Important user information on the regional climate simulation data with CLM (December 2008)

1. The CERA database project 'CLM_regional_climate_model_runs'

The Model and Data Group (M&D) at the Max Planck Institute for Meteorology, Hamburg (referred to hereinafter as data producer) has computed regional climate scenarios, which are made available via the CERA[1] database of the WDCC[2] in the above named project. The simulations were carried out at the request of the Federal Ministry for Education and Research (BMBF) and in consultation with the group of German regional climate modellers. The cooperation project with the BMBF provides free access to the model data to the scientific and application-oriented community. The provision of the data intends to enable the work of climate impact research projects and to stimulate and support the development of adaptation strategies to climate change.

The data were computed at the German Climate Computing Centre (DKRZ[3]) in Hamburg using the regional climate model CLM. The simulations were carried out with support by the CLM developers (BTU Cottbus, GKSS Research Centre Geesthacht, Potsdam Institute for Climate Impact Research), the scientific steering committee of M&D, the shareholders of DKRZ and the BMBF. A technical and visual quality control has been performed by M&D and a scientific validation has been conducted by BTU Cottbus.

CLM climate simulation data	
Data compilation	Model and Data Group (M&D) at MPI for Meteorology, Hamburg
Model	CLM 2.4.11 (Climate mode of the Local Model of the DWD) Dynamic model; drive: ECHAM5/MPIOM, non-hydrostatic
Model region	Europe
Simulation period	from 1960 to 2100
IPCC[4] emission scenarios	A1B, B1 (from 2001)
Resolution	0.165° (data stream 2), 0.2° (data stream 3); approx. 20 km
Structure	Rotated model grid (data stream 2 = DS2) or Regular lat/lon grid (data stream 3 = DS3); extraction of subregions possible.
Data format	netCDF or ASCII format

[1] CERA: Climate and Environmental Data Retrieval and Archive
[2] WDCC: World Data Center for Climate
[3] DKRZ: German High Performance Computing Centre for Climate- and Earth System Research
[4] IPCC: Intergovernmental Panel on Climate Change

2. User information

The CLM data provided in the CERA database have been obtained according to state of the art science and computational engineering. Shortcomings due to parameterisation of small-scale processes, prescribed variations of vegetation parameters, coarse resolution of ocean variables (forcing) and model output handling are unavoidable. The knowledge of these shortcomings is helpful for the interpretation of the model results and they are listed as far as recognized yet (see chap. 3 below). If you find inconsistencies or questionable features in the CLM data please contact SGA (sga(at)dkrz.de). A description of the model dynamics and parameterisations can be found at
http://www.cosmo-model.org/content/model/documentation/core/default.htm.

A comprehensive set of comparisons of model simulations with different forcing were performed by the BTU Cottbus including a comparison of observed and modelled quantities. Selected results are presented in the report 'Ensemble Simulations over Europe with the Regional Climate Model CLM forced with IPCC AR4 Global Scenarios'. Both, the report and the corresponding figures can be found in the CERA database, listed as experiment 'README, Plots and Reports for CLM regional climate model runs'
(see http://cera-www.dkrz.de/WDCC/ui/BrowseExperiments.jsp?proj=-CLM_regional_climate_model_runs).

When using the model data, please note in particular the following points:

- The model results cannot be interpreted as point related exact values: instead, they represent mean values in space and time with the model resolution (grid spacing and time step).
    - For a spatial analysis it is recommended to use a minimum number of 5x5 grid boxes in homogeneous terrain. Evaluation in highly differentiated terrain demands a special approach, depending on the respective issue. In this case, the user should confer with M&D (data(at)dkrz.de) or SGA (sga(at)dkrz.de).
    - Analysis in time should encompass a time period of at least 15 years. No individual time steps should be analysed, only statistical considerations are recommended.
- To analyse a combination of several model variables in a synergetic way, we strongly advise to use the original rotated model grid (DS2). Since the transformation of CLM output parameters from the original grid to the rectangular grid (DS3) is performed either by bi-linear interpolation or by the nearest-neighbour method, there may arise inconsistencies. This may also occur - although with less effect - by the grid transformation itself (e.g.

due to non-linearity of parameters like relative humidity or height heterogeneity).

- We draw specific attention to the fact that the data were transformed to the regular grid (DS3) without any correction in altitude. In a highly differentiated terrain, this can have an impact on the parameters in the respective grid boxes due to the flattened topography.

- The results on the regular grid (DS3) have been generated as a dedicated product for the end-users to facilitate the access to the CLM results. They are not suitable for the investigation of every possible question.

I. Rights of use

The data producer transfers the simple, non-exclusive and non-transferable right to use the provided data (results of the CLM climate simulations) to the user. The user is not permitted to use the data in unprocessed form, i.e. without ensuing independent products such as evaluations, reports or expertises.

1. Any further disclosure of the data to third parties and any unprocessed further duplication of the data is not permitted.
2. The user undertakes to quote the data according to the guidelines of the WDCC (see http://terms.wdc-climate.de/) in publications based on the provided data material. The exact quotation is stored in the meta data for the CLM data in the WDCC.

II. Liability / warranty

1. The data are made available to the user without any warranty. The user is aware that the data have been obtained according to current state-of-the-art science and computational engineering.
2. The data producer must not be taken into any obligation to third parties on the basis of this agreement. Any liability of the data producer for damage of all kinds resulting from the provision and further processing of the data is ruled out. The user indemnifies the data producer from any liability to damaged third parties.
3. The liability disclaimer stated under II (1) and (2) does not apply insofar as the data producer has acted in gross negligence or with wilful intent.

3. Information about single parameter

A description of all CLM output parameters can be found in the CERA-database (metadata description). Here, a list of those parameters is presented, which need some more explanation and hints at the use of the parameters. This list may not be complete.

ALB, ALB_NN
Surface albedo.
ALB was derived by bi-linear interpolation and ALB_NN by the nearest-neighbour assignment. Surface albedo is set to a maximum 0.7 (snow cover and sea ice) and to a minimum of 0.07 (water). Sea ice is derived from the SST of the MPIOM with the result being transformed to the horizontal resolution of ECHAM5, i.e. approximately 200 km. Thus, Baltic Sea basins are poorly resolved, partly by one grid box, and total melting can be observed via the albedo over such large areas.

ASOB_S
Surface net downward shortwave radiation, i.e. the difference between radiation from above (downward) and radiation from below (upward).
The shortwave radiation balance is strongly influenced by atmospheric pollutants - gases or aerosols. For the CLM runs the aerosol content in the atmosphere is prescribed for rural, urban, desert or sea regions and constant with time.

CLCH
High cloud cover.
Fraction of grid-cell area. High clouds are all clouds at pressure levels $< 400hPa$.

CLCL
Low cloud cover.
Fraction of grid-cell area. Low clouds are all clouds at pressure levels $> 800hPa$.

CLCM
Medium cloud cover.
Fraction of grid-cell area. Medium clouds are all clouds at pressure levels between 800 hPa and 400 hPa.

GPH_*pressure_level*
Geopotential on pressure levels.
Geopotential is the sum of the specific gravitational potential energy relative to the geoid and the specific centripetal potential energy. Due to high orography in the model region, the 1000 hPa, 925 hPa level, and partly also the 850 hPa level are often below the orographic height. Thus, the geopotential of such levels is a fictive value (see also PMSL).

LAI
Leaf area index.
LAI (fractional vegetation coverage) is prescribed for each land use type by an annual course with the maximum (representing the vegetation period) and minimum (repre-

senting the resting period) value of LAI fixed for the given land use type. The function, which describes the annual course, depends on latitude and surface height determining the onset and the duration of the vegetation period.

PLCOV
Vegetation area fraction.
PLCOV is prescribed for each land use type by an annual course with the maximum (representing the vegetation period) and minimum (representing the resting period) value of PLCOV fixed for the given land use type. The function, which describes the annual course, depends on latitude and surface height determining the onset and the duration of the vegetation period.

PMSL
Mean sea level pressure.
Sea level pressure is determined in elevated orographic heights by extrapolating the surface pressure using fictive temperature and humidity values. The error increases with increasing orographic height of the grid box. Therefore, the orography is often reflected in maps of sea level pressure (see also GPH).

PRECIP_TOT
Total precipitation.
The annual sum of precipitation over the coastal areas appears to be too large, particularly over northern Portugal / northwestern Spain where a mean annual sum of more than 2500 mm in the period 1961 - 1990 is simulated. Measured annual precipitation in this region slightly exceeds 1200 mm for the period 1949 to 1995 (Rodriguez-Puebla et al., 1998).
The largest daily sum of PRECIP_TOT in the C20_1-run is 479 mm and occurs in the Mediterranean Sea between Sicily and Tunisia. There, the precipitation maximum is related to a small cyclonic structure.

QV_*pressure_level*
Specific humidity on pressure levels.
Very small negative values of QV may occur at higher atmospheric pressure levels (upper troposphere /lower stratosphere). The reason is the interpolation of small humidity values from model levels to pressure levels which may occasionally produce negative values.
One of the most distinct changes in the scenario runs can be found in high level specific humidity, namely in QV200. The values at the end of the 21. century in the A1B scenario rise to the double of those at the beginning of the century. The increase particularly of summer values with time is possibly due to a rise of the tropopause and thus an upward mixing of humidity up to 200 hPa.

ROOTDP
Root depths is prescribed and constant for each grid box.

SNOW_COV
Number of days with snowcover.

Days with W_SNOW $> 0.01 kg/m^2$ are counted. This parameter gives the ensemble size for the averaging of T_SNOW (see T_SNOW, W_SNOW).

SNOW_DAYS
Number of days with snowfall.
Note that this number is important for the assessment of the monthly and yearly value of T_SNOW (see T_SNOW). Only with a sufficiently high number of snowdays the value of T_SNOW is reliable.

TMAX_2M
Maximum air temperature at 2 m height.
Extraordinary high values above $50\ ^{\circ}C$ occasionally occur for TMAX_2M. At those time periods the temperature values in ECHAM5 are also very high, i.e. partly higher than $45^{\circ}C$ in a grid box 200 km x 200 km.

T_2M
Air temperature at 2 m height.
The histogram of 3 h-values of T_2M in moderate climate regions like Germany exhibits a peak at $0^{\circ}C$, which is presumely due to longer periods of melting snow. Nevertheless, the question remains whether a parameter, which is interpolated from values between the surface and the lowest model level by application of a temperature profile, is so close to the temperature of the melting snow.

T_SNOW
Snow surface temperature.
T_SNOW is set to missing value if $W\_SNOW < 0.01\ kg/m^2$ A mean value is only taken over days with snow in time interval, provided by the parameter SNOW_COV. Mean values of T_SNOW should be neglected where the ensemble size is too small.
Occasionally a snow temperature of $> 0^{\circ}C$ occurs in isolated grid boxes. The reason for a snow temperature higher than the melting point is that while a water equivalent $W\_SNOW \geq 0.01 kg/m^2$ is interpreted as snow for a grid box (mostly on the edge of a snow-covered area), the surface temperature is $> 0\ ^{\circ}C$ due to exchange processes. In the next time step the temperature will be corrected to $0\ ^{\circ}C$ by the release of heat of fusion and the snow cover is going to vanish, i.e. W_SNOW gets a quantity $< 0.01\ kg/m^2$. If no further snowfall occurs at this grid box during the averaging period, the monthly or even yearly averages have an apparent snow temperature higher than the melting point. The validity of T_SNOW can only be assessed regarding the parameter W_SNOW and SNOW_COV.

T_SO_*depth*
Soil temperature in soil layers.
Note: soil temperature is the bulk temperature of the soil, not the surface (skin) temperature. The dataset acronym indicates the lower bound of the soil layer while the data values themselves are representative for the whole layer. All soil parameters are stored once a day at midnight. This results in rather low temperature values in the upper layers (_0001, _0004) and partly higher values at medium layers (_0010, _0022, _0046, _0084) where the diurnal temperature cycle is retarded. The vertical profiles

are only typical for night-time hours and do not reflect the day-time heating of the upper layers. This is also true for monthly and yearly mean values.

VMAX_10M
Wind speed of gust at 10 m height.
VMAX_10M is determined as the maximum value of turbulent gusts and convective gusts. Turbulent gusts are "derived from the turbulence state in the atmospheric boundary layer", convective gusts are "estimated from the wind speed which is transported by the downdraft from higher levels to the ground and the wind speed associated to the downdraft itself" (Schulz and Heise, 2003). Convective gusts are restricted to a maximum value of $30m/s$. This value is sometimes reached in CLM scenario simulations.

WIND_SPEED
Wind speed at 10 m height.
Post processed value, calculated from the vector components as the mathematical norm: $WIND\_SPEED = \sqrt{U\_10M^2 + V\_10M^2}$.
The deceleration of wind speed from sea to land appears to be too strong. An example is the mistral which is well represented in the model over the Gulf of Lion. But there is almost no enhancement in wind speed over land, i.e. within the Rhone delta and in the channel between Massif Central and Pyrenees. It is recommended to make more comparisons to ensure that the wind field is realistically simulated in coastal regions, particularly with respect to wind energy issues.

W_ICE_*depth*
Soil frozen water content in soil layers
Frozen subset of the total water content of the soil W_SO, i.e. the thickness of the liquid water equivalent of the frozen water content of each soil layer. The *'depth'* part of the acronym indicates the lower bound of the soil layer while the data values themselves are representative for the whole layer.
The general warming predicted in the scenarios is clearly manifest in the strong reduction of frozen soil water content in the model region at 94 cm depth.

W_SNOW
Surface snow amount.
Liquid water equivalent of the vertical extent of snow having the same mass per unit area. Surface amount refers to the amount on the ground, excluding that on the plant or vegetation canopy.
$W\_SNOW < 0.01\ kg/m^2$ (liquid water equivalent of 1 cm snow cover in grid box) is used to mask out T_SNOW and to count the days with snow (SNOW_COV) in time interval (monthly and yearly). (see T_SNOW, SNOW_COV)