D12 UK Regional Extreme Rainfall Comparison

Malcolm Haylock, UEA

Introduction

Downscaled indices of rainfall extremes using NCEP reanalyses were available for 14 models from 5 institutions. The aim of this document is to determine if one individual or group of models reproduced the observed extreme indices with greater skill than the other models.

Data and Methodology

Seasonal verification statistics were available for 7 extreme indices:

pav	Mean daily rainfall
pq90	90th percentile of rainday amounts (mm/day)
px5d	Greatest 5-day total rainfall
pint	Simple Daily Intensity (rain per rainday)
pxcdd	Max no. consecutive dry days
pfl90	% of total rainfall from events > long-term P90
pnl90	No. of events > long-term 90th percentile of raindays

Data from 3 stations in SE England and 3 stations in NW England were provided.

The models can be grouped as follows:

Direct methods (downscale seasonal indices)

CCA1 (UEA) – CCA of indices using MSLP
CCA4 (UEA) – CCA indices using best combination of MSLP, T700, RH700 and SH700

Indirect methods

Artificial Neural Networks

RBF (KCL) – multi-site radial basis function
IRBF (KCL) – individual-site radial basis function
MLP (KCL) – multi-site multi-layer perceptron
GAM (UEA) – multi-layer perceptron using hybrid Bernoulli/Gamma error metric
SSE (UEA) – multi-layer perceptron using sum of squares error metric

Rescaling

DYN (ETH) – multi-site rescaling of GCM precipitation with dynamical correction
DYNI (ETH) – individual-site rescaling of GCM precipitation with dynamical correction
LOC (ETH) – multi-site rescaling of GCM precipitation
LOCI (ETH) – individual-site rescaling of GCM precipitation

Others

CR (KCL) – single site linear regression with conditional resampling
CWG (DMI) – conditional weather generator
2SA (FIC) – two step analogue method

Results

Three verification statistics were provided: Spearman (rank) correlation; bias; and RMS error. The total number of possible comparisons for each statistic is 4 seasons x 7 indices x 6 stations = 168. While this is too many comparisons for separate analyses, it is desirable to understand the seasonal dependence of the performance of the models. Therefore results are averaged across the 7 indices and 6 stations.

Correlation

Figure 1 shows the Spearman correlation for each model and season averaged across all indices and stations. The figure shows that the correlation varies widely between models and seasons. Generally DJF has the highest correlations and JJA the lowest.

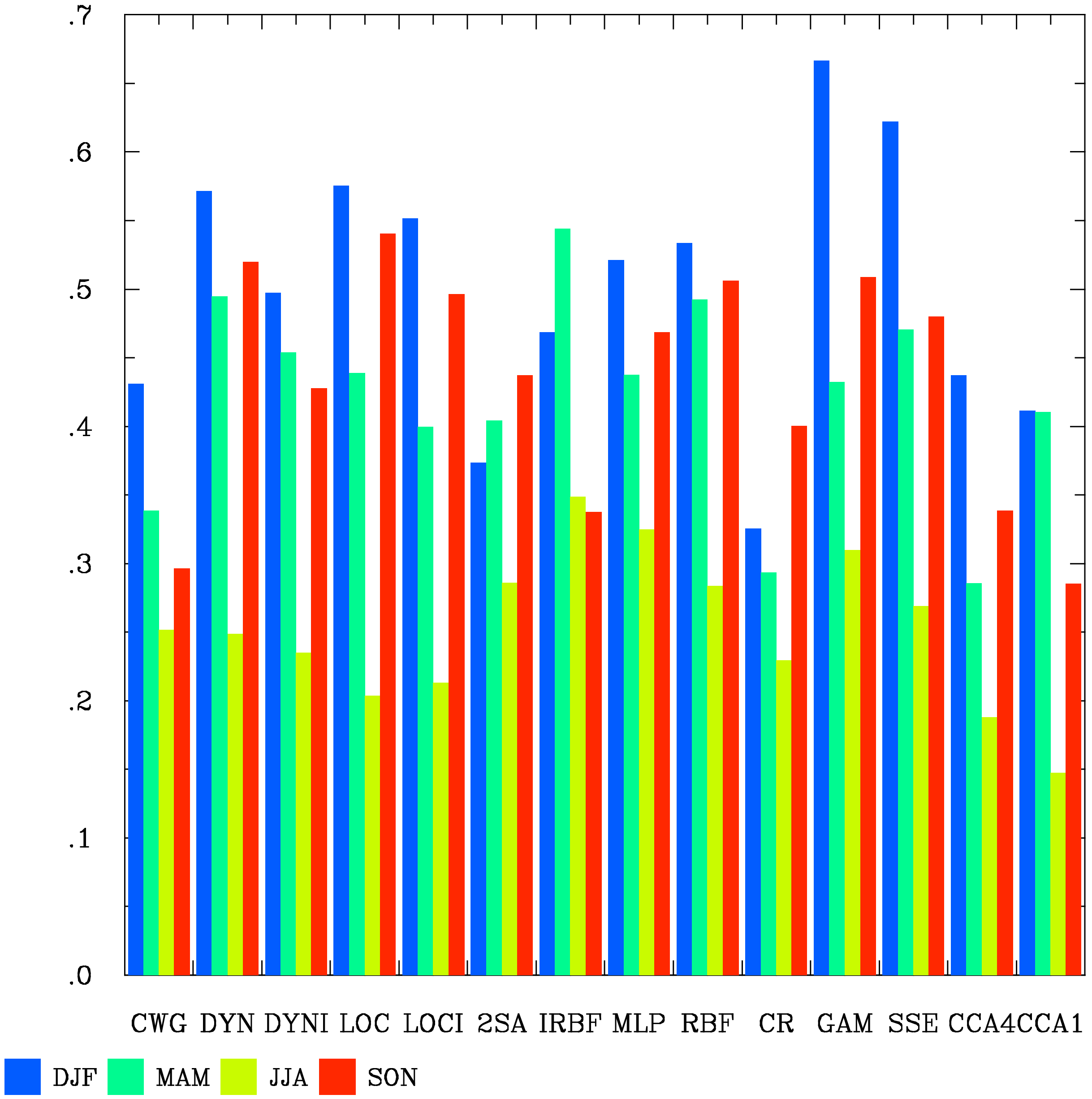

Figure 1: Spearman correlation for each model and season averaged across all indices and stations.

Table 1 shows the correlation performance of the models ranked for each season. Artificial neural network (ANN) and rescaling models generally performed the best with no model consistently outperforming the others.

DJF	MAM	JJA	SON
GAM	IRBF	IRBF	LOC
SSE	DYN	MLP	DYN
LOC	RBF	GAM	GAM
DYN	SSE	2SA	RBF
LOCI	DYNI	RBF	LOCI
RBF	LOC	SSE	SSE
MLP	MLP	CWG	MLP
DYNI	GAM	DYN	2SA
IRBF	CCA1	DYNI	DYNI
CCA4	2SA	CR	CR
CWG	LOCI	LOCI	CCA4
CCA1	CWG	LOC	IRBF
2SA	CR	CCA4	CWG
CR	CCA4	CCA1	CCA1

Table 1: Model performance for each season ranked by correlation. Models with highest correlations are listed first. Names are colour coded according to the type of model: direct methods (blue); ANN models (red); rescaling methods (green); and others (black).

Averaging correlations across seasons gives the results as shown in Fig. 2. The 5th and 95th percentiles of the distribution of correlations are also shown. Note that the spread of correlations within each model is much higher than the differences between the means of the models. The ANN models were the strongest performers followed by the rescaling methods. The linear regression with conditional resampling (CR) and direct methods (CCA1 and CCA4) had the lowest correlations.

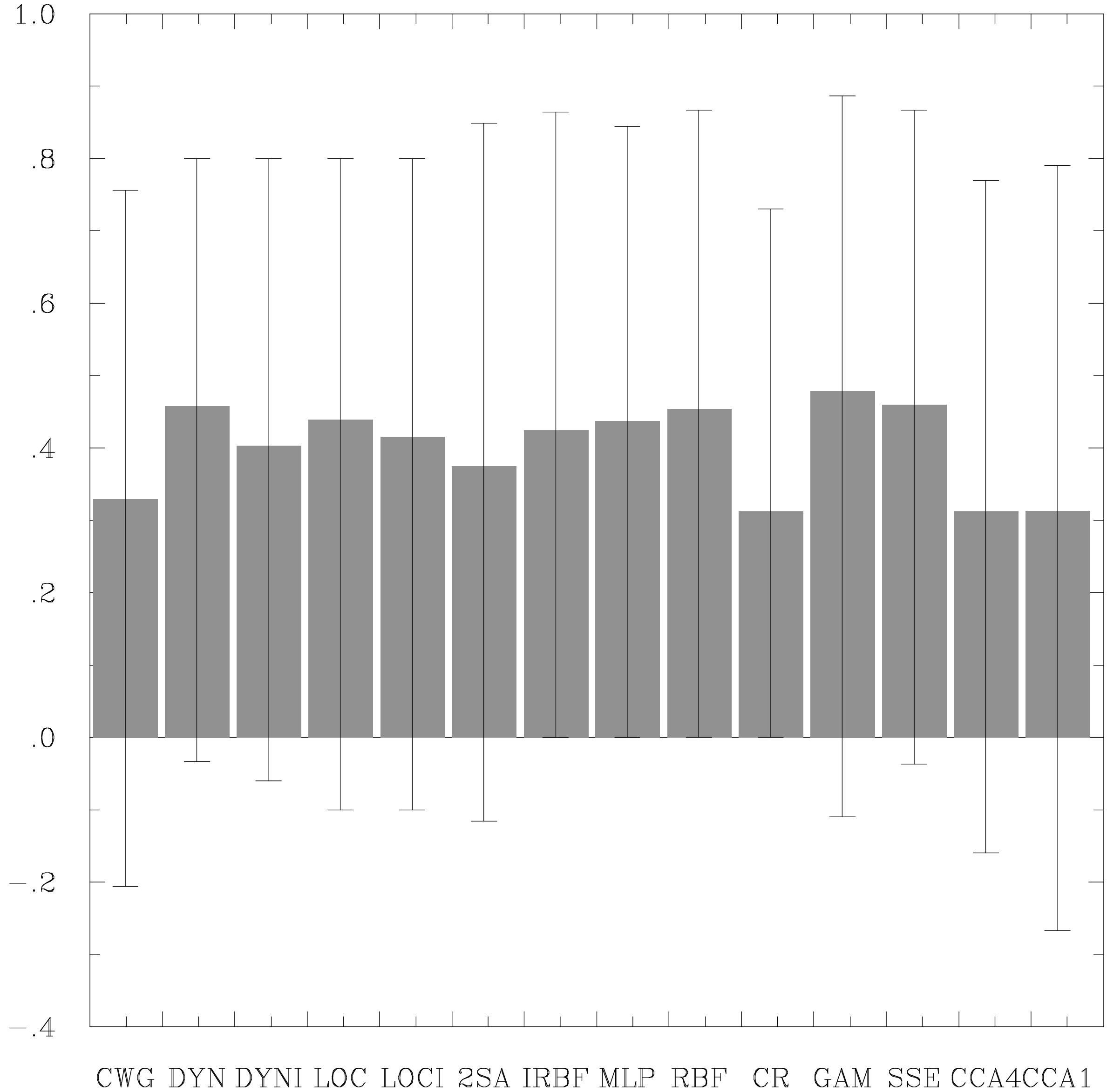

Figure 2: Spearman correlation for each model averaged across all seasons, indices and stations. The 5th and 95th percentiles of the distribution of correlations are also shown.

RMS Error

Since each of the indices has different units, RMS error cannot be averaged across indices. Therefore each RMS error was converted to a rank compared to other models for each index, season and station. The ranks were then averaged across all indices, seasons and stations. Models were given a higher rank for lower RMS errors (better performance).

Figure 3 shows the average rank for each model. There was no single class of models that outperformed the others. The two ANN models GAM and SSE has lower average RMS errors (higher ranks) than the others. All except CR had similar errors with CR having much a much lower average RMS rank score. As for correlation, the spread of correlations within each model (as indicated by the 5th and 95th percentiles) is much higher than the differences between the means of the models.

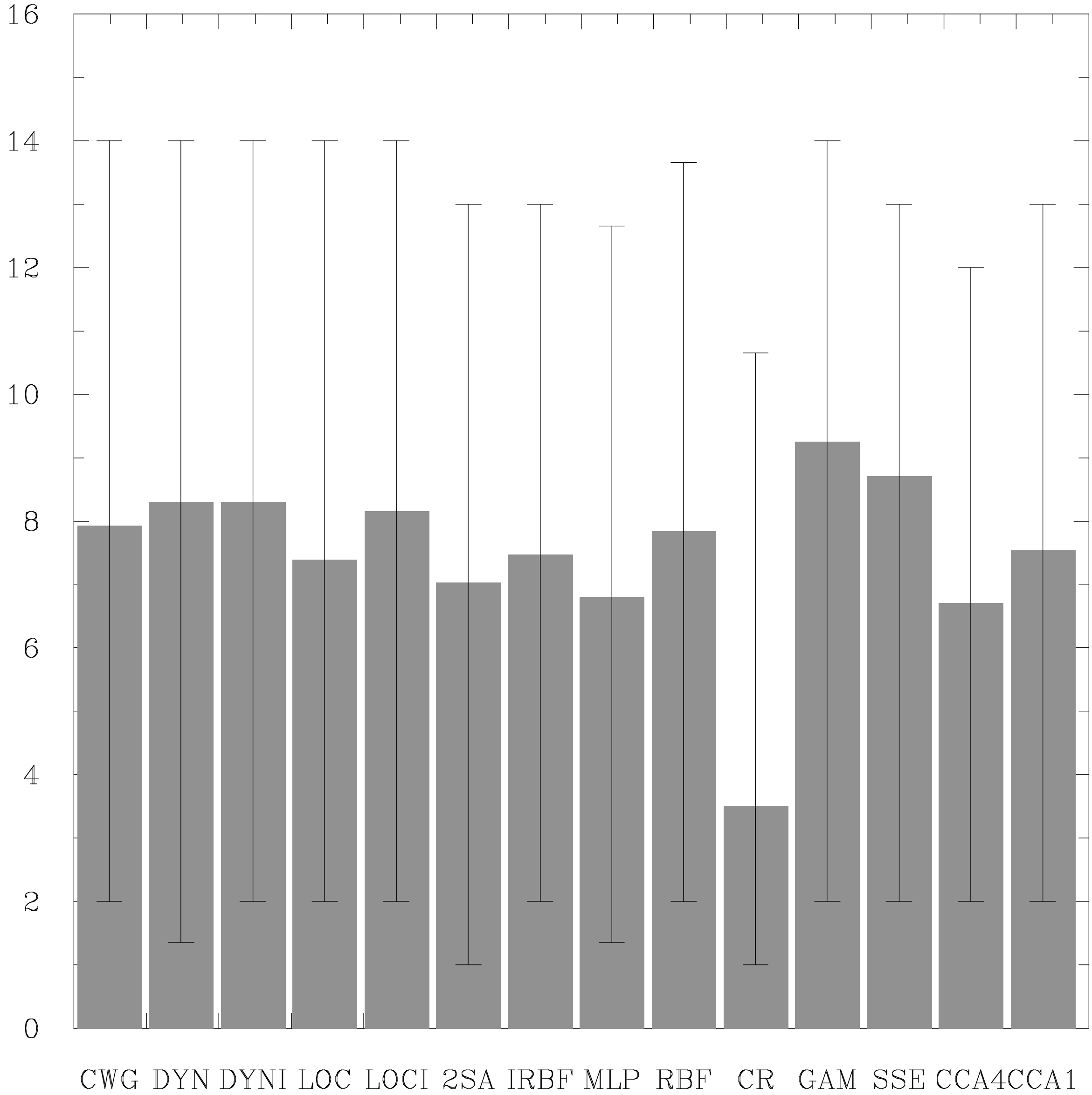

Figure 3: Rank of rms error for each model averaged across all seasons, indices and stations. Lower RMS errors are given a higher rank. The 5th and 95th percentiles of the distribution of correlations are also shown.

Bias

Similarly to RMS error, the bias scores were ranked before averaging, to allow comparison between indices. However, unlike RMS error, bias can be of either sign so the absolute value of the bias was used.

Figure 4 shows the rank of the bias averaged across all seasons, indices and stations. The rescaling methods, the direct methods and the conditional weather generator generally performed the best. The ANN methods and analogue method had the highest biases.

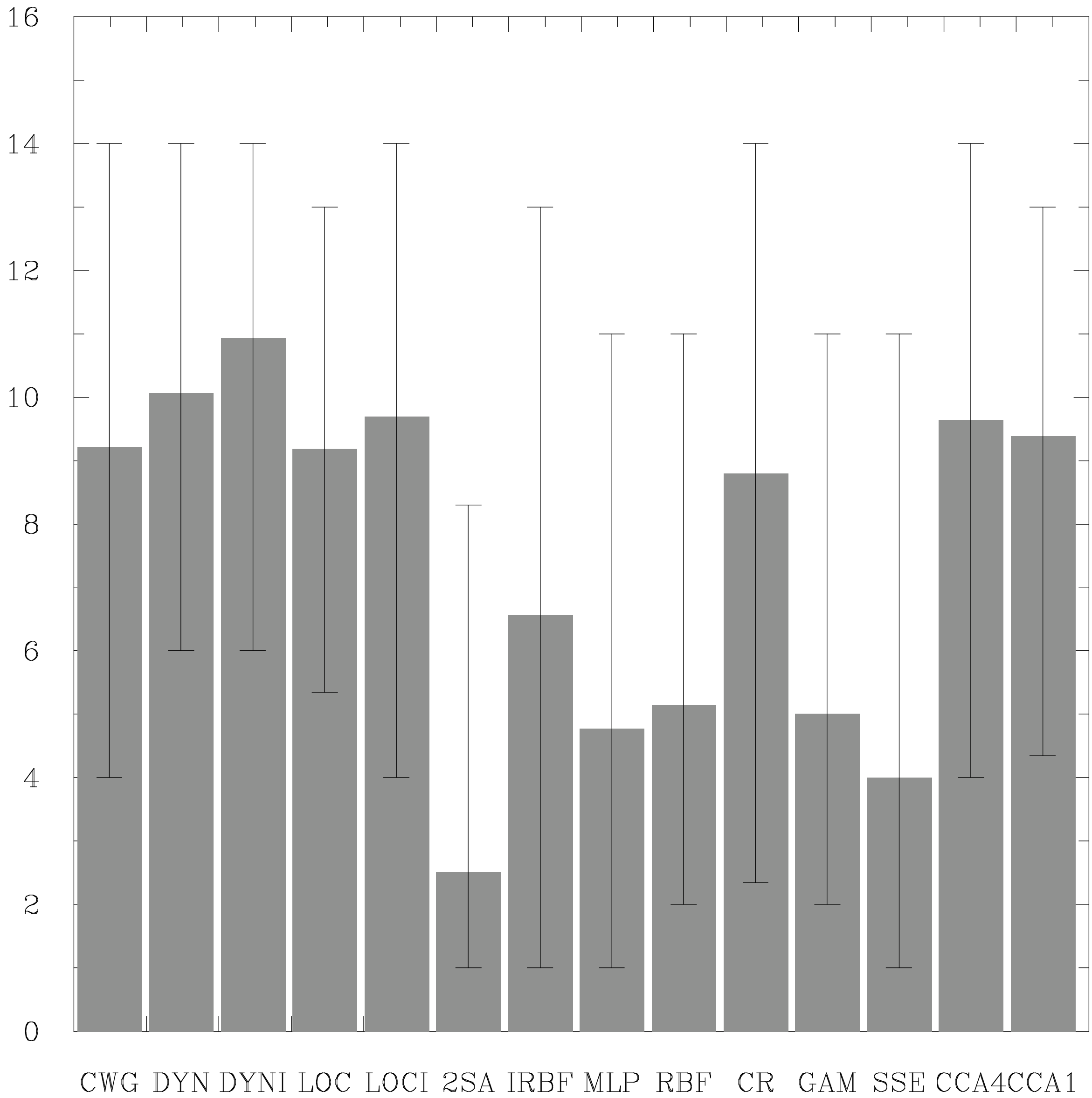

Figure 4: Rank of abs(bias) for each model averaged across all seasons, indices and stations. Lower biases are given a higher rank.

Conclusion

This comparison of downscaling methods examined correlation, rank of RMS error and rank of absolute bias averaged across 6 stations and 7 indices in the UK. A broad grouping of the models by methodology showed a similar performance amongst models from the same group. Generally the artificial neural network models had the highest correlation between observed and modelled indices but also amongst the highest biases. Two of the ANN models also had the lowest average RMS error of any model, however this was not consistent across all the ANN models. The rescaling methods performed only slightly less well for the correlations and also scored well for the RMS error and correlation. While the direct methods had average biases and RMS errors, they had amongst the lowest correlations.