AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2014JD021732

Key Points:

- Comparison between bias-corrected RCMs and bias-corrected GCMs
- The first application of stochastic MOS for GCM precipitation
- It is challenging to demonstrate the value added by RCMs in this setup

Correspondence to:
J. M. Eden,
j.m.eden@bham.ac.uk

Citation:
Eden, J. M., M. Widmann, D. Maraun, and M. Vrac (2014), Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing, *J. Geophys. Res. Atmos.*, *119*, 11,040–11,053, doi:10.1002/2014JD021732.

Received 10 MAR 2014
Accepted 18 AUG 2014
Accepted article online 21 AUG 2014
Published online 6 OCT 2014

Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing

Jonathan M. Eden[1], Martin Widmann[1], Douglas Maraun[2], and Mathieu Vrac[3]

[1]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom, [2]GEOMAR - Helmholtz Centre for Ocean Research Kiel, Kiel, Germany, [3]Laboratoire des Sciences du Climat et de l'Environnement, CEA Saclay, Gif-sur-Yvette, France

Abstract In order to assess to what extent regional climate models (RCMs) yield better representations of climatic states than general circulation models (GCMs), the output of each is usually directly compared with observations. RCM output is often bias corrected, and in some cases correction methods can also be applied to GCMs. This leads to the question of whether bias-corrected RCMs perform better than bias-corrected GCMs. Here the first results from such a comparison are presented, followed by discussion of the value added by RCMs in this setup. Stochastic postprocessing, based on Model Output Statistics (MOS), is used to estimate daily precipitation at 465 stations across the United Kingdom between 1961 and 2000 using simulated precipitation from two RCMs (RACMO2 and CCLM) and, for the first time, a GCM (ECHAM5) as predictors. The large-scale weather states in each simulation are forced toward observations. The MOS method uses logistic regression to model precipitation occurrence and a Gamma distribution for the wet day distribution, and is cross validated based on Brier and quantile skill scores. A major outcome of the study is that the corrected GCM-simulated precipitation yields consistently higher validation scores than the corrected RCM-simulated precipitation. This seems to suggest that, in a setup with postprocessing, there is no clear added value by RCMs with respect to downscaling individual weather states. However, due to the different ways of controlling the atmospheric circulation in the RCM and the GCM simulations, such a strong conclusion cannot be drawn. Yet the study demonstrates how challenging it is to demonstrate the value added by RCMs in this setup.

1. Introduction

It is widely acknowledged that future climates will be associated with changes in global precipitation. While such changes act at all spatial scales, it is at local and regional scales where changes in daily precipitation characteristics, including extreme events, are most important for impact assessment. General circulation models (GCMs) are the most important tool for estimating precipitation for climate change scenarios but do not resolve small spatial scales. The production of high-resolution scenarios from regional climate models (RCMs), nested into GCMs over a limited area, is computationally expensive and is only justified if RCMs improve the representation of regional climate simulated by the driving GCMs. The value added by RCM simulations can be difficult to quantify and has been addressed in a number of recent studies [e.g., *Castro et al.*, 2005; *Feser*, 2006; *Prömmel et al.*, 2010; *Diaconescu and Laprise*, 2013], including those focusing specifically on precipitation [e.g., *Lucas-Picher et al.*, 2012; *Luca Di et al.*, 2012; *Zou and Zhou*, 2013]. As RCMs typically contain some degree of bias, RCM output is often subject to statistical bias correction [e.g., *Engen-Skaugen*, 2007; *Graham et al.*, 2007; *Lenderink et al.*, 2007; *Piani et al.*, 2010a; *Themessl et al.*, 2011]. In recent literature, this has been referred to as Model Output Statistics (MOS) [*Maraun et al.*, 2010], a term originally coined in the context of numerical weather prediction [*Glahn and Lowry*, 1972; *Klein and Glahn*, 1974; *Wilks*, 2006]. Such statistical corrections may also be applied to GCM output but the extent to which MOS-corrected RCMs outperform MOS-corrected GCMs remains unclear.

While RCMs are able to resolve atmospheric processes at sub-GCM grid scales, postprocessing using MOS to correct systematic bias is important in improving the usefulness of model output to impact modelers and other end-user groups. This two-step approach to downscaling is restricted by the availability of RCM simulations and their associated computational expense. An alternative is to calibrate statistical corrections and downscaling directly for precipitation simulated by the driving GCM, thus removing the requirement for an RCM step [e.g., *Schmidli et al.*, 2006]. Statistical correction of GCM-simulated precipitation has been applied in the context of hydrological modeling [*Sharma et al.*, 2007; *Piani et al.*, 2010b] and crop yield [*Ines and Hansen*, 2006] simulations but has been almost entirely limited to "distributionwise" calibration; that is, the statistical relationship underpinning the correction is derived between long-term means or distributions of precipitation intensity. In fitting a distributionwise correction, the predictor distribution is mapped directly onto that of the predictand meaning that the calibration appears to be perfect. Additional validation is required in order to demonstrate the predictive power of the predictors and thus to justify the correction itself. In the case that calibration is based on a simulation in which the day-to-day evolution of large-scale weather states matches that of the real world, it is possible for statistical relationships to be derived between sequences of simulated and observed precipitation events; this is referred to as pairwise correction. This setup provides information about predictive power of the statistical correction either directly from the cost function considered for calibration or by analysis of skill scores. Although this is not a direct measure for how skillful postprocessed climate change simulations are, this type of validation yields information on how well local states are predicted given correct large-scale states, which is a key aspect of statistical downscaling.

When driven by reanalysis fields and thus forced to the temporal evolution of large-scale weather, RCM simulations provide a basis for pairwise correction. However, for GCMs used for climate change scenarios, there are usually no historical simulations available that include assimilation of observational data, meaning the sequences of simulated and observed day-to-day weather are independent and therefore fitting of pairwise MOS is not possible. Following a feasibility study for GCM-MOS based on the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis [*Widmann et al.*, 2003], *Eden et al.* [2012] demonstrated that it is possible to force the sequence of weather into temporal phase with reality using a simulation of the ECHAM5 atmospheric GCM in which the prognostic variables describing circulation and temperature are nudged toward corresponding reanalysis fields. The simulated precipitation field, which was not nudged and solely calculated by model physics, was shown to capture well the temporal variability of observed precipitation in many parts of the extratropics. *Eden and Widmann* [2014] then showed that pairwise MOS correction of monthly mean precipitation from ECHAM5 outperforms traditional statistical downscaling (so-called Perfect Prog) methods across large parts of Europe, North America, and Australia.

The majority of MOS methods that have been applied to RCM output are deterministic and do not account for any noise that is not explained by the predictors [*Maraun*, 2013]. Such methods thus correct only systematic bias. *Wong et al*. [2014] recently proposed a stochastic MOS model for simultaneously correcting and downscaling simulated precipitation. The stochastic model, which was fitted pairwise between sequences of observations and precipitation from an RCM simulation driven with observed boundary conditions, included two regression-based components: a logistic regression for estimating wet day occurrence, and a vector generalized linear model (VGLM) that estimates distribution parameters as linear combinations of a set of predictors. This method was shown to perform generally well across a sample of eight UK stations.

Here we apply a stochastic MOS correction to both RCM- and GCM-simulated precipitation across the whole of the UK. We follow the approach of *Wong et al*. [2014] using a MOS model based on logistic regression and a VGLM to estimate gamma distribution parameters. The model is first of all applied to two RCM simulations driven by observed boundary conditions and, in the case of the second, by spectral nudging within the domain. Second, we apply the model to precipitation from the nudged simulation of the ECHAM5 GCM, described by *Eden et al.* [2012]. This paper thus represents the first development of a pairwise probabilistic correction for GCM-simulated daily precipitation. Our approach provides in principle a basis to compare RCM-MOS and GCM-MOS and to assess value added by postprocessed RCMs relative to postprocessed GCMs. As mentioned, several studies have addressed the question of added value given by RCMs, finding in general that RCMs yield better representation of regional-scale climate, defined by climate indices and other statistics (e.g., precipitation quantiles), than the data used to drive them and particularly so in regions associated with complex physiographic features [e.g., *Feser*, 2006; *Luca Di et al.*, 2012]. However, in regions where large-scale forcings are dominant, an RCM may deteriorate the simulated climate of a strongly performing GCM [*De Sales and Xue*, 2011; *Luca Di et al.*, 2012], and there are subsequent examples of low-lying regions where RCMs add no noticeable value or even weaken the skill of a simulation [*Winterfeldt and Weisse*, 2009]. Some studies have found that RCMs specifically require some form of bias correction in order to add value to precipitation simulations [e.g., *Halmstad et al.*, 2013], but no focus has yet been given to a comparison of RCM- and GCM-simulated precipitation following statistical postprocessing. Such a comparison is potentially an important aspect of validating precipitation downscaling, and the approach used here offers a platform on which to begin a discussion on this topic.

The remainder of the paper is structured as follows. Section 2 describes the RCM and GCM simulations and the observational data to be used, in addition to the statistical model. The performance of the statistical model when applied to RCM and GCM precipitation is evaluated in section 3. A discussion is given in section 4 with conclusions drawn regarding the added value of the additional RCM step in the downscaling and correction process.

2. Data and Methods

2.1. Data and Setup of Simulations

In its most simple form, MOS involves a bias correction of the mean or distribution of precipitation simulated by a free-running numerical model. Such a simulation does not assimilate observations and thus does not match the temporal evolution of atmospheric states in the real world. In this case, fitting a statistical model can only be done distribution wise. Likewise, the sequence day-to-day weather from an RCM driven by a free-running GCM will not be synchronized with observations, and again only a distributionwise correction is possible. An alternative approach is to drive an RCM with an atmospheric reanalysis in order to approximately synchronize the sequence of simulated and observed time series. Such a setup provides the basis for fitting pairwise corrections, including regression-based models.

Wong et al. [2014] noted that driving an RCM at its boundaries alone allowed the RCM the freedom to generate internal variability, the extent of which negatively impacted on the predictive skill of their MOS model. Instead, *Wong et al*. [2014] fitted their MOS model on a simulation of COSMO-CLM version 4.8 [*Rockel et al.*, 2008] that is driven by the NCEP-NCAR reanalysis at its boundaries and also incorporates spectral nudging [*von Storch et al.*, 2000] of the large-scale upper level (above 850 hPa) horizontal wind speed components within the model domain [*Geyer and Rockel*, 2013; *Geyer*, 2014]. Perfect boundary RCM simulations are readily available from the data archives of international projects such as ENSEMBLES. Spectrally nudged RCM simulations have been used in the production of climate change projections as part of the North American Regional Climate Change Assessment Program [*Mearns et al.*, 2013] but are less common across the European region and are rarely made available in the public domain. The extent of the benefit of fitting MOS against a spectrally nudged simulation is unclear, and *Wong et al*. [2014] acknowledged that there may be regions of the UK where the performance of stochastic MOS is sufficiently strong when calibrated on a perfect boundary simulation. For these reasons, our MOS model was fitted on precipitation from both the spectrally nudged COSMO-CLM simulation [*Geyer and Rockel*, 2013; *Geyer*, 2014] used by *Wong et al*. [2014] and KNMI-RACMO2 [*van Meijgaard et al.*, 2008] boundary driven by ERA-40. The simulations were carried out over similar Europe-wide domains (see *van Meijgaard et al.* [2008] and *Geyer* [2014] for full details), and output was available at resolutions of approximately 18 × 18 km and 25 × 25 km, respectively. Additionally, the MOS model was fitted on precipitation from the nudged ECHAM5 simulation described by *Eden et al.* [2012] in which the prognostic fields (divergence, vorticity, temperature, and surface pressure) are forced to corresponding daily fields from ERA-40 [*Uppala et al.*, 2005]. Model output is on a T63 Gaussian grid, at an approximate resolution 200 km latitude × 150 km longitude at 45°. Further details about the simulation, including setup and analysis of bias, can be found in *Eden et al.* [2012].

Ideally, the forcing of large-scale weather to reanalysis fields should be undertaken such that the internal variability of the large-scale states is the same in all cases. This is not trivial and would require ensemble simulations and extensive testing of nudging constants. While this complex approach is not possible here, it is nevertheless important to understand what effect the respective nudging technique has on precipitation in each simulation. *Eden et al.* [2012] demonstrated that the nudged ECHAM5 simulation is able to reproduce the interannual variability of monthly and seasonal geopotential height and temperature, and also that the skill is spatially dependent and far weaker in the tropics. A broadly similar pattern exists for precipitation. Figure 1 shows the correlation of observed and simulated seasonal precipitation and sea level pressure from RACMO2, CCLM, and ECHAM5. Correlation in both fields is generally high across all simulations. It is unsurprising that the greater freedom in the boundary-forced KNMI-RACMO2 simulation results in weaker correlation, particularly in Eastern and Central parts of Europe. COSMO-CLM and ECHAM5 produce fairly similar correlation patterns in spite of the different nudging setups used for each simulation. The higher resolution of CCLM is better able to represent temporal variability in mountainous regions than ECHAM5.

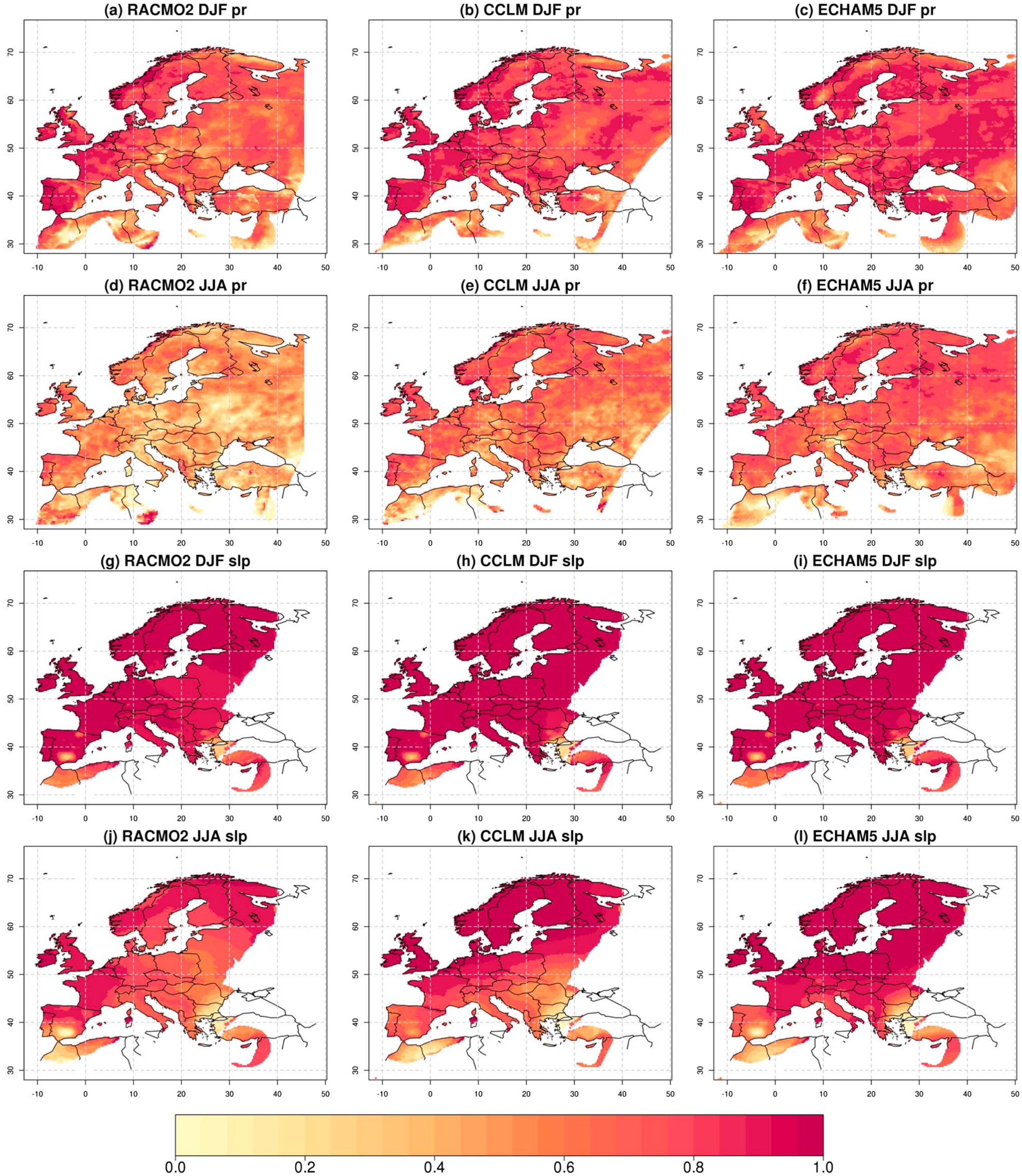

Figure 1. Correlation between observed (E-OBS) and simulated seasonal mean (a–f) precipitation and (g–l) sea level pressure from RACMO2, CCLM (spectrally nudged to reanalysis fields) and ECHAM5 (nudged to reanalysis fields) for the period 1961–2000.

Overall, the strength of the correlation patterns in Figure 1 associated with each model is high, justifying the application of our MOS correction model to both RCM-simulated (KNMI-RACMO2 and COSMO-CLM) and GCM-simulated (ECHAM5) precipitation (hereafter referred to as RCM-MOS and GCM-MOS). All RCM and GCM output is taken for the period 1961–2000, and the MOS model was fitted and validated separately for winter (December-January-February; DJF) and summer (June-July-August; JJA). For fitting, local-scale daily precipitation observations were taken from the Meteorological Office Integrated Data Archive System (MIDAS). A total of 465 stations were chosen based on at least 90% completeness for each season and each 10 year period between 1961 and 2000. Fitting was made between station observations and precipitation from the RCM or GCM grid cell that resides over each station. To account for spatial discrepancies between observed and RCM precipitation, we also fitted our MOS model on averages of simulated precipitation across the three-by-three and five-by-five grid cells centered on the station of interest. The MOS model was cross validated using a leave-one-out framework. A MOS correction is derived separately for each decade based on fitting data for the remaining three decades. For instance, when the validation period is 1991–2000, observed and simulated precipitation from 1961 to 1990 is used for model fitting.

2.2. Stochastic MOS Model

Statistical representation of daily precipitation characteristics requires modeling of the probability density function. The gamma distribution is a good fit for wet day precipitation intensities, at least up to the high quantiles [e.g., *Katz*, 1977]. In a stationary context, a gamma distribution fitted on observed precipitation for a given period provides an estimate for distribution of real world precipitation. By contrast, downscaling requires the distribution to be estimated as a function of a given predictor. In the context of a pairwise stochastic approach, the family of generalized linear models (GLMs) offers an important framework that allows a time-dependent probability distribution to be estimated as a function of a time series of predictors [*McCullagh and Nelder*, 1989; *Dobson*, 2001].

Our method uses two models belonging to the GLM class to downscale precipitation occurrence and intensity as part of a two-step process. First of all, the probability of precipitation occurrence is estimated using a logistic regression [e.g., *Chandler and Wheater*, 2002]. To model the probability p_i of greater than 1 mm of precipitation on a day i, conditional on simulated precipitation x_i, we use

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha x_i + \beta, \tag{1}$$

where α and β are coefficients to be estimated.

Second, precipitation intensity is estimated using a vector generalized linear model (VGLM). VGLMs were developed as an extension to the GLM framework [*Yee and Wild*, 1996; *Yee and Stephenson*, 2007] and allow for multiple distribution parameters to be estimated from the same set of predictors. In our case the rate parameter λ and shape parameter γ of the observed precipitation depend linearly on the simulated precipitation $x(t)$, and the model has the form

$$\lambda_i = \lambda_0 + \beta_\lambda x_i \tag{2}$$

$$\gamma_i = \gamma_0 + \beta_\gamma x_i, \tag{3}$$

where the regression parameters β_λ and β_γ are determined by Maximum Likelihood Estimation.

The probability that observed precipitation on a given day (R_i) is less than or equal to a particular precipitation intensity (r) is given by

$$Pr_{\lambda,\gamma}(R_i \le r) = \Gamma_{\lambda,\gamma}(R_i \le r|W) \times p_i + (1 - p_i), \tag{4}$$

where $\Gamma_{\lambda,\gamma}(R_i \le r|W)$ is the gamma cumulative distribution function and p_i is the probability of that given day being wet.

3. Results

To assess the predictive power of our approach across the UK, we use skill scores that have originally been applied in the verification of weather forecasts [*Joliffe and Stephenson*, 2003; *Wilks*, 2006]. The four 10 year validation periods are merged to produce a 40 year continuous, independently estimated series for which skill scores are calculated.

The two components of our method, the logistic model and the VGLM, are evaluated separately in terms of their ability to estimate precipitation occurrence and intensity, respectively. The Brier score (BS) [e.g., *Wilks*, 2006] is used to assess the performance of the logistic model to estimate dry and wet (i.e., precipitation greater than 1 mm), measuring the mean squared error between N pairs of forecast probabilities f_i and actual observations o_i, where $i = 1, \dots, N$:

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{N}(f_i - o_i)^2. \quad (5)$$

The forecasts f_i are given as probabilities between 0 and 1; the observations o_i are given as 0 and 1 for observed dry and wet days, respectively. Thus, the closer the forecast to observations, the lower the Brier score. The Brier skill score (BSS) gives the improvement over the Brier score of a reference model $\mathrm{BS}_{\mathrm{ref}}$, in this case the climatology:

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{ref}}}. \quad (6)$$

The quantile score (QS) [*Friedrichs and Hense*, 2007; *Thorarinsdottir and Johnson*, 2012] is used to assess the performance of the VGLM to estimate specific quantiles of precipitation. The QS for the α-quantile q_α is defined as the weighted average distance between n pairs of observations o_i and forecasts $q_\alpha(f_i)$:

$$\mathrm{QS}_\alpha = \sum_{i=1}^{N}\rho_\alpha(o_i - q_\alpha(f_i)), \quad (7)$$

where

$$\rho_\alpha(u) = \begin{cases} \alpha u & \text{for } u \geq 0; \\ (\alpha - 1)u & \text{for } u < 0. \end{cases} \quad (8)$$

Similar to the BSS, the quantile skill score (QSS) quantifies the improvement over the estimate from reference model $\mathrm{QSS}_{\mathrm{ref}}$, which in this case is the stationary gamma model:

$$\mathrm{QSS}_\alpha = 1 - \frac{\mathrm{QS}_\alpha}{\mathrm{QS}_{\alpha,\mathrm{ref}}}. \quad (9)$$

3.1. Application to RCM Precipitation (RCM-MOS)

First of all, the dependence of the model performance on the size of the predictor domain was assessed. Climate models typically suffer from location bias due to a large degree of random spatial variability, which, on a daily time scale, may result from misrepresentation of topographical features or the divergence of a simulated weather system from an observed trajectory. This results in poor temporal correlation between precipitation observed at a given station and simulated precipitation at the grid cell over that station. One way of dealing with this when fitting pairwise statistical corrections is to define the predictor as precipitation within a multiple grid cell domain rather than at a single grid cell. For instance, *Wong et al*. [2014] took as a predictor the average of simulated precipitation across an area of 3×3 grid cells centered on a given station. We compared the skill of our method associated with three different predictor domain sizes: single grid cell in addition to 3×3 and 5×5 centered grid cells. Table 1 details the UK average Brier and quantile skill scores associated with different predictor domain sizes. The 3×3 and 5×5 predictor domains perform slightly better than the single cell. For consistency with previous work, the remainder of our analysis of RCM precipitation uses a 3×3 predictor domain.

Second, focus was given to how model performance is influenced when precipitation is taken from an RCM simulation that includes spectral nudging. RCMs are able to produce their own random day-to-day weather and, while nesting an RCM within a reanalysis will force the large-scale weather states into temporal phase with the real world, the random component may become more dominant with distance from the simulation boundaries and at smaller scales. In principle, the addition of spectral nudging forces the large-scale weather state throughout the RCMs spatial domain, thus reducing the mismatch between simulated and observed day-to-day weather. Figure 2 shows observed and simulated daily winter (DJF) precipitation at two locations with contrasting precipitation climatologies: Kinlochewe in North West Scotland and Dover

10.1002/2014JD021732

Table 1. UK Average Seasonal Brier and Quantile Skill Scores When Fitted on CCLM-Simulated Precipitation With Different Predictor Domain Sizes (1961–2000)

	BSS		QSS_{50}		QSS_{90}	
	DJF	JJA	DJF	JJA	DJF	JJA
Single grid cell	0.15	0.11	0.10	0.04	0.14	0.13
3 × 3 predictor domain	0.16	0.13	0.10	0.04	0.15	0.14
5 × 5 predictor domain	0.16	0.14	0.10	0.05	0.16	0.16

in South East England. At Kinlochewe, for the example period shown (1991–1995), both simulations capture the variability in day-to-day precipitation reasonably well and there is little notable difference between them. Winter precipitation in north-west Scotland is dominated by westerly weather systems, the temporal evolution and trajectory of which is likely to be sufficiently represented by the boundary-driven RACMO2 simulation. By contrast, at Dover there are notable differences in the time series of simulated precipitation between RACMO2 and CCLM. In many cases, peaks in observed daily precipitation are matched by CCLM but not RACMO2. Additionally, there are several dry spells that are not correctly simulated by RACMO2. This mismatch in sequence and magnitude of precipitation events is likely to be expected in regions of the UK that are (a) further from the boundaries of the model's domain and (b) influenced to a greater extent by nonwesterly weather systems.

The distribution of skill scores across the UK allows us to quantify the differences between the two RCMs when used for fitting our downscaling model. Brier skill scores are shown in Figure 3. The positive BSS values indicate that the estimation of wet day occurrence from our logistic model is stronger than an estimate simply based on the climatology. Skill scores are generally higher in winter (DJF) than in summer (JJA). During winter, skill scores for both RACMO2 and CCLM are as high as 0.3 in large parts of the western UK with the exception of Northern Ireland. In the central and eastern parts of the UK, skill scores are lower but generally around 0.1 greater for CCLM than RACMO2. This west-east difference reflects the topographical influence on UK precipitation, with daily precipitation occurrence along the wetter west coast proving far easier to

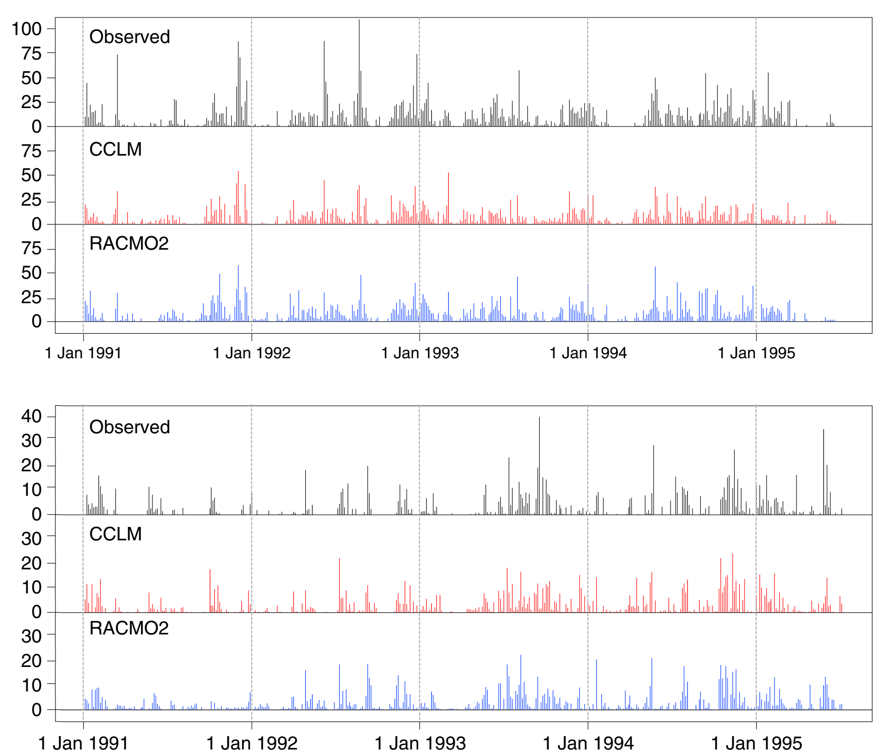

Figure 2. Winter (DJF) observed (black), CCLM (red), and RACMO2 (blue) precipitation at (top) Kinlochewe (−5.308, 57.613) and (bottom) Dover (1.322, 51.130) (1991–1995).

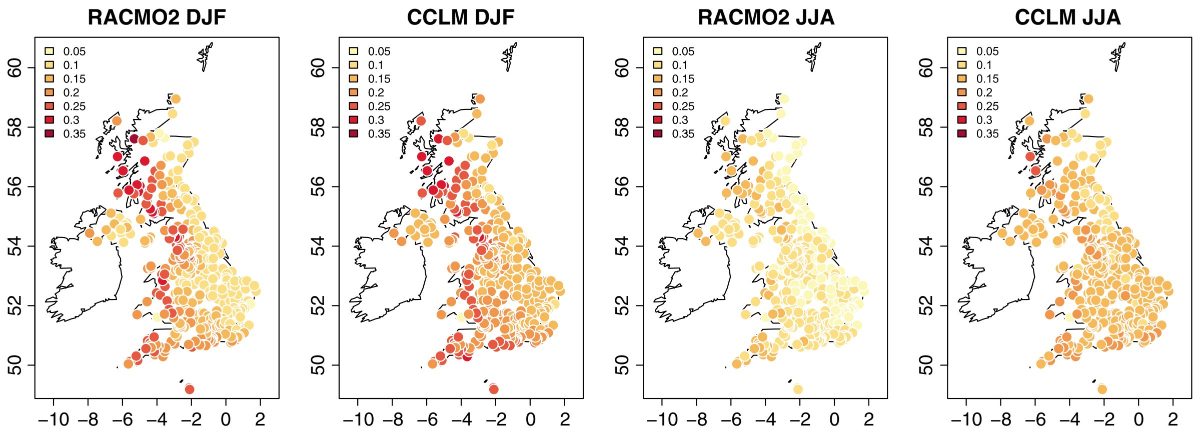

Figure 3. Cross-validated Brier skill scores for MOS fitted on precipitation from RACMO2 and CCLM for winter (DJF) and summer (JJA) for the period 1961–2000.

estimate compared to rest of the country. The advantage of spectral nudging in CCLM is clear in central and eastern UK, but there is little skill to be gained along the west coast. During summer, CCLM produces higher-skill scores throughout the UK.

A summary of quantile skill scores is presented in Figure 4; results for the 50th, 75th, 90th, and 95th percentiles are shown. Again, the skill scores for all quantiles are almost always positive across the whole of the UK, indicating that the VGLM has greater predictive power than a stationary gamma model. The improvement over the stationary model is in general smaller for the median than for the higher percentiles (90th, 95th, and particularly 75th). During winter, the west-east pattern in the BSS results is not only most noticeable for the median but also present at higher percentiles. The VGLM performs strongly even in estimating the 95th percentile, suggesting that our method is capable of predicting events that lead to heavy precipitation. The improvement in predictive power added by spectral nudging is again most apparent in central and eastern UK. The difference in skill scores between RACMO2 and CCLM in these areas is fairly consistent at all quantiles. During summer, skill scores are in general a lot lower. CCLM offers greater predictive power although few stations exhibit scores of greater than 0.25.

As mentioned earlier, winter precipitation along the west coast is dominated by westerly weather systems. The proximity of such systems to the edge of the RCM domain means that their day-to-day variability is sufficiently represented in an RCM with a boundary-driven setup. The influence of the RCM's own internal variability on the position of precipitation-bearing weather systems can be expected to become greater with distance from the domain boundary. For this reason, the addition of spectral-nudging in CCLM produces noticeably higher skill scores in central and eastern UK. During summer (JJA), the dominance of westerlies on precipitation is lesser than during winter and the addition of spectral nudging produces stronger skill scores across all parts of the UK.

3.2. Application to ECHAM5 Precipitation (GCM-MOS)

With our method shown to exhibit good predictive power when applied to RCM precipitation, we now evaluate its skill when applied to precipitation from a nudged GCM simulation. Brier skill scores, presented in Figure 5, are greater than 0.25 across the majority of the UK during winter, and particularly high across southern England and Wales. Skill scores are in general lower during summer, with only a small number of coastal stations associated with skill scores greater than 0.25. Quantile skill scores calculated for the same four percentiles (median, 75th, 90th, and 95th) are presented in Figure 6. During winter, the higher quantiles show stronger skill in the south and east of the UK, and particularly so along the south coast (QSS up to 0.35). For the median, the skill is stronger in the west of the UK with skill scores in the east not much higher than 0.2. During summer, it is in Central England and Wales that the VGLM performs most strongly. The results shown in Figures 5 and 6 clearly demonstrate the good potential of our method when applied to ECHAM5 precipitation. The high skill indicates that ECHAM5 sufficiently resolves the weather events leading to precipitation events of different magnitudes, despite a much coarser resolution than that used in RCM simulations.

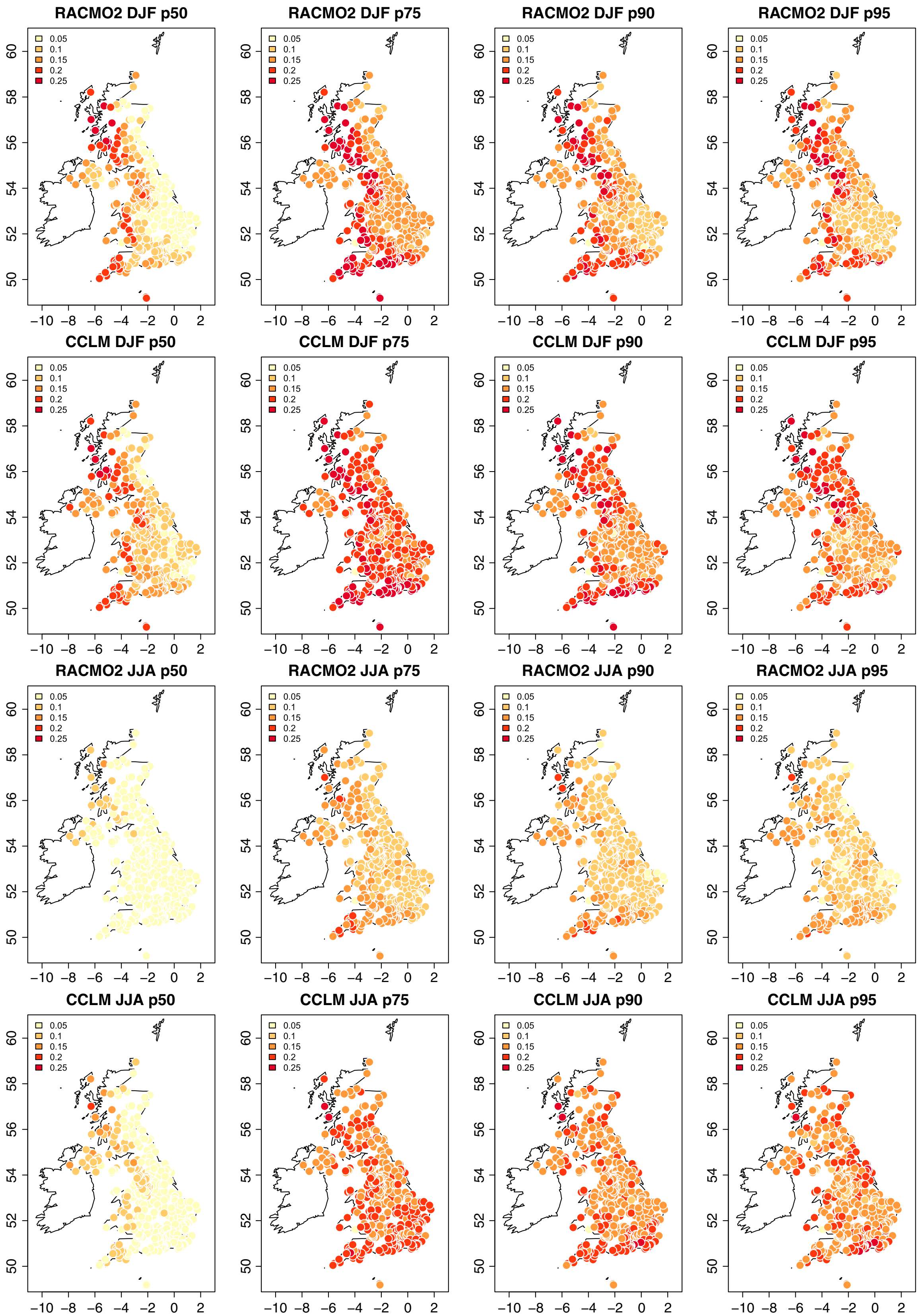

Figure 4. Cross-validated quantile skill scores for MOS fitted on precipitation from RACMO2 and CCLM for winter (DJF) and summer (JJA) for the period 1961–2000. Quantile skill scores are presented for the 50th (p50), 75th (p75), 90th (p90), and 95th (p95) percentiles.

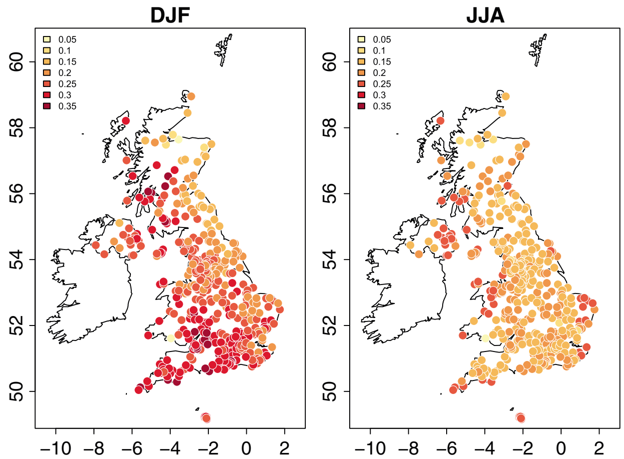

Figure 5. Cross-validated Brier skill scores for MOS fitted on precipitation from ECHAM5 for winter (DJF) and summer (JJA) for the period 1961–2000.

A natural next step is to compare the performance of RCM-MOS and GCM-MOS. The results in Figures 3-6 show that, in general, the skill scores are higher when fitted on ECHAM5 precipitation, but there are notable exceptions. Table 2 shows the average Brier and quantile skill scores for models fitted on CCLM and ECHAM5 precipitation within nine regions of the UK. In Scotland, particularly during winter, there is very little difference in skill. In Northern Scotland, CCLM actually performs slightly better than ECHAM5. The dominance of frontal and orographic processes on precipitation in the northern parts of the UK may lead to it being well captured by both nudged simulations. ECHAM5 produces higher skill scores in Northern Ireland, possibly due to the smaller role of topography in determining precipitation distribution. ECHAM5 consistently produces Brier and quantile (above the median) skill scores of 7–10% greater than CCLM in southern, central, and eastern parts of the UK during winter. In summer, the difference is smaller and only apparent in South West England and Central and Eastern England; both models are indistinguishable in South East England. Interestingly, there is little

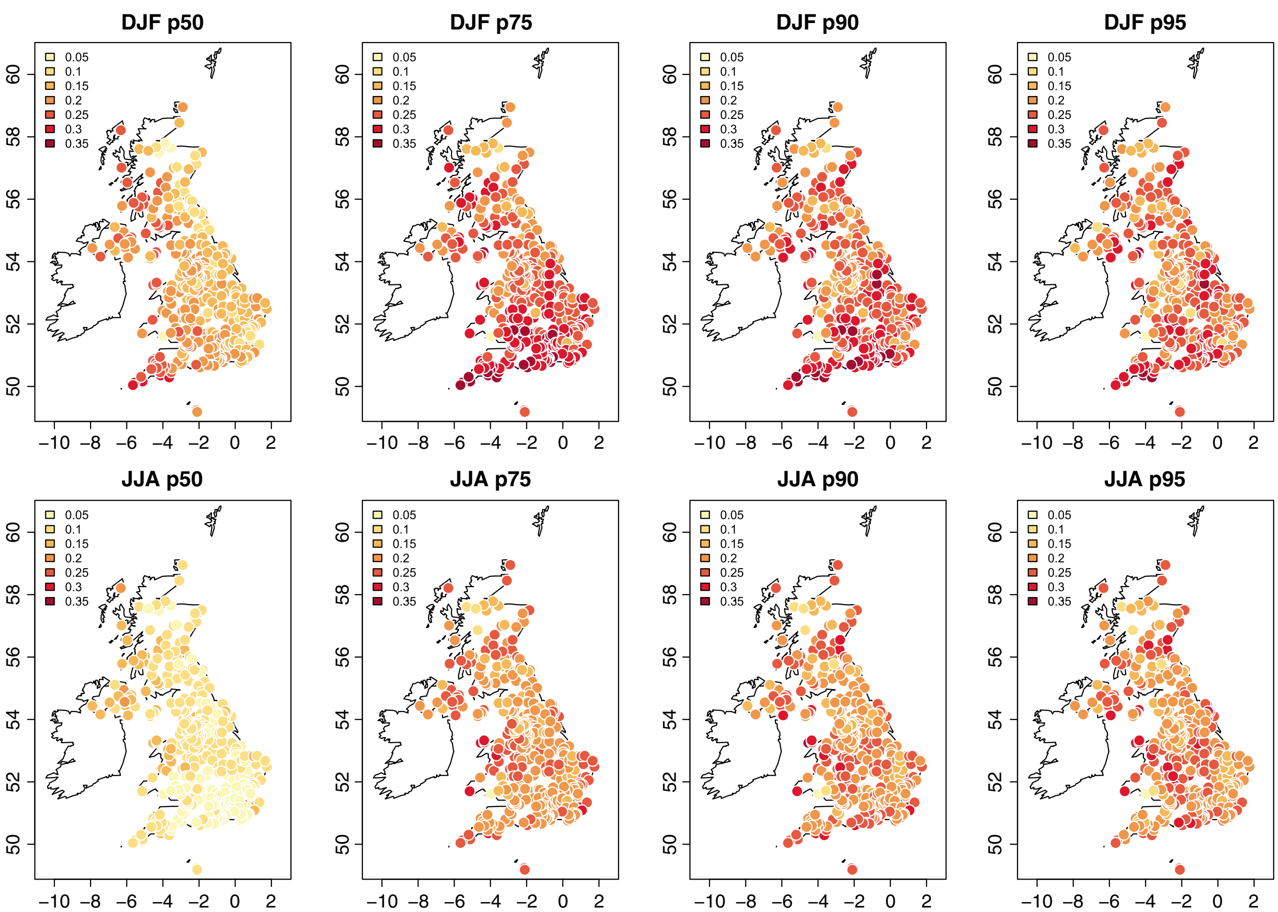

Figure 6. Cross-validated quantile skill scores for MOS fitted on precipitation from ECHAM5 for winter (DJF) and summer (JJA) for the period 1961–2000. Quantile skill scores are presented for the 50th (p50), 75th (p75), 90th (p90), and 95th (p95) percentiles.

Table 2. Differences in Regionally Averaged Seasonal Brier and Quantile Skill Scores for Models Fitted on CCLM and ECHAM5 Precipitation (1961–2000)

	Season	BSS	QSS_{25}	QSS_{50}	QSS_{75}	QSS_{90}	QSS_{95}
Northern Scotland	DJF	−0.03	−0.01	−0.02	−0.01	0.00	−0.03
	JJA	0.00	0.01	0.00	0.00	0.01	0.01
East Scotland	DJF	0.02	0.01	0.03	0.03	0.04	0.05
	JJA	0.04	0.02	0.04	0.03	0.05	0.05
South Scotland	DJF	0.03	0.03	0.03	0.03	0.02	0.02
	JJA	0.02	0.01	0.02	0.02	0.03	0.04
North West England	DJF	0.04	0.03	0.04	0.04	0.04	0.05
	JJA	0.02	0.01	0.02	0.02	0.04	0.05
North East England	DJF	0.05	0.02	0.04	0.05	0.06	0.06
	JJA	0.03	0.02	0.03	0.02	0.04	0.04
Northern Ireland	DJF	0.09	0.05	0.08	0.08	0.09	0.09
	JJA	0.09	0.04	0.06	0.07	0.07	0.06
South West England	DJF	0.08	0.04	0.07	0.08	0.09	0.09
	JJA	0.02	0.01	0.02	0.03	0.05	0.07
Central and Eastern England	DJF	0.09	0.03	0.07	0.08	0.10	0.10
	JJA	0.04	0.01	0.03	0.03	0.06	0.07
South East England	DJF	0.10	0.03	0.07	0.08	0.10	0.09
	JJA	0.02	0.00	0.01	0.02	0.03	0.03

difference in Brier skill scores between models during summer (except in Northern Ireland). It is important to note that the smaller number of wet days during summer is likely to be more difficult to estimate, and the stronger nudging setup used in ECHAM5 does not appear to produce a better performance.

4. Discussion and Conclusions

We have applied a stochastic Model Output Statistics (MOS) method to simultaneously correct and downscale RCM- or GCM-simulated precipitation to the point scale across the United Kingdom. In contrast to deterministic MOS methods that only correct systematic bias, the stochastic approach explicitly accounts for unexplained variability and produces probabilistic estimates for precipitation at the point scale. A similar approach has been previously applied to downscale RCM-simulated precipitation at eight stations in the UK; our work assesses skill over a dense network of stations and represents the first application of this approach to precipitation from a GCM simulation. Furthermore, comparison of MOS corrected output from each class of numerical model provides a basis to assess the added value of RCMs in this setup.

Our method includes two component models: a logistic regression for estimating daily precipitation occurrence; and a VGLM for estimating precipitation intensity. Both models required pairwise fitting between temporally coherent sequences of simulated and observed precipitation events. To achieve this, we used two simulations (RACMO2 and CCLM) driven by reanalysis fields using a perfect boundary setup and spectral nudging, respectively, and a GCM simulation (ECHAM5) nudged to ERA-40. The predictive power of our method for the period was assessed in a leave-one-out cross validation framework for the period 1961–2000 using verification skill scores, which have been developed in the context of weather forecasting.

When applied to RCM output, our method performs substantially better when fitted on a simulation that includes spectral nudging, which corroborates the findings of *Wong et al*. [2014]. The CCLM simulation used here is nudged only to upper level winds; it is uncertain to what extent nudging to other variables would improve model performance. For instance, nudging to atmospheric circulation at different (particularly lower) levels would limit the random variability of the RCM at the surface. The application of our method to RCM output provides important clarification of the potential impact of simulation setup and predictor domain size on model performance. However, the strong performance of our method when applied to GCM output constitutes a potentially more important finding of this study. Previous work demonstrated that deterministic MOS performs well for downscaling monthly mean precipitation from a simulation of ECHAM5 nudged to ERA-40 [*Eden and Widmann*, 2014], and our results show that a strong performance also exists in a stochastic framework for downscaling daily precipitation. GCMs are known to generally underestimate high intensities of daily precipitation, particularly in comparison to RCMs [e.g., *Jacob et al.*, 2014]. The fact that our method represents precipitation events up to the 95th percentile suggests that given realistic large-scale circulation and temperature, ECHAM5 is able to simulate grid cell precipitation that contains useful information about actual episodes of heavy precipitation. It is possible to optimize the approach for extreme precipitation. The VGLM developed by *Wong et al*. [2014] was used to estimate six parameters of a mixture distribution [*Frigessi et al.*, 2002; *Vrac and Naveau*, 2007] that combined both gamma and generalized Pareto distributions in order to represent both the core and extreme tail of the distribution. This method has not yet been applied to GCM-simulated precipitation and is an avenue for future research.

Although GCM-MOS has previously been implemented, a direct comparison has not yet been made with RCM-MOS by other work seeking to quantify the added value of RCMs. For the setup used in this study, GCM-MOS generally produces higher Brier and quantile skill scores than RCM-MOS and particularly so across central and southern parts of the UK. This leads to an important question: does applying a stochastic correction to higher-resolution output from an RCM produce better results and, if so, to what extent? More specifically, precisely what value is added by the additional RCM step in the downscaling process? Our approach permits, at least in principle, a comparison of RCM and GCM following MOS correction but the lesser performance of RCM-MOS is perhaps contrary to what might be expected: that calibrating a statistical model on high-resolution simulated output would produce better results.

It is important to highlight that the differences in skill between RCM-MOS and GCM-MOS may be partly due to the different degree of internal variability in each simulation, i.e., to how much the simulated weather states can deviate from those in the driving reanalyses. Different degrees of internal variability are likely because of the different ways of how the simulated weather states are brought in agreement with the reanalyses. RACMO2 is only constrained to the reanalysis at the lateral boundaries of the model domain, whereas CCLM and ECHAM5 are nudged to the reanalysis everywhere. Moreover, the nudging techniques used in CCLM and ECHAM5 are different; in CCLM only the upper level winds are nudged, while circulation and temperature fields throughout the troposphere are nudged in ECHAM5. The more comprehensive nudging in ECHAM5 is likely to allow less internal variability than in the RCM simulations. In addition, the variability that is not controlled by the reanalyses can be expected to be larger on smaller spatial scales; thus, it is likely to be larger in RCM than in GCM simulations even if the internal variability on the same spatial scales was similar. As shown in Figure 1, correlations between simulated and observed precipitation and sea level pressure are indeed marginally stronger in ECHAM5 across Europe with the exception of regions of complex topography. In general, however, it appears that the internal variability in all simulations is fairly similar, at least on monthly and seasonal time scales. In order to fully quantify the internal variability ensemble simulations are required, which are beyond the scope of this paper.

Although we cannot exclude that the potential differences in the similarity of simulated and observed weather states affect the performance of the MOS models to some extent in our setup, our study demonstrates that the predictive power of GCM precipitation for estimating point-scale daily precipitation is high and similar to that of RCM precipitation. Whether this predictive power extends to other regions, particularly to those characterized by complex topography that are known to be poorly represented in GCMs, is an important question for subsequent research. Our findings also highlight the difficulties of demonstrating the value added by RCMs in terms of predictive power. As discussed in previous work addressed the concept of added value [e.g., *Luca Di et al.*, 2012], it is clear that added value should not simply be defined by greater detail at local scales. We have shown that such detail can be added stochastically; GCMs have potentially high predictive value at local scales, and the predictive skill of an RCM must be greater in order to add value.

This work has clearly demonstrated that stochastic MOS is a useful tool for downscaling simulated precipitation from both RCM and GCM simulations to the point scale. The method used here performs well during both winter and summer in large parts of the UK with different precipitation climatologies. In the context of application to climate change studies, a key benefit of precipitation downscaling with MOS is that the simulated precipitation, in principle, comprehensively captures the different factors that might contribute to precipitation changes. As for all statistical correction and downscaling methods the usefulness of applying MOS in a future climate depends on the stationarity of the underpinning statistical relationships. The extent to which MOS may be transferable to climate change scenarios is an important question, although results from a pseudo-reality study indicate that MOS relationships for precipitation might indeed be stationary under climate change [*Maraun*, 2012]. Future application of pairwise models needed for stochastic MOS is constrained by the availability of multiple climate simulations that are forced to reanalyses. The constraint is a particular issue for GCM-MOS; the majority of GCM simulations made available for phase 5 of the Coupled

Model Intercomparison Project (CMIP5) are free running, meaning that pairwise fitting of statistical correction models is not possible. It is likely that nudged simulations could be undertaken using the CMIP5 suite of models without great additional effort, and we believe that the results here highlight the potential value that such simulations would bring.

Acknowledgments
This work has been undertaken with the PLEIADES project, funded by the Volkswagen Foundation (Az.: 85 423 and 85 425). ECHAM5 provided by the Model and Data group at the Max Planck Institute for Meteorology in Hamburg, Germany. ECHAM5 nudging modules were developed by I. Kirchner. MIDAS precipitation data are made available by the British Atmospheric Data Centre (http://badc.nerc.ac.uk/home/index.html). We also thank S. Rast and D. Grawe for assistance in setting up the local running of the ECHAM5 simulations. The ENSEMBLES data used in this work were funded by the EU FP6 Integrated Project ENSEMBLES (contract 505539) whose support is gratefully acknowledged. We thank B. Geyer and colleagues from the Helmholtz Centre Geesthacht for kindly providing the CCLM simulations.

References

Castro, C. L., R. A. Pielke, and G. Leoncini (2005), Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS), *J. Geophys. Res.*, *110*, D05108, doi:10.1029/2004JD004721.

Chandler, R. E., and H. S. Wheater (2002), Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, *Water Resour. Res.*, *38*(10), 1192, doi:10.1029/2001WR000906.

De Sales, F., and Y. Xue (2011), Assessing the dynamic-downscaling ability over South America using the intensity-scale verification technique, *Int. J. Climatol.*, *31*(8), 1205–1221.

Diaconescu, E. P., and R. Laprise (2013), Can added value be expected in RCM-simulated large scales?, *Clim. Dyn.*, *41*(7-8), 1769–1800.

Dobson, A. J. (2001), *An Introduction to Generalized Linear Models*, Chapman and Hall, London, U. K.

Eden, J. M., and M. Widmann (2014), Downscaling of GCM-simulated precipitation using Model Output Statistics, *J. Clim.*, *27*(1), 312–324.

Eden, J. M., M. Widmann, D. Grawe, and S. Rast (2012), Skill, correction and downscaling of GCM-simulated precipitation, *J. Clim.*, *25*(11), 3970–3984.

Engen-Skaugen, T. (2007), Refinement of dynamically downscaled precipitation and temperature scenarios, *Clim. Change*, *84*(3-4), 365–382, doi:10.1007/s10584-007-9251-6.

Feser, F. (2006), Enhanced detectability of added value in limited-area model results separated into different spatial scales, *Mon. Weather Rev.*, *134*(8), 2180–2190.

Friedrichs, P., and A. Hense (2007), Statistical downscaling of extreme precipitation events using censored quantile regression, *Mon. Weather Rev.*, *135*, 2365–2738.

Frigessi, A., O. Haug, and H. Rue (2002), A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, *5*(3), 219–245.

Geyer, B. (2014), High-resolution atmospheric reconstruction for Europe 1948-2012: CoastDat2, *Earth Syst. Sci. Data*, *5*, 147–164.

Geyer, B., and B. Rockel (2013), coastDat-2 COSMO-CLM Atmospheric Reconstruction, *Tech. Rep.*, World Data Center for Climate, Hamburg, Germany, doi:10.1594/WDCC/coastDat-2_COSMO-CLM.

Glahn, H. R., and D. A. Lowry (1972), The use of Model Output Statistics (MOS) in objective weather forecasting, *J. Appl. Meteorol.*, *11*, 1203–1211.

Graham, L. P., J. Andreasson, and B. Carlsson (2007), Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods—A case study on the Lule River basin, *Clim. Change*, *81*(Suppl. 1), 293–307, doi:10.1007/s10584-006-9215-2.

Halmstad, A., M. R. Najafi, and H. Moradkhani (2013), Analysis of precipitation extremes with the assessment of regional climate models over the Willamette River Basin, USA, *Hydrol. Process.*, *27*, 2579–2590, doi:10.1002/hyp.9376.

Ines, A. V. M., and J. W. Hansen (2006), Bias correction of daily GCM rainfall for crop simulation studies, *Agric. For. Meteorol.*, *138*, 44–53.

Jacob, D., et al. (2014), EURO-CORDEX: New high-resolution climate change projections for European impact research, *Regional Environ. Change*, *14*, 563–578.

Joliffe, I. T., and D. B. Stephenson (2003), *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Wiley, Chichester, U. K.

Katz, R. (1977), Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, *16*, 671–676.

Klein, W. H., and H. R. Glahn (1974), Forecasting local weather by means of model output statistics, *Bull. Am. Meteorol. Soc.*, *55*, 1217–1227.

Lenderink, G., A. Buishand, and W. van Deursen (2007), Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach, *Hydrol. Earth Syst. Sci.*, *11*(3), 1143–1159.

Luca Di, A., R. de Elia, and R. Laprise (2012), Potential for added value in precipitation simulated by high-resolution nested regional climate models and observations, *Clim. Dyn.*, *38*(5-6), 1229–1247.

Lucas-Picher, P., M. Wulff-Nielsen, J. H. Christensen, G. Adalgeirsdottir, R. Mottram, and S. B. Simonsen (2012), Very high resolution regional climate model simulations over Greenland: Identifying added value, *J. Geophys. Res.*, *117*, D02108, doi:10.1029/2011JD016267.

Maraun, D. (2012), Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums, *Geophys. Res. Lett.*, *39*, L06706, doi:10.1029/2012GL051210.

Maraun, D. (2013), Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue, *J. Clim.*, *26*, 2137–2143.

Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, *48*, RG3003, doi:10.1029/2009RG000314.

McCullagh, P., and J. A. Nelder (1989), *Generalized Linear Models,* 2nd ed., Chapman and Hall, London, U. K.

Mearns, L. O., et al. (2013), Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP), *Clim. Change*, *120*, 965–975.

Piani, C., J. O. Haerter, and E. Coppola (2010a), Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, *99*(1-2), 187–192, doi:10.1007/s00704-009-0134-9.

Piani, C., G. P. Weedon, M. Best, S. M. Gomes, P. Viterbo, S. Hagemann, and J. O. Haerter (2010b), Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models, *J. Hydrol.*, *395*, 199–215.

Prömmel, K., B. Geyer, J. M. Jones, and M. Widmann (2010), Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature, *Int. J. Climatol.*, *30*, 760–773.

Rockel, B., A. Will, and A. Hense (2008), The regional climate model COSMO-CLM (CCLM), *Meteorol. Z.*, *17*(4), 347–348.

Schmidli, J., C. Frei, and P. L. Vidale (2006), Downscaling from GCM precipitation: A benchmark for dynamical and statistical downscaling methods, *Int. J. Climatol.*, *26*(5), 679–689.

Sharma, D., A. D. Gupta, and M. S. Babel (2007), Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand, *Hydrol. Earth Syst. Sci.*, *11*(4), 1373–1390.

Themessl, M. J., A. Gobiet, and A. Leuprecht (2011), Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *Int. J. Climatol.*, *31*(10), 1530–1544, doi:10.1002/joc.2168.

Thorarinsdottir, T. L., and M. S. Johnson (2012), Probabilistic wind gust forecasting using nonhomogeneous Gaussian regression, *Mon. Weather Rev.*, *140*, 889–897.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*(612, Part B), 2961–3012, doi:10.1256/qj.04.176.

van Meijgaard, E., L. H. van Ulft, W. J. van de Berg, F. C. Bosveld, B. J. J. M. van den Hurk, G. Lenderink, and A. P. Siebesma (2008), The KNMI regional atmospheric climate model RACMO, version 2.1., *KNMI Tech. Rep., 302*, 43 pp., Available from KNMI, Post bus 201, 3730 AE, De Bilt, Netherlands.

von Storch, H., H. Langenberg, and F. Feser (2000), A spectral nudging technique for dynamical downscaling purposes, *Mon. Weather Rev.*, *128*(10), 3664–3673.

Vrac, M., and P. Naveau (2007), Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resour. Res.*, *43*, W07402, doi:10.1029/2006WR005308.

Widmann, M., C. S. Bretherton, and E. P. Salathe (2003), Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor, *J. Clim.*, *16*(5), 799–816.

Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 627 pp., Elsevier, London, U. K.

Winterfeldt, J., and R. Weisse (2009), Assessment of value added for surface marine wind speed obtained from two regional climate models, *Mon. Weather Rev.*, *137*, 2955–2965, doi:10.1175/2009MWR2704.1.

Wong, G., D. Maraun, M. Vrac, M. Widmann, J. M. Eden, and T. Kent (2014), Stochastic model output statistics for bias correcting and downscaling precipitation including extremes, *J. Clim.*, *27*, 6940–6959.

Yee, T. W., and A. G. Stephenson (2007), Vector generalized and additive extreme value models, *Extremes*, *10*, 1–19.

Yee, T. W., and C. J. Wild (1996), Vector generalized additive models, *J. R. Stat. Soc.*, *58*, 481–493.

Zou, L., and T. Zhou (2013), Near future (2016-40) summer precipitation changes over China as projected by a Regional Climate Model (RCM) under the RCP8.5 emissions scenario: Comparison between RCM downscaling and the driving GCM, *Adv. Atmos. Sci.*, *30*(3), 806–818.

