Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

# D2B.31: Statistical Downscaling of ENSEMBLES RCM simulations

Due date of deliverable: February 2009
Actual submission date: September 2009

Start date of project: 1 September 2004 Duration: 60 Months

Met Éireann

Revision 1

| Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) | | |
|---|---|---|
| **Dissemination Level** | | |
| **PU** | Public | x |
| **PP** | Restricted to other programme participants (including the Commission Services) | |
| **RE** | Restricted to a group specified by the consortium (including the Commission Services) | |
| **CO** | Confidential, only for members of the Consortium (including the Commission Services) | |

# Statistical downscaling of ENSEMBLES RCM simulations

**D2B.31**

**Tido Semmler, Eoin Sherlock, Ray McGrath**

**Met Éireann**

# 1 Introduction

Statistical downscaling is usually applied to GCM simulations with coarse resolutions. These GCM simulations generally have horizontal resolutions of 100 to 300 km which are too coarse to capture local effects and extreme events. Wilby et al. (2004) provide a comprehensive background on the statistical downscaling approach which is a well-established method to predict mean and extreme values of near-surface parameters through statistical relationships between large-scale meteorological parameters and local near-surface parameters. Weather typing, weather generators and regression methods are three commonly used groups of statistical downscaling approaches. Schmith and Goodess (2005) have investigated the performance of different statistical downscaling methods. Both Wilby et al. (2004) and Schmith and Goodess (2005) find that none of the investigated methods can be identified as the single best method. Depending on the application, one or the other method can have slight advantages over other methods. One major concern about statistical downscaling is the consistency of these methods in different climate regimes. While the methods are usually developed for a comparatively short time period with only limited climate change, e. g. the second half of last century, they are often applied to a time period with a stronger climate signal from the GCM simulations. Frias et al. (2006) and Timbal et al. (2008) address this concern by investigating the consistency of statistical downscaling methods with GCM results using past and future climate simulations. The conclusion is that the choice of predictors is important for the consistency.

In this study, the possibility of statistical downscaling of dynamically downscaled GCM simulations is explored. Usually GCM data are either dynamically or statistically downscaled to capture local effects. The quality of results of dynamically and statistically downscaled GCM data is very similar (Christensen et al., 2007). The statistical downscaling is much less computationally demanding while requiring a careful selection of the predictors (Schmith and Goodess, 2005). While dynamical and statistical downscaling methods are usually seen as competing methods, they are combined in this study: the GCM data are first dynamically downscaled (DD) using an RCM and then statistically downscaled (SD). Since the RCM simulations are already on a high resolution (25 km) grid, and therefore capture local effects to some degree, the aim of the statistical downscaling in this study is focused on extreme events which tend to be underestimated in RCM simulations. It is explored if the additional statistical downscaling can correct the biases of RCM simulations.

# 2 Methods

Seven different ENSEMBLES RCM simulations are statistically downscaled: the ECHAM5-driven C4I, KNMI, MPI and SMHI simulations, the BCM driven METNO and SMHI simulations and the ARPEGE-driven DMI simulation. The 10 major STARDEX indices (Table 1, with indices taken from Schmith and Goodess, 2005, their Table 2) are analysed from all DD and SD simulations for present-day and future climate conditions for each complete season individually, i.e. for a dataset extending from January 1961 to December 2000 the indices are calculated for all seasons starting from spring 1961 and ending in autumn 2000. These STARDEX indices represent modest or moderate extremes rather than very extreme events to avoid sample sizes that are too small.

| Designation | Description |
|---|---|
| txq90 | Tmax 90$^{th}$ percentile (°C) |
| tnq10 | Tmin 10$^{th}$ percentile (°C) |
| tnfd | Number of frost days Tmin < 0°C |
| txhw90 | Heat wave duration (days) |
| pq90 | 90$^{th}$ percentile of rain day amounts (mm/day) |
| px5d | Greatest 5-day total rainfall |
| pint | Simple daily intensity (rain per rain day) |
| pxcdd | Max no. of consecutive dry days |
| pfl90 | % of total rainfall from events > long-term 90$^{th}$ percentile |
| pnl90 | No. of events > long-term 90$^{th}$ percentile of rain days |

***Table 1: STARDEX diagnostic extreme indices analysed in the study. Table taken from Schmith and Goodess (2005), their Table 2. Tmax and Tmin are the daily maximum and minimum 2 m temperatures. Rain days are days with equal to or more than 1 mm/day precipitation. Dry days are days with less than 1 mm/day precipitation. The heat wave duration is determined from the number of days with a mean temperature of more than the 90$^{th}$ percentile of the season. The long-term 90$^{th}$ percentile is calculated from the time period 1961-1990. All indices are calculated for each season.***

Results of two different methods of statistical downscaling are compared to the direct model output of the RCM simulations. For this development phase, the RCM simulations driven by ERA-40 rather than GCM data are usually used to avoid too strong deviations of the simulations from the observed weather. Six ERA-40 driven RCM simulations are used in this study for the development phase: C4I, DMI, KNMI, METNO, MPI and SMHI. In the case of SMHI, the regressions developed for the ERA-40 driven simulation are applied to two GCM-driven simulations (BCM and ECHAM5) while in the other cases the regressions are applied to one GCM-driven simulation. RCM simulations driven by ERA-40 data are constrained at their lateral boundaries to the observed weather while this is not the case for the RCM simulations driven by GCM data. One of the statistical downscaling methods, developing multiple linear regressions using large-scale predictors from reanalysis data and near-surface local predictands from observation data and applying the found regressions to the output of GCM simulations, is a commonly used method. It is varied in this study because of the fact that we are downscaling RCM simulations. Instead of using rather coarse-resolution reanalysis data, multiple regressions are built from a selection of predictors taken from the ERA-40 driven RCM simulations:

seasonal means of daily values of

- atmospheric temperature at 300, 500, 850 hPa
- geopotential height at 300, 500, 850 hPa
- vertical velocity at 300, 500, 850 hPa
- u-component of the wind at 300, 500, 850 hPa
- v-component of the wind at 300, 500, 850 hPa
- relative humidity at 300, 500, 850 hPa
- minimum 2 m temperature
- maximum 2 m temperature

- precipitation

seasonal standard deviations of daily values of

- minimum 2 m temperature
- maximum 2 m temperature
- precipitation
- vertical velocity at 850 hPa

In addition to these commonly used large-scale predictors, the values of near-surface parameters are used as predictors, which is possible with the high-resolution RCM data. However, even though the RCM simulations are constrained at the lateral boundaries by ERA-40 data, the RCM still has some freedom to develop its own day-to-day weather. Therefore, we cannot expect daily values of precipitation and temperature to correlate very well with daily values of observed data. For this reason seasonal means and standard deviations are used as predictors to directly predict the 10 major STARDEX indices (Table 1).

The second method is a statistical correction method which is applied to four meteorological parameters: daily values of minimum 2 m temperature, maximum 2 m temperature, mean 2 m temperature and rainfall. From both the observations and the ERA-40 driven RCM simulations, the values are sorted in ascending order and linear regressions are built using the simulated parameter as predictor and the observed parameter as predictand. Due to the sorting of the values in ascending order, daily, as opposed to seasonal, values can be used for the development of the regression equations. The timing of events does not play any role; the focus is on the distribution of values, which is corrected towards the observed distribution according to this method. The regressions are calculated for the ERA-40 driven RCM simulation for 1961-2000 in order to apply them to the GCM driven RCM simulations for 1961-2000. The 10 STARDEX indices are calculated from the statistically corrected meteorological parameters.

Because the timing of specific events does not play any role in this statistical correction method, it is even possible to build regressions using sorted values from the GCM-driven rather than the ERA-40 driven RCM simulations as predictors, and sorted observation values as predictands. With this method the bias introduced by the combination of a GCM and an RCM can be corrected as opposed to the correction of the bias introduced by the RCM only.

There is one important limiting factor in the application of both of these statistical approaches to future simulations. It is assumed that the derived relationships do not change under future climate conditions. This stationary assumption may not be valid under future climate conditions. To investigate the robustness of the relationships in a changing climate, the regressions are first only developed for 1961-1980 and applied to 1981-2000. Root mean square errors (RMSE) and correlations (COR) of the two 20-year periods are compared to each other.

Four representative Irish observation stations covering the whole 40-year period of 1961-2000 have been selected for this study. Table 2 gives details of these stations.

| Station (WMO ID) | Name | Latitude | Longitude | Elevation (m) |
|---|---|---|---|---|
| 03976 | Belmullet | 54°14'N | 10°00'W | 11 |
| 03962 | Dublin | 53°26'N | 06°15'W | 71 |
| 03960 | Kilkenny | 52°40'N | 07°16'W | 66 |
| 03957 | Valentia | 51°56'N | 10°15'W | 11 |

***Table 2: Meteorological observation stations***

# 3 Evaluation of the methods

## *3.1 Multiple linear regression*

In the development phase, combinations of the predictors listed in the previous section have been tested for each of the predictands in Table 2. For most predictands the most significant regressions are found through combining the data for all four seasons. However, for the number of frost days (tnfd) and the heat wave duration (txhw90) it is advantageous to separate the data by season. Frost days do not occur in summer and are therefore investigated for winter, spring and autumn, while the heat wave duration is calculated for summer.

Table 3 shows the best set of predictors for each of the predictands.

| Predictand | Predictors |
|---|---|
| txq90 | Seasonal mean and std. dev. of daily maximum 2 m temperature |
| tnq10 | Seasonal mean and std. dev. of daily minimum 2 m temperature |
| tnfd | Seasonal mean and std. dev. of daily minimum 2 m temperature |
| txhw90 | Seasonal mean and std. dev. of daily maximum and minimum 2 m temperature and daily precipitation |
| pq90 | Seasonal mean and std. dev. of daily precipitation and vertical velocity at 850 hPa |
| px5d | Seasonal mean and std. dev. of daily precipitation and vertical velocity at 850 hPa |
| pint | Seasonal mean and std. dev. of daily precipitation and vertical velocity at 850 hPa |
| pxcdd | Seasonal mean and std. dev. of daily precipitation and vertical velocity at 850 hPa |
| pfl90 | Seasonal mean and std. dev. of daily precipitation and vertical velocity at 850 hPa |
| pnl90 | Seasonal mean and std. dev. of daily precipitation and vertical velocity at 850 hPa |

***Table 3: Best predictor sets for each of the STARDEX diagnostic extreme indices***

Comparing the statistically downscaled output with the direct model output, the found regressions show a clear improvement of the root mean square error (RMSE) for all predictands and for all stations. Improvements in the RMSE are more pronounced for the temperature indices compared to the precipitation indices. The correlation improves in most cases, and the improvement is most pronounced in cases in which the correlation is low for the direct model output. As an example, the RMSEs and correlations for Dublin for the ERA-40 driven C4I-RCM simulation are given in Table 4.

| | txq90 | tnq90 | tnfd DJF | tnfd MAM | tnfd SON | txhw90 JJA | pq90 | px5d | pint | pxcdd | pfl90 | pnl90 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RMSE model | 2.41 | 2.98 | 16.0 | 7.33 | 4.51 | 2.36 | 4.19 | 18.1 | 1.32 | 7.74 | 0.165 | 3.97 |
| RMSE stat. | 0.647 | 0.744 | 4.34 | 3.65 | 2.12 | 1.78 | 3.38 | 15.5 | 1.01 | 4.68 | 0.140 | 1.78 |
| COR model | 0.984 | 0.972 | 0.700 | 0.200 | 0.0 | 0.793 | 0.212 | 0.270 | 0.284 | 0.445 | 0.158 | 0.152 |
| COR stat. | 0.984 | 0.975 | 0.870 | 0.590 | 0.554 | 0.722 | 0.359 | 0.438 | 0.453 | 0.513 | 0.297 | 0.258 |

***Table 4: Root mean square errors (RMSE) and correlations (COR) for the 10 STARDEX indices for Dublin for direct model output and statistically downscaled output based on the 159 seasons from spring 1961 to autumn 2000 for the ERA-40 driven C4I-RCM simulation. For tnfd and txhw90 the multiple regressions have been developed for each of the four seasons separately; therefore RMSE and COR are based on 39 seasons for winter and 40 seasons for spring, summer and autumn.***

To test if these statistical relationships are robust, regressions have been derived for 1961-1980 (development time period) and applied to 1981-2000 (application time period). They have not been derived for tnfd and txhw90 for the 20-year periods as the regressions for these indices are derived for each of the four seasons separately leading to a sample size that is too small. From Table 5, which shows average scores over all simulations and all stations, it can be seen that for the application time period, the quality of the results is only slightly decreasing compared to the development time period. In some cases the COR is smaller, in some other cases it is larger in the application period compared to the development period, while the RMSE is always somewhat larger, typically by about 10 to 20% and on average over the 8 parameters by 16%. But in the application time period the RMSE is still always smaller and the correlation similar in the statistically downscaled output compared to the direct model output.

| | txq90 | tnq10 | pq90 | px5d | pint | pxcdd | pfl90 | pnl90 |
|---|---|---|---|---|---|---|---|---|
| RMSE model 1961-1980 | 1.98 | 3.04 | 3.87 | 18.9 | 1.33 | 5.19 | 0.147 | 2.69 |
| RMSE model 1981-2000 | 2.28 | 2.79 | 3.74 | 19.6 | 1.34 | 5.60 | 0.141 | 2.99 |
| RMSE stat. 1961-1980 | 0.851 | 0.664 | 2.99 | 14.9 | 1.00 | 4.32 | 0.124 | 1.86 |
| RMSE stat. 1981-2000 | 1.04 | 0.972 | 3.22 | 16.2 | 1.10 | 4.74 | 0.131 | 2.21 |
| COR model 1961-1980 | 0.967 | 0.976 | 0.438 | 0.402 | 0.503 | 0.595 | 0.281 | 0.417 |
| COR model 1981-2000 | 0.965 | 0.966 | 0.438 | 0.440 | 0.520 | 0.538 | 0.353 | 0.496 |
| COR stat. 1961-1980 | 0.967 | 0.980 | 0.483 | 0.485 | 0.536 | 0.532 | 0.324 | 0.512 |
| COR stat. 1981-2000 | 0.960 | 0.967 | 0.443 | 0.491 | 0.527 | 0.481 | 0.326 | 0.485 |

***Table 5: Root mean square errors (RMSE) and correlations (COR) for 8 STARDEX indices averaged over the six ERA-40 driven RCM simulations and over the four stations for direct model output and statistically downscaled output. The multiple regressions have been developed using data for 1961-1980 and applied to the data for 1981-2000.***

While the maximum and minimum values of the temperature indices over the 40-year period are usually captured better in the statistically downscaled output than in the direct model output, this is not the case for the precipitation indices. The statistically downscaled values of the precipitation indices tend to show a smaller variability compared to the direct model output and the observations. This certainly is a disadvantage as the most extreme values of the indices are of special interest. As an example, Figure 1 shows scatter plots for txq90 and px5d in Kilkenny for the ERA-40 driven DMI-RCM simulation.

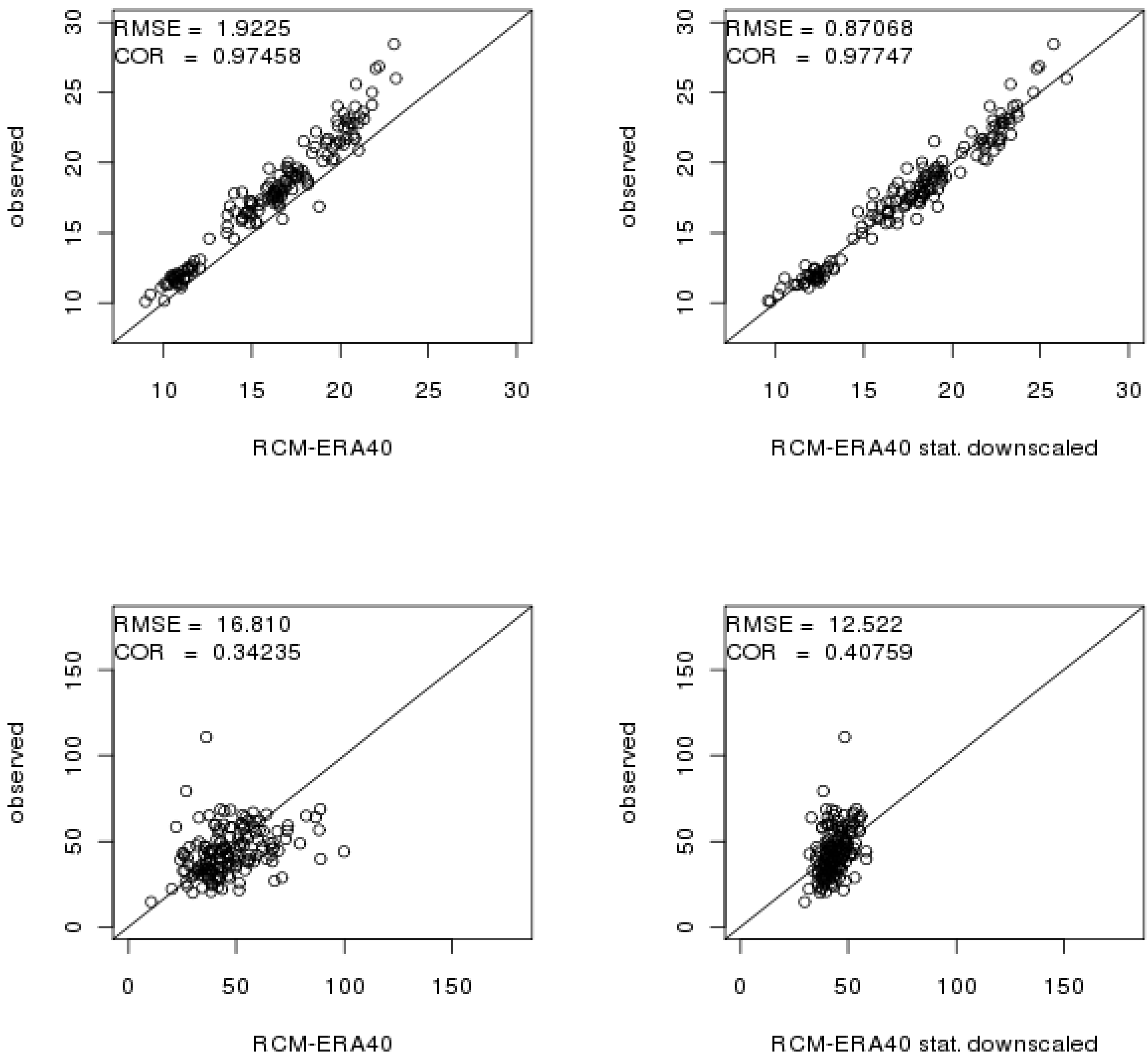


***Figure 1: Scatter plots of the 90% percentile of the daily maximum temperature [°C] per season (txq90, upper row) and of the maximum 5-day total rainfall [mm] per season (px5d, lower row) for Kilkenny from spring 1961 to autumn 2000 between station observations and the ERA-40 driven DMI-RCM simulation (left column) and between station observations and statistically downscaled ERA-40 driven DMI-RCM output (right column). The root mean square error (RMSE) and the correlation (COR) are given in the upper left corner of each plot.***

## *3.2 Statistical correction method*

The statistical correction method differs from the multiple linear regression method as it is applied to daily values of maximum, minimum and mean 2 m temperature and precipitation. Therefore, it is worth investigating the fits of these parameters before examining the derived STARDEX indices.

As an example, Figure 2 shows observed, model simulated and statistically corrected precipitation for 1961-1980 and 1981-2000 from the ERA-40 driven SMHI-RCM simulation for Valentia. The regression has been developed for 1961-1980 and applied to 1981-2000 to investigate its robustness.

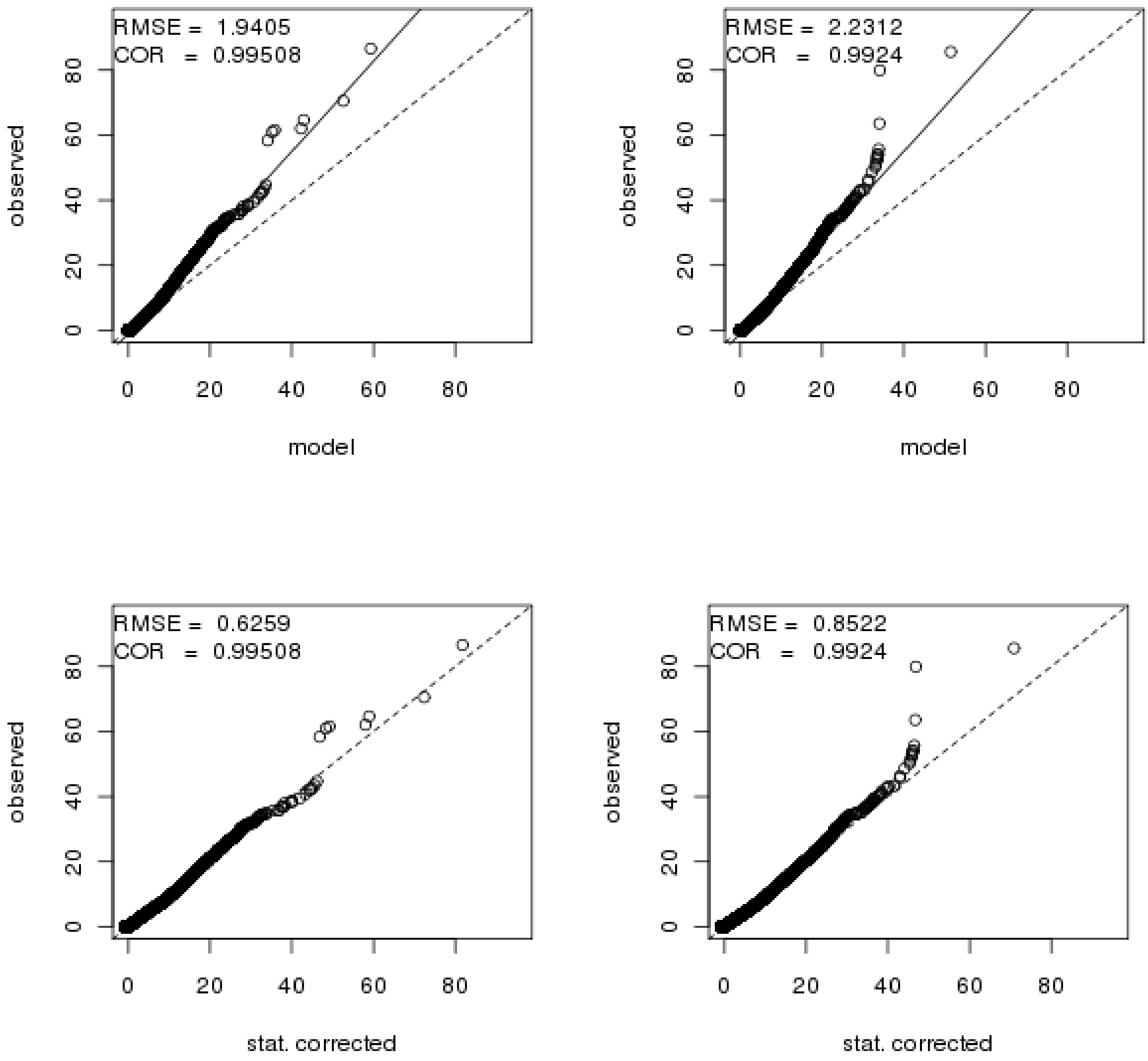


***Figure 2: Sorted daily precipitation values [mm/day] from station observations against ERA-40 driven SMHI-RCM simulation data (upper row) and from station observations against statistically corrected ERA-40 driven SMHI-RCM simulation data (lower row) for Valentia for 1961-1980 (left column) and 1981-2000 (right column). The solid line in the upper row plots indicates the found regression equation for 1961-1980. The root mean square error (RMSE) and the correlation (COR) are given in the upper left corner of each plot.***

The statistical correction method reduces the RMSE and leaves the COR unchanged compared to the direct model output. However, the RMSE increases from the development time period 1961-1980 to the application time period 1981-2000. While the values up to 40 mm/day are also well represented in the application time period, the higher values show some stronger deviations. Nevertheless, the statistical correction method still gives better results than the direct model output for the application time period. It is reassuring that the regression equations are rather similar if we use only data for the period 1961-1980 or the whole time period 1961-2000 to develop the regression – as an example we give the regression equations for the ERA40-driven SMHI simulation for Valentia:

$p_{statcor}$=-0.8597 + 1.3939*$p_{model}$ (1961-1980)

$p_{statcor}$=-0.9250 + 1.4051*$p_{model}$ (1961-2000)

Generally the regressions for 2 m temperature (daily maximum, minimum and mean temperature) are as robust as for precipitation.

After applying the statistical correction to the simulated daily values of maximum, minimum and mean temperature as well as precipitation, the STARDEX indices can be calculated from these bias corrected values. Comparing RMSEs and correlations of the statistical correction method with the multiple regression statistical downscaling method shows that the multiple regression statistical downscaling usually gives better performances than the statistical correction method. However, when sorting the values of the extreme indices in ascending order from both the statistically downscaled output and from the statistically corrected output and comparing these to the sorted extreme indices from observations, it becomes clear that the statistical correction method in most cases performs better than the multiple regression statistical downscaling method (Table 6). Furthermore it has the advantage of capturing the extreme events better than both the direct model output and the statistically downscaled output (Figure 3). Sorting of the extreme indices can be justified since we are not interested in the timing of the extreme events but in their distribution and intensity.

| | txq90 | tnq10 | pq90 | px5d | pint | pxcdd | pfl90 | pnl90 |
|---|---|---|---|---|---|---|---|---|
| RMSE model | 2.00 | 2.81 | 1.80 | 7.86 | 0.692 | 3.09 | 0.029 | 1.33 |
| RMSE stat. | 0.479 | 0.281 | 1.81 | 9.23 | 0.595 | 2.87 | 0.086 | 1.13 |
| RMSE s. c. | 0.573 | 0.697 | 0.935 | 4.39 | 0.310 | 1.53 | 0.024 | 0.738 |
| COR model | 0.990 | 0.996 | 0.992 | 0.980 | 0.992 | 0.986 | 0.994 | 0.977 |
| COR stat. | 0.989 | 0.997 | 0.991 | 0.969 | 0.908 | 0.924 | 0.983 | 0.980 |
| COR s. c. | 0.990 | 0.996 | 0.992 | 0.980 | 0.990 | 0.988 | 0.994 | 0.976 |

***Table 6:Root mean square errors (RMSE) and correlations (COR) for 8 STARDEX indices averaged over the six ERA-40 driven RCM simulations and the four stations for sorted direct model output (model), sorted statistically downscaled output (stat.) and sorted statistically corrected output (s. c.) based on the 159 seasons from spring 1961 to autumn 2000.***

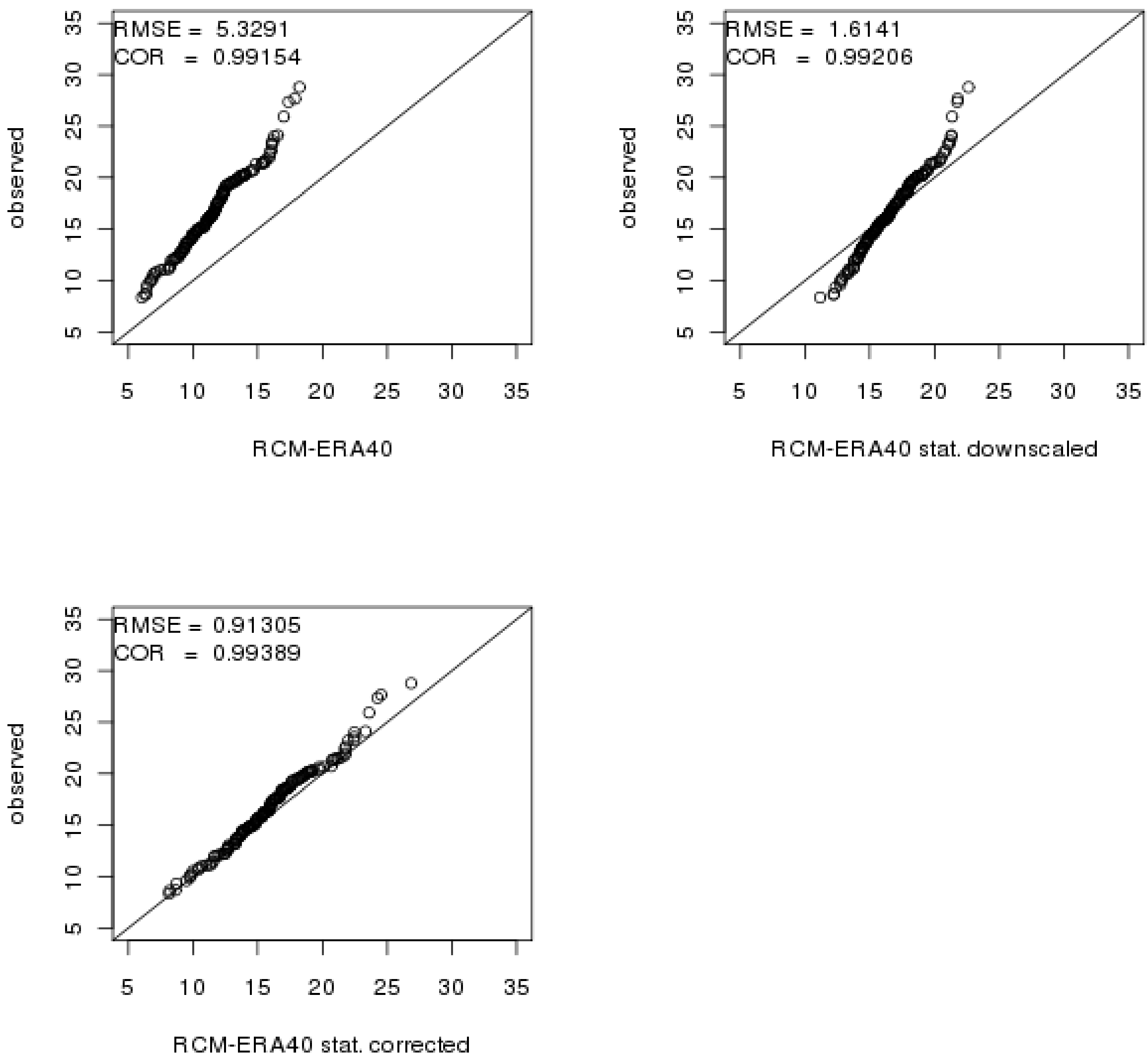


***Figure 3: Sorted values of the seasonal 90%-percentile of precipitation (pq90) for Valentia for (a) direct model output, (b) statistically downscaled output and (c) statistically corrected output against station observations for the ERA-40 driven SMHI-RCM simulation.***

While the multiple regressions for the standard statistical downscaling have to be derived for ERA-40 driven RCM simulations and then applied to the corresponding GCM driven RCM simulations, the proposed statistical correction method can also be applied to the GCM driven RCM simulations to correct the bias of the combination of GCM and RCM rather than simply the RCM. This will be referred to as the direct statistical correction method.

To compare the performance of the three methods for the GCM driven RCM simulations, the regressions of the statistical downscaling and the statistical correction method, which were developed for the ERA-40 driven RCM simulations, can be applied to the GCM driven RCM simulations and compared to the results of the direct statistical correction method (Table 7). It can be seen that the direct statistical correction method usually gives the best results, which is not surprising as this is the only method which considers biases in the GCM as well.

| | txq90 | tnq10 | pq90 | px5d | pint | pxcdd | pfl90 | pnl90 |
|---|---|---|---|---|---|---|---|---|
| RMSE model | 3.05 | 2.70 | 2.16 | 12.2 | 0.856 | 5.00 | 0.042 | 1.79 |
| RMSE stat. | 1.55 | 1.31 | 2.33 | 11.5 | 0.789 | 3.51 | 0.092 | 1.55 |
| RMSE s. c. | 1.69 | 1.60 | 1.65 | 9.56 | 0.851 | 3.57 | 0.036 | 1.08 |
| RMSE dir. s. c. | 0.798 | 0.922 | 1.21 | 6.27 | 0.337 | 2.52 | 0.036 | 0.757 |
| COR model | 0.982 | 0.989 | 0.990 | 0.978 | 0.990 | 0.978 | 0.992 | 0.975 |
| COR stat. | 0.981 | 0.990 | 0.989 | 0.969 | 0.991 | 0.935 | 0.985 | 0.981 |
| COR s. c. | 0.982 | 0.989 | 0.990 | 0.978 | 0.991 | 0.981 | 0.990 | 0.976 |
| COR dir. s. c. | 0.982 | 0.989 | 0.991 | 0.978 | 0.991 | 0.983 | 0.976 | 0.976 |

***Table 7: Root mean square errors (RMSE) and correlations (COR) for 8 STARDEX indices averaged over the seven GCM driven RCM simulations and the four stations for sorted direct model output, sorted statistically downscaled output, sorted statistically corrected output and sorted directly statistically corrected output based on the 159 seasons from spring 1961 to autumn 2000.***

However, Table 8 shows that the RMSEs are slightly less stable than for the standard multiple regression statistical downscaling technique when dividing the 40-year period into a development period of 1961-1980 and an application period of 1981-2000. While the COR shows only minor changes, the RMSE generally increases by 0 to 45% from the development period to the application period. Averaged over the 8 parameters, the increase in RMSE amounts to 21%. Nevertheless, in the application period the RMSEs are in all cases still lower for the direct statistical correction method than for the direct model output, while the CORs are comparable.

| | txq90 | tnq10 | pq90 | px5d | pint | pxcdd | pfl90 | pnl90 |
|---|---|---|---|---|---|---|---|---|
| RMSE model 1961-1980 | 2.85 | 2.86 | 2.41 | 12.6 | 0.944 | 5.06 | 0.045 | 2.03 |
| RMSE model 1981-2000 | 3.28 | 2.58 | 2.07 | 12.6 | 0.849 | 5.09 | 0.054 | 1.68 |
| RMSE dir. s. c. 1961-1980 | 0.830 | 0.953 | 1.52 | 6.70 | 0.348 | 2.55 | 0.039 | 0.840 |
| RMSE dir. s. c. 1981-2000 | 0.978 | 0.968 | 1.80 | 9.67 | 0.438 | 2.62 | 0.052 | 1.06 |
| COR model 1961-1980 | 0.978 | 0.988 | 0.984 | 0.972 | 0.984 | 0.969 | 0.979 | 0.964 |
| COR model 1981-2000 | 0.980 | 0.985 | 0.983 | 0.965 | 0.982 | 0.972 | 0.985 | 0.968 |
| COR dir. s. c. 1961-1980 | 0.978 | 0.988 | 0.985 | 0.972 | 0.988 | 0.972 | 0.973 | 0.965 |
| COR dir. s. c. 1981-2000 | 0.980 | 0.985 | 0.981 | 0.965 | 0.983 | 0.980 | 0.969 | 0.967 |

***Table 8: Root mean square errors (RMSE) and correlations (COR) for 8 STARDEX indices averaged over the seven GCM driven RCM simulations and the four stations for sorted direct model output and sorted directly statistically corrected output. The multiple regressions have been developed using data for 1961-1980 and applied to the data for 1981-2000.***

# 4 Summary and conclusions

Three different methods of statistical downscaling have been applied and evaluated using regional climate model datasets. Previous studies have usually focused on statistical downscaling of output of global climate models. In this study the synergistic use of dynamical and statistical downscaling has been investigated.

The first method is a standard method of developing multiple linear regressions between seasonal means and standard deviations of regional climate model output and 10 STARDEX extreme indices calculated from observations. The second and third methods are statistical correction methods in which the distributions of precipitation and 2 m temperature from either ERA-40 driven or global climate model driven regional climate model output are corrected towards the observed distributions. The 10 STARDEX extreme indices are afterwards calculated from the statistically corrected model output.

Based on the three investigated statistical downscaling approaches applied to four investigated stations, the statistical correction method which corrects the distributions of the global climate model driven regional climate model output towards the observed distributions gives the best results when comparing root mean square errors and correlations of the seasonal STARDEX extreme indices. Both statistical correction methods have the advantage over the standard statistical downscaling method of being able to better capture the highest and lowest values of the STARDEX extreme indices. However, dividing the 40 years of data into a development period of 1961-1980 and an application period of 1981-2000, the regressions for the standard statistical downscaling method are slightly more robust than for the other two methods, i.e. root mean square errors and correlations degrade slightly less from the development to the application period.

While the standard multiple linear regression statistical downscaling method is well-established, the statistical correction method should be further tested. While all of the methods lead to improvements in the root mean square errors and correlations for the STARDEX extreme indices compared to the direct model output, the most extreme events of the STARDEX extreme precipitation indices are only captured better in the statistical correction methods compared to the direct model output, while the standard multiple regression method does not capture these extreme events. Because of the slightly stronger degradation of root mean square errors from the development period to the application period, it might be more risky to apply these methods to investigate future climate change.

## References

Christensen, J. H., B. Hewitson, A. Busuioc, A. Chen, X. Gao, I. Held, R. Jones, R. K. Kolli, W.-T. Kwon, R. Laprise, V. Magana Rueda, L. Mearns, C. G. Menendez, J. Räisänen, A. Rinke, A. Sarr and P. Whetton, 2007: Regional climate projections. In: *Climate Change 2007. The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Frias M. D., E. Zorita, J. Fernandez and C. Rodriguez-Puebla, 2006: Testing statistical downscaling methods in simulated climates. *Geophysical Research Letters,* **33,** L19807. doi:10.1029/2006GL027453.

Schmith, T. and C. Goodess (eds.), 2005: STARDEX: Statistical and regional dynamically downscaling of extremes for European regions. Deliverable D12: Downscaled extremes based on NCEP reanalysis data (1958-2000). *Available online* at http://www.cru.uea.ac.uk/projects/stardex/deliverables/D12/D12_synthesis_v21.pdf

Timbal, B., P. Hope and S. Charles, 2008: Evaluating the consistency between statistically downscaled and global dynamical model climate change predictions. *Journal of Climate*, **21,** 6052-6059. doi:10.1175/2008JCLI2379.1.

Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton and L. O. Mearns, 2004: Guidelines for use of climate scenarios developed from statistical downscaling methods. *Available online* at http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf