution of the grid and the computed time interval and its temporal resolution, because the temporal resolution is determined by how many time steps are simulated within an experiment's interval.

Interpolation of the input data to increase the information of the resulting classifications before the computation is justified because the input variables are represented as a continuous field. The input data represents monthly mean values for precipitation and temperature, thus we have already accumulated values as an input, which remove the larger spatial variations in smaller regions for shorter time periods. For example, climate front systems can cause large differences in precipitation and temperature in small regions for short time intervals, but on larger time intervals (in this case month or years), these patterns do not exist anymore.

The interpolation of both input fields for *pr* and *tas* adds information to the resulting classification, especially in regions of the field where high gradients exist in at least one of the two input variables between two input cells in their original resolution. The interpolation increases the information of the resulting classification because the interpolated field lowers the gradient between the values of the now smaller neighboring grid cells. Thus, the location of the threshold boundaries will most likely be different from the exact center of two neighboring grid cells containing a classification threshold boundary in the original resolution. This results in higher detail, which is supported by the data but would not be visible in a classification of the original non-interpolated resolution.

Climate classification maps for paleoenvironmental simulations as provided by PMIP III are a valuable resource for researchers concerning paleoenvironmental questions of any kind. The tool to calculate Köppen-Geiger classifications from CMIP/PMIP standardized climate model simulations, as presented in this work, may also be used for the evaluation and comparison of different climate models. The resulting classification datasets also deliver valuable input data for paleoenvironmental modeling, such as Ecological Niche Modelling or Site Catchment Analysis, as an additional environmental input layer.

Because Köppen (1900) developed the classification system on the basis of vegetational occurrence patterns, it is applicable for evaluating climate models, as shown in the works presented in Section 2, with vegetation occurrences in geoarchaeological archives and records.

Source data sets as well as the resulting classification in vector and raster GIS data formats, including the source code of the classification algorithm, are available through the CRC806-Database (Willmes et al. 2014e), the research data management infrastructure of the research project the here presented research was conducted for.

References

- Braconnot P, Harrison S P, Kageyama M, Bartlein P J, Masson-Delmotte V, Abe-Ouchi A, Otto-Bliesner B, and Zhao Y 2012 Evaluation of climate models using palaeoclimatic data. *Nature Climate Change* 2: 417-24
- Braconnot P, Harrison S P, Otto-Bliesner B, Abe-Ouchi A, Jungclaus J, and Peterschmitt J-Y 2011 The Paleoclimate Modeling Intercomparison Project contribution to CMIP5. *CLIVAR Exchanges* 16: 15-19
- Clark P U and Mix A C 2002 Ice sheets and sea level of the Last Glacial Maximum. *Quaternary Science Reviews* 21: 1-7
- COHMAP-MEMBERS 1988 Climatic changes of the last 18,000 years: Observations and model simulations. *Science* 241: 1043-52
- Ehlers J, Gibbard P L, and Hughes P D 2011 Introduction. In Gibbard P L, Ehlers J and Hughes P D (eds) *Quaternary Glaciations: Extent and Chronology, A Closer Look*. Oxford, UK, Elsevier: 1-14
- GDAL Development Team 2013 *GDAL: Geospatial Data Abstraction Library, Version 1.10.1*. Wayland, MA, Open Source Geospatial Foundation (available at <http://www.gdal.org>)
- Geiger R and Pohl W 1954 Eine neue Wandkarte der Klimagebiete nach W. Köppens Klassifikation. *Erdkunde* 8: 58-61
- GEBCO (General Bathymetric Chart of the Oceans) 2010 The GEBCO 08 Grid: Version 20100927. WWW document, <http://www.gebco.net>
- Giorgetta M A, Jungclaus J, Reick C H, Legutke S, Bader J, Böttinger M, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H-D, Ilyina T, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Mueller W, Notz D, Pithan F, Raddatz T, Rast S, Redler R, Roeckner E, Schmidt H, Schnur R, Segschneider J, Six K D, Stockhause M, Timmreck C, Wegner J, Widmann H, Wieners K-H, Claussen M, Marotzke J, and Stevens B 2013 Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project Phase 5. *Journal of Advances in Modeling Earth Systems* 5: 572-97
- Gnanadesikan A and Stouffer R J 2006 Diagnosing atmosphere-ocean general circulation model errors relevant to the terrestrial biosphere using the Köppen climate classification. *Geophysical Research Letters* 33: 22
- GRASS Development Team 2014 *Geographic Resources Analysis Support System (GRASS GIS) Software*. Beaverton, OR, Open Source Geospatial Foundation (available at <http://grass.osgeo.org>)
- Helffrich G 1996 GEBCO digital atlas. *Terra Nova* 8: 659-61
- Hey T and Payne M C 2015 Open science decoded. *Nature Physics* 11: 367-69
- Kalvova J, Halenka T, Bezpalcova K, and Nemesova I 2003 Köppen climate types in observed and simulated climates. *Studia Geophysica et Geodaetica* 47: 185-202
- Kelso N V and Patterson T 2010 Introducing natural earth data: NaturalEarthData.Com. *Geographia Technica* 2010, Special Issue: 82-89
- Kleidon A, Fraedrich K, and Heimann M 2000 A green planet versus a desert world: Estimating the maximum effect of vegetation on the land surface climate. *Climatic Change* 44: 471-93
- Köppen W and Geiger R 1936 Das geographische System der Klimate. In Köppen W and Geiger R (eds) *Handbuch der Klimatologie*. Berlin, Verlag von Gebrüder Borntraeger, Vol. 1, Part C: 1-44
- Köppen W 1900 Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographische Zeitschrift* 6: 593-679
- Köppen W 1918 Klassifikation der Klimate nach Temperatur, Niederschlag und Jahreslauf. *Petermanns Geographische Mitteilungen* 64: 193-248
- Kottek M, Grieser J, Beck C, Rudolf B, and Rubel F 2006 World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15: 259-63
- Lohmann U, Sausen R, Bengtsson L, Cubasch U, Perlwitz J, and Roeckner E 1992 *The Köppen Climate Classification as a Diagnostic Tool for General Circulation Models*. Hamburg, Germany, Max-Planck-Institut für Meteorologie Technical Report (available at <http://centaur.reading.ac.uk/31838/>)

- Neteler M, Bowman M H, Landa M, and Metz M 2012 GRASS GIS: A multi-purpose open source GIS. *Environmental Modelling and Software* 31: 124-30
- Neteler M and Mitasova H 2004 *Open Source GIS: A GRASS GIS Approach* (Second Edition). Dordrecht, The Netherlands: Kluwer
- Peel M C, Finlayson B L, and McMahon T A 2007 Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences* 11: 1633-44
- QGIS Development Team 2014 *QGIS Geographic Information System*. WWW document, <http://qgis.osgeo.org>
- Raddatz T, Reick C, Knorr W, Kattge J, Roeckner E, Schnur R, Schnitzler K-G, Wetzel P, and Jungclaus J 2007 Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century? *Climate Dynamics* 29: 565-74
- Richter J, Melles M, and Schäbitz F 2012 Temporal and spatial corridors of Homo sapiens sapiens population dynamics during the Late Pleistocene and early Holocene. *Quaternary International* 274: 1-4
- Russell R 1931 *Dry Climates of the United States: I, Climatic Map*. Berkeley, CA, University of California Press
- Shennan I and Horton B 2002 Holocene land- and sea-level changes in Great Britain. *Journal of Quaternary Science* 17: 511-26
- Taylor K E, Stouffer R J, and Meehl G A 2012 An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* 93: 485-98
- Warmerdam F 2008 The Geospatial Data Abstraction Library. In Hall G and Leahy M (eds) *Open Source Approaches in Spatial Data Handling*. Berlin, Springer: 87-104
- Willmes C, Becker D, Brocks S, Hütt C, and Bareth G 2014a *Köppen-Geiger Classification of MPI-ESM-P LGM Simulation*. Köhn, University of Köhn, Collaborative Research Centre 806 Our Way to Europe Database
- Willmes C, Becker D, Brocks S, Hütt C, and Bareth G 2014b *Köppen-Geiger Classification of MPI-ESM-P Mid-Holocene Simulation*. Köhn, University of Köhn, Collaborative Research Centre 806 Our Way to Europe Database
- Willmes C, Becker D, Brocks S, Hütt C, and Bareth G 2014c *Köppen-Geiger Classification of MPI-ESM-P Pre-Industrial Simulation*. Köhn, University of Köhn, Collaborative Research Centre 806 Our Way to Europe Database
- Willmes C, Becker D, Brocks S, Hütt C, and Bareth G 2014d *pyGRASS Implementation of Köppen-Geiger Classifications from CMIP5 Simulations*. Köhn, University of Köhn, Collaborative Research Centre 806 Our Way to Europe Database
- Willmes C, Kürner D, and Bareth G 2014e Building research data management infrastructure using open source software. *Transactions in GIS* 18: 496-509
- Zambelli P, Gebbert S, and Ciolli M 2013 Pygrass: An object oriented Python Application Programming Interface (API) for Geographic Resources Analysis Support System (GRASS) Geographic Information System (GIS). *ISPRS International Journal of Geo-Information* 2: 201-19