Contribution to D12 (ARPA-SMR,AUTH and ADGB)
Downscaling of extreme events in Emilia-Romagna and Northern Italy-regional analysis

ARPA-SMR (Downscaling Methods – MLR and CCA)
C. Cacciamani, R. Tomozeiu, A. Morgillo, V. Pavan

ARPA-SMR has been developed two statistical downscaling models that consist on the application of the relationships identified between the large-scale and smaller-scale climate. These models downscale seasonal indices of extreme rainfall and temperature at eight stations from Emilia-Romagna. Then, the methods has been applied on eight stations from Greece, that has been designed as secondary study region for ARPA-SMR

Data

<u>Predictors and Predictands</u>

Both methods are based on the multiple linear regression with predictors derived from NCEP reanalysis: geopotential height at 500hPa (Z500), mean sea level pressure (MSLP), temperature at 850hPa (T850) and specific humidity at the levels: 1000hPa, 950hPa, 850hPa, 750hPa.
In order to find an optimum area of predictors such as the skill of the statistical models to have high performance, it has been tested different windows:

- 90°W-90°E; 0°N-90°N (Z500,MSLP,T850)- area A
- 60°W-60°E; 20°N-90°N (Z500,MSLP,T850)- area B
- 12.5°W-30°E; 30°N-55°N (Z500,MSLP,T850)- area C
- 35°W-35°E; 30°N-60°N (Z500,MSLP,T850)- area D
- 5°E-20°E; 37.5°N-47.5°N (specific humidity)- area E

The predictands are the seasonal “core” extreme indices (7 indices for precipitation and 6 indices for temperature) computed for the period 1958-2000 at eight stations from Emilia-Romagna:

- mean daily rainfall (pav)
- 90th percentile of rainday amounts (pq90)
- greatest 5-day total rainfall (px5d)
- simple daily intensity (pint)
- maximum number of consecutive dry days (pxcdd)
- % of total rainfall from events greater than long term pq90 (pfl90)
- number of events greater than long-term 90th percentile of raindays (pnl90)
- mean maximum temperature (txav)
- 90th percentile of maximum temperature (tmax90)
- Mean minimum temperature (tnav)
- Number of frost days (tnfd)
- Heat wave duration (txhw90)

Methods

The methods tested are:

1. Multiple Linear Regression based on the principal components of the data sets used in the analysis (MLR);

2. Multivariate regression based on Canonical Correlation Analysis (CCA)

The models have been calibrated on the period 1958-1978 & 1994-2000 and validated on the period 1979-1993.The Pearson correlation coefficient, BIAS and RMSE were used as the skill measures of the models. A description of the methods is presented in the following.

Statistical Downscaling model based on MLR

Multiple Linear Regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. The predictors used in this case were the Principal Components (PCs) of the Z500, MSLP, T850 while the predictands were the time series of the extreme indices of temperature and precipitation at eight stations from Emilia-Romagna. The number of PCs used in the model vary, firstly it has been used only the first 4PCs, then it has been added the number of PCs such as the explained variance to be 97.5% from the total variance. The skill of the model has been slightly improved in the second situation.

Statistical Downscaling model based on Canonical Correlation Analysis (CCA)

The statistical model based on the Canonical Correlation Analysis was introduced in the climate research by Barnett and Preisendorfer (1987). This technique finds pairs of patterns such as the correlation between two corresponding pattern coefficient is maximized. In order to determined these Canonical Correlation Patterns (CCP) and the canonical correlation coefficients, firstly it is has been calculated the covariance matrices and the cross covariance between the fields. From the products of these matrices the adjoint patterns are derived as eigenvectors. In order to reduce the noise of the field involved, before the CCA, the data sets are projected on EOFs (empirical orthogonal functions) and only those explaining the most of the total observed variance are retained. The most important CCA pairs are then used in a multivariate linear model in order to estimate the predictand anomalies from the predictor anomaly field.

Results

The skill of the statistical model is dependent on the station, predictors and the numbers of EOFs predictors retained in the CCA analysis. The optimum areas of predictors are: “C/D” and “B”area.
Figures 1 and 3 present the Spearman correlation for hindcast precipitation/temperature indices while the figures 2 and 4 present the BIAS of each extreme indices. The comparison between the skill of both downscaling methods, presented in figures 1, 2, 3, 4 (CCA_ARPA and MLR_ARPA) emphasis that for some extreme indices, the performance of the model based on CCA is a little higher than those based on MLR.
The season with best skill of extreme precipitation and temperature is winter for both methods (Figure 1, 2, 3, 4-CCA_ARPA and MLR_ARPA) followed by spring and autumn, while summer has a lower skill especially for extreme precipitation.
The mean daily precipitation, mean minimum and maximum temperature are the indices best downscaled followed by the number of consecutive dry days, number of frost days, 10th percentile of minimum temperature, 90th percentile of maximum temperature, and heat wave duration index (see Figure 1, 2, 3, 4-CCA_ARPA and MLR_ARPA). The indices with low skill are: number of events greater than 90th percentile of precipitation, intensity of precipitation, 90th percentile of precipitation and greatest 5 days total rainfall. Similar conclusions has been obtained when the same methods (CCA and MLR) has been applied to another data set, represented by the time series of extreme events computed at eight Greek stations, that represent for ARPA–SMR the second region to test (results not shown).
AUTH (University of Thessaloniki-Greece) has been downscaled the extreme indices at eight stations from Emilia-Romagna using as methods: Multiple Linear Regression, Canonical

Correlation Analysis(CCA) and Artificial Neural Network (Nnet).The performance of the models are displayed in the figures 1, 2, 3, 4 (abbreviations: MLR_AUTH,CCA_AUTH, NNet_AUTH). The results provided by AUTH and obtained with the methods CCA and MLR are similar with those obtained by ARPA –SMR.

References

Barnett, T. P. and Preisendorfer, R. 1987. Origin and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis". *Mon. Wea. Rev*., 1825-1850.

von Storch, H., Zorita, E. and Cubasch, U. 1993. Downscaling of global climate change estimates to regional scale: An application to Iberian rainfall in wintertime', *J. Clim.* **6**, 1161-1171

Wilks D. 1995: Statistical method in the atmospheric Sciences. – Academic Press.

ADGB (Downscaling Method - HYPER4)

E. Tosi and S. Alberghi

Characteristics:
- a "type 2" method (daily downscaling, then calculation of indices)
- downscales precipitation
- calculates number of extremes over 90th perc. of rainy days in the whole period (index Pnl90)
- Calibration period: 1966-1990. Verification period: cross validation on 1966-1990
- not seasonally-based, but a unique full-year calibration (to eventually allow annual cycle shifts)
- gives a unique index of precipitation for all Northern Italy, not indexes on single stations

The method uses precipitation data from MAP. A "rainy day" on Northern Italy is defined as a day in which at least 1mm of precipitation falls in at least 25 neighbouring grid points, excluding thunderstorms, i.e. peaks in the precipitation pattern. Using daily maximum precipitation values on Northern Italy a list of observed extreme events days (with precipitation greater than 95th percentile of the entire distribution) is drawn.

Afterwards EOF analysis is performed and a daily series of the first two precipitation PCs is used in the subsequent DS method. An index is computed from the "observed" PCs (sum of squares of adimensional PCs), that is well correlated (>0.85) with max. precipitation value on Nortern Italy. 90th percentile of this index in rainy days is used as a threshold for Pn90 calculation.

NCEP parameters used are:
- 500hPa geopotential height anomaly (2 grid points)
- 500hPa geostrophic wind direction
- 700hPa relative humidity
- 500-850hPa thickness gradient direction
- precipitable water (calculated from specific humidity at 500, 700 and 850hPa)

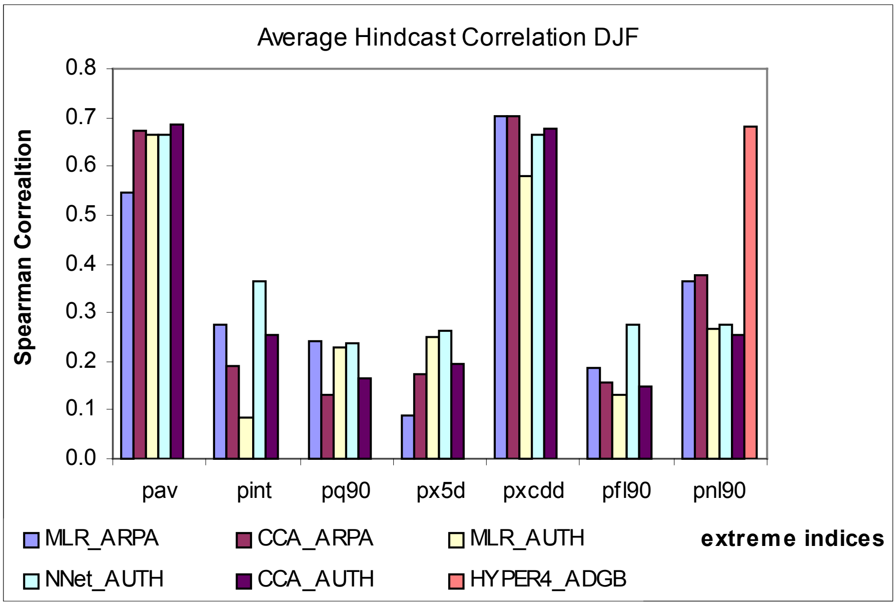
Except that geopotential height anomaly, other parameters are considered in only 1 grid point (the most selective, or the most "predictive" one).

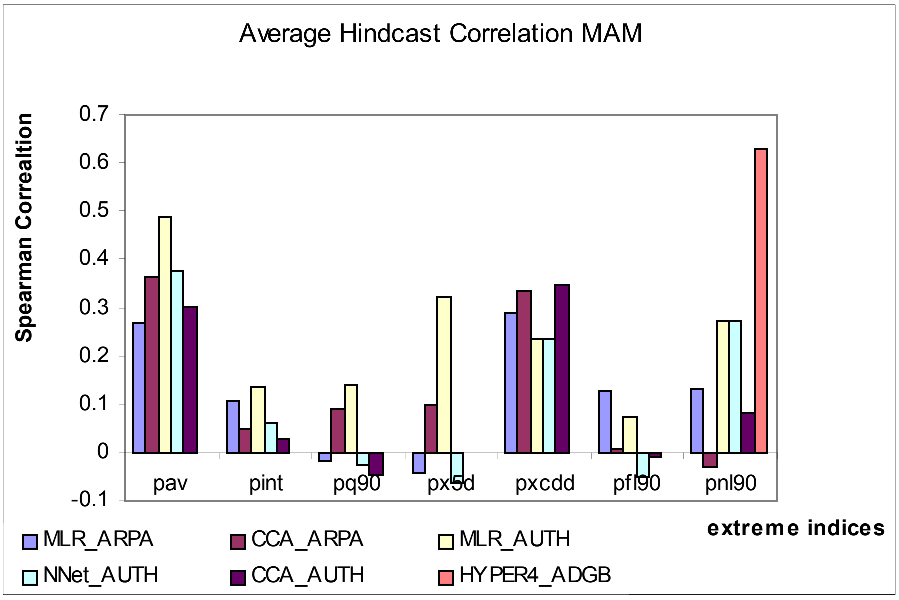
The method consists of a two-steps algorithm:
- First, a preselection of "potentially extreme" days, based on the range of values that large-scale NCEP parameters undertake in the observed extreme events list. This "pre-selections" allows to discard about 60% of days

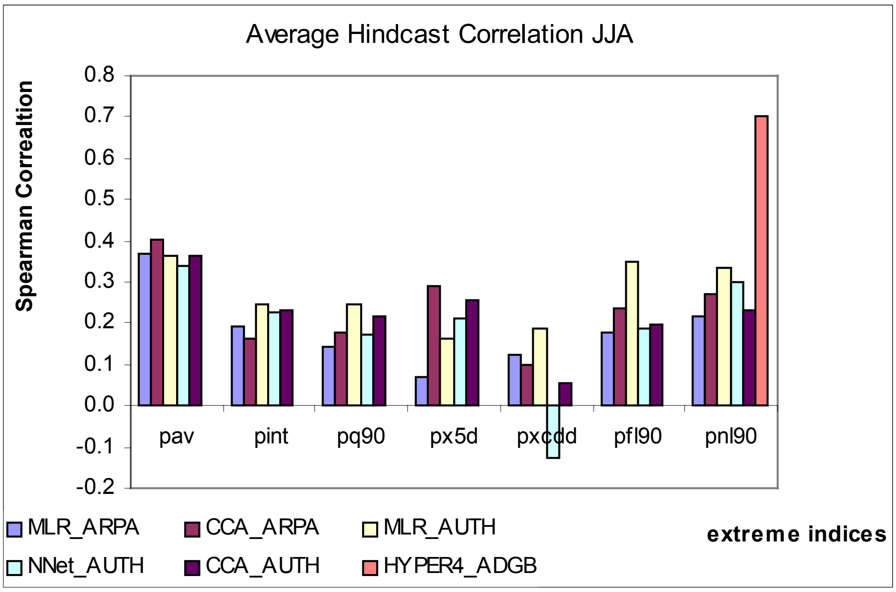
- Second, a resampling (random) procedure in the 4-dimensional hyper-space of parameters (thickness is not used) to reproduce correctly the statistics of precipitation events for a given set of large-scale parameters values. PC1 and PC2 are separately downscaled for each pre-selected day, then the 'sum of squares of PCs' index is computed. Finally Pn90 is calculated. The procedure is iterated 100 times to have a better stability of results. Mean values are considered.

Note that:
- rainy days on Northern Italy (as defined above) are about 55% of total days
- a 90th percentile calculated on rainy days only is used
- BIAS displayed in figure (2 abbreviation: HYPER4_ADGB) and RMSE are dimensional quantities, thus comparison with other methods for this region is not really fair (on a single Northern Italy station, rainy days are less than one third of total)

Figure 1 Spearman Correlation for hindcast precipitation indices –average over all stations from Emilia-Romagna (by ARPA-SMR and AUTH) and N-Italy (by ADGB) and different downscaling methods







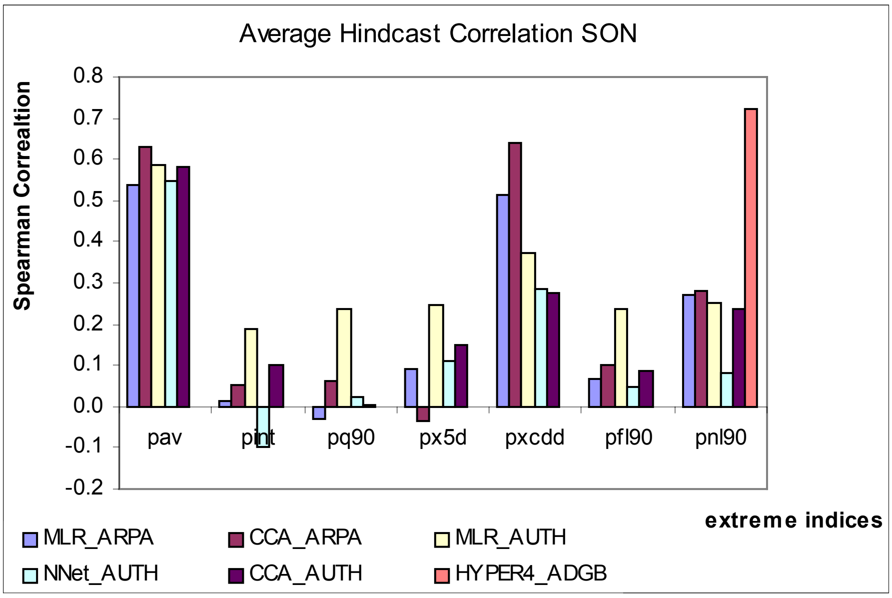
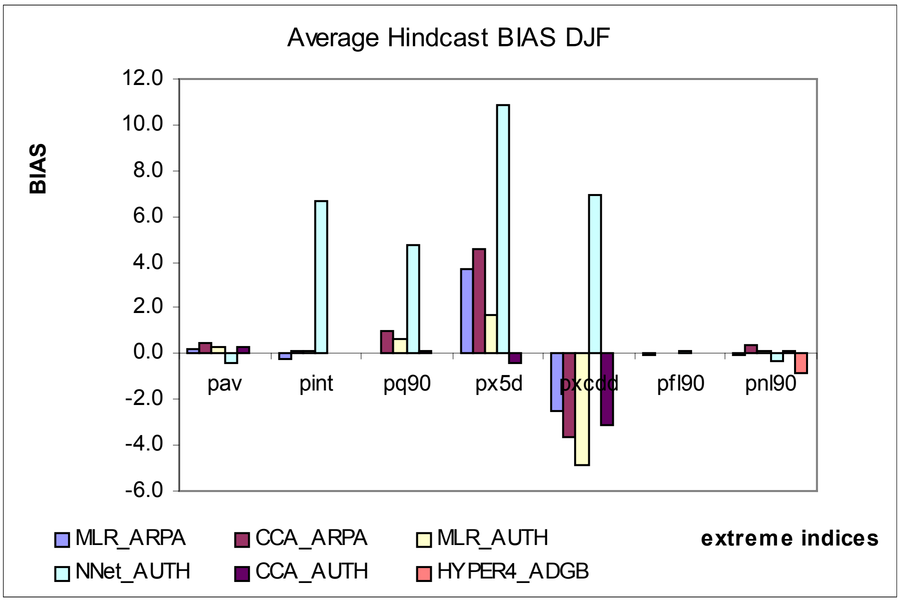
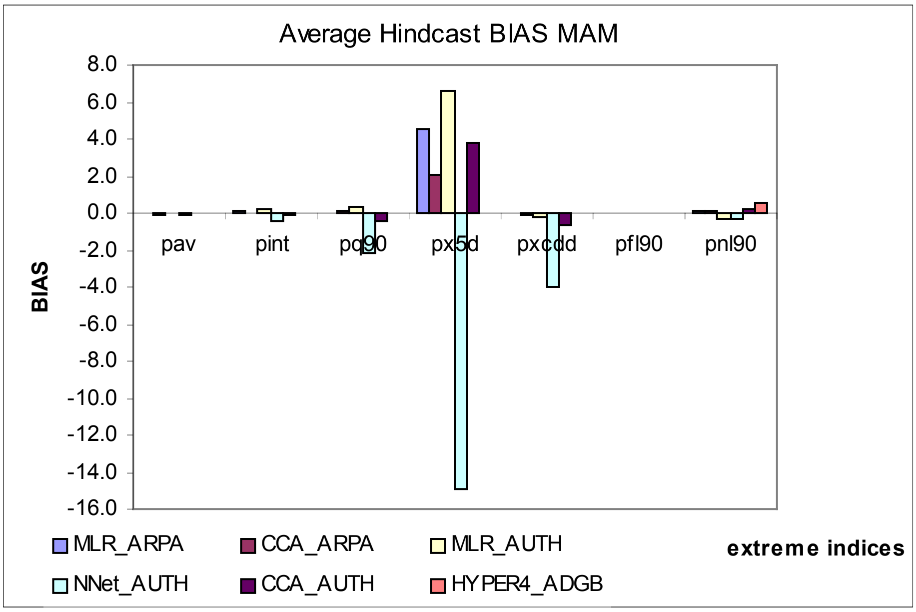
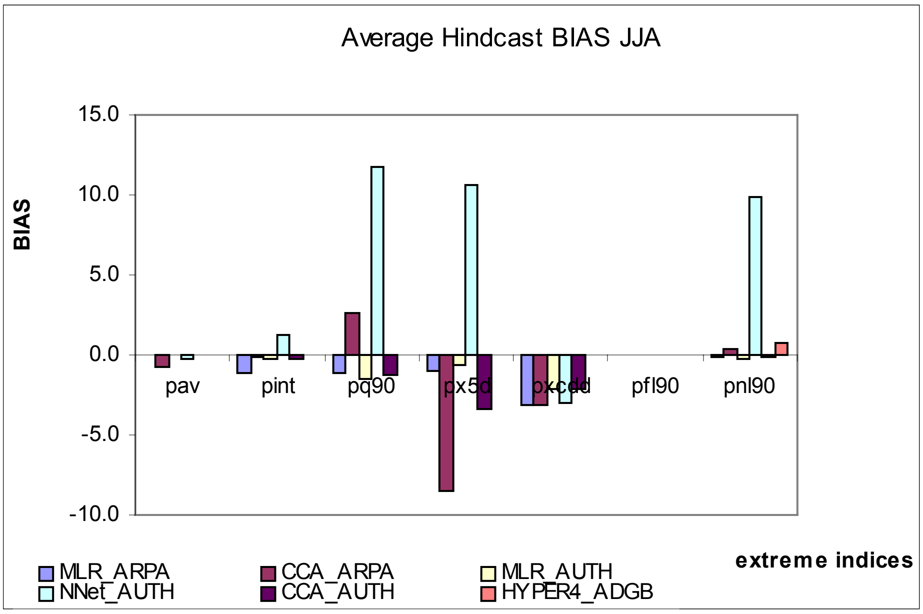


Figure 2 BIAS for hindcast indices –average over all stations from Emilia-Romagna (by ARPA-SMR and AUTH) and N-Italy (by ADGB) and different downscaling methods




Average Hindcast BIAS MAM
BIAS
8.0
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0
-8.0
-10.0
-12.0
-14.0
-16.0
pav
pint
pq90
px5d
pxcdd
pfl90
pnl90
MLR_ARPA
CCA_ARPA
MLR_AUTH
NNet_AUTH
CCA_AUTH
HYPER4_ADGB
extreme indices


Average Hindcast BIAS JJA
BIAS
15.0
10.0
5.0
0.0
-5.0
-10.0
pav
pint
pq90
px5d
pxcdd
pfl90
pnl90
MLR_ARPA
CCA_ARPA
MLR_AUTH
NNet_AUTH
CCA_AUTH
HYPER4_ADGB
extreme indices

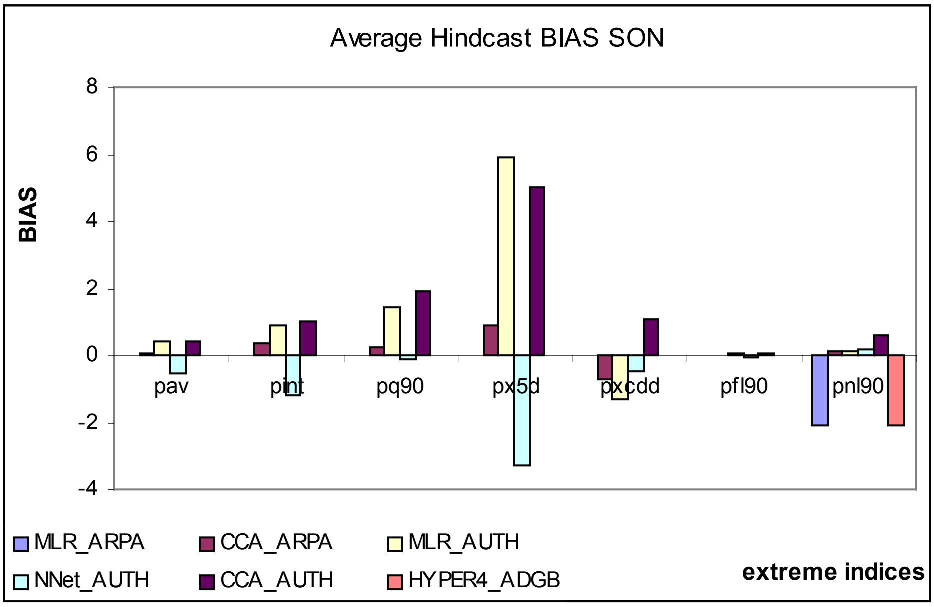
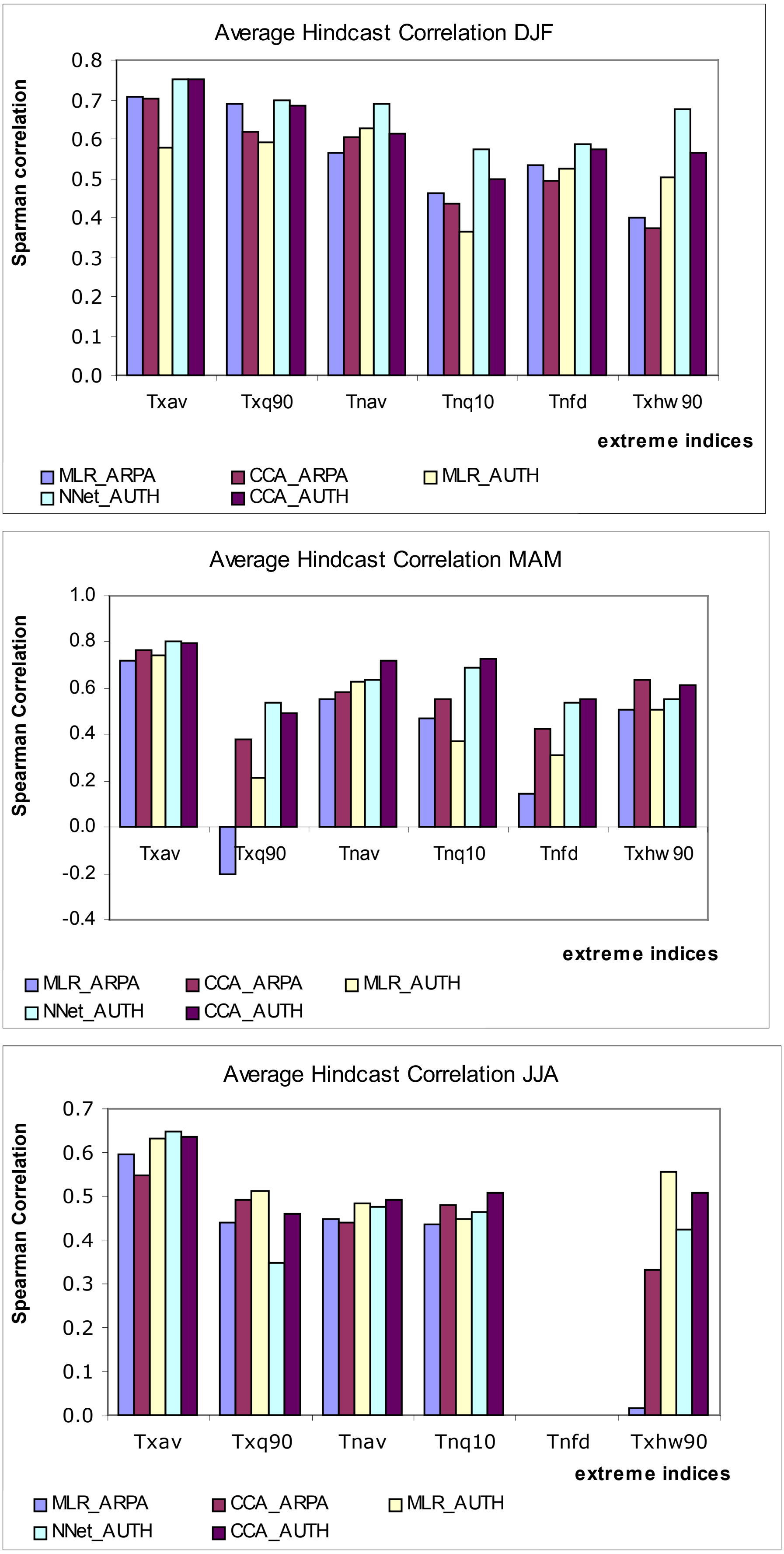

Average Hindcast BIAS SON
BIAS
8
6
4
2
0
-2
-4
pav
pint
pq90
px5d
pxcdd
pfl90
pnl90
MLR_ARPA
CCA_ARPA
MLR_AUTH
NNet_AUTH
CCA_AUTH
HYPER4_ADGB
extreme indices

Figure 3 Spearman Correlation for hindcast temperature indices –average over all stations from Emilia-Romagna (by ARPA-SMR and AUTH) and N-Italy (by ADGB) and different downscaling methods



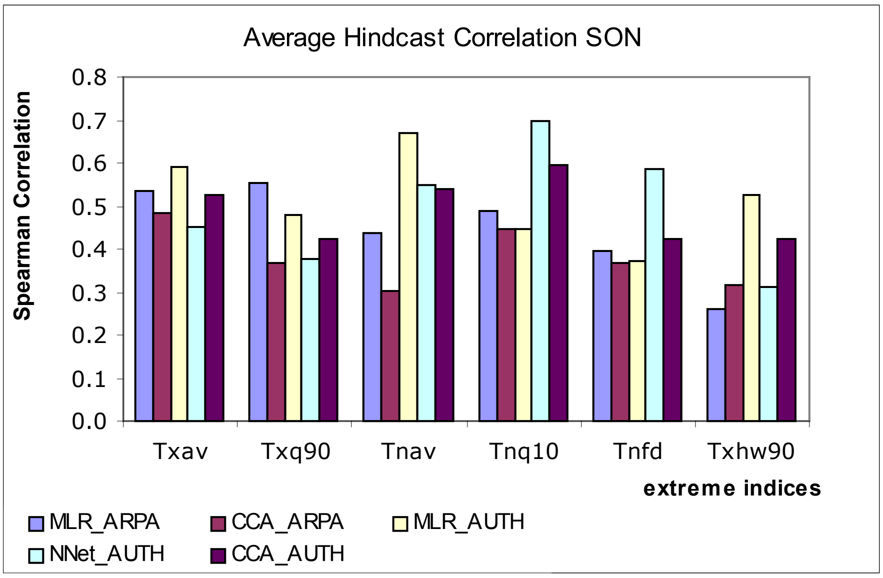
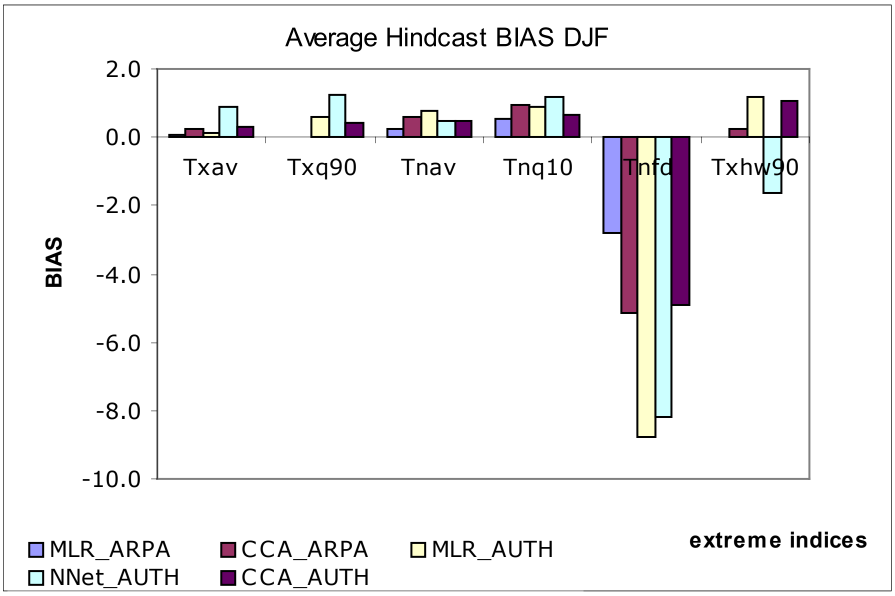
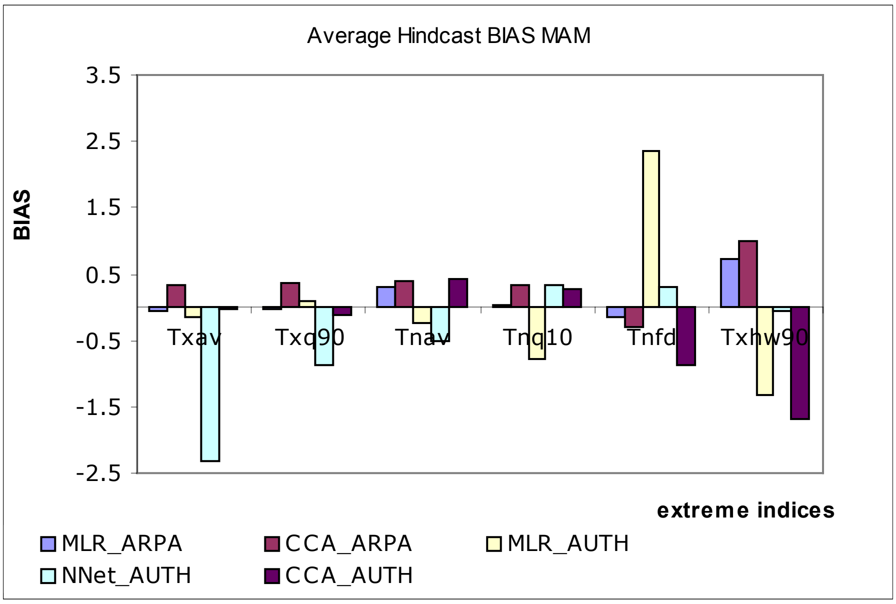
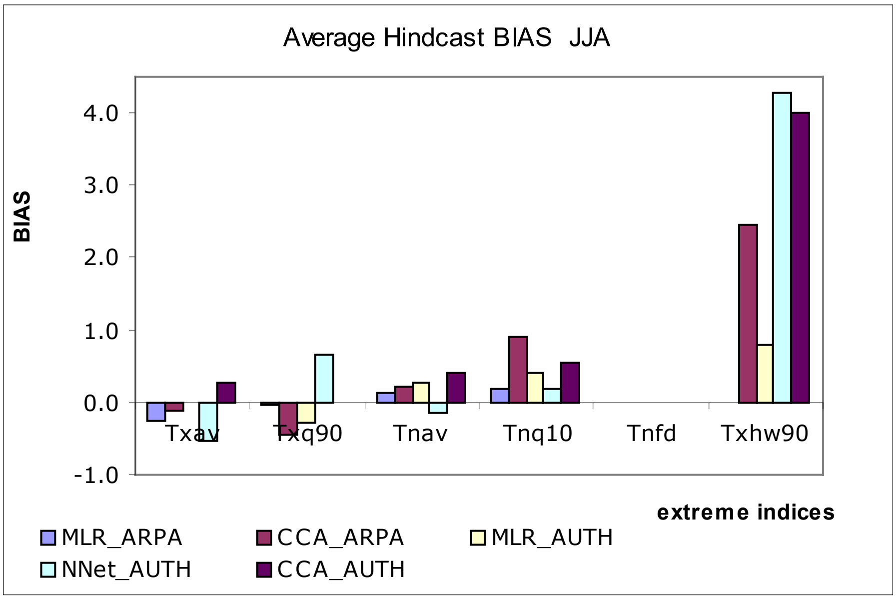


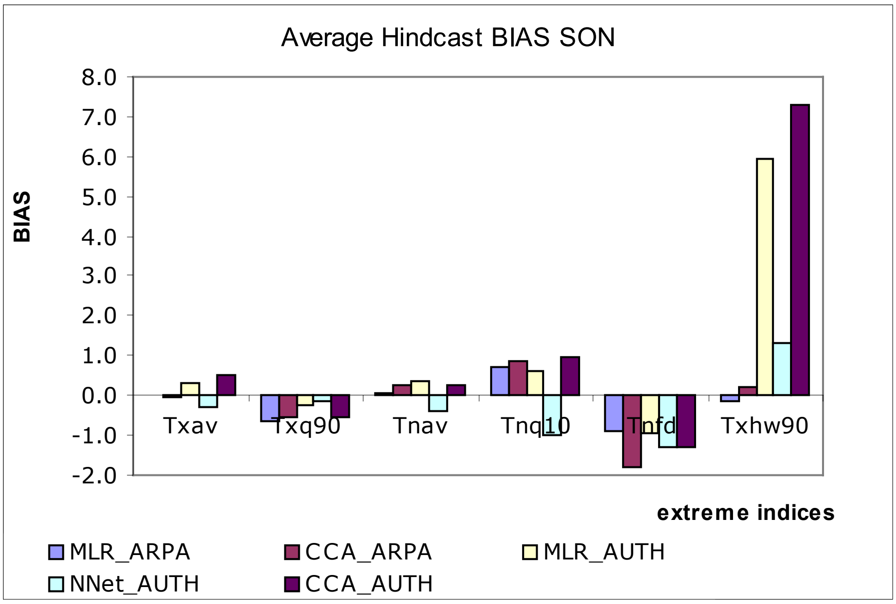
Figure 4 BIAS for hindcast temperature indices –average over all stations from Emilia-Romagna (by ARPA-SMR and AUTH) and N-Italy (by ADGB) and different downscaling methods







Average Hindcast BIAS JJA
BIAS
4.0
3.0
2.0
1.0
0.0
-1.0
Txav
Txq90
Tnav
Tnq10
Tnfd
Txhw90
extreme indices
MLR_ARPA
CCA_ARPA
MLR_AUTH
NNet_AUTH
CCA_AUTH



Average Hindcast BIAS SON
BIAS
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
Txav
Txq90
Tnav
Tnq10
Tnfd
Txhw90
extreme indices
MLR_ARPA
CCA_ARPA
MLR_AUTH
NNet_AUTH
CCA_AUTH