Assignment: Module 6

Learning Goals

Subject knowledge:

- Bias correction
- Statistical downscaling

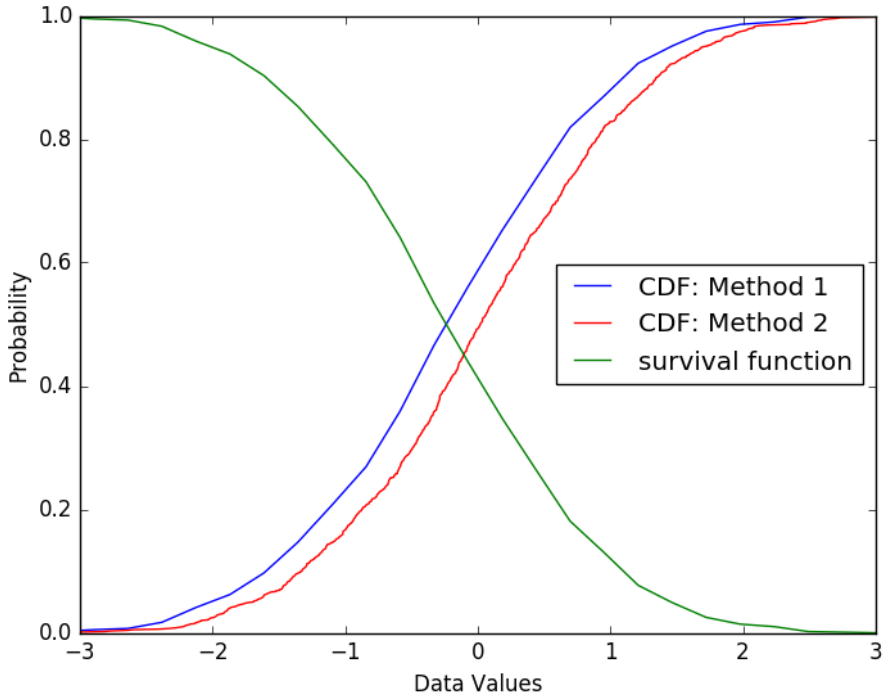
Analysis skills:

- Get familiar with CFS reforecasts and operational forecasts
- Regrid (interpolate) data to a meshgrid of different spatial resolution
- Carry out simple bias correction for mean and variance
- Carry out quantile mapping bias correction
- Estimate the cumulative distribution function
- Carry out statistical downscaling using an analogue method

Example: Calculate CDF

Copy and run a sample script, "/data/zhuowang/a/zhuowang/ATMS521/Sample_Scripts/cdf_2methods.py", and you should generate the following plot.

The sample script shows two methods to estimate the cumulative distribution function (CDF). The first method uses the histogram function and the second method employs a simple sorting function. The two methods produce CDF curves with slight differences.



Overview

Several datasets will be used in this assignment. We will apply different statistical downscaling techniques to monthly mean 2-m air temperature (T2m) in December 2021, which is derived from the CFSv2 operational forecast initialized on November 2, 2021. T2m from the CFSv2 reforecasts and CFS Reanalysis (CFSR) will be used for bias correction, in which the CFS reanalysis will be used as "truth" (or observations).

In Problem 1, we will interpolate T2m from the CFSv2 operational forecasts to a finer-resolution, regional mesh grid.

In Problem 2 and Problem 3, we will apply the simple mean-variance bias correction method and the quantile mapping bias correction method, respectively, to the interpolated T2m derived from Problem 1.

In Problem 4, we will apply an analogue forecast method to forecast T2m in Dec 2021. The best matches will be searched for selected variables from the CFSv2 operational forecasts in Dec 2021 and the CFS reanalysis during Nov-Feb, 1979-2010.

Problem 1

Please interpolate the monthly mean forecast of T2m in December 2021 from the CFSv2 operational forecast to a mesh grid that has the same resolution as the CFR reanalysis. Please follow the steps below. We will focus on a regional domain (220°E -320°E, 20°N -70°N).

1) Read the CFSv2 operational forecasts (6-hourly data) initialized at 0000 UTC Nov 2 2021 and calculate the monthly mean T2m forecast in Dec 2021 over a regional domain (220°E -320°E, 20°N -70°N).
2) Read latitude and longitude information of T2m over a regional domain (220°E -320°E, 20°N -70°N) from the CFSR.
3) The horizonal resolution of the CFSv2 operational forecasts is 1°X1°, and the horizontal resolution of the CFSR is about 0.31°X0.31°. Please interpolate the monthly mean CFSv2 forecast of T2m in Dec 2021 to a mesh grid of the same horizontal resolutions as the CFSR over the domain (220°E -320°E, 20°N -70°N), and plot the original and interpolated T2m forecasts over North America side by side.

Input data:

- CFSR: /data/zhuowang/c/zhuowang/Data/CFSR/flxf01.gdas.${YYYYMM}.grb2 where "flx" denotes the surface data, and YYYY is the four-digit year and MM the two-digit month.
- CFSv2 operational forecasts (6 hourly data): /data/zhuowang/c/zhuowang/Data/CFSv2/2021/flxf2021${MMDDHH}.01.2021110200.grb2 where ${MMDDHH} specifies the forecast validation time, and 2021110200 is the forecast initialization time. For simplicity, we will only consider one ensemble member initialized at 0000 UTC Nov 2, 2021. You will need average all 6-hourly data for Dec 2021 to get the monthly mean.

Problem 2

Please apply the simple mean and variance bias correction to the CFSv2 forecast of monthly mean T2m in Dec 2021 following the steps below. All analysis will be done for a regional domain (220°E -320°E, 20°N -70°N).

Steps:

- Read the monthly mean T2m in December during 1982-2010 from the CFSR and save in a 3D array t2m_o, which is a function of year, latitude and longitude.
- Calculate the monthly mean T2m in December using the 6-hourly data from the CFSv2 reforecasts initiated at 0000 UTC Nov 2 of the same year for each year from 1982 to 2010, and interpolate the data to the same resolution as the CFSR data. Denote the 3D array t2m_f, which has the same dimension as t2m_o and is a function of year, latitude and longitude.
- Calculate the monthly mean T2m forecast in December, 2021 using the 6-hourly data from the CFSv2 operational forecasts initiated at 0000 UTC Nov 2, 2021, and interpolate the data to the same resolution as the CFSR data. Denote the 2D array as T_{2021}, which is a function of latitude and longitude (same as what you have done in problem 1).
- Calculate the mean and standard deviation of t2m_o at each grid point during 1982-2010, denoted as $\bar{T}_o$ and σ_o, respectively.
- Calculate the mean and standard deviation of t2m_f at each grid point during 1982-2010, denoted as $\bar{T}_f$ and σ_f, respectively.
- Bias correct T_{2021} at each grid point using the equation below:

  $$\hat{T}_{2021}=\bar{T}_o + \left(T_{2021} - \bar{T}_f\right) \cdot \sigma_o/\sigma_f$$

  Where $\hat{T}_{2021}$ is the bias-corrected T2m forecast.
- Plot $\hat{T}_{2021}$ and the monthly mean T2m in Dec 2021 from the CFSR side by side, and calculate the RMSE and pattern correlation between the two.

Input data:

- CFSR: /data/zhuowang/c/zhuowang/Data/CFSR/flxf01.gdas.${YYYYMM}.grb2
- CFSv2 operational forecasts in 2021 (6 hourly data):

/data/zhuowang/c/zhuowang/Data/CFSv2/2021/flxf2021${MMDDHH}.01.2021110200.grb2

where “flx” denotes surface data

- CFSv2 reforecasts during 1982-2010 (6 hourly data):

/data/zhuowang/c/zhuowang/Data/CFSv2/${YYYY}/flxf*

The files follow the same name convention as the operational forecasts.

Problem 3

Please apply the quantile mapping bias correction method to the CFSv2 forecast of monthly mean T2m in Dec 2021 following the steps below (the first three steps are the same as Problem 2). All analysis will be done for a regional domain (220°E -320°E, 20°N - 70°N).

Steps:

- Read the monthly mean T2m in December during 1982-2010 from the CFSR and save in a 3D array t2m_o, which is a function of year, latitude and longitude.
- Calculate the monthly mean T2m in December using the 6-hourly data from the CFSv2 reforecasts initiated at 0000 UTC Nov 2 of the same year for each year from 1982 to 2010, and interpolate the data to the same resolution as the CFSR data. Denote the 3D array t2m_f, which has the same dimension as t2m_o and is a function of year, latitude and longitude.
- Calculate the monthly mean T2m forecast in December, 2021 using the 6-hourly data from the CFSv2 operational forecasts initiated at 0000 UTC Nov 2, 2021, and interpolate the data to the same resolution as the CFSR data. Denote the 2D array as T_{2021}, which is a function of latitude and longitude (same as what you have done in problem 1).
- Construct the CDF for each grid point for t2m_r.
- Construct the CDF for each grid point for t2m_f.
- Carry out the quantile mapping at each grid point for T_{2021}, and $\hat{T}_{2021}$ denotes the bias-corrected T2m forecast.
- Plot $\hat{T}_{2021}$ and the monthly mean T2m in Dec 2021 from the CFSR side by side, and calculate the RMSE and pattern correlation between the two.

In the quantile mapping method, forecast values are replaced with observation values with the same percentiles as shown in the figure below.

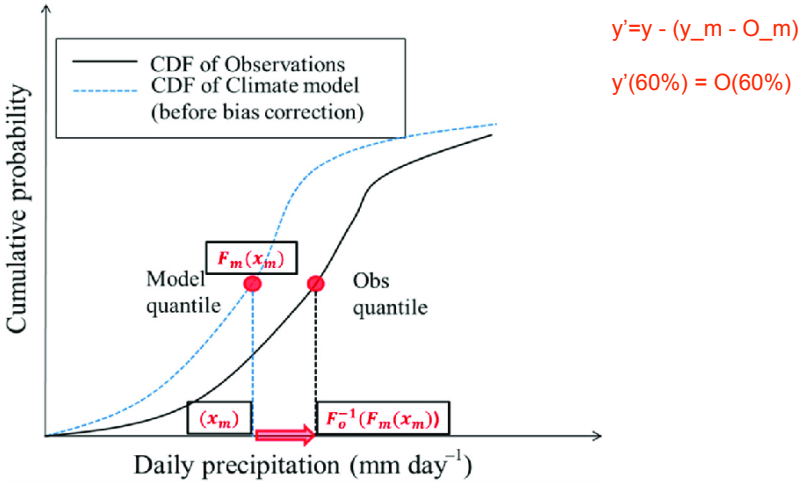


Figure 1 The demonstration of quantile mapping bias correction approach (From Gupta et al. 2019, https://doi.org/10.3390/w11051102)

Problem 4

We will apply an analogue method for statistical downscaling of T2m in Dec 2021.

The CFSv2 operational forecasts of monthly mean **anomalies** of 500-hPa geopotential height (H500) and surface pressure (Ps) in Dec 2021 (forecasts initialized at 0000 UTC Nov 2, 2021) will be compared to the monthly mean anomalies of H500 and Ps from the CFS reanalysis. The month that has strongest pattern correlations with the CFSv2 forecasts is chosen as its analog. The observed T2m from the CFSR in that month is then taken as the forecast for Dec 2021. To increase the sample size, we will search for the best match in Jan, Feb, Nov and Dec during 1979-2010. The data are deseasonalized by removing the corresponding long-term mean month mean.

We will focus on the regional domain (220°E -320°E, 20°N -70°N) for T2m forecasts (same as Problems 1-3), but will examine a slightly larger domain (180°E -340°E, 15°N -75°N) to search for the analog for H500 and Ps.

Please follow the steps below.

- Calculate the long-term mean monthly mean H500 and Ps in Dec during 1982-2010 using the 6-hourly data from the CFSv2 reforecasts initiated at 0000 UTC Nov 2 in each year from 1982 to 2010.
- Calculate the monthly mean H500 and Ps in December, 2021 using the 6-hourly data from the CFSv2 operational initiated at 0000 UTC Nov 2, 2021, and calculate the monthly mean anomalies by removing the corresponding long-term monthly mean in Dec derived from the CFSv2 reforecasts derived in the previous step.
- Calculate the monthly mean H500 and Ps in Jan, Feb, Nov and Dec during 1979-2010 from the CFS reanalysis (there are 32*4=128 months). Calculate the monthly mean anomalies by removing the corresponding long-term mean monthly mean over 1979-2010.
- Please calculate the pattern correlations of H500 and Ps monthly mean anomalies between the CFSv2 forecast in Dec 2021 and the CFS reanalysis in each of the 128 months. Although one can assign different weights to H500 and Ps, we will take the sum of the H500 and Ps pattern correlations for simplicity.
- Please select three months with the highest pattern correlation sums. The averaged T2m over those months from the CFSR can be regarded as the forecast for Dec 2021. Please plot this average (i.e., your forecast) along with the T2m in Dec 2021 from the CFSR (i.e., the corresponding observation), and calculate the RMSE and pattern correlation between the two.

Input data:

- CFSv2 operational forecasts in 2021 (6 hourly data):
/data/zhuowang/c/zhuowang/Data/CFSv2/2021/pgbf2021${MMDDHH}.01.2021110200.grb2
where “pgb” denotes the pressure-level data
- CFSv2 reforecasts during 1982-2010 (6 hourly data):
/data/zhuowang/c/zhuowang/Data/CFSv2/${YYYY}/pgbf*
The files follow the same name convention as the operational forecasts.

- CFS reanalysis monthly mean pressure-level data (pgb*) and surface data: (flx*)

  /data/zhuowang/c/zhuowang/Data/CFSR/pgbhnl.gdas.${YYYYMM}.grb2
  /data/zhuowang/c/zhuowang/Data/CFSR/flxf01.gdas.${YYYYMM}.grb2
  /data/zhuowang/c/zhuowang/Data/CFSR/pgbhnl.gdas.202112.grib2
  /data/zhuowang/c/zhuowang/Data/CFSR/flxf01.gdas.202112.grib2

The CFSv2 analysis was used to extend the CFS reanalysis beyond 2010. Please note that the CFSv2 analysis in 2021 has different horizontal resolution from the CFS reanalysis during 1979-2010. You need to regrid the data to the same resolution before calculating the anomalies, pattern correlation and RMSE in Problem 4. More information on the CFSv2 products after 2010 can be found at https://rda.ucar.edu/datasets/ds094.2/