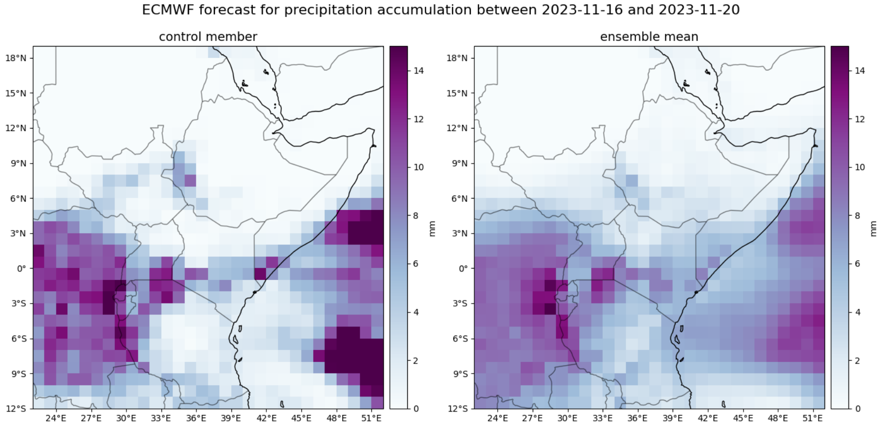
Bias correction of long-range ensemble precipitation forecasts over the Greater Horn of Africa via quantile mapping

Michael Scheuerer

November 2023

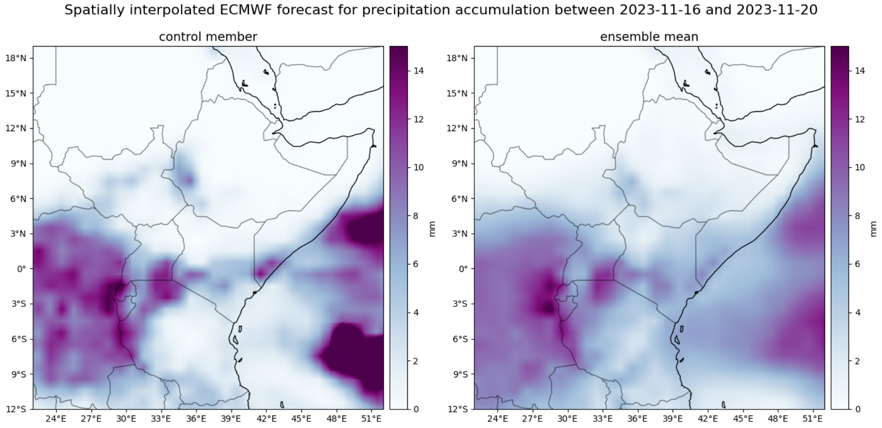
Bias correction of long-range ensemble forecasts

Long-range forecast systems like SEAS5 from ECMWF make predictions several months into the future. To represent the large uncertainty in these forecasts, an ensemble of different predictions is generated.



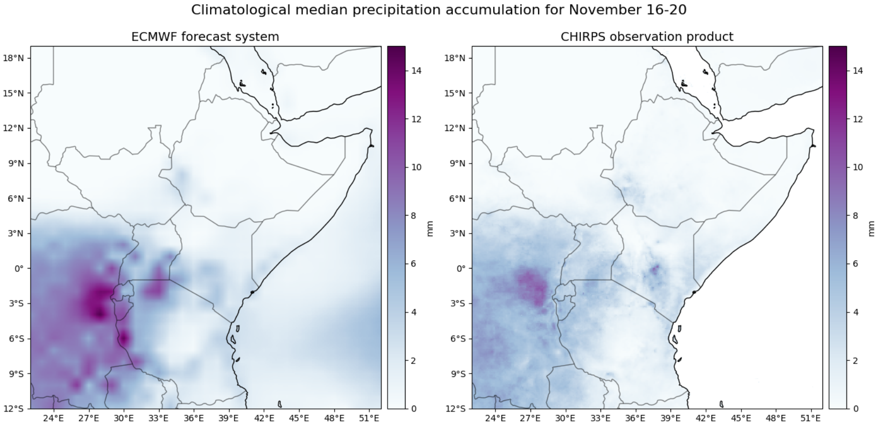
Bias correction of long-range ensemble forecasts

The horizontal resolution of these long-range forecast systems is often rather coarse, and therefore some form of downscaling to a higher resolution is required.



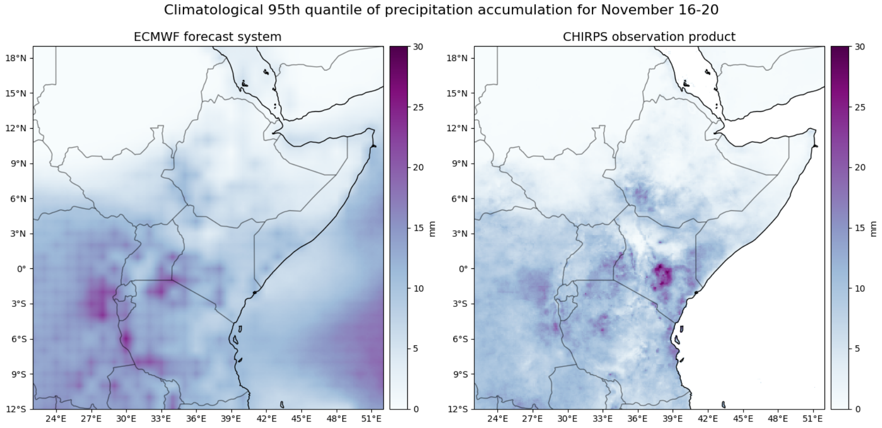
Bias correction of the long-range ensemble forecasts

The climatology of the forecasts (downscaled, or simply interpolated as done here) often differs from the climatology the 'truth' (here: CHIRPS). That means, the forecasts are subject to systematic biases:



Bias correction of the long-range ensemble forecasts

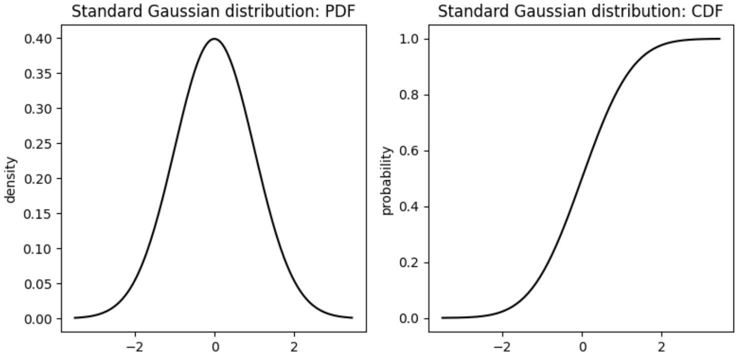
The climatology of the forecasts (downscaled, or simply interpolated as done here) often differs from the climatology the 'truth' (here: CHIRPS). That means, the forecasts are subject to systematic biases:



Quick recap of probability theory: PDFs and CDFs

Let's quickly recap two ways of characterizing a probability distribution:

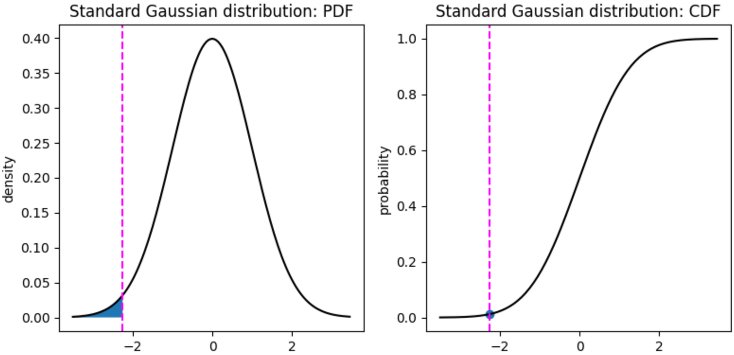
- probability density function (PDF)
- cumulative distribution function (CDF)



Quick recap of probability theory: PDFs and CDFs

Let's quickly recap two ways of characterizing a probability distribution:

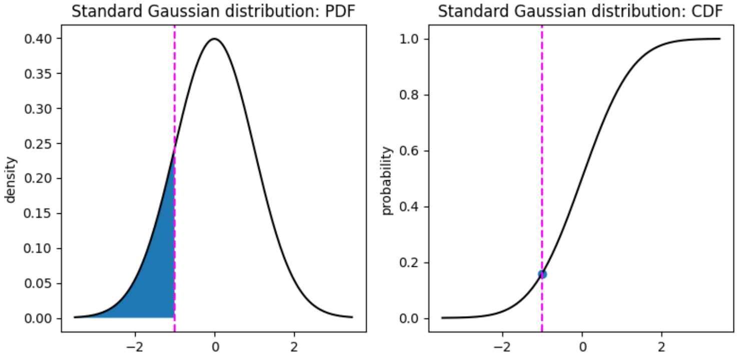
- probability density function (PDF)
- cumulative distribution function (CDF)



Quick recap of probability theory: PDFs and CDFs

Let's quickly recap two ways of characterizing a probability distribution:

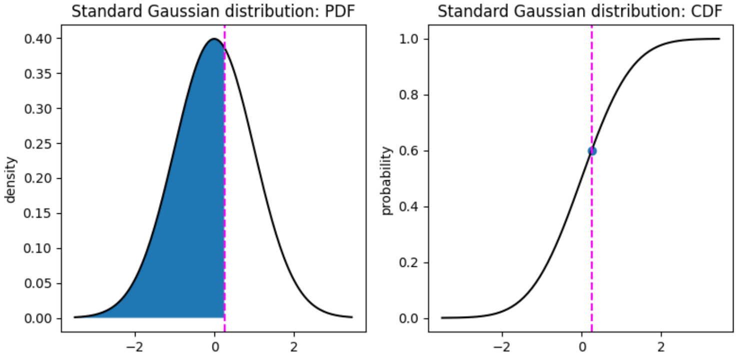
- probability density function (PDF)
- cumulative distribution function (CDF)



Quick recap of probability theory: PDFs and CDFs

Let's quickly recap two ways of characterizing a probability distribution:

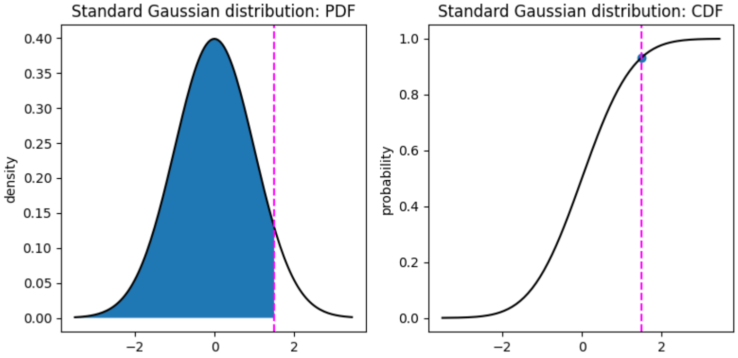
- probability density function (PDF)
- cumulative distribution function (CDF)



Quick recap of probability theory: PDFs and CDFs

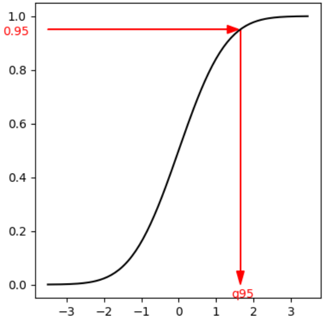
Let's quickly recap two ways of characterizing a probability distribution:

- probability density function (PDF)
- cumulative distribution function (CDF)



Quick recap of probability theory: Quantiles

The CDF is closely related to the quantiles of a probability distribution:



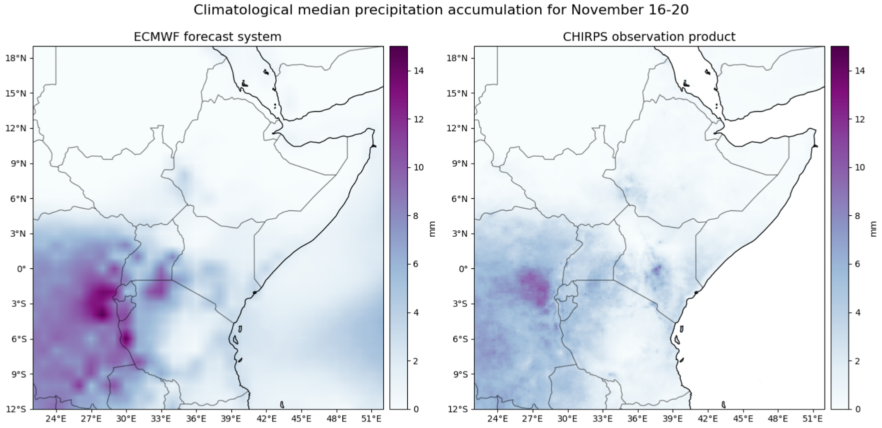
To find the α-quantile, e.g. $\alpha = 0.95$:

- draw a horizontal line at level α
- find the value on the x-axis for which the line intersects the CDF curve

This relation will be useful when we discuss statistical bias correction.

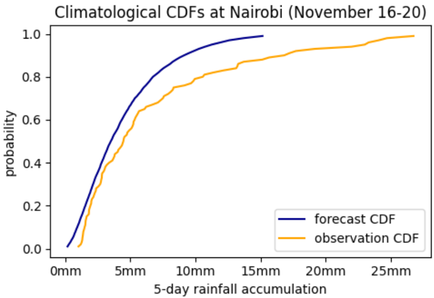
Back to out biased ensemble forecasts

Above, we had already compared the median (i.e. 0.5-quantile) and the 0.95-quantile of the forecasts and observation climatology at each grid point in our map:



Back to out biased ensemble forecasts

If we focus on a single grid point (here: near the centre of Nairobi), we can look at the entire CDFs that describe the climatological distribution of forecast and observed rainfall accumulations for November 16-20.



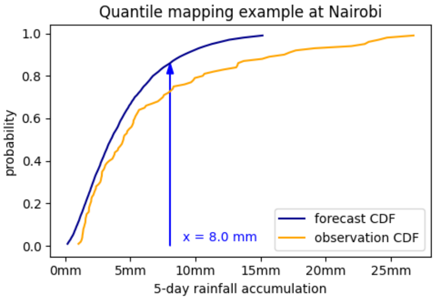
The CDFs are estimated from historical forecast and CHIRPS data.

They can be used to bias-correct future forecasts via quantile mapping.

How does quantile mapping work?

Let's denote the CDF of the forecast climatology by F_{fcst} and the CDF of the CHIRPS climatology by F_{obs}. A bias-corrected version $\tilde{x}$ of a new forecast x is obtained by:

$$\tilde{x} = F_{obs}^{-1}\left(F_{fcst}(x)\right)$$

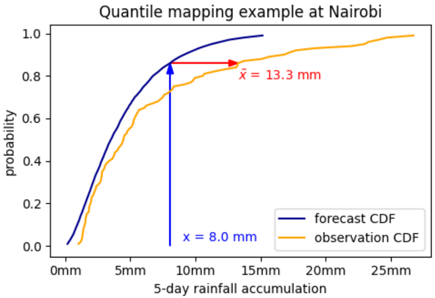


For a given value of x, we first find the probability on the forecast CDF ...

How does quantile mapping work?

Let's denote the CDF of the forecast climatology by F_{fcst} and the CDF of the CHIRPS climatology by F_{obs}. A bias-corrected version $\tilde{x}$ of a new forecast x is obtained by:

$$\tilde{x} = F_{obs}^{-1}\left(F_{fcst}(x)\right)$$

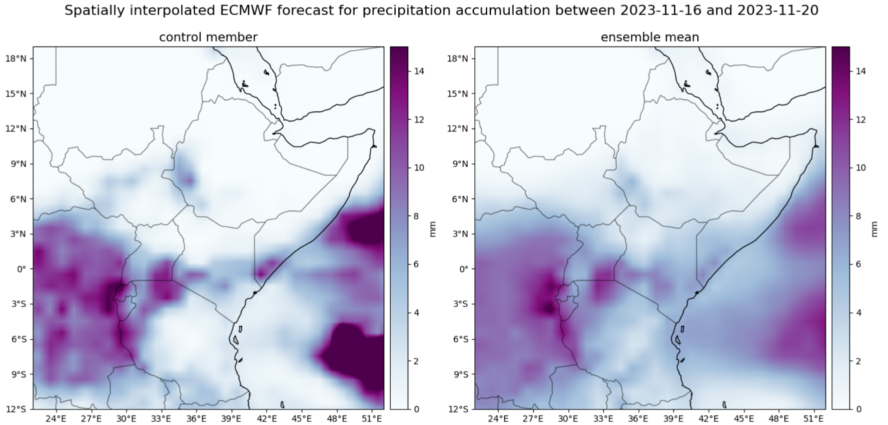


For a given value of x, we first find the probability on the forecast CDF ...

... and then find the quantile of the observation CDF for that probability level.

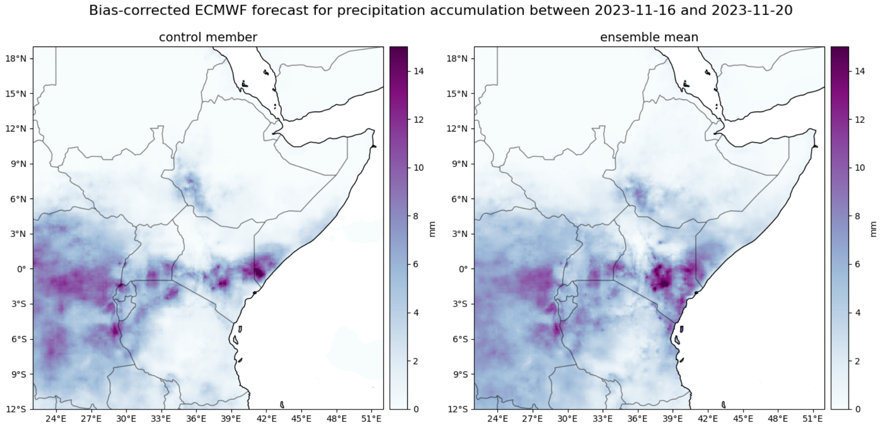
Bias corrected ECMWF forecasts for 11-16 November 2023

After applying the quantile mapping procedure to our example, the bias-corrected forecasts have similar characteristics as the CHIRPS observation data, e.g. heavier precipitation in south-central Kenya:



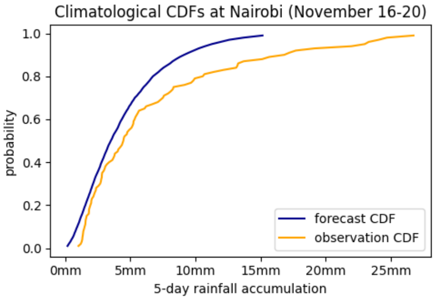
Bias corrected ECMWF forecasts for 11-16 November 2023

After applying the quantile mapping procedure to our example, the bias-corrected forecasts have similar characteristics as the CHIRPS observation data, e.g. heavier precipitation in south-central Kenya:



Some more technical details on quantile mapping

In practice, we don't have estimates of the full CDFs, but we approximate them through estimates of climatological percentiles, i.e. quantiles for levels $\alpha = 0.01, 0.02, \ldots, 0.99$:

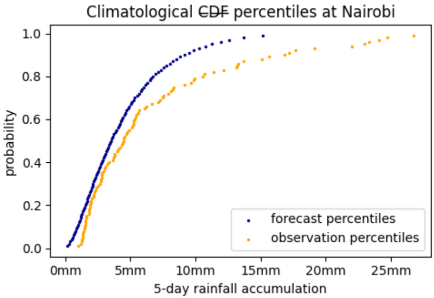


The quantile mapping procedure explained above still works when we only have these points.

It only requires an additional linear interpolation step.

Some more technical details on quantile mapping

In practice, we don't have estimates of the full CDFs, but we approximate them through estimates of climatological percentiles, i.e. quantiles for levels $\alpha = 0.01, 0.02, \ldots, 0.99$:



The quantile mapping procedure explained above still works when we only have these points.

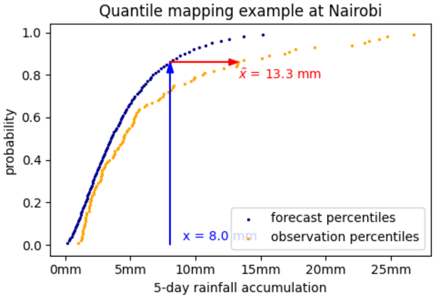
It only requires an additional linear interpolation step.

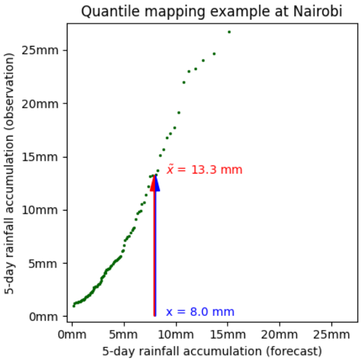
Alternative view on quantile mapping

Instead of plotting the forecast and observation percentiles as an approximation of the respective CDFs, we can just plot them against each other. The mapping

$$\tilde{x} = F_{obs}^{-1}(F_{fcst}(x))$$

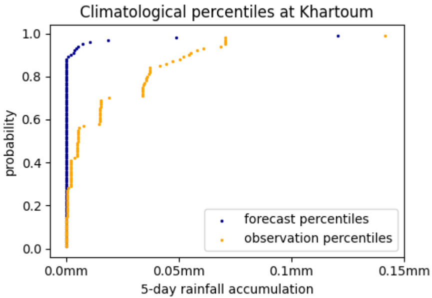
can then be performed in a single step:

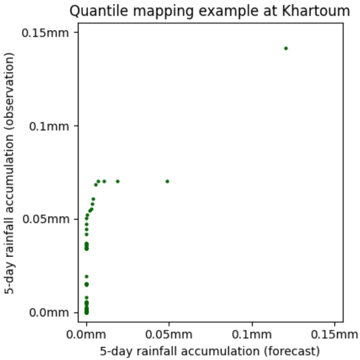




Alternative view on quantile mapping

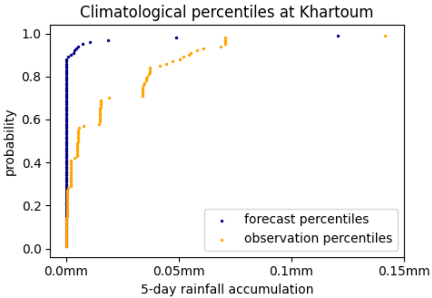
This alternative illustration makes it easier to come up with solutions for challenging situations, e.g. very dry climates where it is difficult to get good estimates of the climatological percentiles.

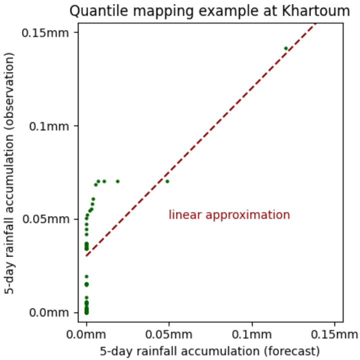




Alternative view on quantile mapping

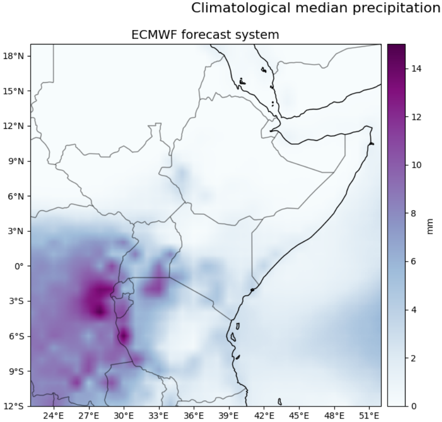
This alternative illustration makes it easier to come up with solutions for challenging situations, e.g. very dry climates where it is difficult to get good estimates of the climatological percentiles.





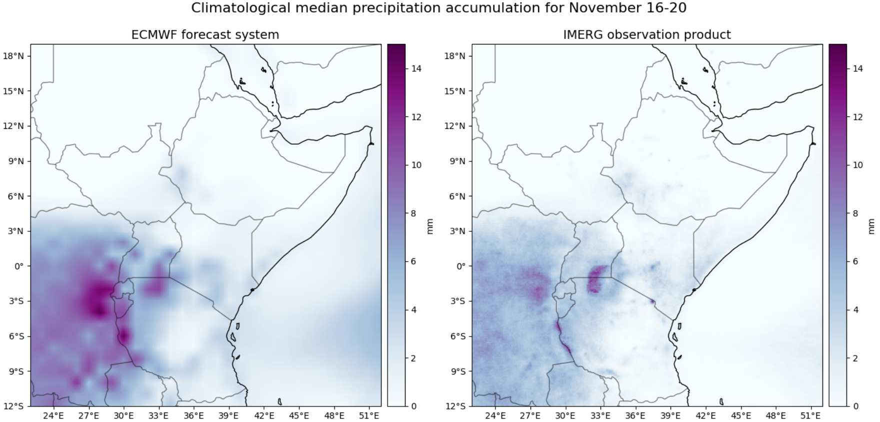
The role of the 'ground truth' data set

So far, we have used the CHIRPS data set as an approximation of the 'true' rainfall amounts. There exist other data sets that could be more suitable for particular regions and applications.



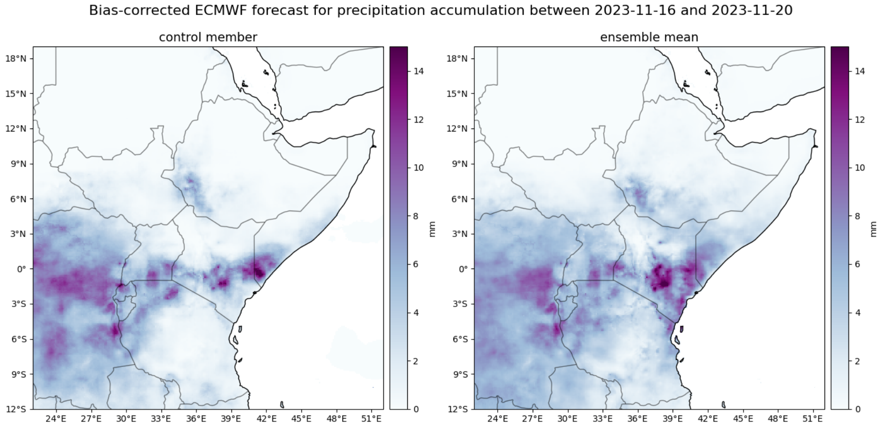
The role of the 'ground truth' data set

So far, we have used the CHIRPS data set as an approximation of the 'true' rainfall amounts. There exist other data sets that could be more suitable for particular regions and applications.



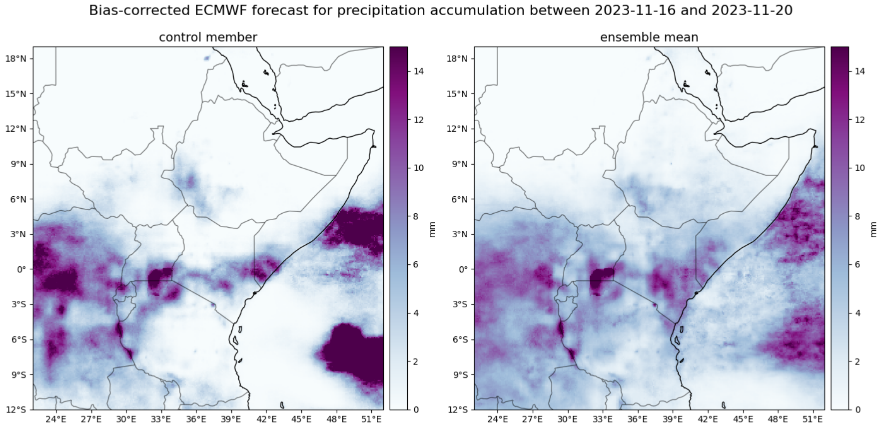
The role of the 'ground truth' data set

Keep in mind that 'bias correction' always means bias-correction against a particular ground truth data set. The resulting forecasts can look very different, depending on the chosen ground truth data set.



The role of the 'ground truth' data set

Keep in mind that 'bias correction' always means bias-correction against a particular ground truth data set. The resulting forecasts can look very different, depending on the chosen ground truth data set.



Exercises

You are now ready to run your own forecast bias-correction.

Exercises can be found at

`https://github.com/ScheuererNR/FoundationalTraining-2023/tree/main`

If the scripts have not already been downloaded to your account, go to your home directory and type

git clone `https://github.com/ScheuererNR/FoundationalTraining-2023.git`

You can either use a Jupyter Notebook 'downscaling_exercise.ipynb', or work with a standard Python script 'downscaling_exercise.py'.

In either case, you will need several Python libraries to run the scripts: standard libraries like numpy, pandas, matplotlib, etc., and the more special libraries xarray and cartopy, introduced in the Python session.