STARDEX

STAtistical and Regional dynamical Downscaling of EXtremes for European regions

EVK2-CT-2001-00115

Deliverable D12

Downscaled extremes based on NCEP Reanalysis data (1958-2000)

FOREWORD

The STARDEX project on STAtistical and Regional Dynamical downscaling of Extremes for European regions is a research project supported by the European Commission under the Fifth Framework Programme and contributing to the implementation of the Key Action "global change, climate and biodiversity" within the Environment, Energy and Sustainable Development.

STARDEX will provide a rigorous and systematic inter-comparison and evaluation of statistical and dynamical downscaling methods for the construction of scenarios of extremes. The more robust techniques will be identified and used to produce future scenarios of extremes for European case-study regions for the end of the 21st century. These will help to address the vital question as to whether extremes will occur more frequently in the future.

For more information about STARDEX, contact the project co-ordinator Clare Goodess (c.goodess@uea.ac.uk) or visit the STARDEX web site:
http://www.cru.uea.ac.uk/projects/stardex/

STARDEX is part of a co-operative cluster of projects exploring future changes in extreme events in response to global warming. The other members of the cluster are MICE and PRUDENCE. This research is highly relevant to current climate related problems in Europe. More information about this cluster of projects is available through the MPS Portal:
http://www.cru.uea.ac.uk/projects/mps/

STARDEX is organised into five workpackages including Workpackage 4 on 'Inter-comparison of improved downscaling methods with emphasis on extremes' which was responsible for the production of this deliverable (D12). Workpackage 4 is co-ordinated by Torben Schmith from the Danish Meteorological Institute, Denmark.

STARDEX PROJECT MEMBERS

UEA	University of East Anglia, UK
KCL	King's College London, UK
FIC	Fundación para la Investigación del Clima, Spain
UNIBE	University of Berne, Switzerland
CNRS	Centre National de la Recherche Scientifique, France
ARPA-SMR	Servizio Meteorologico Regionale, ARPA-SMR Emilia-Romagna, Italy
ADGB	University of Bologna, Italy
DMI	Danish Meteorological Institute, Denmark
ETH	Swiss Federal Institute of Technology, Switzerland
FTS	Fachhochschule Stuttgart – Hochschule für Technik, Germany
USTUTT-IWS	Institut für Wasserbau, Germany
AUTH	University of Thessaloniki, Greece

CO-ORDINATORS NOTE

This D12 synthesis report on 'Downscaled extremes based on NCEP Reanalysis data' incorporates preliminary results from partners. However, as this report clearly shows, the large scatter in performance across stations makes it very difficult to draw conclusions from results presented in this box diagram format. Thus the STARDEX partners agreed to undertake more detailed analyses of the results for each region. Specific partners were given responsibility for co-ordinating each study and guidance was circulated by the scientific steering group. These analyses were completed by autumn 2004 and the resulting regional reports are available from the public web site. Based on these regional analyses, it was possible to draw conclusions about the performance of the methods and make recommendations. The D12 regional analyses are also described and summarized in a paper entitled 'An intercomparison of statistical downscaling methods for Europe and European regions – assessing their performance with respect to extreme temperature and precipitation events', with myself as lead author. This paper was submitted to the PRUDENCE special issue of *Climatic Change* in March 2005. Pre-prints are available on request from **c.goodess@uea.ac.uk**.

Clare Goodess
STARDEX co-ordinator
September 2005

D12 AUTHORS AND VERSION HISTORY

Lead author:
Torben Schmith, DMI

Contributing authors:
Stefano Alberghi, ADGB
Christina Anagnostopoulou, AUTH
András Bárdossy, USTUTT
Carlo Cacciamani, ARPA-SMR
Gavin Cawley, UEA
Bo Christiansen, DMI
Christoph Frei, ETH
Clare Goodess, UEA
Dimitrios Gyalistras, UNIBE
Malcolm Haylock, UEA
Colin Harpman, KCL
Yeshewatesfa Hundecha, USTUTT
Panagiotis Maheras, AUTH
Valentina Pavan, ARPA-SMR
Guy Plaut, CNRS
Jürg Schmidli, ETH
Eric Simonnet, CNRS
Konstantia Tolika, AUTH
Rodica Tomozeiu, ARPA-SMR
Ennio Tosi, ADGB
Rob Wilby, KCL

Version 1.x:	15. July 2004
Version 2.0:	1 November 2004
Version 2.1:	30 September 2005

CONTENTS

1. Introduction

This deliverable deals with the application of the different statistical methods for downscaling extremes to the NCEP/NCAR reanalysis data for the period 1958-2000. During this period we have observational data and therefore this deliverable describes a verification of the different downscaling methods, which will enable an inter-comparison of these.

For the sake of making a meaningful inter-comparison, it was decided to adopt standard procedures to the widest possible extent. Therefore, a standard set of regions and a standard set of stations (precipitation and temperature) within the regions was agreed on for the inter-comparison. Predictors varied among methods, but it was agreed that all predictors should be based on the NCEP/NCAR reanalysis dataset. Finally, a standard set of extreme indices for precipitation and temperature was defined for the verification procedure and also for use in other deliverables.

2. Methodology

Downscaling of precipitation and temperature extremes was carried out for selected sub-regions of Europe by STARDEX partners according to table 1 below

Table 1: Regions where each partners methods are applied in the inter-comparison

	Iberia	Greece	Alps	Germany	UK	Italy
UEA	x	x	x	x	x	x
KCL	x				x	
FIC	x	x	x	x	x	x
UNIBE			x			
CNRS	x		x			
ARPA-SMR		x				x
DMI	x	x	x	x	x	x
ETH			x		x	
USTUTT/FTS			x	x		
AUTH		x				x

The NCEP reanalyses are regarded as more reliable from 1958 onwards because of the radiosondes then being more effectively assimilated. On the other hand, many predictor series will not be updated for the latest years. Therefore, data from the period 1958-2000 are considered in the comparison.

In STARDEX deliverable D10, recommendations for circulation parameters suitable as predictors were given and these will be used throughout this deliverable. Predictors will be taken from NCEP reanalysis in 2.5x2.5 deg resolution as downloadable from the STARDEX website.

Predictand time series will be taken from the 'FIC dataset'. This dataset has been homogeneity tested at the monthly level. Preferably homogeneous series and series without any data gaps or with only a few data missing should be used. For some regions, a higher density predictand dataset will be additionally used.

As the study was focused mainly on the changes in extremes, a number of extreme temperature and precipitation indices were defined. Many of the indices are based on thresholds defined on the basis of statistical quantities such as the 90th or the 10th percentiles. The base period for the calculation of such quantities was set between 1961 and 1990. This makes the indices applicable to a wide variety of climates as no arbitrary threshold values are used. The only exception is a fixed threshold value of 0°C used to define frost days; which is, of course, applicable to all climates. The indices used in the study are listed in Table 2.

The seasonal and annual time series of each index were computed for all the stations used in the study using the STARDEX extreme indices software (available from the STARDEX web site: http://www.cru.uea.ac.uk/projects/stardex/). The magnitudes and directions of their trends were computed. The Kendall-tau test was used to test the statistical significance of trends. All trends were considered significant at the 95% level.

Table 2: STARDEX Diagnostic Extreme Indices analysed in the study

Designation	Description
	a) Precipitation related indices
pav	Precipitation average (mm/day)
pq90	90th percentile of rainday amounts (mm/day)
px5d	Greatest 5-day total rainfall
pint	Simple Daily Intensity (rain per rainday)
pxcdd	Max no. of consecutive dry days
pfl90	% of total rainfall from events > long-term 90th percentile
pnl90	No. of events > long-term 90th percentile of raindays
	b) Temperature related indices
txav	Average Tmax (deg. C)
tnav	Average Tmin (deg. C)
txq90	Tmax 90th percentile (deg. C)
tnq10	Tmin 10th percentile (deg. C)
tnfd	Number of frost days Tmin < 0 °C
txhw90	Heat wave Duration (days)

3. Description of downscaling methods

UEA

Canonical Correlation analysis (CCA)

A downscaling method is being developed to be applied to the entire European regional dataset which models the STARDEX indices of rainfall extremes using large scale patterns of circulation. Rather than modelling the daily rainfall itself, then calculating the indices, this method will use seasonal measures of large-scale circulation variability to model the seasonal indices of extremes directly. The predictands will be the six STARDEX core indices of rainfall extremes calculated seasonally. Predictors will be selected from circulation variables, calculated over the entire European and East Atlantic region. From our work preparing D10, the best predictor seems to be sea level pressure but other variables will be considered including temperature, geopotential height and relative humidity at three atmospheric levels as well as sea surface temperature. The model uses canonical correlation analysis. Two methods to select predictors will be tested: one using just the variable with the best correlation with the indices (MSLP); and a cross-validation of all possible predictor combinations.

Artificial Neural Networks

Statistical methods for downscaling daily precipitation using Bayesian multilayer perceptron artificial neural networks are being developed at UEA. Parameters in the model are dependent on circulation predictors (zonal/meridional airflows, vorticity, and divergence at multiple pressure heights) and moisture predictors (relative humidity and specific humidity at multiple pressure heights). In addition to the usual sum-of-squares error metric (**ANN-SSE**), we investigated other data misfit terms corresponding to more realistic assumptions regarding the actual noise process. We began by evaluating the hybrid Bernoulli/Gamma error metric (**ANN-GAMMA**). The distribution of the amount of precipitation is modelled by the (upper) incomplete Gamma function. The model is then trained to approximate the conditional probability of rainfall and the scale and shape parameters of the Gamma distribution modelling the predictive distribution of the amount of precipitation.

KCL

Artificial Neural Networks (ANN)

The first downscaling method uses two artificial neural network models (Radial Basis Function (**RBF**) and Multi Layer Perceptron (**MLP**)) to model daily rainfall for two UK regions and two Iberian regions. All the sites for each region are modelled simultaneously, however individual RBF (**IRBF**) models are further applied to three designated sites per region, providing a comparison with same sites extracted from the multi-site results. The predictors for the modelling processes are determined by stepwise linear regression from the available 2.5 x 2.5 NCEP reanalysis grid boxes. These consist of circulation predictors (zonal/meridional airflows, vorticity, and divergence at multiple pressure heights) and moisture predictors (relative humidity and specific humidity at multiple pressure heights). For the multi-site predictor selection the predictand used is the area average and for individual sites the daily precipitation series at that station is used. The poor performance of neural networks in the literature for the downscaling of precipitation data is likely to be attributed to the large number of dry days (i.e. zeros), which heavily bias the training process. The proposed method attempts to circumvent this problem by adopting a two-stage approach analogous to weather generation techniques, where precipitation is downscaled using separate occurrence and amount processes.

Conditional Resampling (CR)

The second downscaling method uses a conditional resampling technique which first uses a single site linear regression model (SDSM) to generate a series of daily precipitation amounts at a "marker" site, in this instance the area average. Wet–day amounts are resampled from the empirical distribution of area averages conditional on the large–scale atmospheric forcing and a stochastic error term. The actual wet-day amount is determined by mapping the modelled normal cumulative distribution value onto the observed cumulative distribution at the marker site. The marker site amount is then cross-referenced to the actual amount at each station. The predictors for this method are determined in the same way as the previous models.

FIC

The two step analogue method (2SA)

This method estimates daily precipitation amount and daily maximum and minimum temperature for a site for day "X" in two steps:

- The "n" most similar (analogous) days (according to the geostrophic fluxes at 1000 and 500 hPa) to the day "X" are selected from a reference dataset
- In a second step, precipitation and temperatures are obtained applying further analyses that search (only in the "n" days population) for relationships between the predictands and some more predictors. For precipitation, averaging the 100 most similar days (analogous) precipitation. For temperature, a multiple linear regression (only in the "n" days population) with forward and backward stepwise selection of predictors, using as potential predictors low troposphere thickness, the averaged temperatures of the previous days, and a sinusoid function of the day of the year.

CNRS

Potential Precipitation Circulation Index (PPCI)

At a first stage, a ``Potential Precipitation Circulation Index" (ppci) is looked for through a linear regression of daily precipitation against the anomaly pattern correlation (apc) of an day's Large Scale Circulation (LSC) pattern (its Z700 height field on a large euro-atlantic sector) with the Intense Precipitation Events (IPE) LSC clusters of [Plaut et al., 2001] which we now call ``Precipitation Regimes" (PR), and with the ``Weather Regimes" (WR) patterns. The acronym "PR" is a natural extension of "WR", the commonly used one when all daily LSCs are classified. Since the ppci is the result of a linear regression, there exists no deterministic link between its value on a given day and the corresponding precipitation; however, for the same reason, there exists a statistical link, and, the higher its value, the higher the probability of occurrence of an IPE (see Figure 4 of the STARDEX deliverable D10_CNRS report).

The main steps are then the followings: we first classify the (learning period) ppci values into 20 categories. To generate precipitation series from LSCs, we start from the LSC of each day and compute its corresponding ppci. We then randomly choose an analog within the set of (learning period) days having the same ppci category. We assign to the involved day the precipitation (or the dry character) of its (learning period) analog. Seasonal accumulations as well as STARDEX extreme indices may be easily computed from the generated precipitation series. Several precipitation series may be generated, given the stochastic character of the algorithm.

Two characteristics of the model are its robustness (there are less than 10 regressors for thousands of observations) and its only dependence on large scale circulation (may be an advantage or a drawback according to applications).

UNIBE

Canonical correlation analysis (CCA)

The statistical downscaling procedure was based on Canonical Correlation Analysis (CCA) in the space spanned by the first few Empirical Orthogonal Functions (EOFs) of the predictor and predictand fields on a seasonal basis. In the present study, varying numbers of predictor fields were used, according to the method proposed by Gyalistras *et al*. (1994). The predictors were given by large-scale fields for sea-level pressure (slp) and geopotential height, temperature, specific humidity and relati1ve humidity at the 1000, 850, 700, 500 and 300 hPa levels.

ARPA-SMR

Linear regression

The statistical downscaling method is based on multiple linear regression (**MLR**). The presence of significant links between the occurrence of temperature/precipitation extremes and particular large-scale Euro-Atlantic circulation patterns designs our predictors. In particular, a significant correlation has been found between North Atlantic Oscillation /European Blocking and extreme indices. Then a multiple linear regression has been developed using as predictors the first 4 PCs of 500mb geopotential height anomalies that explained together more than 70% of the total variance (M**LR-Z500**) or 850mb temperature (**MLR-T850**). The work has been done for all indices and seasons for the Italian stations selected to be representative.

Canonical correlation analysis

Optimum statistical downscaling models for the extreme events at 8 representative stations from Emilia Romagna region were investigated. The statistical model used is based on the Canonical Correlation Analysis (**CCA**). The CCA finds the optimum linear combination of two multidimensional vectors (predictand and predictor). Before the CCA the data sets are projected on EOFs (empirical orthogonal functions) and those that explain 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.

The skill of the statistical model is dependent on the station, predictors and the numbers of EOFs predictors retained in the CCA analysis. Mean sea level pressure (SLP), specific humidity (shmed) at 850mb and temperature at 850mb (T850) were considered as predictors, while the predictands are the core indices of extreme precipitation and temperature. All the predictors time series were extracted from the NCEP reanalysis, with resolution 2.5°x2.5°.

The skill of the downscaling model shows that SLP is a good predictor especially for winter extreme precipitation while the combination SLP+shmed or SLP+T850 is a good predictor for extreme precipitation from the other seasons. The T850mb is a good predictor for spring, summer and autumn extreme temperature while the combination SLP+T850mb is a good predictor for winter temperature.

DMI

Conditional weather generator (CWG)

We apply a weather generator approach based on daily values of an index describing the large scale circulation in which the probabilistic characteristics of the precipitation are quantified. The relation between the circulation and the precipitation characteristics is calculated for each station and for each season.

For calibrating the model, a surface pressure pattern is obtained as the average pressure difference between rainy days and dry days measured at a given station. The circulation index is then calculated by regressing the daily surface pressure field on this pattern. We then divide the circulation index into a number of quantiles, usually between 5 and 10, and for each quantile the following precipitation characteristics are calculated: the probability for wet/dry days, the probabilities for a wet/dry day following a dry/wet day, and the two parameters describing the gamma-distribution that best approximates the probability density of the rain amount (only wet days).

When applying the model, the daily circulation index is calculated by regressing the daily surface pressure field on the pattern found above. Using the dependence of the probabilities on the circulation index a two-state Markov process is used to obtain the sequence of dry/wet days. Then, for each wet day the rain amount is drawn from a gamma-distribution with the parameters corresponding to the circulation index of that day.

ETH

Local rescaling

As our main predictor we use GCM simulated precipitation. Two statistical downscaling methods are investigated: (i) local rescaling of GCM simulated precipitation (LOC), and (ii) local rescaling with a dynamical correction (DLOC) (Widmann and Bretherton, 2003). The intended application for these two methods is to downscale from the GCM scale to the regional scale in order to produce mesoscale gridded precipitation fields. However, for the purpose of this intercomparison, the methods are applied directly to station data. The first method uses GCM simulated precipitation as its only predictor. The second method requires in addition to the GCM simulated precipitation a proxy of the large-scale flow, for instance the first three principal components of the geopotential height field. While the first method is calibrated on a monthly basis, the second method is calibrated seasonally.

USTUTT/FTS:

Regression model from circulation patterns (MREG)

In this method, the seasonal extreme indices are directly downscaled from seasonal circulation predictors. A range of potential predictor variables are selected from the NCEP reanalysis data, including humidity, temperature, and geopotential heights at different pressure levels and mean sea level pressure. The mean sea level pressure is classified into objective circulation patterns using a fuzzy rule-based classification approach. Other derived potential predictors, such as moisture flux, wind direction, vorticity and divergence will also be considered. Once the potential predictor variables are selected, the extreme indices will be modelled using a regression model. The ‘best’ predictors will be further selected using a screening regression. Two different models will be tested. One is a regression model without taking the objective sea level pressure circulation patterns in to consideration and the other one will be a model conditioned on circulation patterns.

MV autoregressive (MVA)
A conditional multivariate autoregressive model is implemented to generate daily timeseries of precipitation at multiple locations simultaneously. Because of the high probability of occurrence of dry days and a continuous distribution of the rainfall amount on wet days, random variables with mixed distribution are used to describe the distribution of rainfall at a given location. A power transformed and truncated normal distribution is used to model the distribution of precipitation at a given location and time. Truncation is made to account for dry days and power transformation is applied as the distribution of precipitation amounts is usually skewed. The expected value of precipitation at a given location and time is modelled as a variable conditioned to the circulation type and moisture flux as well as the time of the year corresponding to that day. That means the annual cycle of the expected values is taken in to account. The annual cycle of the lag-1 day autocorrelation function is also taken into account. But it is not conditioned to the circulation pattern type. In addition, the model takes the spatial variability of precipitation into account by incorporating the covariance structure of the precipitation at the different locations, which is dependent on the circulation pattern type. The model parameters related to the annual cycle of the mean precipitation are estimated using the maximum likelihood and those related to the spatial covariance function are estimated using the least square technique.

AUTH
Multiple Regression Analysis using the circulation types approach (MREG)
According to the results of D10, the classification of the circulation types for the 500hPa has proven to be the most appropriate of the construction of the Multiple Regression Model for the precipitation indices. On the other hand, the classification of the circulation types for the thickness (1000 – 500hPa) has proven to be the most appropriate for the construction of the model, in order to simulate the temperature indices. For every index, season and station, a different model has been constructed, according to the relationship of the circulation types, the precipitation indices and the temperature indices. The selection of the most efficient model has been done by combining the correlation coefficient values of the observed with the simulated as well as the RMSE values.

Artificial Neural Network (ANN)
Artificial Neural Network models have a structure based on human neural system. Acting as black box models permit the estimation of a number of parameters from predictor parameters, the two groups been connected by non-linear relations. These models have three layers (input – hidden – output) or a more complex structure. The main advantage of ANNs is that no initial transfer function is needed to be chosen and the system is self-training. As for the previous models, for the construction of the model for the simulation of the precipitation indices, SLP and 500hPa geopotential values have been used, while for the simulation of the temperature indices the thickness (1000-500hPa) values have been used.

Canonical Correlation Analysis (CCA)
This method isolates linear combinations of multiple predictor variables and linear combinations of multiple predictand variables that have maximum correlation coefficients. This approach assumes that the relationship between predictor and predictand will remain stable in the future climate, an assumption based on physical interpretability of the relationship found and in the ability of the statistical model to reconstruct local climate anomalies from the large scale observations in an independent historical period. For the simulation of the precipitation indices, SLP and 500hPa geopotential values have been used,

while for the simulation of the temperature indices the thickness (1000-500hPa) values have been used.

4. Principles for verification

The models will be verified based on the four standard seasons. They will be calibrated on the period 1958-1978 & 1994-2000 (regarded as one period). Subsequently, the models should for each season produce a time series of annual values of the relevant indices for the period 1979-1993. This should be compared with the index values calculated from observations. Comparisons will be made in terms of root mean square error and linear correlation coefficient.

5. Preliminary results of verification and discussion

Intercomparing about twenty different methods applied to six different regions and four seasons easily becomes overwhelming even when limited to the parameters listed in Table 2. Therefore, limitations have to be applied. We will focus on some well-defined questions, such as:

- Is there any difference in performance of the methods between winter and summer?
- Is there any systematic difference in performance between regions?
- Almost all the methods (the ETH methods are the exceptions) are variants of either 'type 1' where the seasonal index is downscaled using seasonal circulation indices or 'type 2' in which daily precipitation/temperature values are generated and the seasonal indices are calculated from these. We will ask whether there is any systematic difference between the performance of type 1 (direct) and type 2 (indirect) methods.
- Do the locally developed methods perform better than the European-widely applied methods?

5.1 Precipitation

We begin our evaluation in the UK region, where a range of different methods have been applied. The index with the highest performance is, as expected, pav for all methods. Going to pint and pq90 (see Figure 1) the most striking feature is the difference in performance, measured by CORR, of the methods between summer and winter. In summer, values of CORR are centered below 0.3, which is hardly significant, while in winter CORR is centered around 0.5. Within the two seasons we focus on, the variation among stations for a given method seems larger than the variations between methods. Therefore, one can hardly point to one 'contest-winning' method. These conclusions are supported by looking at the similar plots for the other verification indices.

There is a quite big overlap between methods applied to UK and Iberia. The same conclusions hold here: Better performance during winter than during summer. Note, however, that some of the methods are not able to calculate some of the extreme indices during summer (Figure 2). Also note a large variation in performance between stations for a given method, which again makes it difficult to point to the best method. Also no significant difference between type 1 and type 2 methods can be seen.

In the Germany, Alps, Italy and Greece regions it becomes even more difficult to draw any conclusions, there is even no significant difference between summer- and winter performance (Figure 3).

If we now compare the performance of the models in these three regions, it appears that the models compare equally well in UK and Iberia, whereas they compare significantly worse in Greece.

5.2 Temperature

Fewer methods have been applied to temperature. The largest number of different methods has been applied to the Italy and Greece regions. In Figure 4 is, as an example, shown the box plot of the index txq90 for the Greece region. As for precipitation, there seems to a large variation in performance, measured in terms of CORR and RMSE, within each method. Therefore, again, it does not seem reasonable to point to the 'best' method. Similar conclusions hold when looking at the other indices and regions.

6. Concluding remarks

From this analysis it is evident that the large scatter (across the different stations) in performance of each method and region hampers the drawing of definite conclusions. It is impossible to point to the 'best' method for a given region from the box plots used here and similarly it is impossible to tell which of the type 1 or 2 methods are superior. The only thing we can conclude is that the potential for statistical downscaling is larger during winter than during summer in the western UK and Iberia regions.

From this analysis, we cannot tell the reason for the large scatter in the performance of the downscaling methods and, in particular, whether it is real or artificial. One cause for artificial scatter could be the difference in the quality of the individual data series (inhomogeneities). Another one could be the short verification period (15 years) which makes the variance of CORR large.

References

Gyalistras, D., Von Storch, H., Fishlin, A., and Beniston, M., 1994: Linking GCM simulated climatic changes to ecosystem models: case studies of statistical downscaling in the Alps. *Clim. Res.* **4**, 167-189.

Widmann, M. L., and C. S. Bretherton, 2003: Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor. *J. Climate*, **16**, 799-816.

Table 3: Stations used for each region in the intercomparison

Region	Station	Country
Iberia	Albacete/Los Llano	Spain
	Valencia	Spain
	Alicante Ciudad Ja	Spain
	Murcia/Alcantarill	Spain
	Murcia/San Javier	Spain
	Beja	Portugal
	Coimbra	Portugal
	Lisboa Geofisica	Portugal
	Santarem	Portugal
	Pegoes	Portugal
	Alvega	Portugal
	Mora	Portugal
	Penhas Douradas	Portugal
	Portalegre	Portugal
	Badajoz/Talavera	Spain
	Alcuescar	Spain
Greece	Ioannina	Greece
	Agrinio	Greece
	Kalamata	Greece
	Alexandroupoli	Greece
	Mytilini	Greece
	Samos	Greece
	Rodos	Greece
Alps	Innsbruck-Univ.	Austria
	Nice	France
	Montelimar	France
	Muenchen	Germany
	Bologna	Italy
	Lazzaro Alberoni	Italy
	Bobbio	Italy
	Arosa	Switzerland
	Locarno-Monti	Switzerland
	Zuerich	Switzerland
Germany	Feldberg/Schw.	Germany
	Karlsruhe	Germany
	Mannheim	Germany
	Deuselbach	Germany
	Koeln-Wahn	Germany
	Giessen	Germany
	Wuertzburg	Germany
	Saarbruecken-E.	Germany
	Kahler Asten	Germany
	Nuernberg-Kra.	Germany
UK	Cambridge	United Kingdom
	Goudhurst	United Kingdom
	Oxford	United Kingdom
	Eskdalemuir	United Kingdom
	Ringway	United Kingdom
	Shawbury	United Kingdom
Italy	Bobbio	Italy
	Lazzaro Alberoni	Italy
	Bedonia	Italy
	Bologna	Italy
	Alfonsine	Italy

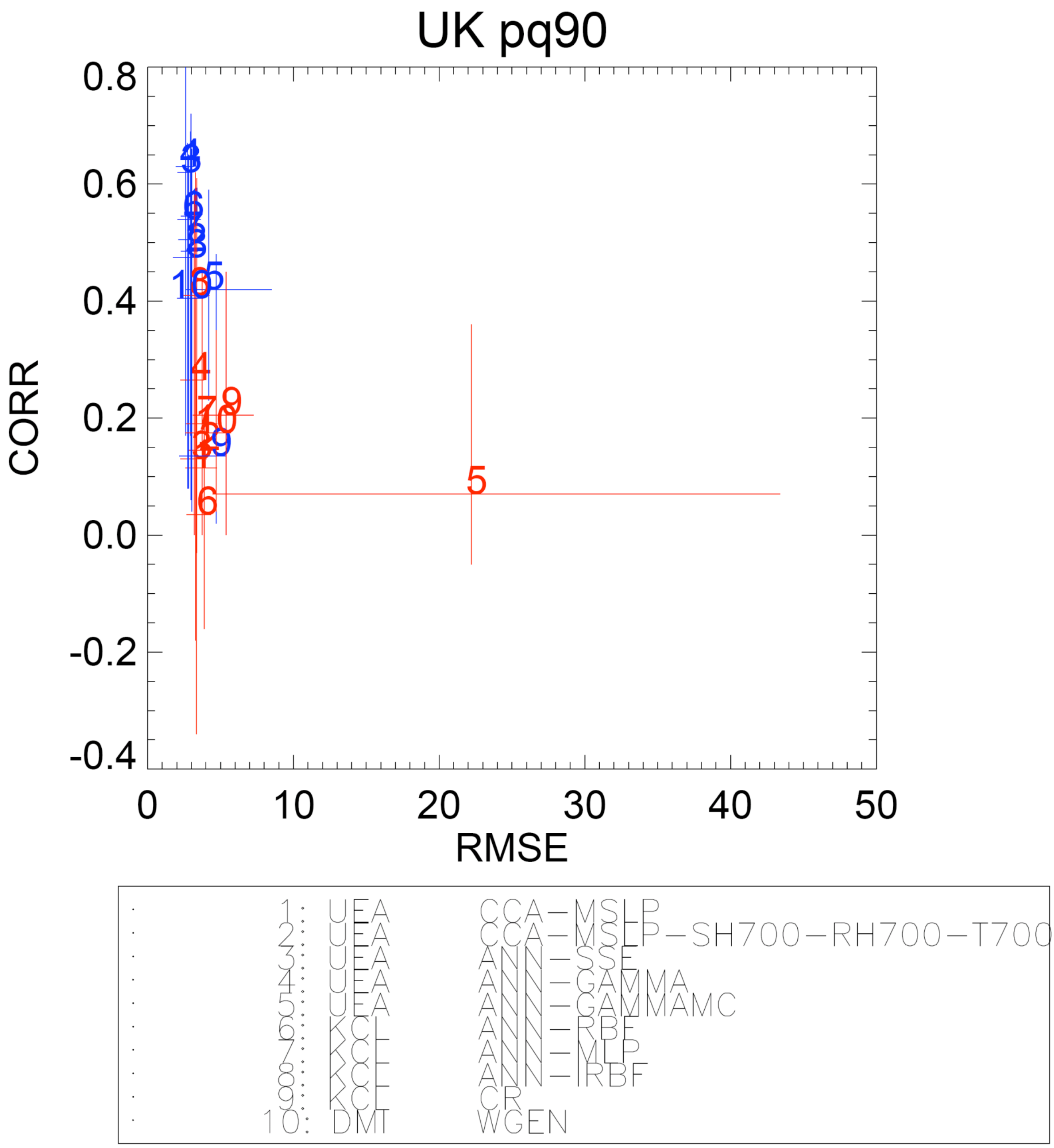

Figure 1: Box diagram showing median values of RMSE and CORR for each method. The horizontal lines are the range of values of RMSE for all stations and the vertical lines are the similar range for CORR. Red is for summer and blue is for winter. This diagram is for the index pq90 and the UK region.

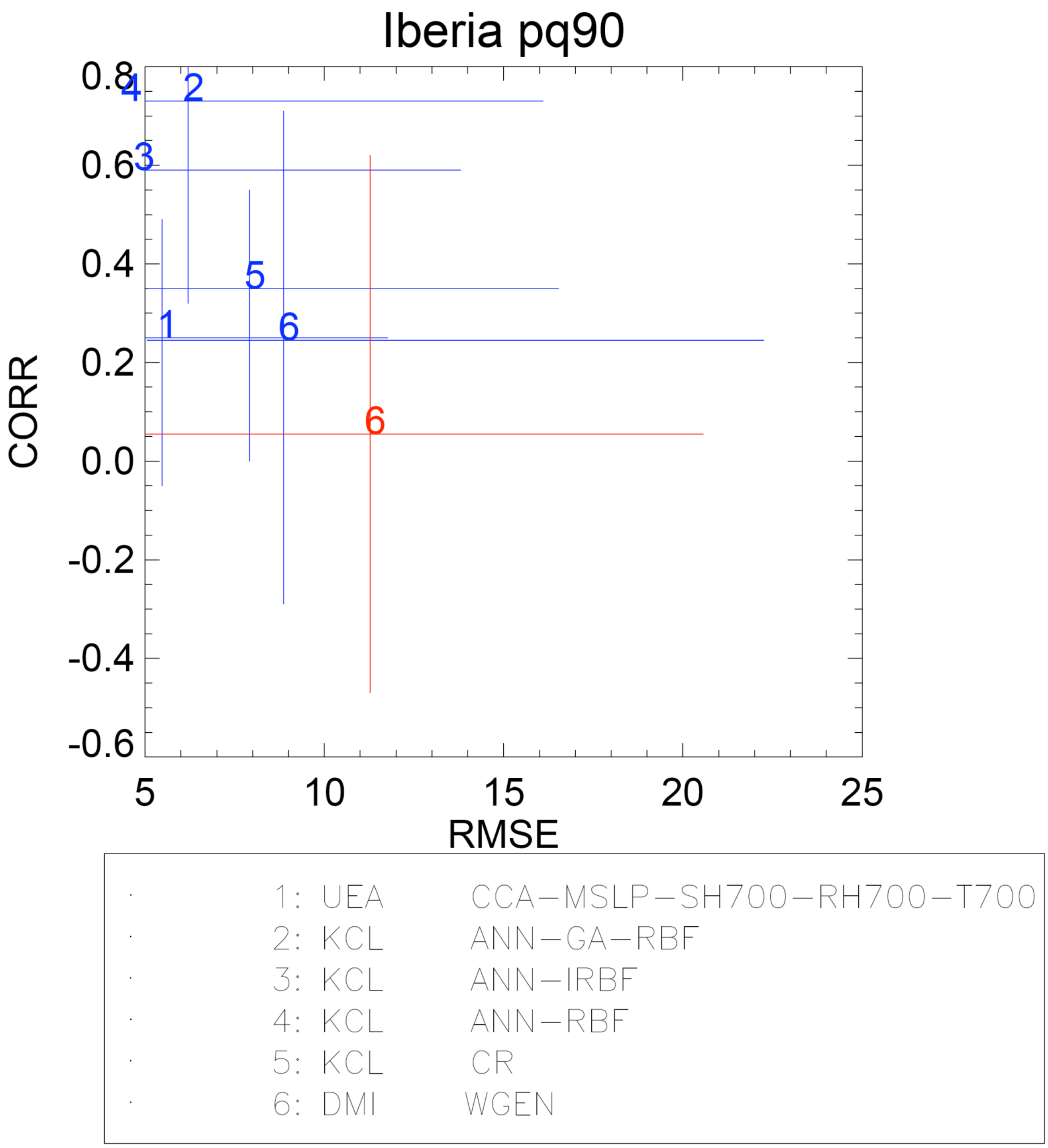

Figure 2: Box diagram showing median values of RMSE and CORR for each method. The horizontal lines are the range of values of RMSE for all stations and the vertical lines are the similar range for CORR. Red is for summer and blue is for winter. This diagram is for the index pq90 and the Iberia region.

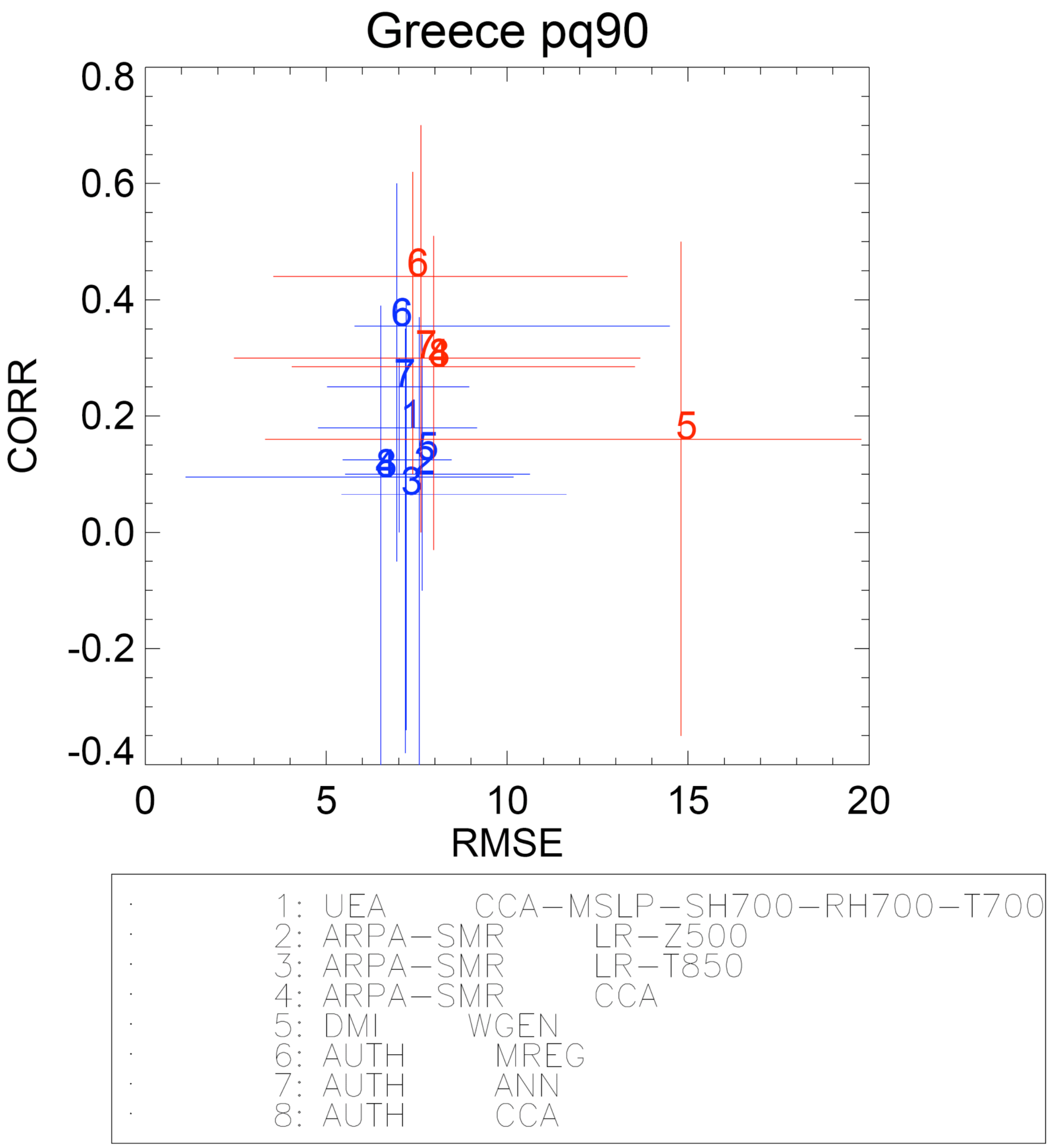

Figure 3: Box diagram showing median values of RMSE and CORR for each method. The horizontal lines are the range of values of RMSE for all stations and the vertical lines are the similar range for CORR. Red is for summer and blue is for winter. This diagram is for the index pq90 and the Greece region.

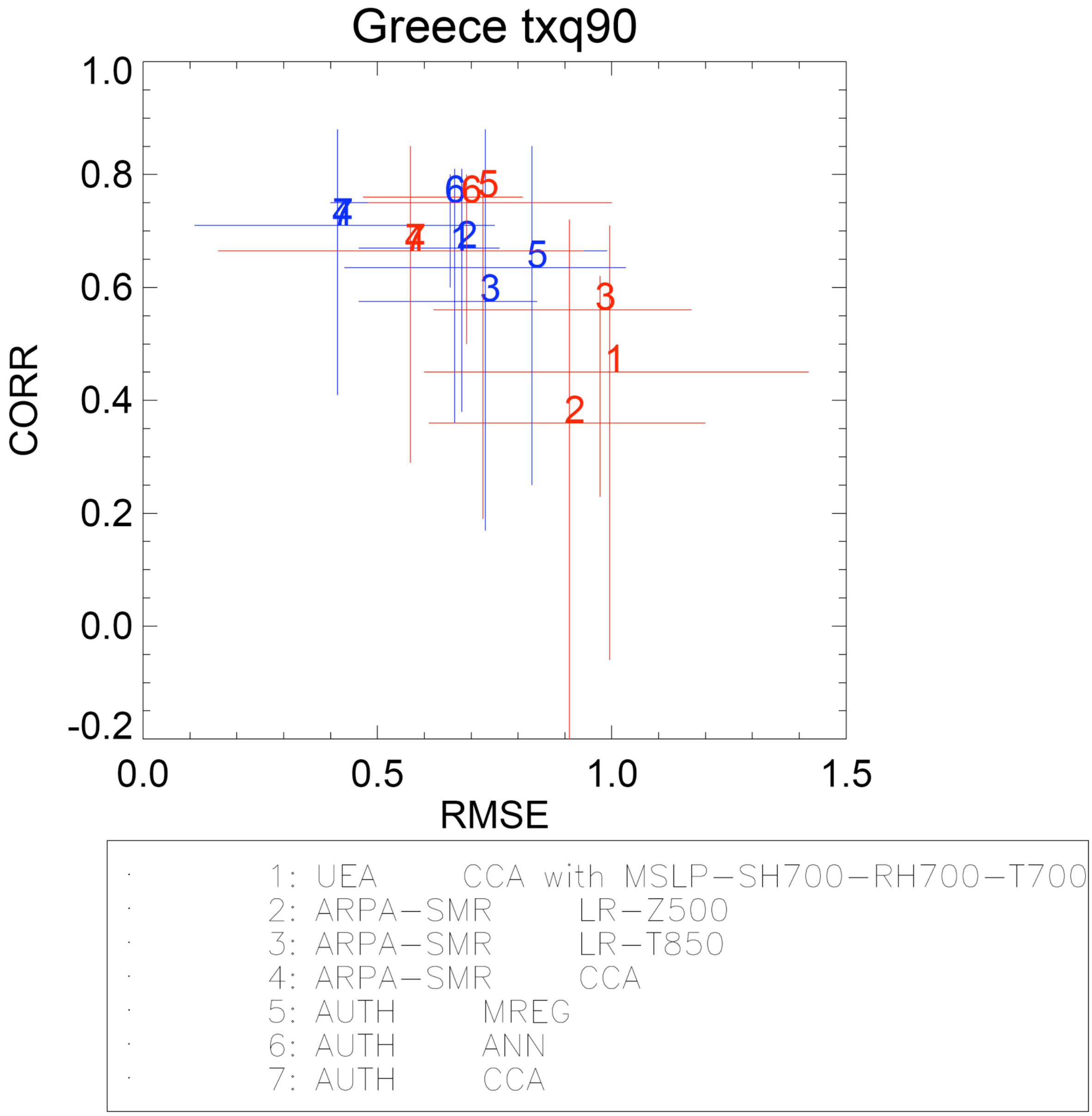

Figure 4: Box diagram showing median values of RMSE and CORR for each method. The horizontal lines are the range of values of RMSE for all stations and the vertical lines are the similar range for CORR. Red is for summer and blue is for winter. This diagram is for the index txq90 and the Greece region.