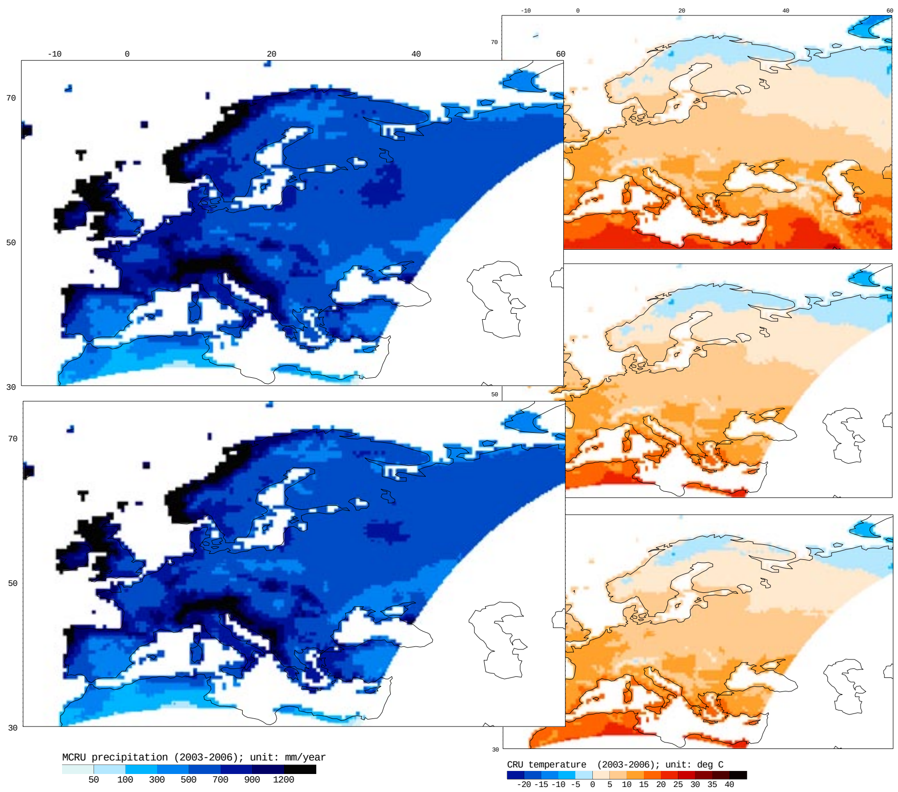
Max-Planck-Institut für
Biogeochemie

TECHNICAL REPORTS
15



Constructing a consistent historical climate data set for the European domain

by
Y. Chen, G. Churkina, M. Heimann

ISSN 1615-7400

MAX-PLANCK-GESELLSCHAFT

Technical Reports - Max-Planck-Institut für Biogeochemie 15, 2009

Max-Planck-Institut für Biogeochemie
P.O.Box 10 01 64
07701 Jena/Germany
phone: +49 3641 576-0
fax: + 49 3641 577300
http://www.bgc-jena.mpg.de

Constructing a consistent historical climate data set for the European domain

Youmin Chen, Galina Churkina and Martin Heimann
Max-Planck-Institute for Biogeochemistry

Constructing a consistent historical climate data set for the European domain

Youmin Chen, Galina Churkina and Martin Heimann
Max-Planck-Institute for Biogeochemistry

1. Introduction

The meteorological environment represents one of the most important factor of ecosystem functioning and constitutes a key input driver for ecological models. Therefore, in order to perform continental scale ecosystem model simulations, time series of gridded meteorological fields are needed. Since in situ meteorological observations are heterogeneous in space and time, state of the art analyses in conjunction with dynamical general circulation models (weather forecast models) can be used to produce internally consistent gridded fields of the relevant meteorological variables. Here we first compare several of such meteorological data products of various origins in view of using them as drivers of gridded ecosystem models at the continental scale. In particular we compare:
(1). CRU: i.e. CRU TS 2.1 dataset, which consist of the monthly meteorological fields compiled by the Climate Research Unit of the University of East Anglia, Norwich, UK. (http://www.cru.uea.ac.uk/cru/data/hrg.htm)
(2). ECHAM5: Meteorological fields are simulated by the global climate model ECHAM5 driven by observed changes in greenhouse gas and aerosol forcing (http://www.mpimet.mpg.de/en/wissenschaft/modelle/echam/echam5.html).
(3). ECMWF and NCEP: Reanalysis data compiled by the weather forecasting centers of the European Center for Medium-Range Weather Forecast (ECMWF, 1997) and the National Center for Environmental Prediction (Kalnay, E., et al. 1996,) respectively.
(4). REMO: High-resolution meteorological fields from a regional mesoscale model simulation forced at the boundaries by NCEP reanalysis (Feser et al., 2001).

Through the data comparison it is found that no any single data set could satisfy all our purpose as ecosystem modeling driver. Normally for ecosystem model simulation the daily data covering long time period is needed. However, the data sets with daily data available usually cover the short time periods such as REMO (1948-2007) and ECMWF (1958-2001). CRU data set covers relatively long time period (1901-2002) but only the monthly data is available. The data set ECHAM5 could supply longer time daily data (1860-2000), but it is purely model simulation and the data is not well comparable with the in-situ observation. Therefore our goal is to develop climate data sets satisfying the following conditions: (1) both daily and monthly data are available; (2) the data is comparable with the in-situ observation; (3) the data cover long time period.

Below are work steps we followed in making the new data set called “MCRU”:

1. **Based on available CRU variables calculate new variables**: Relative Humidity, Daylight Temperature and Daylight vapor pressure deficit and Solar radiation

2. **Re-map** variables from CRU dataset and variables obtained at step 1 from 0.5 degree to 0.25 degree grids
3. **Define common grids** between variables obtained at step 2 and REMO variables
4. **Establish regression** between variables obtained in step 3 and REMO variables and extend variables obtained in step 3 until 2007
5. **Construct monthly values** for three periods: 1861-1900; 1901-2002; and 2003-2007.
6. **Calculate daily values** from monthly values obtained at step 5 using ECHAM5 daily variability for 1861-1947 and REMO daily variability for 1948-2007

A new data set was generated following the above mentioned steps, the name of which is "MCRU", e.g. the Modified CRU data set. The monthly values of MCRU variables are equal to the respective values from CRU data set when the CRU data is available (1901-2002). The MCRU monthly values for 1861-1900 and 2003-2007 are calculated from other sources but consistent with CRU data. The early MCRU monthly values (before 1901) are computed from ECHAM5 data set and later MCRU monthly values (after 2002) are the regression results between REMO and CRU data sets. In addition, we have to manage to generate some variables, which are not directly available from CRU data set but could be calculated through the existing CRU variables, which we still call CRU variables such as relative humidity, daylight temperature, vapor pressure deficit and radiation.

2. Step 1: Calculating relative humidity, daylight temperature, vapor pressure deficit and radiation

From the CRU archive there are variables as temperature, precipitation, cloudiness and vapor pressure. Other variables such as relative humidity, daylight temperature, daylight vapor pressure deficit as well as radiation are not directly available. They have to be calculated from the existing variables based on the standard formula or on algorithm used in ecosystem models.

Relative Humidity (RHY)

Formula: RHY=100*(e/es)

e=es0*exp(lv/Rv*(1/T0-1/Td)) -vapor pressure

es=es0*exp(lv/Rv*(1/T0-1/T)) -saturated vapor pressure

T- absolute temperature [K]

Td - absolute dewpoint temperature [K]

T0: T0=273.15 [K] – temperature constant

es0: es0=6.11 [hPa] – vapor pressure constant

lv: lv=2.5e6=2.5*10^6 - gas constant for air

Rv: Rv=461.5 [joules*Kelvin/kilogram] - gas constant for water vapor

Daylight Temperature (tday)

tday=((tmax-tmean)*0.45)+tmean

tmax: daily maximum temperature [◦C]

tmean: daily mean temperature [◦C)]

Daylight Vapor Pressure Deficit (VPD)
Formula: VPD=610.7*(exp(17.38*tday/(239.0+tday))-exp(17.38*tdew/(239.0+tdew))
tday=((tmax-tmean)*0.45)+tmean [◦C]
tmax - daily maximum temperature [◦C]
tdew - daily dewpoint temperature [◦C]
tmean - daily mean temperature [◦C]

Radiation (RAD: surface solar radiation downwards)
The radiation variable is calculated using algorithm of ORCHIDEE or LPJ models based on the existing CRU variables such as the temperature, cloudiness and so on. After the comparison of the radiation calculated both models, radiation calculated by ORCHIDEE model is chosen as the part of the “MCRU” data sets since the radiation pattern from ORCHIDEE looks more similar to REMO radiation pattern (figure 12).

Step 2: Re-map climate variable from 0.5 grid to 0.25 grid
The CRU data has 0.5/0.5 degree spatial grids over the land only. The REMO and ECHAM5 data have been interpolated into 0.25/0.25 degree spatial grids. There is a mismatch between REMO grids and CRU grids. We remap the CRU data from 0.5/0.5 grids to 0.25/0.25 grids in a simple way, i.e. simply divide one CRU pixel into four pixels with same value so that each of the sub-pixels has 0.25/0.25 degree interval.

3. calculating MCRU variables by data combination

(3.1) Temperature

(3.1.1) spatial patterns

Figure 1a and 1b show temperature patterns of 20 year average from 1980 to 1999, respectively for CRU data and MCRU data. The slight differences in the temperature patterns between CRU and MCRU datasets are explained by different grid sizes, one is 0.5 degree and another one is 0.25 degree. The figure 1(c) shows the MCRU temperature pattern for the period from 2003 to 2006, when the MCRU data was generated by regression between CRU and REMO. The similarity of MCRU patterns between early time (1980-1999) and later time (2003-2006) indicates that the regression of temperature from CRU data and REMO data gives a quite good result in term of the multiple averages

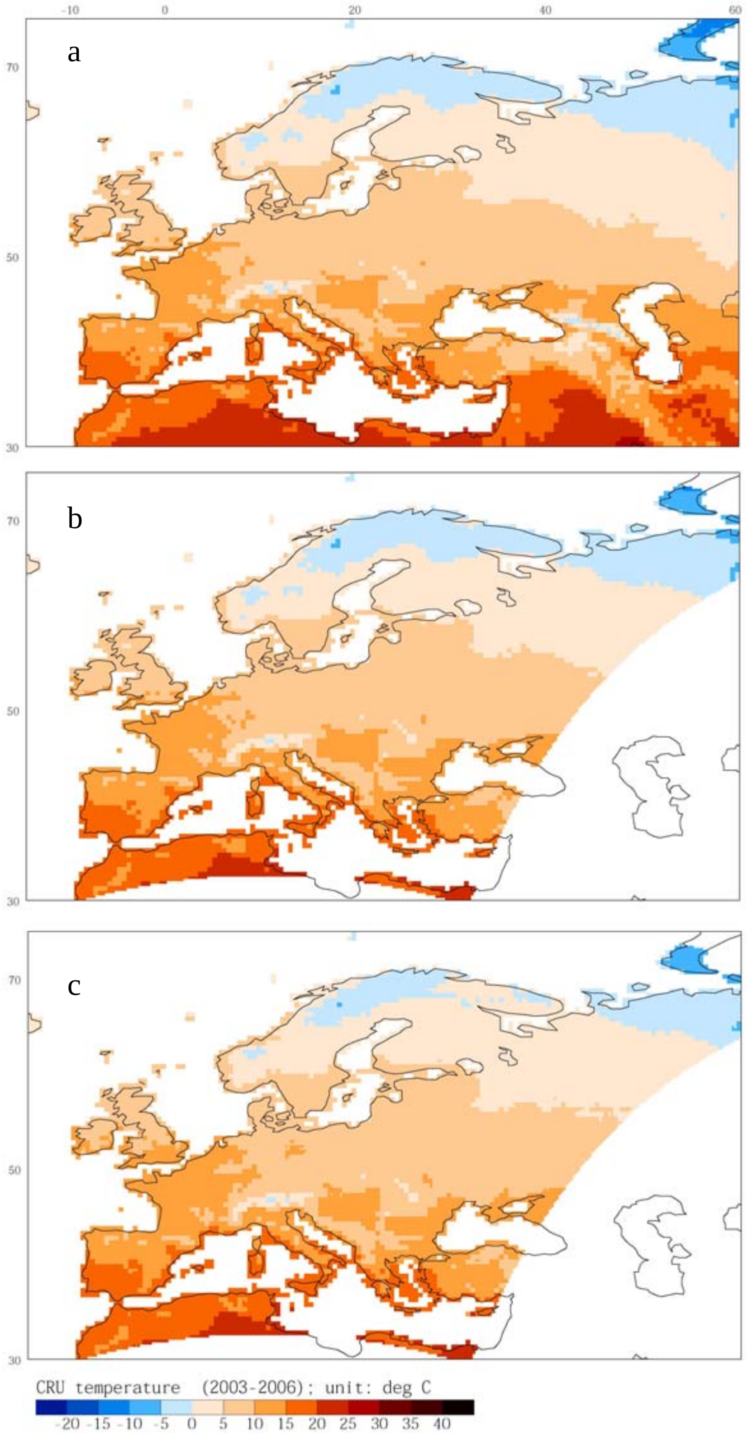


Figure 1: Spatial patterns of temperatures from (a) CRU averaged over 20 years (1980-1999); (b) MCRU averaged over 20 years (1980-1999); (c) MCRU averaged over four years (2003-2006)

(3.1.2) regression validation

The MCRU variables after the year 2002 have been obtained as the result of regression between REMO and CRU variables. We used 30 years of climate data (1971-2000) to develop the regression equations. The regression was performed for each pixel and each month separately. Therefore there are total 24531 valid pixels for MCRU data sets, which lead to 294372 (23531*12 months) regression equations. Generally the regression correlation coefficient is used to analyze these regression equations. So we have the same number of correlation coefficients as for that of regression equations for regression validation. We used the histogram to see how the regression correlation coefficients are distributed (Figure 2). Most of the correlation coefficients are larger than 0.8, which tell that the regression of temperature give good results. Finally the monthly MCRU data is constructed, in which the data from 1901-2002 are exactly the CRU data; the data before 1901 are modified ECHAM5 data (see about radiation variable above?); and the data after 2002 are regression results between REMO and CRU.

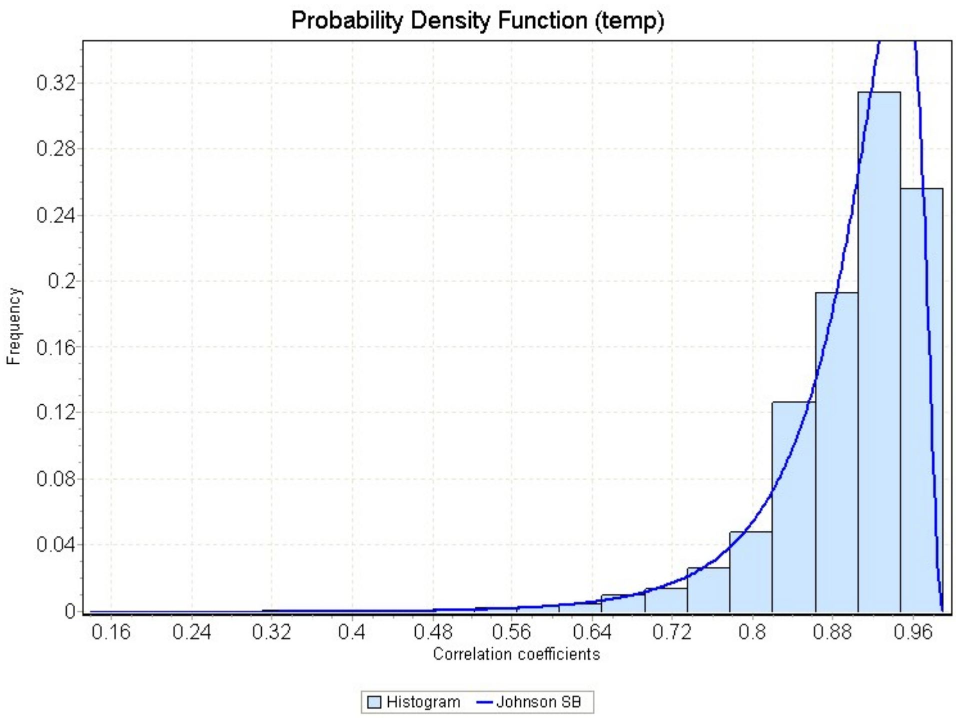


Figure 2: The histogram of correlation coefficients for temperatures for 2003-2007. The smooth line was fitted by Johnson SB method.

(3.1.3) daily data construction

Based on the monthly MCRU climate data we construct the MCRU daily data, in which the REMO or the ECHAM5 daily data will be used together. When we use REMO daily data to construct MCRU daily data, first the difference between monthly CRU and REMO variables is defined. Then the REMO daily data will add up the monthly difference, which leads to the MCRU daily data. The daily data in specific month will add up the same monthly difference and the neighbor months will have the different monthly values to be added up. Therefore it could make abrupt change between the neighbor months. We assume that these sorts of abrupt changes are small for three reasons: the scale of monthly difference is much smaller than the daily variability; the monthly differences between different datasets are small; the monthly differences between the neighbor months are small. As long as one reason is reasonable we would generate the reasonable MCRU daily data.

The daily data construction is illustrated below:

$$diff = monthly_{CRU} - monthly_{REMO(ECHAM5)}$$

$$daily_{MCRU} = daily_{REMO(ECHAM5)} + diff$$

Figure 3 shows the weighted annual average of temperature for Europe. There are two curves close to each other: one is from the original MCRU monthly data and another one is from the daily data. The red curve is from REMO data which shows the difference between monthly REMO and CRU (same as MCRU) temperatures is quite small.

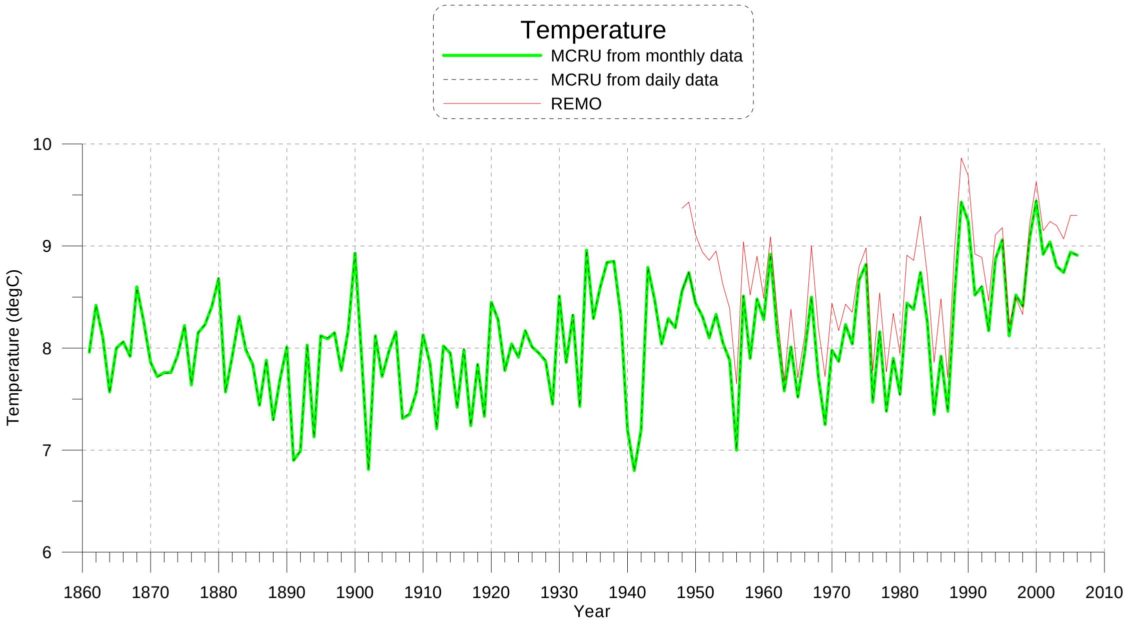


Figure 3: The area weighted average of MCRU (green and dash lines) and REMO (red) temperatures for Europe. The dash line is the weighted average from the constructed daily data.

(3.1.4) data comparison and annual variation

Figure 4 shows the weighted average temperature of Europe as in the figure 3, but with more data sets such as CRU, ECHAM5, ECMWF, NCEP and REMO. ECHAM5 shows substantial inconsistence with all other data sets. The annual temperatures of Europe from other data sets are comparable with each other.

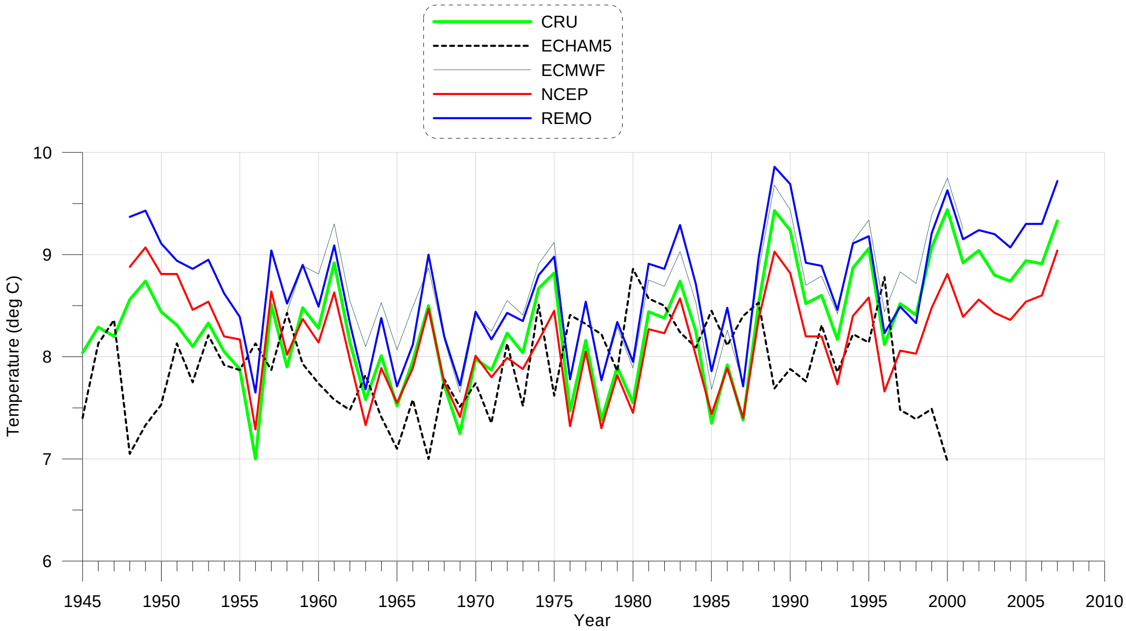


Figure 4: The comparison of area weighted average temperatures from CRU, ECHAM5, ECMWF, NCEP and REMO data set for Europe.

(3.1.5) daily data illustration

Validation of spatial climate datasets with site observations

The data validation above is based on monthly data; however the final data set is at the daily scale. It is difficult to validate all the daily data since it is too much of work because of the large amount of data. On the other hand, the observed daily data is not always available for validating the simulated data set. Here we use 10 year daily data (1980-1989) from meteorological station in Jena to compare data with the corresponding REMO and MCRU data.

First we used the quartile-quartile plot (q-q plot) for comparing the daily distribution of temperature (figure 5). The x-direction is about REMO or observation (OBS) data and y-direction is about MCRU or REMO data. These three q-q plots are going along the diagonal, which means the data from MCRU, REMO and observations have similar distribution.

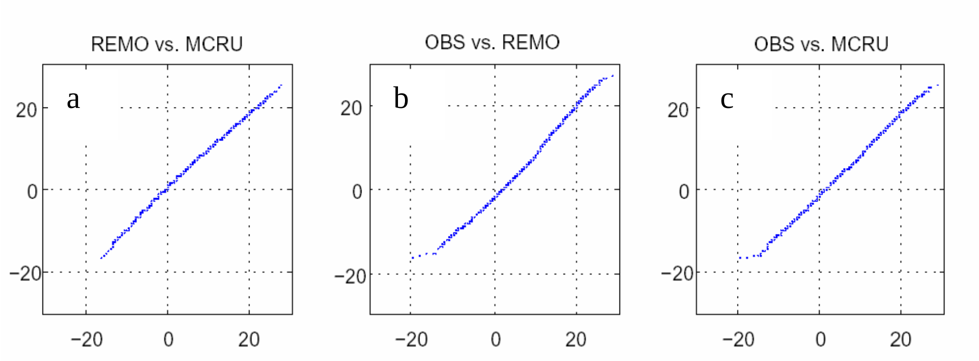


Figure 5: The quartile-quartile plot (q-q plot) between REMO, MCRU and OBS (observation) based on 10 year daily temperature data from 1980 to 1989. (a) REMO and MCRU; (b) OBS and REMO; (c) OBS and MCRU.

Based on the 10 year daily data we made 10 year average for each day (figure 6) which shows the mean seasonal variation of the variables. There are a bit different values for the 10 year average but the temporal variation is quite similar between OBS, REMO and MCRU. Three data sets show the regular temperature drop in the beginning of January (day 10) and the end of February (day 50).

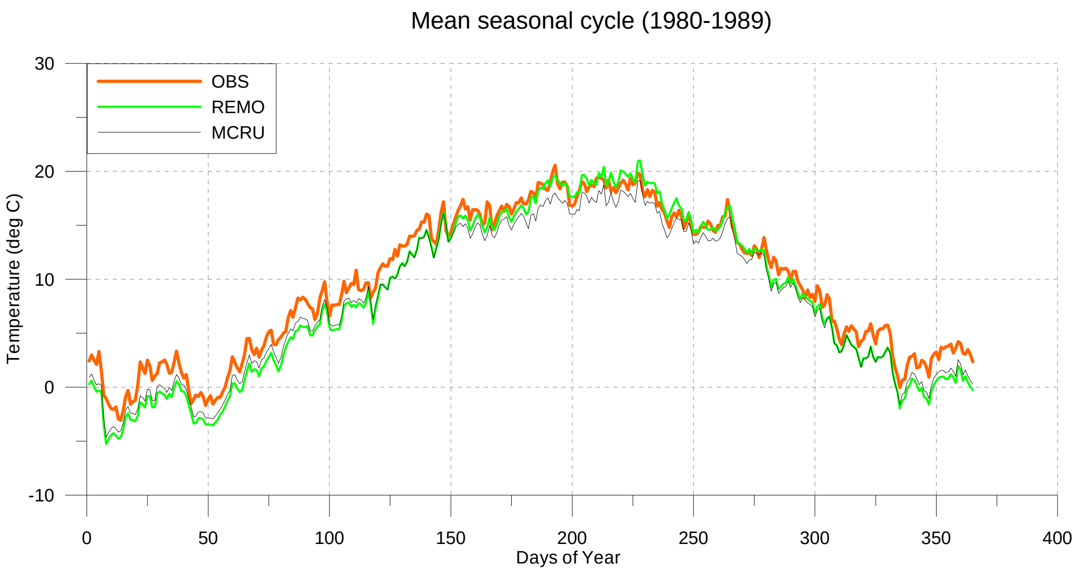


Figure 6: The seasonal cycle of temperatures from OBS, REMO and MCRU data sets. Each data point is calculated as the ten year daily averages from 1980 to 1989.

(3.2) precipitation

Precipitation is one of the most complicated variables because it has discontinuous spatial patterns and the sum instead of average is often used for analysis. When making precipitation data, one cannot simply add up component as it is done for temperature. Instead, the ratio between different components is used to construct the precipitation data.

(3.2.1) spatial patterns

Figure 7a shows the MCRU precipitation pattern computed as 20 year average (1980-1999) over European domain. It shows several larger precipitation centers, such as Norwegian coast, along Alps, UK coast and Portugal coast. Figure 7b depicts also the average precipitation pattern but for 2003 - 2006, when the regression was used to extend CRU data set as MCRU data. Since precipitation patterns on Figure 7a and Figure 7b look similar, we conclude that using regression to extend precipitation data provides rather good spatial pattern results.

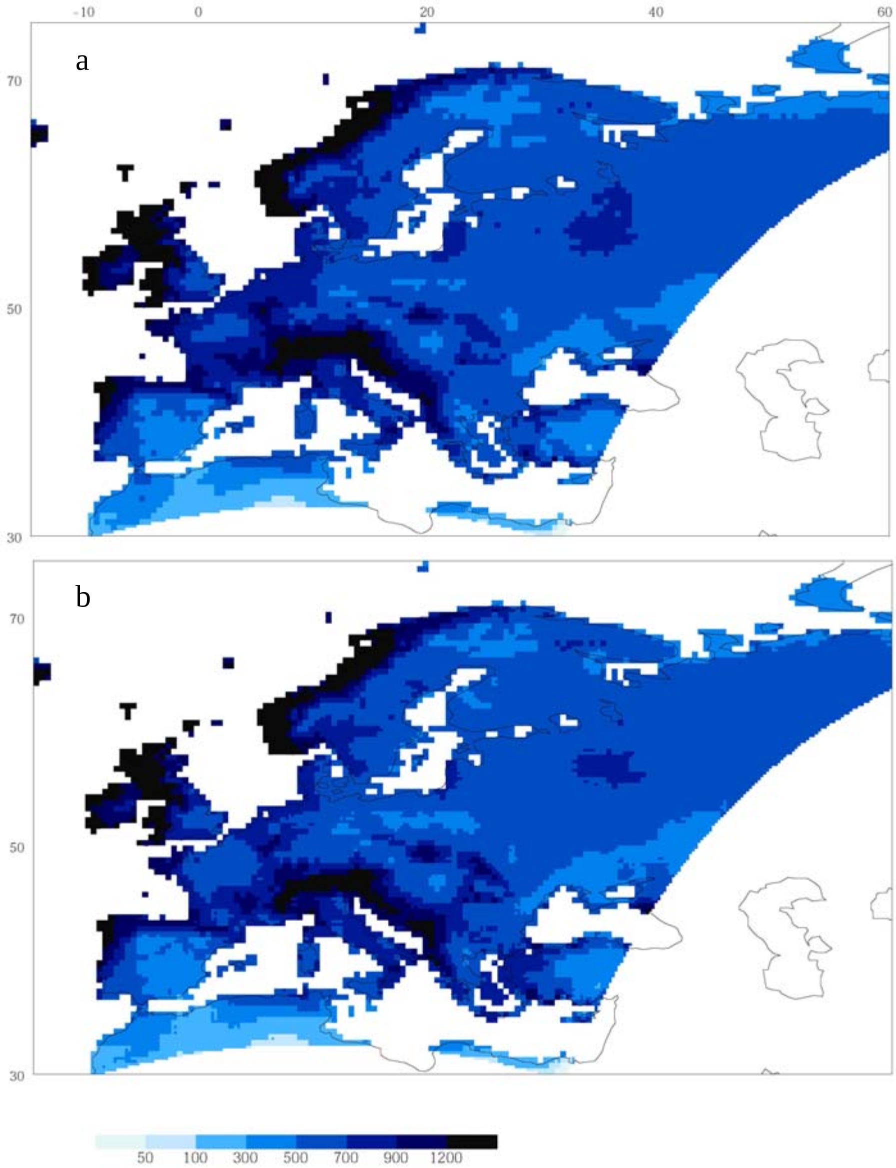


Figure 7: Spatial patterns of annual precipitation from MCRU dataset: (a) averaged over 20 years (1980-1999); (b) averaged over four years (2003-2006).

(3.2.2) regression validation

To check the quality of precipitation data for 2003-2006, the correlation coefficients are analyzed for validating the regression equations. Similarly as for temperature variable, there are total 294372 correlation coefficients, and the histogram is used to see the distribution of the correlation (figure 8). Most of the correlation coefficients are concentrated around 0.7 and 0.8, that indicates that the regression gives good results for majority of the pixels. However, for some pixels the correlation coefficients close to zero. When the correlation coefficient is very small, or near zero, the regression equation will take the value of the multiple year average as the regression results. Thus, it is assured that the average pattern from regression results will always be similar to the original average pattern, no matter how good are the regression coefficients. In case the regression coefficients are small, there will be no or little temporal variation since every year's values were taken from multiple year average.

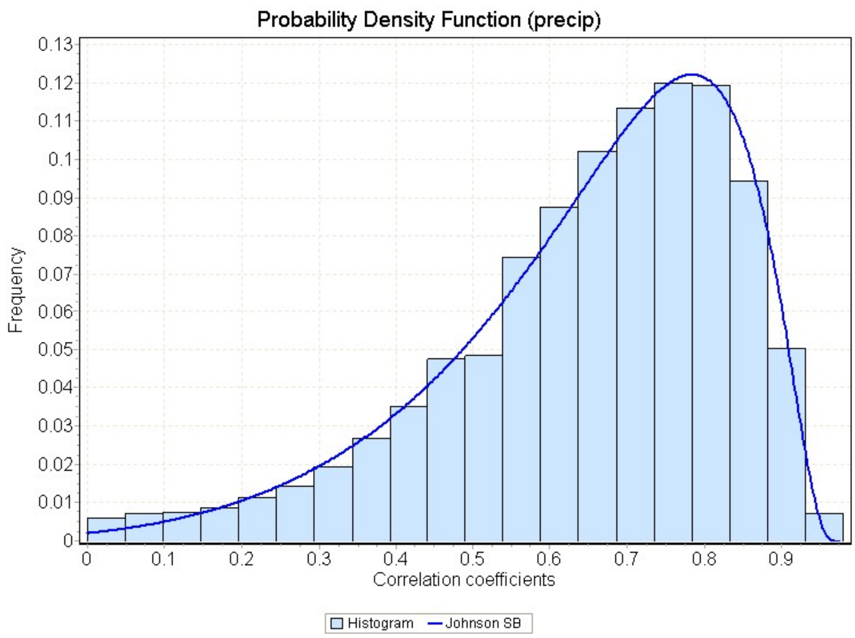


Figure 8: The histogram of correlation coefficients of regression for extending precipitation data for 2003-2007. The smooth line is the fitting by Johnson SB method.

(3.1.6) daily data construction

As it is mentioned above that the ratio will be used to calculate the precipitation. The *ratio* is defined as:

$$ratio = (monthly_{CRU} - monthly_{REMO(ECHAM5)}) \Big/ monthly_{REMO(ECHAM5)}$$

$monthly_{CRU}$ – monthly precipitation from CRU data set.
$monthly_{REMO\ (ECHAM\ 5)}$ - monthly precipitation data from REMO (or ECHAM5) data set.

The monthly precipitation are often very small in the arid and semiarid regions, which lead to the very large ratio. In this case an unrealistically large precipitation would be calculated. To prevent it from happening, an artificial control is used to calculate the ratio, which is as:

if $monthly_{\ REMO\ (ECHAM\ 5)}$ <= 1mm then *ratio*=0

It constrains the amount of precipitation with a reasonable value. Once the ratio based on monthly data is calculated, the MCRU daily data can be computed as:

$$daily_{MCRU} = daily_{REMO(ECHAM\,5)}(1 + ratio)$$

$daily_{MCRU}$ – daily precipitation from MCRU dataset
$daily_{REMO(ECHAM5)}$ – daily precipitation from REMO or ECHAM5 dataset

When the computed daily precipitation is negative, it is forced to zero, i.e.

$$(daily_{MCRU} \geq 0)$$

Although the area weighted average for precipitation from MCRU monthly data and from constructed daily MCRU data mostly coincide well, some mismatch exists sometimes (Figure 9). The precipitation mismatch occurs because of the artificial control of monthly ratio and negative daily values. The red curve is from REMO data, which shows significant disagreement of precipitation data for CRU and REMO data sets.

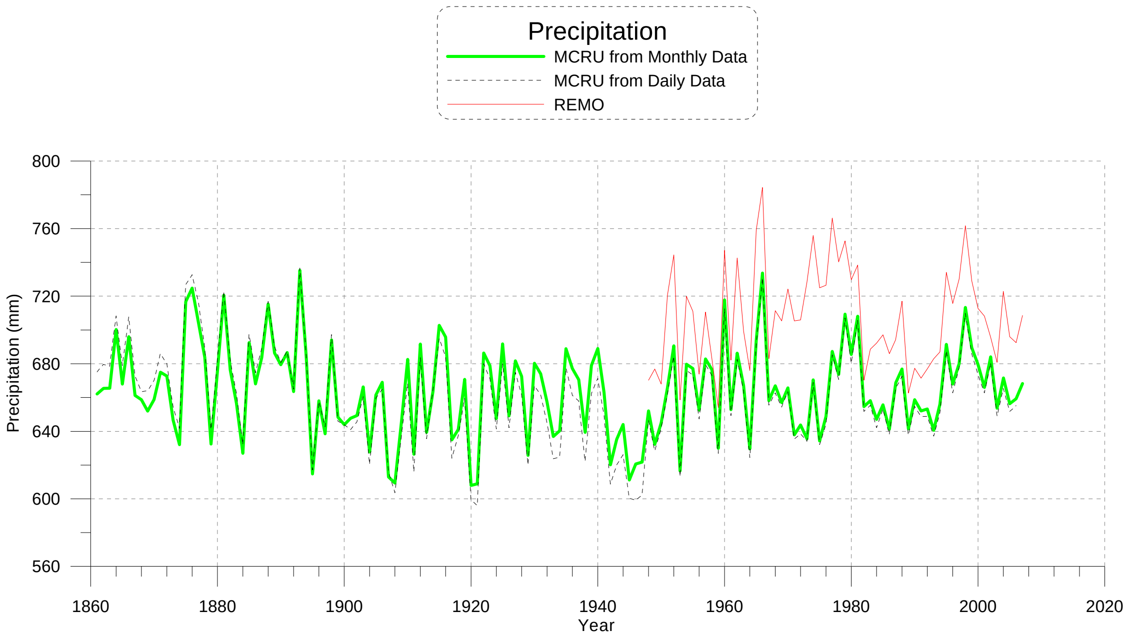


Figure 9: The area weighted average of MCRU (green and dash lines) and REMO (red) precipitation. The dash line is the weighted average from the constructed daily data.

(3.2.3) data comparison and annual variation

Contrary to temperature data comparison, the comparison of area weighted precipitation data from CRU, ECHAM5, ECMWF, NCEP and REMO displays much large difference between different data sets (Figure 10). NCEP data shows the decline of precipitation. Annual precipitation from ECMWF is substantially smaller (up to 100 mm per year) than from the other data sets. CRU precipitation is located in the middle of all the data sets and shows stable temporal variation without clear trend and abrupt changes during the displayed time period. Nevertheless the inter-annual variability of precipitation from most datasets is similar, except for ECHAM5 data set.

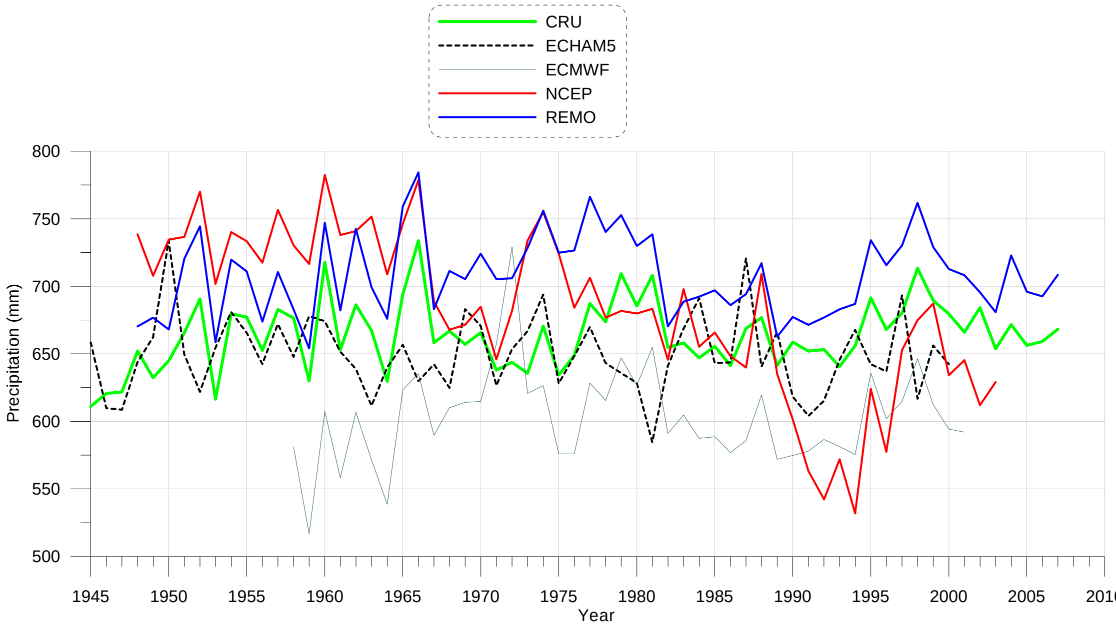


Figure 10: The precipitation data comparison with the area weighted average from CRU, ECHAM5, ECMWF, NCEP and REMO data sets respectively.

(3.2.4) Detrending of ECHAM5 precipitation

To produce MCRU data set for 1861-2002 we combined the precipitation from ECHAM5 for 1861-1900 and CRU for 1901-2002. First, monthly precipitation from ECHAM5 was shifted based on the difference between precipitation from CRU and ECHAM5 for 30 year average (1971-2000). This shift revealed inconsistencies between precipitation from ECHAM5 and CRU data sets. The area weighted average precipitation from ECHAM5 had negative trend, while precipitation from CRU had positive trend. Although both trends were not significant in term of 5% significance, when one combined these two data sets together, there was a “V” shape before and after the connection point, which may not be realistic. Since we had more confidence in precipitation from CRU data sets, precipitation from ECHAM5 was detrended (Figure 11a). Apparently, the precipitation for 1901-2007 was the same and the resulting precipitation for 1861-1900 was different

(Figure 11b). The detrended ECHAM5 precipitation is better, though the “V” shape is still noticeable in year 1900.

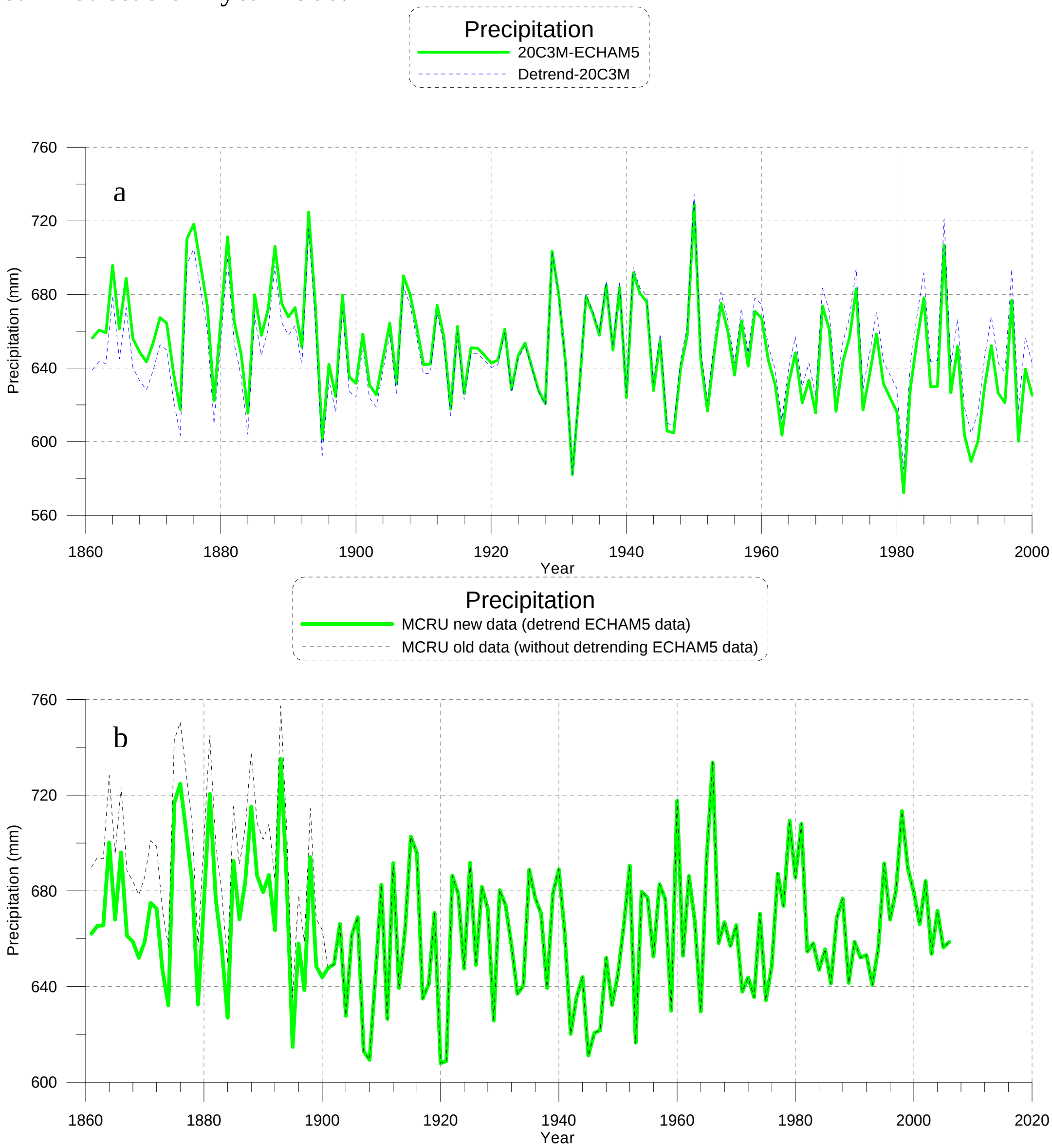


Figure 11: (a) The comparison of ECHAM5 raw data and detrended data in area weighted average and (b) the comparison of MCRU precipitation data constructed with raw (dish line) and detrended ECHAM5 data (green line).

(3.3) Solar radiation

The radiation data used in this study is called the “surface solar radiation downwards (SSRD)”, which is one of the important forcing factors for ecosystem model. It is equal to the net solar radiation plus reflected short wave radiation. There is no radiation data directly available from CRU data. However, some ecosystem models could calculate the radiation based on the existed CRU variables, e.g. temperature and cloudiness. We used algorithms of two ecosystem models to calculate radiation, i.e. ORCHIDEE and LPJ, with the corresponding radiation data sets as ORCHIDEE_CRU and LPJ_CRU respectively. The unit of radiation data is W/m^2.

(3.3.1) spatial patterns of solar radiation

The ORCHIDEE_CRU and REMO radiation have similar spatial patterns (figure 12a and 12c) and the also similar temporal variation (figure 17 and figure 18). The LPJ-CRU radiation has erroneous spatial patterns, with radiation decreasing from south-west to north-east (Figure 12b). We used ORCHIDEE algorithm for estimation of radiation in MCRU dataset.
The MCRU radiation averaged over four years (2003-2006) is the regression result between ORCHIDEE_CRU radiation and REMO radiation. The MCRU radiation patterns averaged over four years had similar structure to the 20 year (1980-1999) average pattern (Figures 12d and 12e).

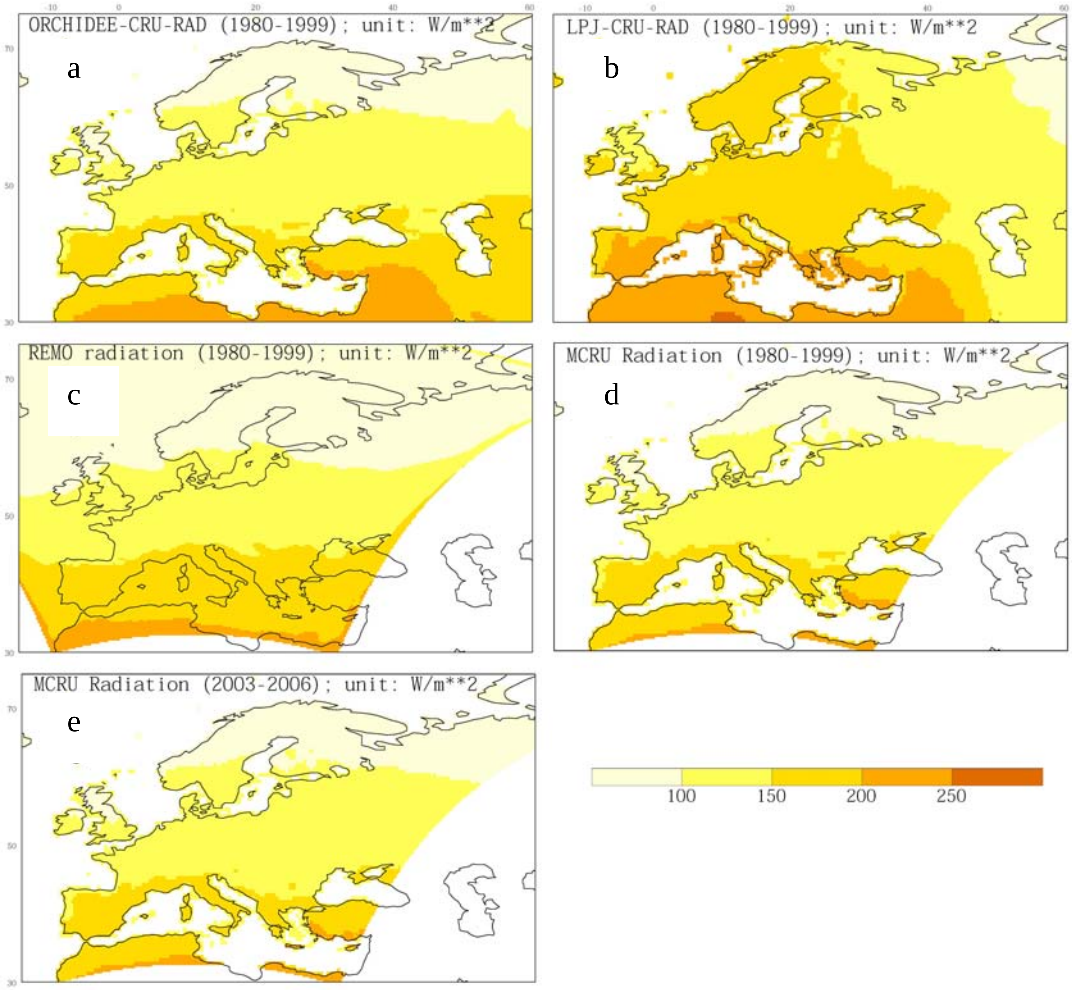


Figure 12: The radiation spatial patterns in 20 year average of ORCHIDEE-CRU (a); LPJ-CRU (b); REMO (c); MCRU (d), as well as MCRU pattern (e) with 4 year average (2003-2006).

(3.3.2) validation of regression

The MCRU radiation after 2003 is the result of regression between REMO radiation and ORCHIDEE_CRU radiation (294372 regression equations).The regression results for radiation (Figure 13) are worse than for temperature and precipitation (Figure 2 and 8). Though most of the correlation coefficients concentrated around 0.7, there are many pixels with correlation coefficients close to zero.

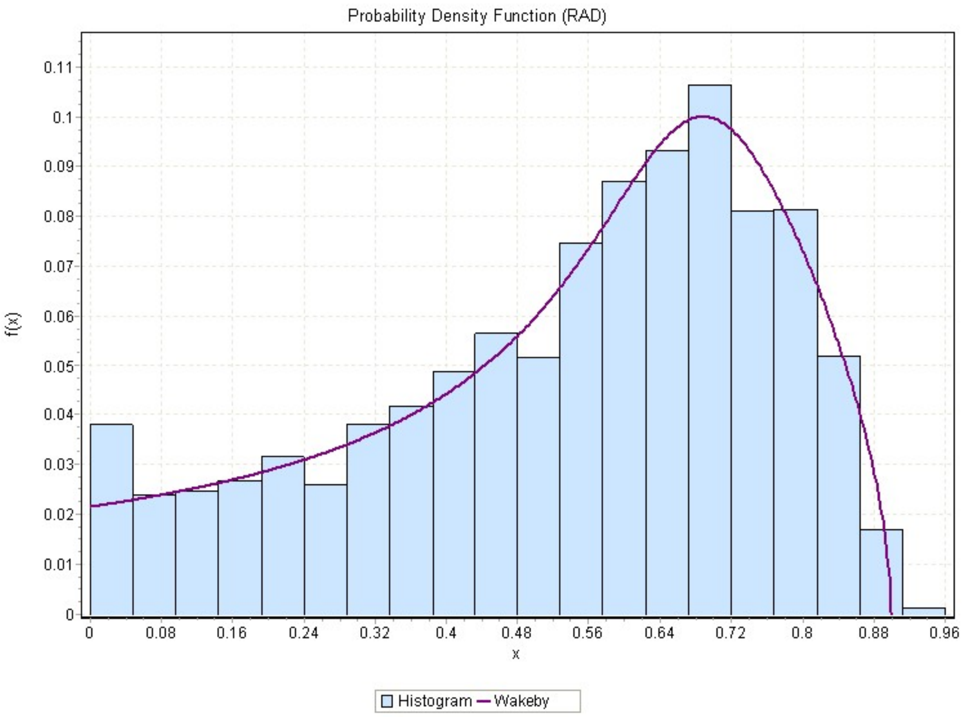


Figure 13: The histogram of correlation coefficients of regression for extending radiation data for 2003-2007. The smooth line is the fitting by Wakeby method.

(3.3.3) Step 5: Modifying ECHAM5 data

To construct the MCRU monthly data, we modified ECHAM5 data for the time period from 1861 to 1900. We use radiation variable as an example to explain how the ECHAM5 data was modified. Figure 14 shows of MCRU, REMO and ECHAM5 radiation data. The difference between the area weighted average radiation from MCRU and ECHAM5 is quite large (Figure 14). The MCRU data for the period form 1861 to 1900 was calculated from the modified ECHAM5 data, so both MCRU and ECHAM5 data have the same annual variation during that period. First we calculate difference (CRU-ECHAM5) between monthly CRU (MCRU) and ECHAM5 variables based on the time period from 1971 to 2000. Afterwards, all the ECHAM5 data from 1861-1900 would be added up the difference (CRU-ECHAM5), which will become the MCRU data.

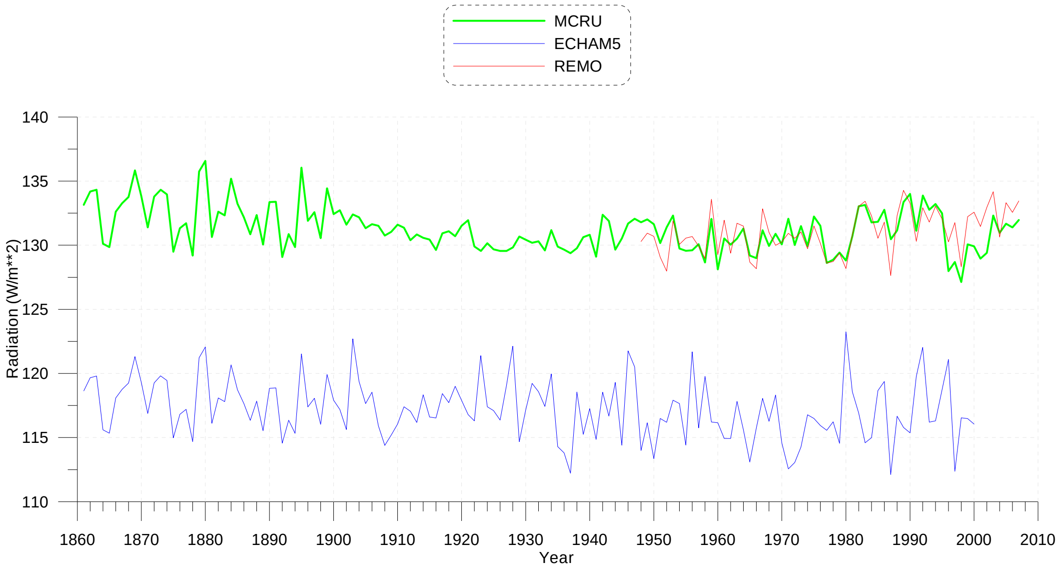


Figure 14: The radiation data comparison between MCRU, ECHAM5 and REMO, which show the early MCRU data (1861-1900) is the shifted result of ECHAM5 data.

(3.3.4) Step 6: daily data construction

Like the temperature variables, radiation is also a continuous variable, so that similar method as for temperature is used to construct the daily radiation data. First the difference of monthly values is calculated as:

$$diff = monthly_{CRU} - monthly_{REMO(ECHAM5)}$$

$monthly_{CRU}$ - monthly temperature or radiation from CRU dataset
$monthly_{REMP(ECHAM5)}$- monthly temperature or radiation from REMO or ECHAM5 datasets

Then the daily values is as

$$daily_{MCRU} = daily_{REMO(ECHAM5)} + diff$$

$daily_{CRU}$ - daily temperature or radiation from CRU dataset
$daily_{REMP(ECHAM5)}$- daily temperature or radiation from REMO or ECHAM5 datasets

The only difference with temperature calculation is that daily values have to be larger than zero, i.e. $daily_{MCRU}$ >=0.

MCRU radiation data from monthly data and from constructed daily data coincide very well (Figure 15). The red curve shows annual radiation from REMO, which has good agreement with MCRU radiation.

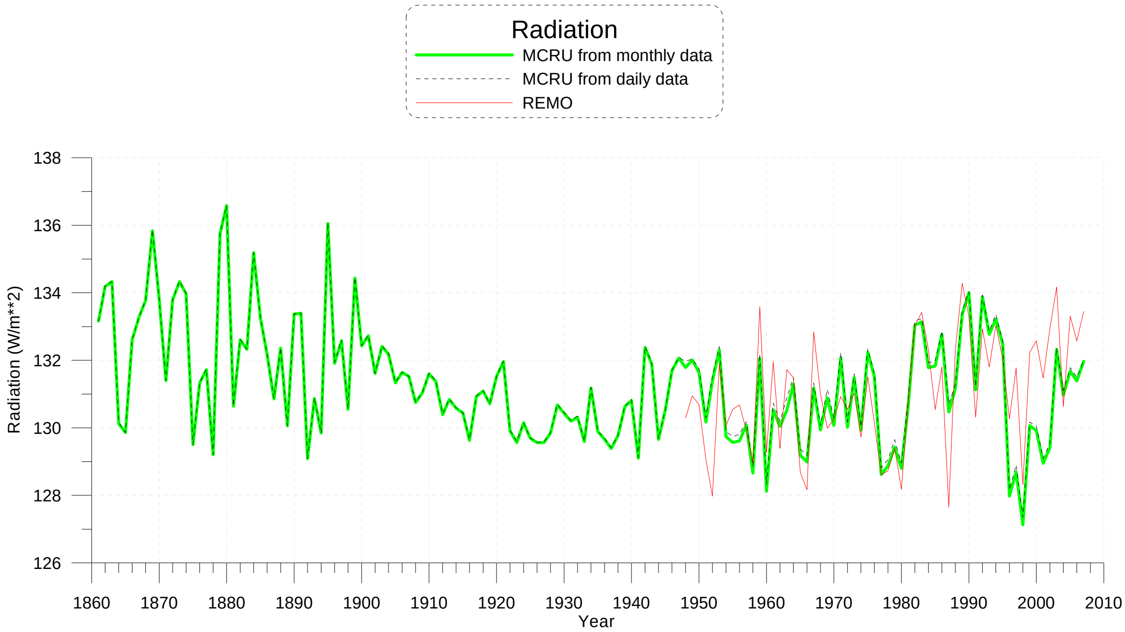


Figure 15: The area weighted average of MCRU (green and dish lines) and REMO (red) radiation. The dish line is the weighted average from the constructed daily data.

From 1901 to 1940 one can see the unusual small variability of the annual radiation variation (figure 14). To find the possible reason, we made a similar plot with cloudiness data (figure 16). The MCRU radiation data is calculated based on CRU cloudiness data using algorithm from ORCHIDEE model, so that the radiation data characteristics come from CRU cloudiness data Radiation data has small variability similar to cloudiness data between 1901 and 1940. Radiation and cloudiness show the opposite variation between them, i.e. higher (lower) cloudiness corresponds to lower (higher) radiation.

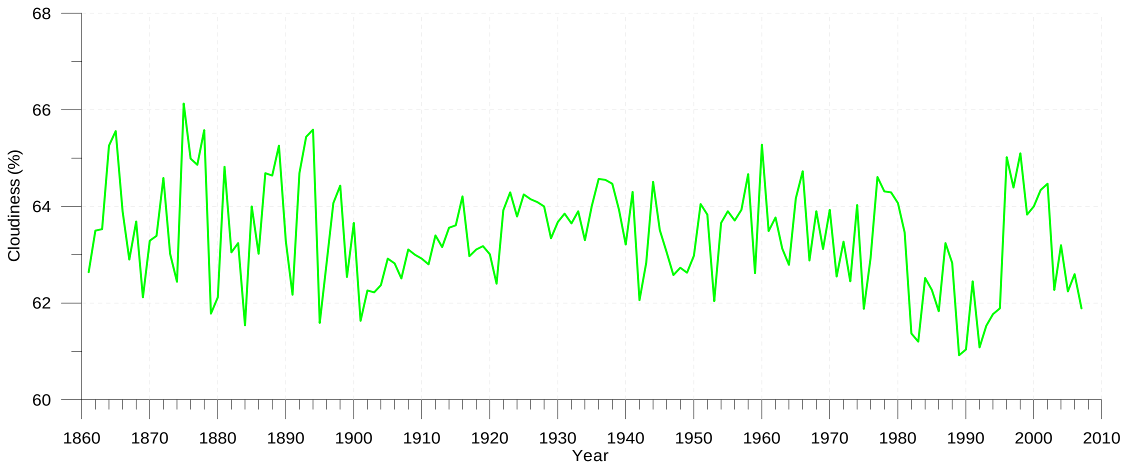


Figure 16: The area weighted average of MCRU cloudiness, which shows the opposite variation with radiation data.

(3.3.5) data comparison and annual variation

Annual radiations from CRU (ORCHIDEE_CRU) and from REMO compare well (Figure 17). The radiation data from ECHAM5 has the smallest values and ECMWF radiation is a bit smaller than REMO and CRU radiation data. The ECMWF radiation has quite similar annual variation as the radiation from CRU and REMO. Figure 18 displays the radiation comparisons for longer time period, where the LPJ_CRU radiation data has the similar inter-annual variation with ORCHIDEE_CRU radiation, but the absolute value radiation from LPJ_CRU is much larger.

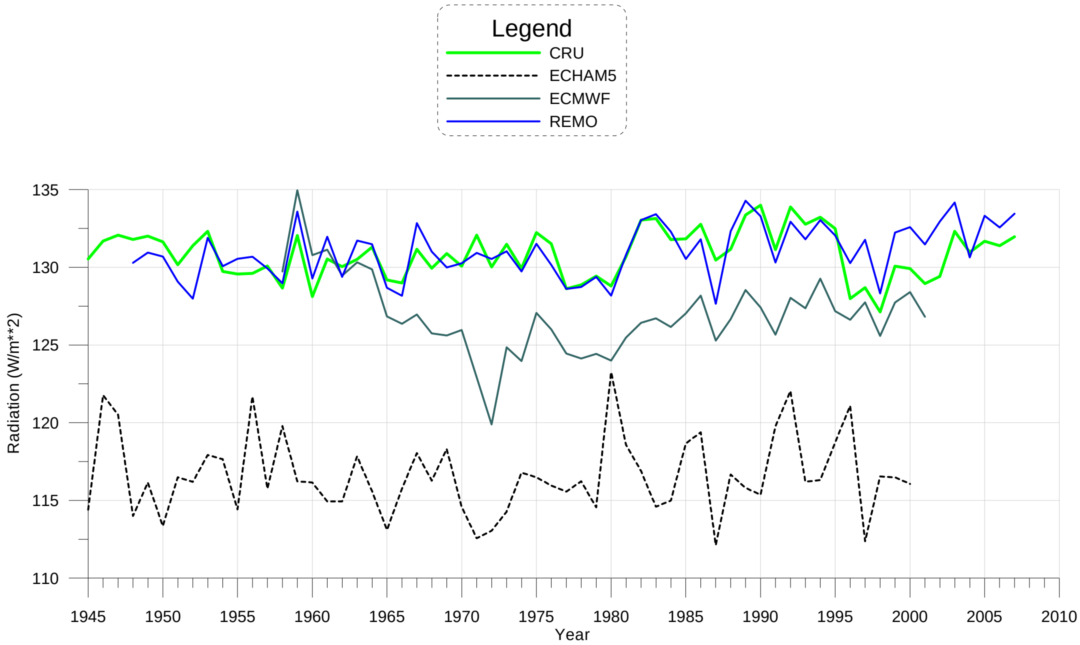


Figure 17: The radiation data comparison with the area weighted average from CRU, ECHAM5, ECMWF, and REMO data sets respectively.

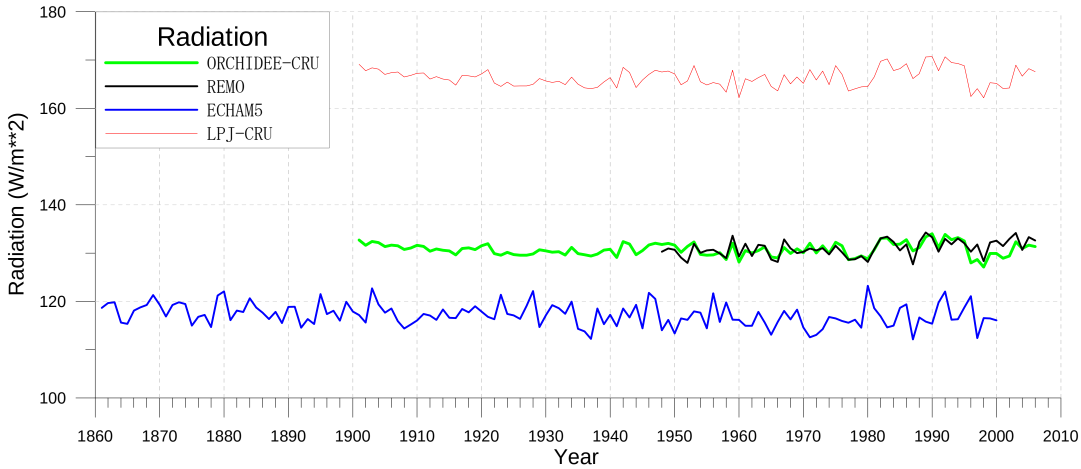


Figure 18: The radiation data comparison with the area weighted average from ORCHIDEE-CRU, REMO, ECHAM5, and LPJ-CRU data sets respectively.

3.4 Vapor Pressure Deficit

(3.4.1) Regression validation

The “daylight vapor pressure deficit (VPD)” is calculated from daylight temperature and dew point temperature, which supply the data from 1901 to 2002. Similarly, the data after 2002 have to be generated through regression with REMO’s VPD data as the predictor. Figure 19 shows the distribution of correlation coefficients of regression equations for extending the VPD data up to 2007. The diagram graph has the similar shape as that for radiation regression, which could be considered that the regression skill for VPD and radiation are at the similar level.

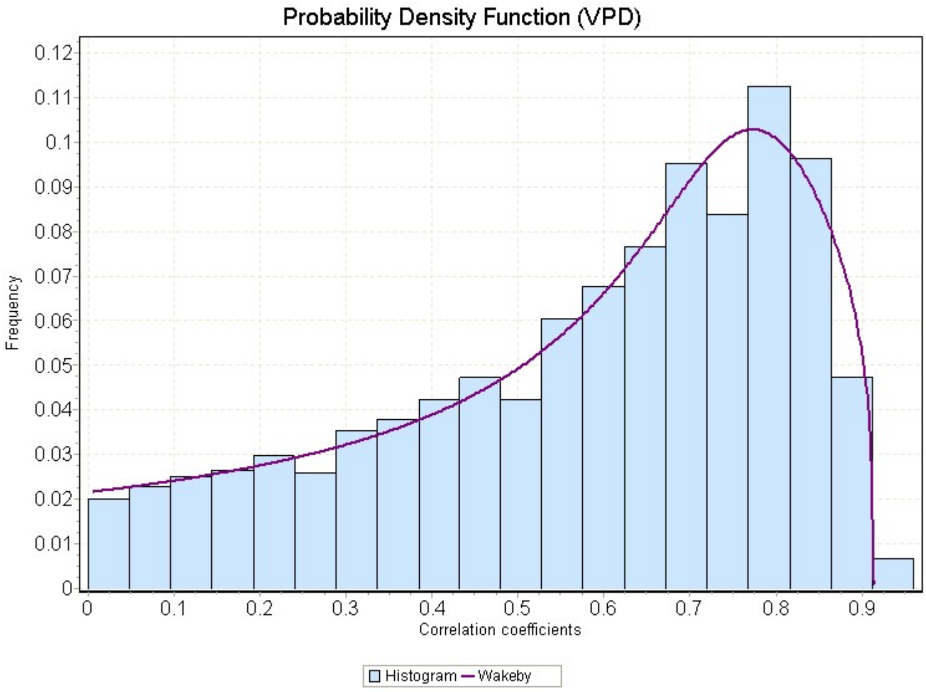


Figure 19: The histogram of correlation coefficients of regression for extending VPD data from 2003 to 2007. The smooth line is the fitting by Wakeby method.

(3.4.2) Daily data construction

Once the monthly VPD data is available we could construct the daily data in the same way as for radiation, i.e. the daily values have to be controlled that it must be larger or equal to zero. if the calculated VPD is negative, it is forced zero. Figure 20 shows the MCRU’s VPD data (both from monthly and daily data) and the REMO’s VPD data. The REMO’s VPD has much larger values than MCRU’s VPD, though their temporal variations are quite similar.

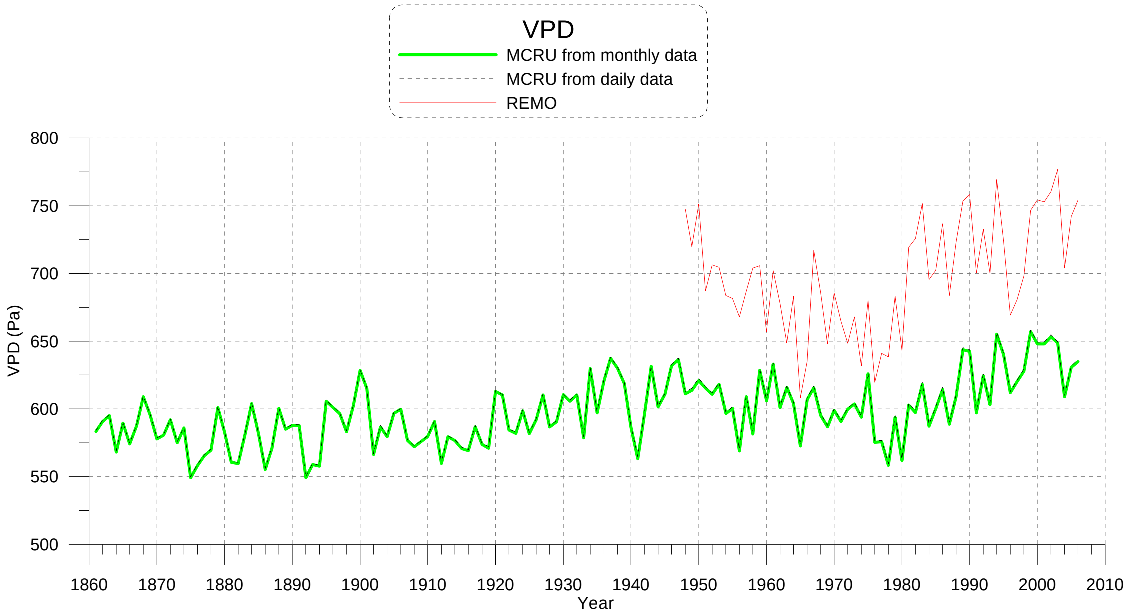


Figure 20: The area weighted average of MCRU (green and dish lines) and REMO (red) vapor pressure deficit (VPD). The dish line is the weighted average from the constructed daily data.

(3.4.3)Data comparison and annual variation
REMO data showed larger VPD than CRU (MCRU) data set, however the CRU's VPD has small difference with the corresponding data from ECMWF (ERA40 data) (figure 21). The ECHAM5 data shows the smallest VPD. Except for ECHAM5 data set, all other data sets show similar temporal variation of VPD.

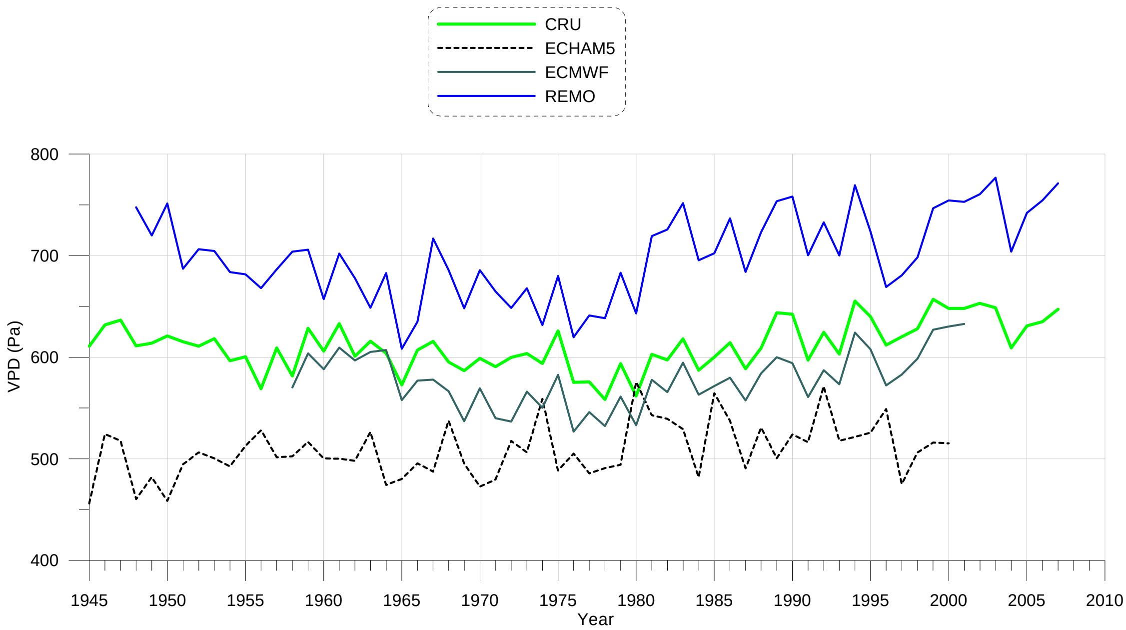


Figure 21: The VPD data comparison with the area weighted average from CRU, ECHAM5, ECMWF, and REMO data sets respectively.

3.5 Relative Humidity

The relative humidity (RHY) is generally not directly available from the climatic data sets. RHY can be calculated from temperature and dew point temperature (or from temperature and vapor pressure for CRU data case). Because … the calculated RHY reached over 400% for in some pixels of the CRU data. In this case the RHY was forced to 100%.

(3.5.1) Regression validation

The correlation coefficients of regression equations show much worse results for relative humidity than for other variables (figure 22). It means that there is a weak relationship between these data sets. When we extend the RHY for the time periods during 2003-2007, for quite a lot pixel it only takes the values from the multiple year average.

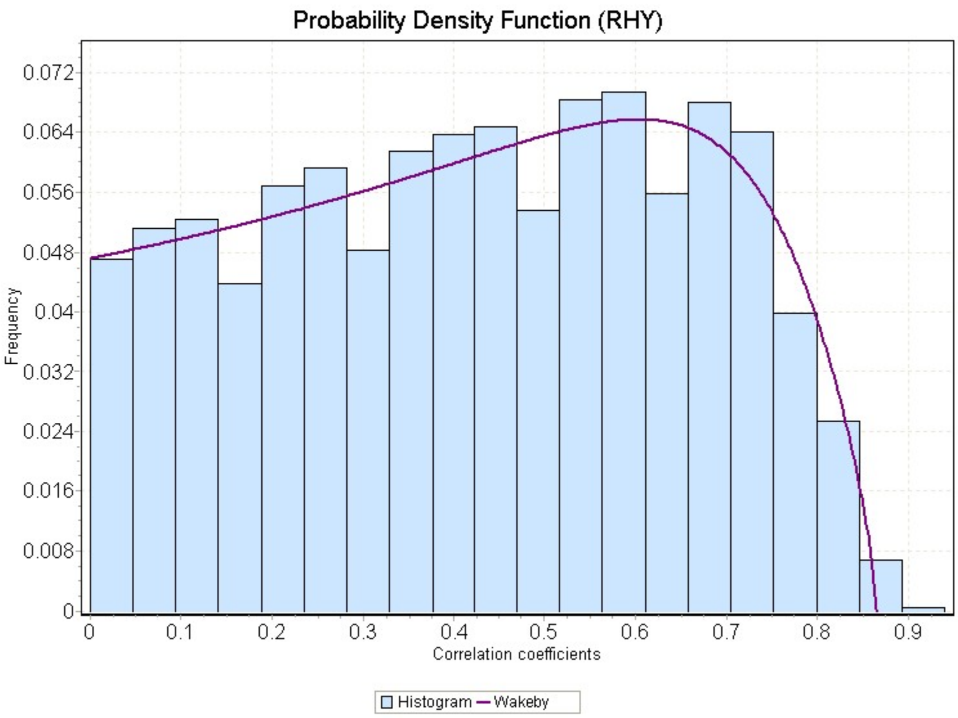


Figure 22: The histogram of correlation coefficients of regression for extending RHY data from 2003 to 2007. The smooth line is the fitting by Wakeby method.

(3.5.2) Daily data construction

The daily data is constructed in the similar way as for VPD, but unlike the VPD data in which only the lower limit (zero) is constrained, both lower limit (zero) and higher limit (100) are constrained for RHY. Thus there is obvious mismatch in weighted average plot (figure 23) for the curves from raw monthly data and from constructed daily data. The

RHY from REMO data set (red curve) shows smaller values than one from MCRU data, nevertheless, the temporal variations of RHY from REMO and MCRU have many similarities.

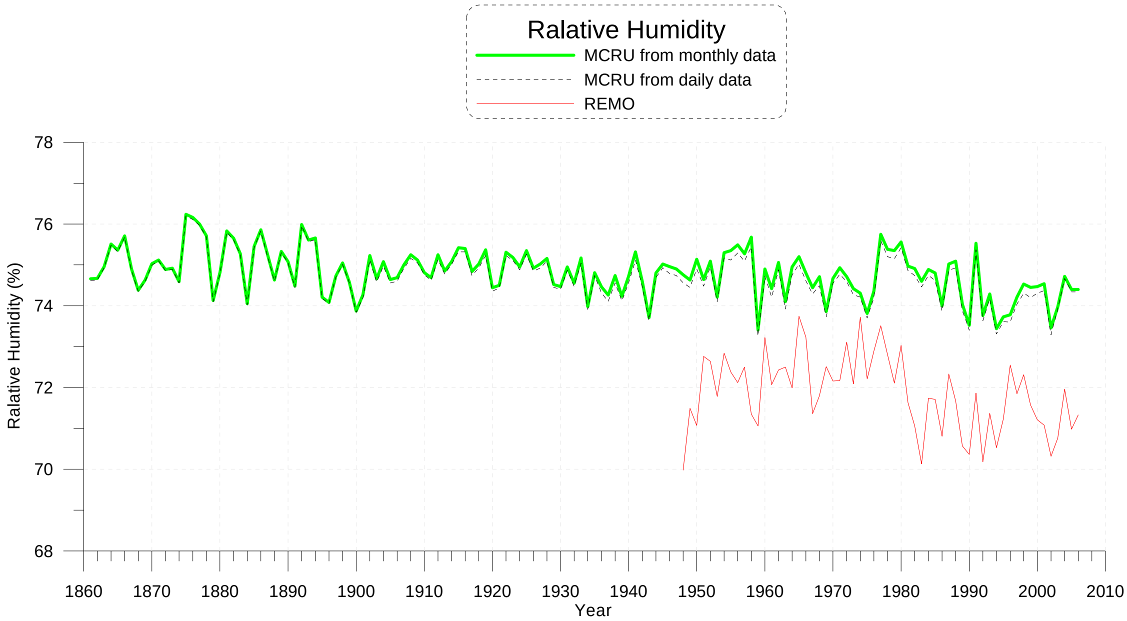


Figure 23: The area weighted average of MCRU (green and dish lines) and REMO (red) relative humidity (RHY). The dish line is the weighted average from the constructed daily data.

3.6 Cloudiness

Since in the CarboEurope-IP project, only monthly cloudiness was required, we constructed only monthly cloudiness data.

(3.6.1)Regression validation

The cloudiness data was directly available from CRU data set. To extend the cloudiness data to the period from 2003 to 2007, a regression equation is employed. The regression result was the worse than result than temperature and precipitation (Figures 2 and 8) but better result than RHY (Figure 22).

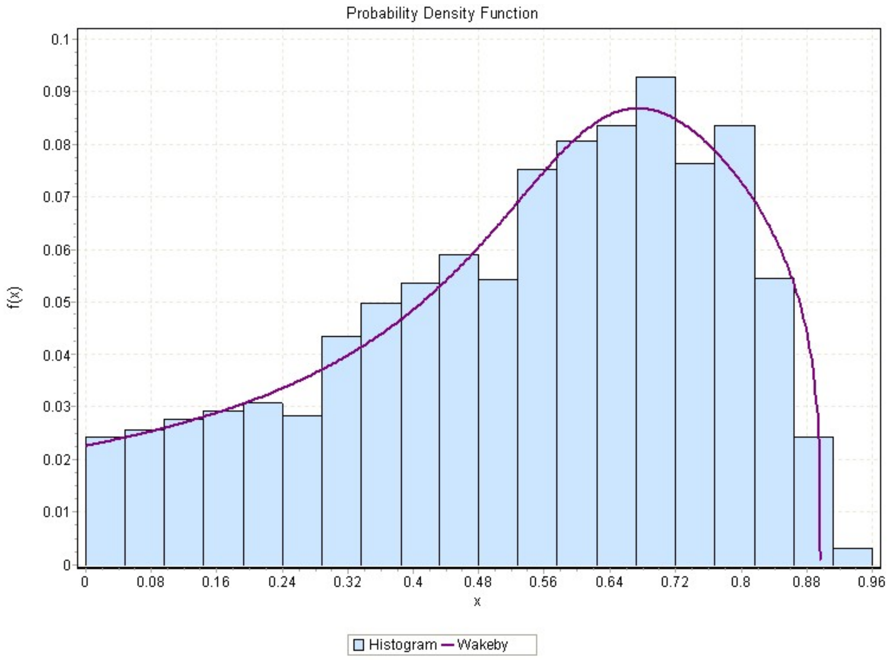


Figure 24: The histogram of correlation coefficients of regression for extending cloudiness data from 2003 to 2007. The smooth line is the fitting by Wakeby method.

(3.6.2)Data comparison and annual variation

Although the radiation data for CRU and REMO are quite similar, the cloudiness of CRU and REMO are quite different (Figure 25). REMO data set shows the lowest cloudiness. The annual cloudiness from ECHAM5 has absolute values similar to CRU data set, but different inter-annual variability. Cloudiness from ECMWF has a bit smaller value, but similar temporal variation as the CRU data set.

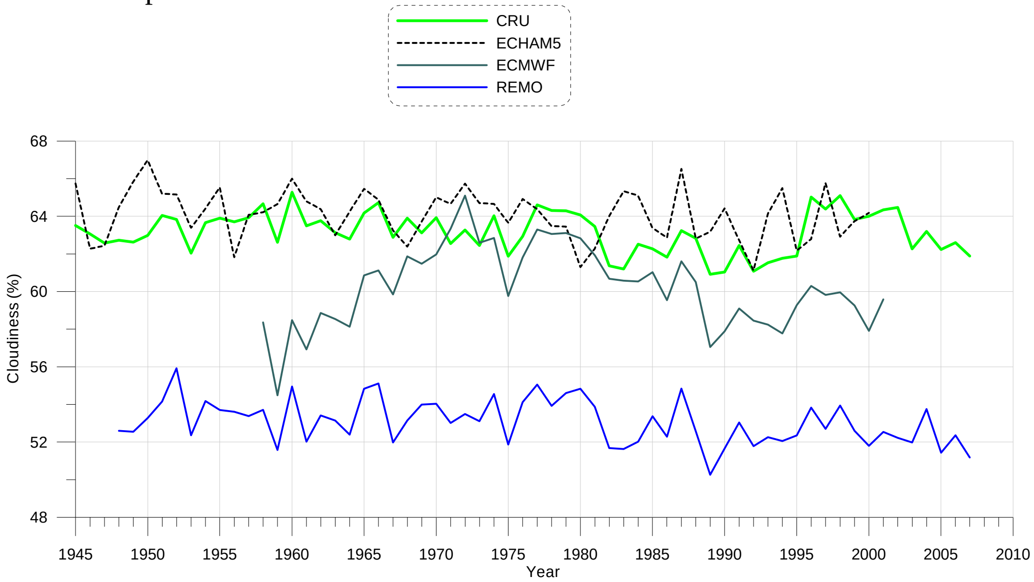


Figure 25: The cloudiness data comparison with the area weighted average from CRU, ECHAM5, ECMWF, and REMO data sets respectively.

4. Climate characteristics shown by MCRU data set

MCRU data set was produced for the CarboEurope-IP research project. For the period from 1901 to 2002, MCRU and CRU variables are the same at a monthly scale. The daily variables of MCRU holds the characteristics of the daily datasets used for data combination, i.e. the characteristics of REMO and ECHAM5 daily data.

The objective here is to analyze the climate trends in temperature and precipitation from MCRU dataset.

In most of Europe temperature was increasing between 1901 and 2000. The exceptions are two small regions in southern Norway and Western Mediterranean sea (Figure 26). The negative trends in southern Norway region are however not significant at 95% confidence (Figure 26b).

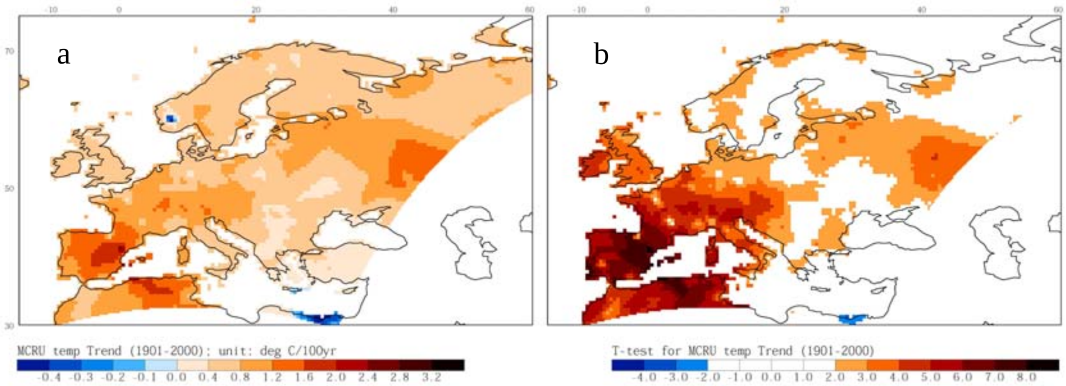


Figure 26: Trends in annual temperature (a) and the corresponding T-test (b), in which the area with only significant trends are plotted. For each pixel trends were calculated over 1901 and 2000.

A clear positive trend (red curve) in European annual temperature is observed for the period from 1901 to 2000 (figure 27). The trend is 0.79 degree C per 100 years, the corresponding t test is 3.98, much larger than 1.96 (threshold at 95% confidence). For the period from 1901 to 2007 the trend is even stronger. In addition, two abrupt changes in temperature were detected at 1933 and 1987 respectively.

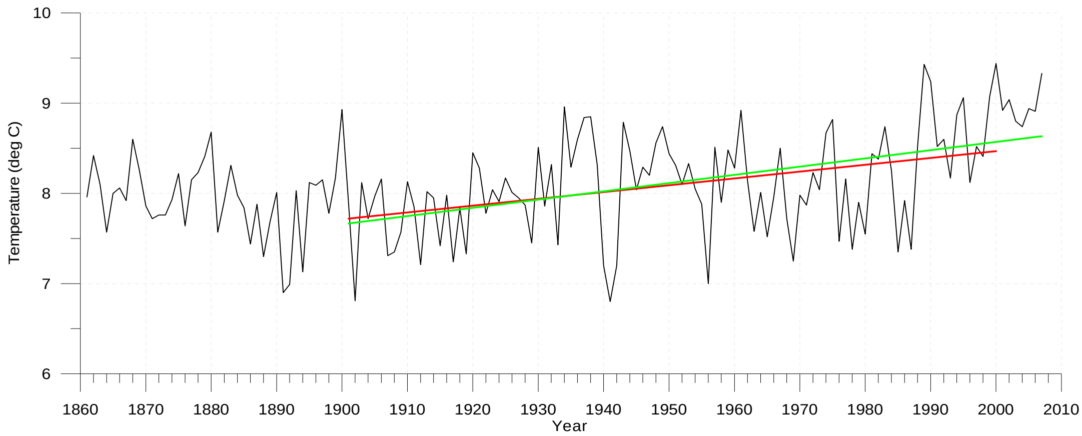


Figure 27: The area weighted average of MCRU temperature. It is showing a positive trend in temperatures for 1901 - 2000 (red line) and a much stronger trend for 1901 - 2007 (green line).

In comparison with the temperature trends, the area with significant trend for precipitation is much smaller (Figure 28a and b). Precipitation has positive trend in northern Europe and negative trend in southern Europe. Therefore, in southern Europe the climate is becoming warmer and drier, so that the drought may occur more often.

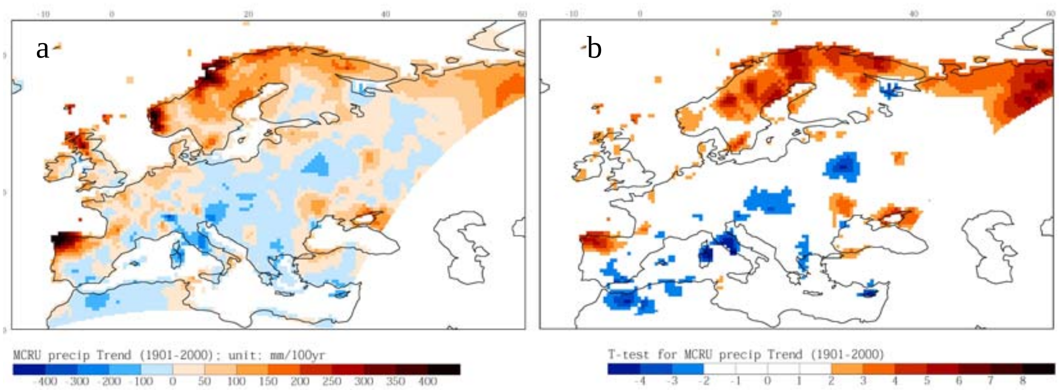


Figure 28: The precipitation trends (a) and the corresponding T-test (b), in which the area with only significant trends are plotted. For each pixel trends were calculated over 1901 and 2000.

Annual precipitation of Europe also shows a positive trend for the period from 1901 to 2000, but the trend is not significant with 95% confidence (Figure 29). It is because that the positive trend in the north and negative trend in the south offset each other.

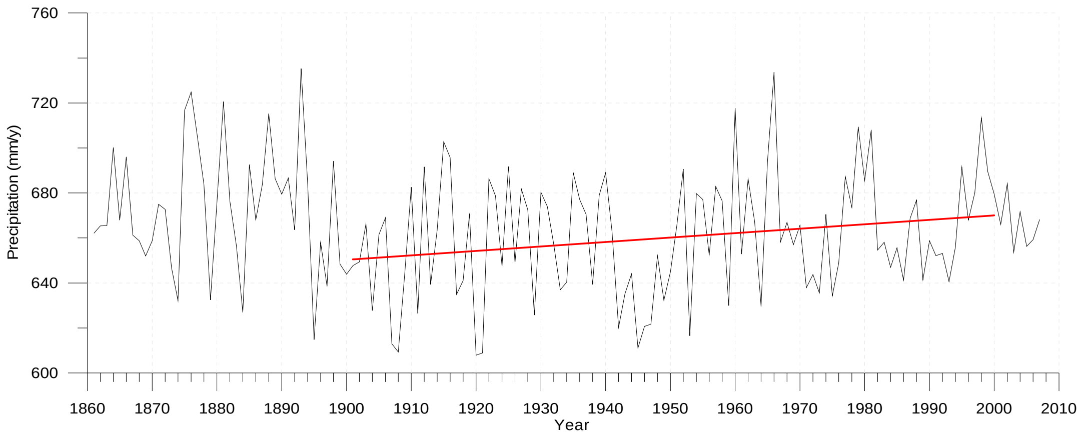


Figure 29: The area weighted average of MCRU precipitation. It is showing a positive trend from 1901 to 2000 but not significant (red line) according to 95% confidence.

5. Summary and conclusion

There are two important aspects of any data: availability and accuracy. Although nowadays there are many different data sources, none of them could claim that its data has 100% accuracy. For any modeling however the data accuracy is important since the wrong data would lead to wrong modeling result and make it difficult to give the correct interpretation. Therefore, we attempt to develop an accurate data set for the ecosystem modeling. After a lots of data comparison, the CRU data at a monthly scale is considered to be relatively accurate. We generated the daily data based on the monthly CRU data as well, which is a typical work for weather generator (WG). Normally WG use many mathematic methods such as auto-regression, random number, Markov Chains, and probability transformation. Here we used an alternative method to generate daily data from monthly data, which we called data combination. The data combination uses the simple arithmetic methods. The important prerequisite for data combination is the daily data availability from other data sets. We used temporal variability and spatial consistence of data from different sources to compute daily climate variables. Therefore, the data combination is the type of method that is comparable to WG in result, but based on much simple mathematic methods.

Nine climatic variables are needed to drive the ecosystem models. These variables could be classified in three groups:
(1) Temperature related variables:
Mean temperature; maximum and minimum temperature, daylight temperature
(2) Humidity related variables:
Precipitation, cloudiness, relative humidity and vapor pressure deficit
(3) Radiation related variable:
Surface solar radiation downwards

Based on the results of regression for climate data extended until 2007, we could classify the variables into four groups with decreasing regression performance: (1) temperature; (2) precipitation; (3) cloudiness, radiation and vapor pressure deficit; (4) relative humidity.

Here we didn’t describe the maximum and minimum temperature, as well as the daylight temperature since all these temperature-related variables have the similar result to the daily mean temperature in the aspect to regression and daily data construction. Generally various data sets have comparable temperatures. The data consistence for other variables is worse.

The analysis of the trends in temperature and precipitation data indicates that there is a positive temperature trend for most areas of Europe with stronger positive trends in the south than in the north. Precipitation has positive trend in the north and negative trend in the south. We conclude that in southern Europe the drought will become more and more serious in the future. However, in northern Europe the warmer temperature and more precipitation could benefit the vegetation growth .

References

Frauke Feser, Ralf Weisse, and Hans von Storch: (draft) Multi-decadal Atmospheric Modeling for Europe Yields Multi-purpose Data.

Geels, C., M. Gloor, P. Ciais, P. Bousquet, P. Peylin, A. T. Vermeulen, R. Dargaville, T. Aalto, J. Brandt, J. H. Christensen, L. M. Frohn, L. Haszpra, U. Karstens, C. Rödenbeck, M. Ramonet, G. Carboni, and R. Santaguida: Comparing atmospheric transport models for future regional inversions over Europe. Part 1: Mapping the CO2 atmospheric signals

ECMWF, The Description of the ECMWF Re-analysis Global Atmospheric Data Archive, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, 1997.

Kalnay, E., et al., The NCEP/NCAR 40-year reanalysis project, Bull. Am. Meteorol. Soc., 77, 437-471, 1996.

Katrin Schöllhammer1, Christian Marquardt2, Karin Labitzke1: Comparison of Three Different Meteorological Datasets (ECMWF, Met Office and NCEP)

Zhao, M., Running, S. W., & Nemani, R. R. (2006). Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses. Journal of Geophysical Research-Biogeosciences 111.