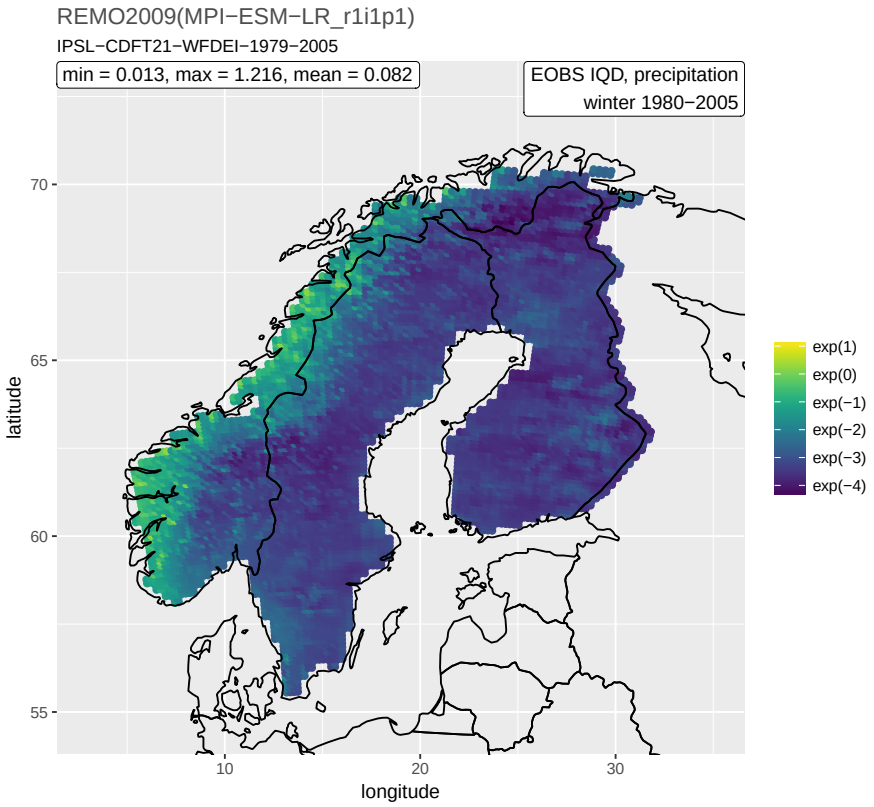
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Evaluation of bias corrected precipitation output from the EURO-CORDEX climate ensemble



Note no 1047

Authors Silius M. Vandeskog
Marion Haugen
Thordis L. Thorarinsdottir

Date 28th January 2020

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

The authors

Silius M. Vandeskog is a master's student in industrial mathematics at the Norwegian University of Science and Technology, Marion Haugen is Senior Research Scientist and Thordis L. Thorarinsdottir is Chief Research Scientist at Norwegian Computing Center.

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modelling. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Evaluation of bias corrected precipitation output from the EURO-CORDEX climate ensemble**
Authors	**Silius M. Vandeskog** <siliusv@gmail.com> **Marion Haugen** <marionh@nr.no> **Thordis L. Thorarinsdottir** <thordis@nr.no>
Date	28th January 2020
Publication number	1047

Abstract

Global circulation models (GCMs) are used for projecting climate changes on a global scale. However, when we need information for local climate changes, a dynamical downscaling through a regional climate model (RCM) may be used to gain more precise information. Therefore it is important to make good RCMs that are unbiased when projecting climate changes. This note investigates the skill of precipitation projections from five combinations of global and regional climate models from EURO-CORDEX and four bias correction methods applied to some of these. This is performed by comparing the model outputs with data from the E-OBS and NGCD data products using integrated quadratic distance.

Keywords	Climate model evaluation, precipitation, dry-days, integrated quadratic distance
Target group	Climate scientists, users of climate information
Availability	Open
Project	PostClim
Project number	220778
Research field	Statistics, climate science
Number of pages	20

Contents

1 Introduction

Projecting future precipitation amounts and dry periods are important aspects, for example when making constructions that should withstand a 100 year flood or when making sure that a city has enough drinking water supplies for the future. To estimate future climate changes, global circulation models (GCMs) are used to simulate changes on a global scale. However, to project climate changes locally with some precision, regional climate models (RCMs) with finer resolution are constructed with boundary conditions from a GCM. These regional models often suffer from various biases, and it is useful to develop bias correction methods (BCMs) that can be applied to the local model outputs.

In this note we will examine properties of the precipitation projections over Fennoscandia (Finland, Sweden and Norway) for the years 1980 to 2005. This will be examined for five combinations of GCMs and RCMs from EURO-CORDEX and four BCMs applied to some of these. We will focus on using the integrated quadratic distance (IQD) for evaluating the fit of the models against data from three different data products, namely E-OBS, NGCD type 1 and NGCD type 2 (see Section 3.3). The two main aspects we will examine are the daily precipitation data and the number of dry-day periods and their lengths. This will be performed separately for the summer and winter seasons.

Previously, the IQD between the same combination of RCMs and BCMs has been calculated against the E-OBS data product (Vandeskog et al. (2017)). The results showed that the BCM that had been trained on the E-OBS data product, LSCE-IPSL-CDFt-EOBS10-1971-2005, performed significantly better than all other correction methods. It is therefore of interest to compare the model output using data products that none of the BCMs have been trained on. This is the purpose of comparing the model outputs against NGCD type 1 and 2 data products.

The remainder of the note is organized as follows. Section 2 describes the data used for all evaluations. The theory behind IQD, the different BCMs and the interpolation for the different data products are explained in Section 3. Section 4 explains how we obtained our results, which are presented and discussed in Section 5. Finally, the conclusion is provided in Section 6.

2 Data

2.1 Climate models

A total of five different combinations of GCMs and RCMs from EURO-CORDEX are chosen for the evaluation. These can be seen in Table 1. Each model output contains daily precipitation data from various start and end dates. The precipitation is measured in $\mathrm{kg\,m^{-2}\,s^{-1}}$. There has also been conducted several bias corrections to the climate model outputs. The correction methods are listed in Table 2.

Table 1. Five global/regional climate model combinations from EURO-CORDEX used in our testing.

Model nr.	Global climate model	Ensemble member	Regional climate model	Institute	Institution name
1	CNRM-CERFACS-CM5	r1i1p1	CCLM4-8-17	CLMcom	Climate Limited-area Modelling Community
2	ICHEC-EC-EARTH	r1i1p1	RACMO22E	KNMI	Royal Netherlands Meteorological Institute
3	ICHEC-EC-EARTH	r2i1p1	CCLM4-8-17	CLMcom	Climate Limited-area Modelling Community
4	MPI-ESM-LR	r1i1p1	CCLM4-8-17	CLMcom	Climate Limited-area Modelling Community
5	MPI-ESM-LR	r1i1p1	REMO2009	MPI-CSC	Helmholtz-Zentrum Geesthacht, Climate Service Center, Max Planck Institute for Meteorology

Table 2. Four bias correction methods applied to the EURO-CORDEX simulations.

Method nr.	Bias correction method	Calibration data set	Years	Institution name	Applied to model nr.
1	LSCE-IPSL-CDFt	EOBS10	1971-2005	Institut Pierre Simon Laplace	1,2,3,4,5
2	METNO-QMAP	MESAN	1989-2010	Norwegian Meteorological Institute	1,2,3,4,5
3	SMHI-DBS45	MESAN	1989-2010	Swedish Meteorological and Hydrological Institute	1,2,4,5
4	IPSL-CDFT21	WFDEI	1979-2005	Institut Pierre Simon Laplace	2,3,5

2.2 Preparation of data

The data for all model outputs and data products are stored in different NetCDF-files and have to be prepared before they are analysed. Several files may have to be merged in order to get data for the entire period from 1980 to 2005. All unwanted data must also be removed. This concerns data before 1980, after 2005 and February 29. This day is removed to make sure each year has the same number of measurements for all the data sets. Additionally, only locations that contains precipitation data for all model outputs and data products are interesting for comparisons. All other locations are removed. We end up with a matrix of dimension 140×155 that covers the area seen in the figure on

the front page of the note. Data is then sorted by season, where summer consists of June, July and August, and winter consists of December, January and February. This gives a total of 92 observations per summer and 90 observations per winter.

2.2.1 Climate models

While all the observational data products use mm as a unit for precipitation, the climate models use $\mathrm{kg\,m^{-2}\,s^{-1}}$. This must be converted to mm. Some model outputs contain negative precipitation values, but these have small absolute values and are likely due to rounding errors. The negative values are thus set equal to zero. Testing finds that the number of zeros after the conversion of negative values has the same magnitude as the number of zeros for model outputs with only non-negative data.

2.2.2 E-OBS

The precipitation unit for E-OBS is already in mm, and the E-OBS data product is not in need of any specific preparations.

2.2.3 NGCD

The coordinates of the NGCD data products are represented using a Lambert azimuthal projection, which is different than that of the other model outputs and data products. In order to compare this with all other model outputs, a projection onto the same coordinates is needed. The NGCD data products also have a much higher resolution than that of the model outputs and E-OBS data product. For all other data sets, precipitation is given in a $12 \times 12\mathrm{km}^2$ grid, while NGCD precipitation is given in a $1 \times 1\mathrm{km}^2$ grid. It is therefore necessary to perform an upscaling of the data before comparisons can be initiated. To perform upscaling, the weighted sum of precipitation for all high resolution grids inside a large RCM grid is calculated. Weights are equal to the fraction of each high resolution grid cell that is inside a given large grid cell. If e.g. 40% of a $1 \times 1\mathrm{km}^2$ grid cell is covered by a given $12 \times 12\mathrm{km}^2$ grid cell, then 40% of the small grid cells precipitation will be added to the large grid.

3 Theory

3.1 IQD

We denote a precipitation observation by $y \in \Omega$ where Ω denotes the non-negative real axis $\mathbb{R}_{\geq 0}$. Similarly, for dry-days, we have that y is the length in days of a dry period with $\Omega = \{0, 1, 2, \ldots\}$. A probabilistic prediction for y is given by a distribution function with support on Ω denoted by $F \in \mathcal{F}$ for some appropriate class of distributions $\mathcal{F}$, with the density denoted by f if it exists.

Scoring rules assess the accuracy of probabilistic predictions by assigning a numerical penalty to each prediction-observation pair. Specifically, a scoring rule is a mapping

$$S : \mathcal{F} \times \Omega^d \to \mathbb{R} \cup \{\infty\} \tag{1}$$

where, in our notation, a smaller penalty indicates a better prediction. A scoring rule is *proper* relative to the class $\mathcal{F}$ if

$$\mathbb{E}_G S(G, Y) \leq \mathbb{E}_G S(F, Y) \tag{2}$$

for all probability distributions $F, G \in \mathcal{F}$, that is, if the expected score for a random observation Y is optimized if the true distribution of Y is issued as the prediction. The scoring rule is *strictly proper* relative to the class $\mathcal{F}$ if (2) holds with equality only if $F = G$. Propriety will encourage honesty and prevent hedging, which coincides with Murphy's first type of goodness (Murphy, 1993).

In some cases, in particular in climate modelling, it is of interest to compare the predictive distribution F against the true distribution of the observations which is commonly approximated by the *empirical distribution function* of the available observations $y_1, \ldots, y_n$,

$$\widehat{G}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left(y_i \leq x\right). \tag{3}$$

The two distributions, F and $\widehat{G}_n$, can be compared using a *divergence*

$$D : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_{\geq 0} \tag{4}$$

where $D(F, F) = 0$.

Assume that the observations $y_1, \ldots, y_n$ forming the empirical distribution function $\widehat{G}_n$ are independent with distribution $G \in \mathcal{F}$. A propriety condition for divergences corresponding to that for scoring rules (2) states that the divergence D is *k-proper* for a positive integer k if

$$\mathbb{E}_G D(G, \widehat{G}_k) \leq \mathbb{E}_G D(F, \widehat{G}_k) \tag{5}$$

and *asymptotically proper* if

$$\lim_{k \to \infty} \mathbb{E}_G D(G, \widehat{G}_k) \leq \lim_{k \to \infty} \mathbb{E}_G D(F, \widehat{G}_k) \tag{6}$$

for all probability distributions $F, G \in \mathcal{F}$ (Thorarinsdottir et al., 2013). While the condition in (6) is fulfilled by a large class of divergences, only score divergences have been shown to fulfil (5) for all integers k. A divergence D is a *score divergence* if there exists a proper scoring rule S such that $D(F, G) = \mathbb{E}_G S(F, Y) - \mathbb{E}_G S(G, Y)$. A score divergence that assesses the full distributions is the *integrated quadratic distance* (Thorarinsdottir et al., 2013):

$$\mathrm{IQD}(F, G) = \int_{-\infty}^{+\infty} (F(x) - G(x))^2 \, \mathrm{d}x \tag{7}$$

In the following, we will apply the IQD to compare empirical distributions of climate model output and observations.

3.2 Bias correction methods

All the BCMs tested in this note work mostly in the same way. They use a data set with precipitation data as calibration, then the distribution from a raw model output is transformed to fit better with the calibration data. The calibration data sets and the various transformations, however, differ in the correction methods we examine.

3.2.1 QMAP

This popular BCM approach is based on the quantile-mapping technique (Gudmundsson et al., 2012). The technique maps a model output x with cumulative distribution function (CDF) F_X, to an observation y with CDF F_Y through a function h (Vrac et al., 2016):

$$y = h(x), \text{ such that } F_Y(y) = F_X(x) \tag{8}$$

Bias correction method 2 uses such a quantile-mapping technique. It can be found in the R package `qmap` (R Core Team, 2016).

3.2.2 CDFt

The Cumulative Distribution Function-transform is used in bias correction method 1 and 4, and can be considered a variant of the empirical quantile-mapping, with h given as

$$h(x) = F_Y^{-1}(F_X(x)), \tag{9}$$

where F^{-1} is the inverse function of the CDF F (Gudmundsson et al., 2012). It first estimates the CDF F_{Yp} and F_{Xp} over a projection time period before applying the distribution-derived quantile mapping in (9) (Vrac et al., 2016). It is unknown to us whether there are any differences in the techniques of the LSCE-IPSL-CDFt method and the IPSL-CDFT21 method.

3.2.3 DBS45

Bias correction method 3 uses a distribution-based scaling for correcting RCMs. First, the number of wet days is adjusted. This is done by identifying a cut-off value that reduces the percentage of wet days in the simulation to that of the MESAN observations. Then all days with less precipitation than the threshold are considered dry days. The remaining precipitation is then transformed to match the observed frequency distribution, using gamma distributions (Yang et al., 2010).

3.3 Interpolation methods

Three different data products containing daily precipitation over Fennoscandia are used for testing the different model outputs. The first one is the E-OBS data product, and the other two are the NGCD type 1 and the NGCD type 2 data products. All data products are created from a set of observations from different weather stations. Interpolation schemes are then added to the data in order to achieve a gridded data set.

3.3.1 E-OBS

The interpolation scheme for the E-OBS data product is divided into three steps. First, monthly means are interpolated using thin-plate splines. Secondly, kriging is used on anomalies with regard to monthly mean. Lastly, the interpolated anomalies are applied to the interpolated monthly mean (Haylock et al., 2008).

3.3.2 NGCD type 1

In the NGCD type 1 data product, the true precipitation at each measuring location is predicted using a scheme proposed in Førland et al. (1996). Spatial interpolation of precipitation is then carried out by applying two irregular triangular networks (TINs). One is a precipitation TIN based on observed precipitation corrected for wind loss. The other is an elevation TIN based on the altitudes of the observation stations (Tveito et al., 2005).

3.3.3 NGCD type 2

The interpolation scheme in the NGCD type 2 data product is based on Bayesian spatial interpolation methods and scale-separation concepts (Lussana et al., 2018).

4 Methods

4.1 Precipitation

We want to evaluate the IQD for daily precipitation for the observational data products and the chosen climate models, separately for each grid point. To perform this we make 18 comparisons per model. One for each year between 1984 and 2001. We compare one year from one of the data products with nine years of model data: four years before and four years after in addition to the year we want to examine. This gives a total of 90 or 92 observations from a data product and 810 or 828 observations from a climate model output for each comparison. After calculating the IQD for each of the 18 comparisons, we find the mean IQD for the entire period in each grid point. In addition to finding IQD between model outputs and data products, IQD is also calculated between the different data products. To find IQD between two data products, one of the data products simply acts as a model output, and 18 comparisons are found per grid point. This gives nine results when comparing three data sets, all acting as both model output and data product. Data products are also compared against themselves.

4.2 Dry-days

To examine how well the models predict dry periods, we make a vector that counts all dry periods from 1980 to 2005 for each grid point. Data is stored in a way such that the number of dry periods of length n days is stored in entry number n of the vector. This gives an array of size $140 \times 155 \times$ maxLength, where maxLength is equal to 90 or 92, depending on what season we examine. We get two of these arrays per climate model. Since dry periods of small length are not very interesting, we remove the first seven entries of our data and only examine dry periods of length greater than one week. A dry-day is hereby defined as any day without precipitation and where there has been no precipitation in the last seven days. A dry-day period is defined as any period of days without precipitation that contains at least one dry-day. We then calculate the IQD for our data. This time we only have one array per season for the entire period, so we only get one comparison per grid point. Although the IQD of something with the unit "# dry-day periods of length $n \in \{8, 9, 10, ..., \text{maxLength}\}$" provides little physical understanding, it is still useful for comparing the different model outputs against each other. Once again, the data products are also compared with each other. This time there is symmetry between model outputs and data products, meaning that the comparison of three data products only yields four unique results, including zero which represents the IQD between a data set and itself.

5 Results and discussion

5.1 Precipitation

After having found IQD values in a 140 × 155 grid for each combination of RCMs and BCMs against all three data products, the mean IQD over all grid points is calculated. Figure 1 displays the resulting IQD values during both summer and winter. It seems clear from the figure that the bias correction generally improves the IQD values of the raw RCM data. It can also be seen that the IQD between two data products often is lower than that for the model outputs. However, there is not a clear conclusion, as this does not always hold. Some of the RCMs have lower IQD than that of other data products. An example of this can be seen in the two middle plots, where the IQD of NGCD type 1 is found in the middle of a cluster of IQD values for bias corrected model outputs. Also, in the upper left and right plots, some of the raw model outputs are better than their bias corrected variants. This only happens for two of the raw models, though.

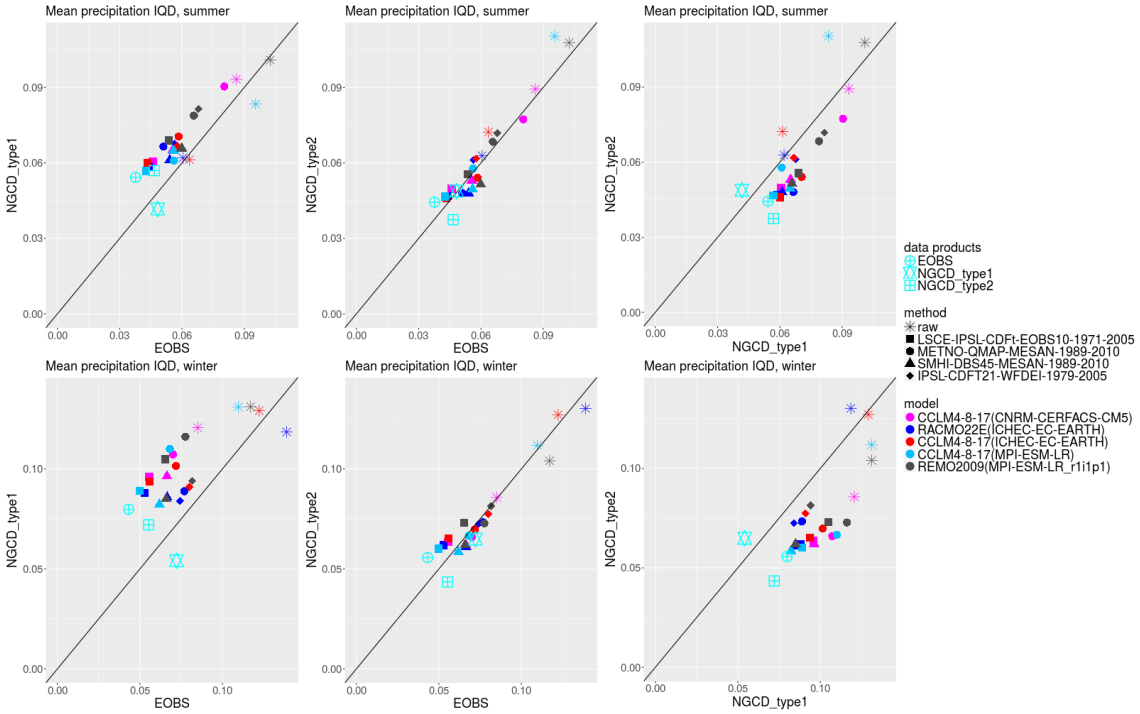


Figure 1. Scatter plot with the mean IQD for precipitation for all available combinations of RCMs and BCMs. Colour indicates climate model, while shape indicates raw data and bias correction methods. The axes present IQD against different types of data products.

We can perform bootstrapping over the IQD across all grid cells to examine the uncertainty in the mean IQD. Figure 2 displays the mean precipitation IQD along with 0.1- and 0.9-quantiles for all RCMs, BCMs and data products. For E-OBS, it is known that the bias correction method trained on the E-OBS data product, displayed as a filled square, performs better than all the other BCMs. This can be seen in the two leftmost plots. The results are not the same for the two other data products. During the summer period, the E-OBS trained correction method still performs better than the others. The differences

from the other BCMs seems to be smaller this time, though. During winter the performance of the E-OBS correction method is much worse, and this time other BCMs gain lower mean IQD. We can see, due to the very small distance between the 0.1- and 0.9-quantiles, that all results are significant.

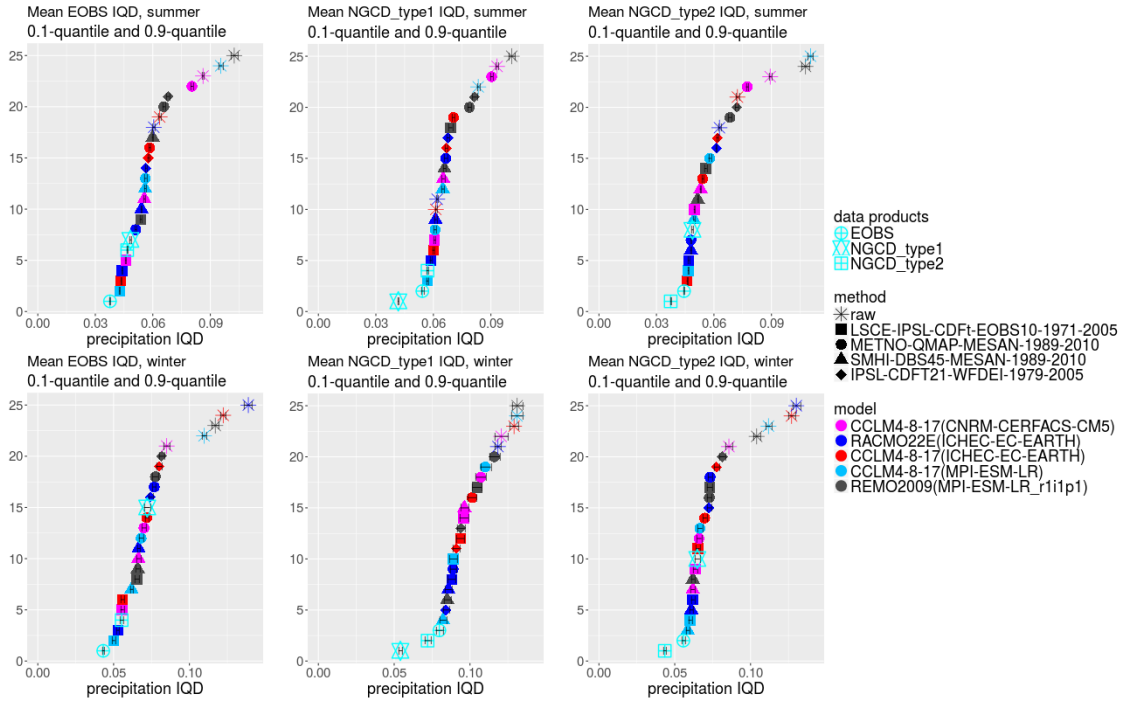


Figure 2. Ranking of the mean precipitation IQD for all RCMs and BCMs. Bootstrapped 0.1- and 0.9-quantiles for the mean are included in the plots. The ranking is performed for IQD against three different data products for summer and winter.

Overall it looks like LSCE-IPSL-CDFt-EOBS10-1971-2005 and SMHI-DBS45-MESAN-1989-2010 are the two BCMs with the highest general skill across the three data products. However, the difference between BCMs seems quite small. One can also see that the difference in IQD values for bias corrected RCMs and data products is small. The IQD between a data product and itself should logically be the lowest score one can obtain. Many of the BCMs come very close to this IQD, indicating that all of the four BCMs work well. We also see that the IQD after bias correction is much lower for almost all the RCMs. This indicates that bias correction improves the model outputs. Clustering of points of the same shape (BCM type) can be seen all over the ranking plots. However, the colours (model type) are almost spread evenly throughout the plots. Thus, it seems likely that the change in IQD after bias correction depends most on which BCM is used, not on the choice of which RCM is used.

5.2 Dry-days

A similar scatter plot to that of Figure 1 is made for the dry-days. The result can be seen in Figure 3. Once again the IQD seems to be somewhat smaller between data products than for the BCMs, and the raw climate model outputs have an even higher IQD than their bias corrected variants. However, these figures are more chaotic than the ones for precipitation. From the "ellipsis-shape" in the four leftmost figures it seems that any model output with a low IQD from the E-OBS data product obtains a high IQD from the NGCD data products, and vice versa.

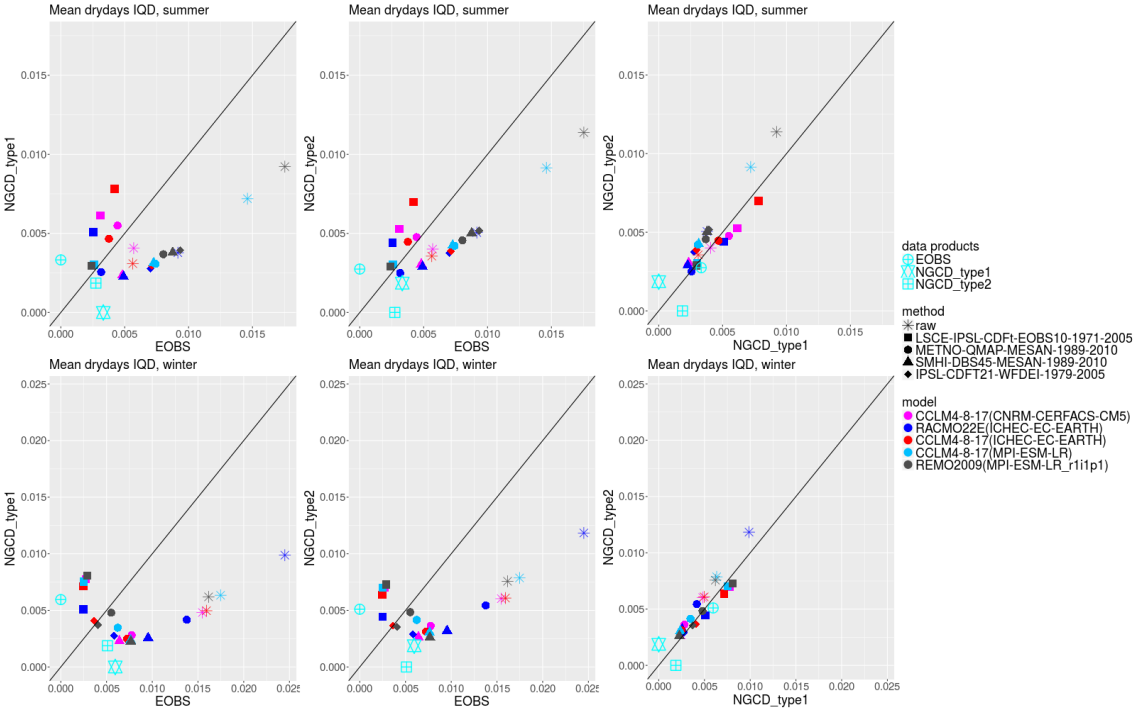


Figure 3. Scatter plot with the mean dry-day IQD for all RCMs and BCMs. Colour indicates climate model, while shape indicates raw data and bias correction methods. The axes present IQD against different types of data products.

The ranking of the different model outputs are examined in Figure 4. We see that the LSCE-IPSL-CDFt-EOBS10-1971-2005 corrected models, which perform very well against the E-OBS data product, obtains among the highest IQD against the NGCD data products. In many cases they have a higher IQD than their raw RCMs. On the other hand, SMHI-DBS45-MESAN-1989-2010 corrected models have a very low IQD against the NGCD data products, but their IQD score against the E-OBS data product is much higher. We can also see from this figure that the IQD for raw climate model outputs sometimes can be found in the lower half of the IQD ranking. Most of the raw model outputs still have a very high IQD, but it is clear by comparisons with Figure 2 that the improvement after bias correction is larger for precipitation IQD. Once again we see that the colours are spread evenly out in the plots, and it seems once more that choice of BCM is much more important than that of RCM.

It seems hard to choose one BCM that is better than the others when we examine dry-

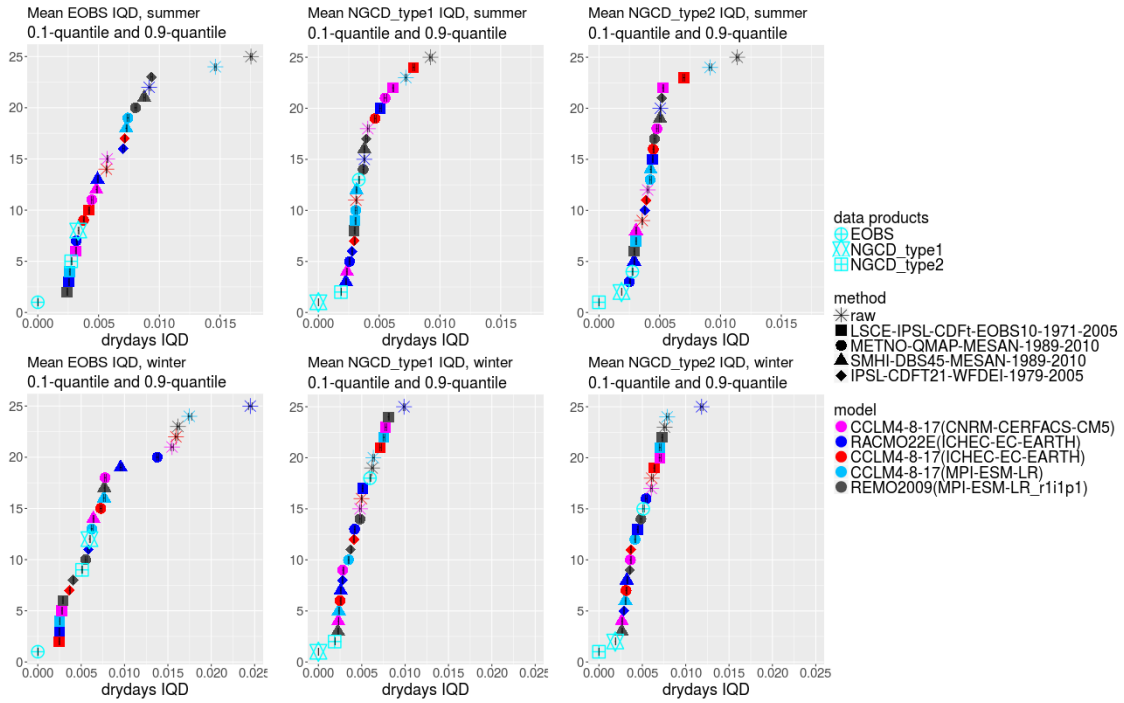


Figure 4. Ranking of the mean dry-day IQD for all RCMs and BCMs. Bootstrapped 0.1- and 0.9-quantiles for the mean are included in the plots. The ranking is performed for IQD against three different data products for summer and winter.

days IQD. However, from Figure 4 we also see that the IQD between data products is often among the highest IQD. Both figures give an impression that the difference in dry-days between E-OBS and NGCD is quite significant. This can indicate that all the RCMs predict dry-days quite well, thus making it look like the IQD between NGCD and E-OBS is very large. Another possibility is that the different data products have different underlying methods for dealing with dry-days, and that either one or all of them are somewhat faulty.

In order to examine the differences in dry-days between E-OBS and the NGCD data products, the histograms in Figures 5 and 6 are created. Figure 5 displays the total length of all dry-day periods occurring at each grid point. It is clear that the number of dry-days per location is much lower for NGCD type 1 than the other two data products. E-OBS has a large number of dry-days at each grid point, and NGCD type 2 can be found somewhere in the middle of the two. However, the shape of the empirical distribution for the sum of all dry-days in the NGCD type 2 data product appears to be more similar to the shape found in the NGCD type 1 data product. This can help explain why the IQD between E-OBS and NGCD type 2 is smaller than the IQD between E-OBS and NGCD type 1 in Figure 4. It does not, however, explain why the difference between the E-OBS data product and the NGCD data products is so large.

In Figure 6 the weighted mean length of all dry-day periods is calculated for each grid point. Weights are set to be the number of occurred dry-day periods of each length i.e. if a location has 5 dry-day periods of length 8 and 4 dry-day periods of length 9, the

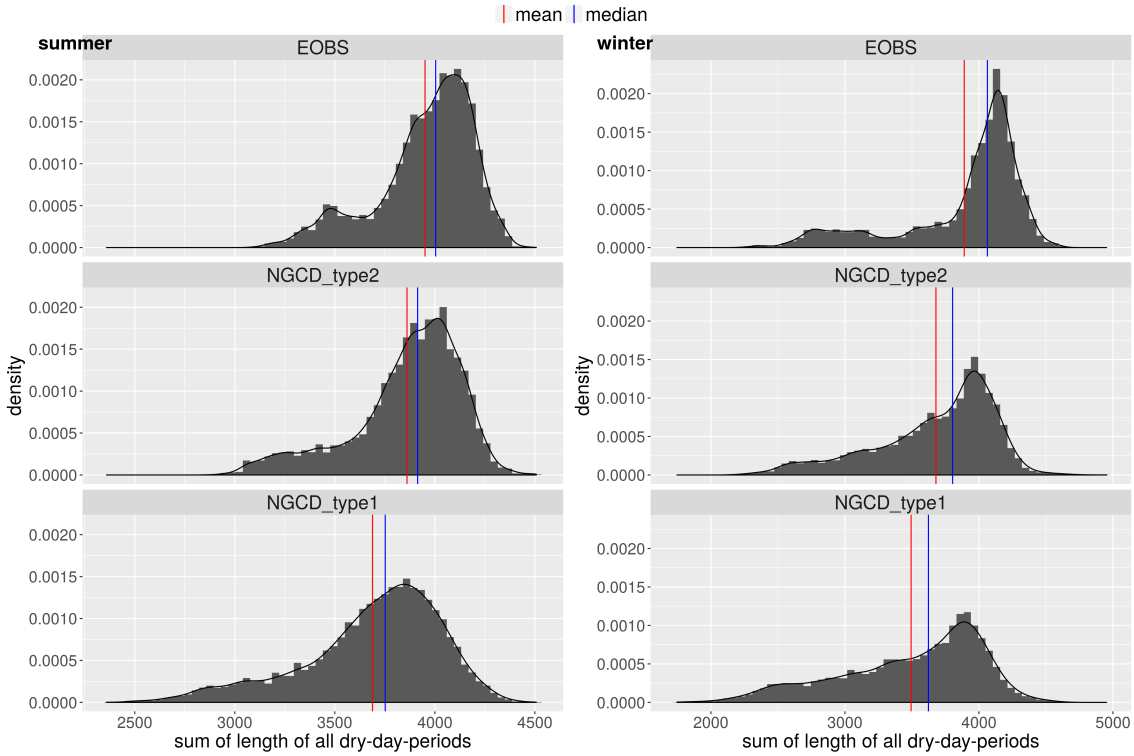


Figure 5. Histograms displaying the sum of the lengths of all dry-day periods at a given grid point, for all grid points. This is performed for all three data products, during summer and winter.

weighted mean length becomes $(8 \cdot 5 + 9 \cdot 4)/(4 + 5) = 8.44$. In this figure there is a clear difference between the NGCD and the E-OBS data products. The weighted mean seems to be half a day larger for the E-OBS data product. It is obvious from this figure that the behaviour of dry-day periods is different in E-OBS than in NGCD. This could mean that there are fewer short and more long dry-day periods in E-OBS.

To examine this further, the empirical distributions for the length of all dry-day periods in the three data products are plotted in Figure 7. One can see that the empirical distributions of the NGCD data products are practically identical, while there is a significant difference compared to the E-OBS data product, which has many more dry-day periods of greater length. In order to understand the importance of this difference, a new plot is made in Figure 8. The same distributions from the data products are displayed, however this time they are plotted on log-scale. We have also added the empirical distributions from the four different BCMs to the plots. One can see that the distributions belonging to the LSCE-IPSL-CDFt-EOBS10-1971-2005 BCM are much closer to the distributions of E-OBS than those of NGCD. The distributions from SMHI-DBS45-MESAN-1989-2010 and IPSL-CDFT21-WFDEII-1979-2005 are also much closer to the distributions from NGCD than that of E-OBS. This fits perfectly well to what was discovered from Figure 4, which means that we possibly have explained the large differences found in dry-day IQD.

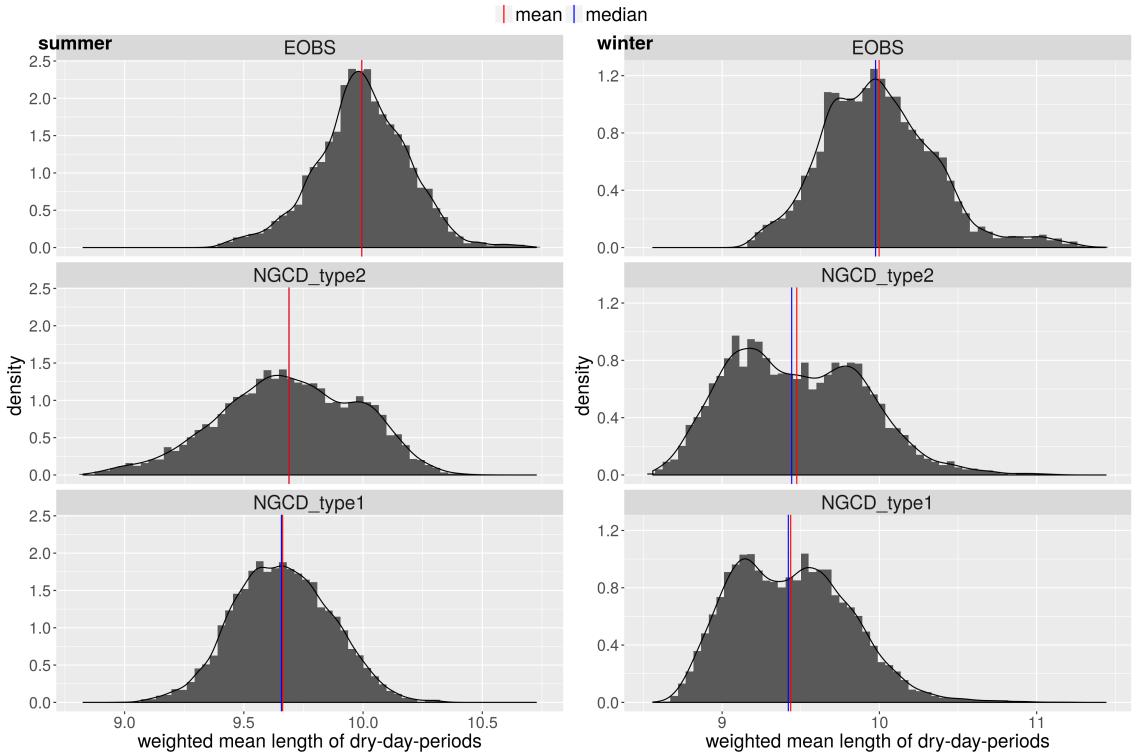


Figure 6. Histograms displaying the weighted mean length of all dry-day periods per grid point. Weights are set to be the number of occurred dry-day periods of each length. The plots display results for all three data products, during summer and winter.

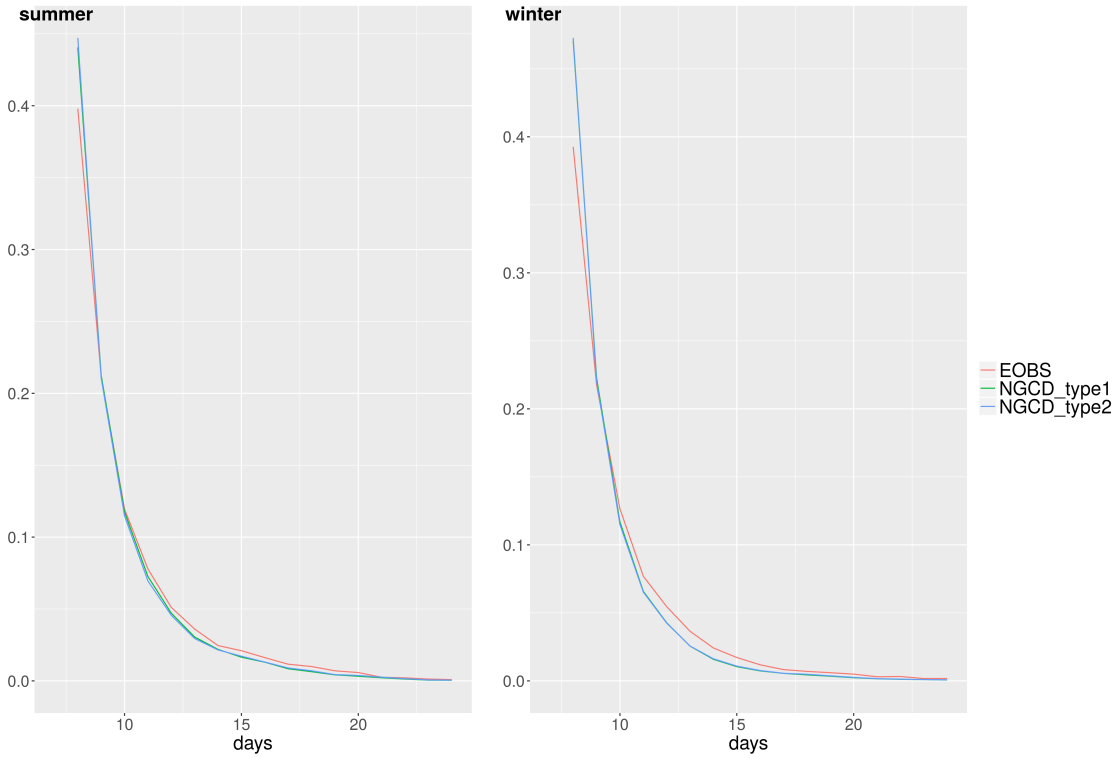


Figure 7. Empirical distributions for the length of all dry-day periods over all grid points. One distribution is made for summer, and one for winter.

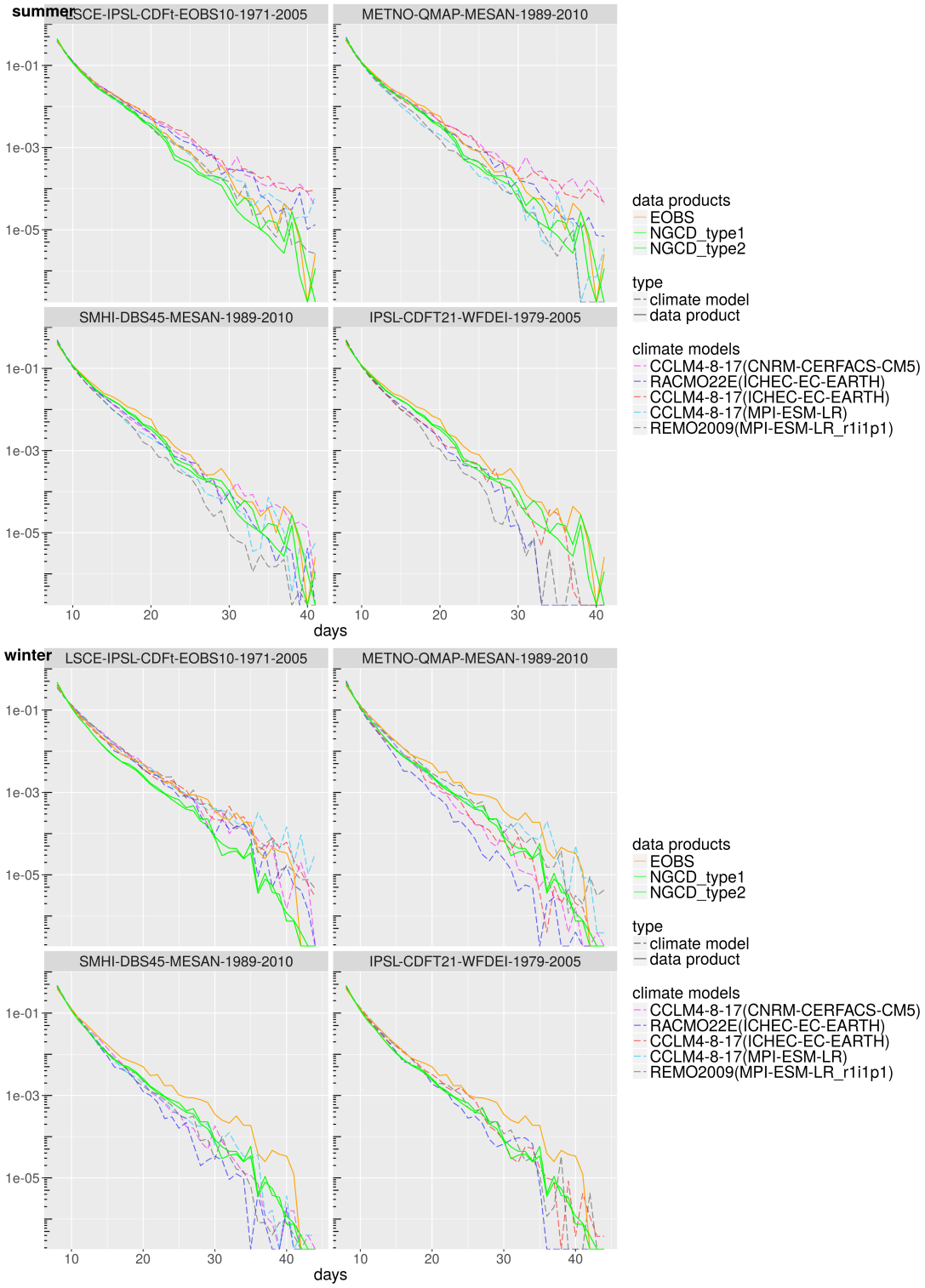


Figure 8. Empirical distributions for the length of all dry-day periods over all grid points. Distributions are plotted in log-scale, and one sub-figure is created for each BCM.

6 Conclusion

We have seen that, for our data, a bias corrected regional climate model output almost always achieves a lower IQD score than the raw model output. We have also found that the change in IQD after bias correction depends the most on which bias correction method is used, not on the choice of which climate model to use.

From previous testing with only the E-OBS data product available, LSCE-IPSL-CDFt-EOBS10-1971-2005 obtained a much lower IQD than all other BCMs and was deemed the best correction method. When testing against the NGCD data products as well we found that LSCE-IPSL-CDFt-EOBS10-1971-2005 still performs well, but now it is often beaten by other BCMs and the overall differences between the BCMs are much smaller. We have found that the data product used in training a BCM greatly affects the result, and the use of the same data product for testing can return overly positive results. Therefore one should always use a completely new data product for testing climate model outputs after bias correction.

When examining precipitation IQD we find that LSCE-IPSL-CDFt-EOBS10-1971-2005 and SMHI-DBS45-MESAN-1989-2010 seems to be the most promising bias correction methods. However, all bias correction methods clearly improve the IQD score, and when comparing with IQD between two data products we can conclude that all four bias correction methods seem to be promising.

When examining dry-days IQD we find that there is a large difference between the E-OBS and the NGCD data products, meaning that one or more of the data products probably is faulty when considering dry-days. The difference between the data products is that E-OBS obtains a higher percentage of very long dry-day periods than the NGCD data products. Because of this difference it is hard to say much about the bias correction methods. LSCE-IPSL-CDFt-EOBS10-1971-2005 agrees with E-OBS, while SMHI-DBS45-MESAN-1989-2010 and IPSL-CDFT21-WFDEII-1979-2005 agree with NGCD. Without further knowledge of the data products we cannot conclude which of these bias correction methods works best when predicting dry-day periods.

References

Førland, E., Allerup, P., Dahlström, B., Elomaa, E., Jónsson, T., Madsen, H., Perälä, J., Rissanen, P., Vedin, H., and Vejen, F. (1996). Manual for operational correction of Nordic precipitation data. *DNMI-Reports*, 24(96):66. 10

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T. (2012). Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods. *Hydrology and Earth System Sciences*, 16(9):3383–3390. 9

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research: Atmospheres*, 113:D20119. doi:10.1029/2008JD010201. 10

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J. (2018). seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth System Science Data*, 10(1):235. 10

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8(2):281–293. 8

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/. 9

Thorarinsdottir, T. L., Gneiting, T., and Gissibl, N. (2013). Using proper divergence functions to evaluate climate models. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):522–534. 9

Tveito, O. E., Bjørdal, I., Skjelvåg, A. O., and Aune, B. (2005). A gis-based agro-ecological decision system based on gridded climatology. *Meteorological Applications*, 12(1):57–68. 10

Vandeskog, S. M., Haugen, M., and Thorarinsdottir, T. L. (2017). Evaluation of precipitation output from the EURO-CORDEX climate ensemble using E-OBS data. NR-notat SAMBA/18/2017:pp. 58. 5

Vrac, M., Noël, T., and Vautard, R. (2016). Bias correction of precipitation through singularity stochastic removal: Because occurrences matter. *Journal of Geophysical Research: Atmospheres*, 121(10):5237–5258. 9

Yang, W., Andréasson, J., Graham, L. P., Olsson, J., Rosberg, J., and Wetterhall, F. (2010). Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies. *Hydrology Research*, 41(3-4):211–229. 9