Manuscript prepared for Geosci. Model Dev.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 23 March 2016

Evaluation of the Plant–Craig stochastic convection scheme (v2.0) in the ensemble forecasting system MOGREPS-R (24km) based on the Unified Model (v7.3)

Richard J. Keane[1,2], Robert S. Plant[3], and Warren J. Tennant[4]

[1]Deutscher Wetterdienst, Frankfurter Strasse 135, 63067 Offenbach, Germany
[2]Meteorologisches Institut, Ludwig-Maximilians-Universität München, Germany
[3]Department of Meteorlogy, University of Reading, UK
[4]Met Office, FitzRoy Road, Exeter, EX1 3PB, UK

Correspondence to: Richard J. Keane (richard.keane@metoffice.gov.uk)

Abstract. The Plant–Craig stochastic convection parameterization (version 2.0) is implemented in the Met Office Regional Ensemble Prediction System (MOGREPS-R) and is assessed in comparison with the standard convection scheme with a simple stochastic scheme only, from random parameter variation. A set of 34 ensemble forecasts, each with 24 members, is considered, over the month of July 2009. Deterministic and probabilistic measures of the precipitation forecasts are assessed. The Plant–Craig parameterization is found to improve probabilistic forecast measures, particularly the results for lower precipitation thresholds. The impact on deterministic forecasts at the grid scale is neutral, although the Plant–Craig scheme does deliver improvements when forecasts are made over larger areas. The improvements found are greater in conditions of relatively weak synoptic forcing, for which convective precipitation is likely to be less predictable.

1 Introduction

Quantitative precipitation forecasting is recognized as one of the most challenging aspects of numerical weather prediction (Ebert et al., 2003; Montani et al., 2011; Gebhardt et al., 2011). While progress is continually being made in improving the accuracy of single forecasts – through improvements in the model formulation as well as increases in grid resolution – a complementary approach is the use of ensembles in order to obtain an estimate of the uncertainty in the forecast (Buizza et al., 2005; Montani et al., 2011; Buizza et al., 2007; Bowler et al., 2008; Thirel et al., 2010; Yang et al., 2012; Zhu, 2005; Abhilash et al., 2013; Roy Bhowmik and Durai, 2008; Clark et al., 2011; Tennant and Beare, 2013). Of course, ensemble forecasting systems themselves remain imperfect, and one of the most important problems is insufficient spread in ensemble forecasts, where the forecast tends to cluster too strongly around rainfall values that turn out to be incorrect.

One reason for lack of spread in an ensemble is that model variability is constrained by the number of degrees of freedom in the model, which is typically much less than that of the real atmosphere. The members of an ensemble forecast may start with a good representation of the range of possible initial conditions, but running exactly the same model for each ensemble member means that the range of possible ways of modelling the atmosphere – of which the model in question is one – are not fully considered. Common ways of accounting for model error are running different models for each ensemble member (e.g. Mishra and Krishnamurti, 2007; Berner et al., 2011), adding random perturbations to the tendencies produced by the parameterizations (e.g. Buizza et al., 1999; Bouttier et al., 2012) and randomly perturbing parameters in physics schemes (e.g. Bowler et al., 2008; Christensen et al., 2015).

Focusing on convective rainfall, and for model grid lengths where convective rainfall is parameterized, another way of accounting for model error is to introduce random variability in the convection parameterization itself (e.g. Lin and Neelin, 2003; Khouider et al., 2010; Plant and Craig, 2008; Ragone et al., 2014). Ideally this should be done in a physically consistent way, so that the random variability causes the parameterization to sample from the range of possible convective responses on the grid scale. A recent overview is given by Plant et al. (2015).

Such "stochastic" convection parameterization schemes have been developed over the last 10 years, and are just beginning to be implemented and verified in operational forecasting setups, with some promise for the improvement of probabilistic ensemble forecasts (e.g Teixeira and Reynolds, 2008; Bengtsson et al., 2013; Kober et al., 2015). The purpose of the present study is to continue this pioneering work of verifying probabilistic forecasts using stochastic convection parameterizations, by investigating the performance of the Plant and Craig (2008) (PC) scheme in MOGREPS, the Met Office ensemble forecasting system (Bowler et al., 2008).

The PC scheme has been shown to produce rainfall variability in much better agreement with cloud resolving model results than for other non-stochastic schemes (Keane and Plant, 2012), and has been shown to add variability in a physically consistent way when the model grid spacing is varied (Keane et al., 2014). It has also been demonstrated that the convective variability it produces, on scales of tens of kilometres, can be a major source of model spread (Ball and Plant, 2008) and further that its performance at large scales in a model intercomparison is similar to that of more traditional methods (Davies et al., 2013).

These are encouraging results, albeit from idealized modelling setups, and it is important to establish whether or not they might translate into better ensemble forecasts in a fully-operational NWP setup. Groenemeijer and Craig (2012) examined seven cases using the COSMO ensemble system with 7km grid spacing and compared the spread in an ensemble using only different realizations of the PC scheme (i.e. where the random seed in the PC scheme was varied but the members were otherwise identical) with that in an ensemble where additionally the initial and boundary conditions were varied. They found the spread in hourly accumulated rainfall produced by the PC scheme to be 25–50% of the total spread, when the fields were upscaled to 35km. The present study investigates the behaviour of the scheme in a trial of 34 forecasts with the MOGREPS-R ensemble, using a grid length of 24km. The mass-flux variance produced by the PC scheme is inversely proportional to the grid box area being used and so it is not obvious from the results of Groenemeijer and Craig (2012) whether the stochastic variations of PC will contribute significantly to variability within an ensemble system operating at the scales of MOGREPS-R. Nonetheless, MOGREPS-R has been shown, in common with most ensemble forecasting systems, to produce insufficient spread relative to its forecast error in precipitation (Tennant and Beare, 2013), suggesting that there is scope for the introduction of a stochastic convection parameterization to be able to improve its performance.

Although the version of MOGREPS used here has now been superseded, the present study represents the first time that the scheme has been verified in an operationally-used ensemble forecasting system for an extended verification period, and provides the necessary motivation for more extensive tuning and verification studies in a more current system. As well as this, the present study aims to reveal more about the behaviour of the scheme itself, building on work referenced above, as well as on recent work by Kober et al. (2015) which focused on individual case studies.

The paper compares the performance of the PC scheme with the default MOGREPS convection parameterization, based on Gregory and Rowntree (1990), in order to seek evidence that accounting for model error by using a stochastic convection parameterization can lead to improvements in ensemble forecasts. Of course, the two parameterizations are different in other ways than the stochasticity of the PC scheme: it is therefore possible that any differences in performance are due to other factors. Nonetheless, the default MOGREPS scheme has benefitted from much experience in developing it alongside the Met Office Unified Model (Lean et al., 2008, UM), whereas relatively modest efforts were made here to adapt the PC scheme to the host ensemble system: thus, any improvements that the PC scheme shows over the default scheme are of clear interest.

2 Methods

2.1 The Plant–Craig stochastic convection parameterization

The Plant and Craig (2008) scheme operates, at each model grid point, by reading in the vertical profile from the dynamical core, and calculating what convective response is required to stabilize that profile. It is based on the Kain-Fritsch convection parameterization (Kain and Fritsch, 1990; Kain, 2004), adapting the plume model used there and also using a similar formulation for the closure, based on a dilute CAPE. It generalizes the Kain-Fritsch scheme by allowing for more than one cloud in a grid box, and by allowing the size and number of clouds to vary randomly. Details of its implementation in an idealized configuration of the UM are given by Keane and Plant (2012); this would be regarded as Version 1.1. The important differences in the implementation for the present study, to produce Version 2.0, are presented here.

The scheme allows for the vertical profile from the dynamical core to be averaged in horizontal space and/or in time before it is input. This means that the input profile is more representative of the large-scale (assumed quasi-equilibrium) environment, and is less affected by the stochastic perturbations locally induced by the scheme at previous time steps. It was decided in the present study to use different spatial averaging extents over ocean and over land, in order that orographic effects were not too heavily smoothed. The spatial averaging strategy implemented was to use a square of 7×7 grid points over the ocean and 3×3 grid points over land; the temporal averaging strategy was to average over the previous 7 time steps (each of 7.5 min) and the current time step. The cloud lifetime was set to 15 minutes. As well as using the averaged profile for the closure calculation, the plume profiles were also calculated for ascent within the averaged environment.

Initial tests showed that the scheme was yielding too small a proportion of convective precipitation over the domain. Two further parameters were adjusted from the study by Keane and Plant (2012), in order to increase this fraction: the mean mass flux per cloud $\langle m \rangle$ and the root mean square cloud radius $\sqrt{\langle r^2 \rangle}$. Similar changes were made for the same reason by Groenemeijer and Craig (2012) in their mid-latitude tests over land, and reflect the fact that the original settings in Plant and Craig (2008) and Keane and Plant (2012) were chosen to match well with cloud-resolving model simulations of tropical oceanic convection. Specifically, the mean mass flux per cloud was reduced here from $2 \times 10^7\,\mathrm{kg s^{-1}}$ to $0.8 \times 10^7\,\mathrm{kg s^{-1}}$ in order to increase the number of plumes produced by the scheme. The entrainment rates used in the scheme are inversely proportional to cloud radius, and a pdf of cloud radius is used characterized by the root mean square cloud value $\sqrt{\langle r^2 \rangle}$. This was increased from 450 m to 600 m, in order to produce less strongly entraining plumes. This had some impact on the convective precipitation fraction, but the scheme still yielded a relatively low proportion of convective rain: $12\,\%$ in these tests, as compared with $50\,\%$ for the standard scheme. The overall amount of rainfall was similar for the two schemes, with the dynamics compensating for the reduction in convective rain produced, and ensuring that the instability was suitably removed by the dynamics and convection scheme combined in both cases. There is no correct answer for the convective fraction, which is both model and resolution dependent in current operational practice. For example, the current ECMWF model has a global average of about $60\,\%$ (Bechtold, 2015). Doubtless the convective precipitation fraction produced by the Plant–Craig scheme in MOGREPS-R could be increased further with stronger changes to parameters and we remark that Groenemeijer and Craig (2012) set $\sqrt{\langle r^2 \rangle}$ to 1250 m for their tests, which would likely have such an effect. We attempted only minimal tuning here and were deliberately rather conservative about the parameter choices made, with the intention that the results can reasonably be considered to represent a lower limit of the possible impact of a more thoroughly adapted scheme.

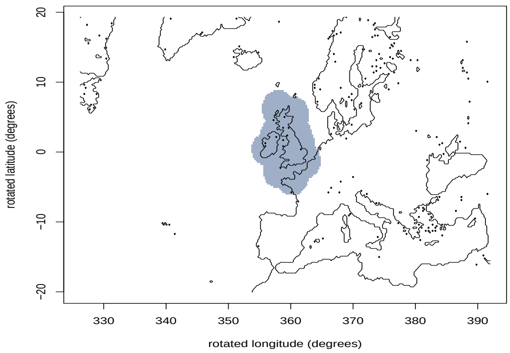

Figure 1. An outline of the MOGREPS NAE domain, with its rotated latitude-longitude grid. The contours are for reference, and are derived from the dataset used in the present study to separate the domain into land and ocean areas. The grey shading shows the region for which radar-derived precipitation data were available.

2.2 Description of MOGREPS

The Met Office Global and Regional Ensemble Prediction System (MOGREPS) has been developed to produce short-range probabilistic weather forecasts (Bowler et al., 2008). It is based on the UM (Davies et al., 2005) with 24 ensemble members, and is comprised of global and regional ensembles. In the present study, the regional ensemble MOGREPS-R was used, with a resolution of 24km and 38 vertical levels. This covers a North Atlantic and European (NAE) domain, which is shown in Figure 1. The model was run on a rotated latitude-longitude grid, with real latitude and longitude locations of the north pole and the corners of the domain given in Table 1. The regional ensemble was driven by initial and boundary conditions from the global ensemble, as described by Bowler et al. (2008). The operational system has been upgraded since these tests and so the present study represents a ‘proof of concept’ for a stochastic convection scheme in a full–complexity regional or global ensemble prediction system, rather than a detailed technical recommendation for the latest version of MOGREPS.

Table 1. Locations of the north pole and the corners of the domain of the NAE rotated grid, in terms of real latitude and longitude.

Location	latitude (°N)	longitude (°E)
north pole	37.5	177.5
bottom-left	16.3	-19.8
top-left	72.7	-80.0
bottom-right	16.5	14.2
top-right	73.2	74.1

Stochastic physics is already included in the regional MOGREPS, in the form of a random parameters scheme, where a number of selected parameters are stochastically perturbed during the forecast run (Bowler et al., 2008). This scheme was retained for the present study, given that the Plant–Craig scheme is intended to account only for the variability in the convective response for a given large-scale state, and as such its design does not conflict with the inclusion of a method to treat parameter uncertainty within other parameterization schemes. The MOGREPS random parameter scheme does introduce variability in parameters that appear within the standard UM convection scheme, which is based on the Gregory and Rowntree (1990) scheme with subsequent developments as described by Martin et al. (2006). No stochastic parameter variation is applied for any of the parameters appearing in the Plant–Craig scheme. Thus, there is no “double counting” of parameterization uncertainty in these tests but rather we are comparing different methods of accounting for convective uncertainties in a framework which also includes a simple stochastic treatment of uncertainties in other aspects of the model physics.

The forecasts using the Plant–Craig scheme were obtained by rerunning the regional version of MOGREPS, with the standard convection scheme replaced by the Plant–Craig scheme, and driven by initial and boundary conditions taken from the same archived data that were used for the operational forecasts. These are compared with the forecasts produced operationally during the corresponding period, so that the only difference between the two sets of forecasts is in the convection parameterization scheme. The study used the UM at version 7.3. The model timestep was 7.5 minutes, within which the convection scheme was called twice, and the forecast length was 54 hours.

2.3 Time period investigated

The time period investigated was from the 10th until the 30th July 2009. This length of time was chosen as being sufficient to obtain statistically meaningful results, but without requiring a more lengthy experiment that would only be justified by a more mature system. The particular month was chosen partly for convenience and partly as a period that subjectively had experienced plentiful convective rain over the UK, therefore providing a good test of a convective parameterization scheme.

Experimental forecasts with the Plant–Craig scheme were generated twice daily (at 06:00 and 18:00 UTC) for comparison with the operational forecast which was taken from the archive. On some days the archive forecast was missing and so no experimental forecast was generated. In total 34 forecasts were generated, with start times shown in Table 2.

2.4 Validation

A detailed validation was carried out against Nimrod radar rainfall data (Harrison et al., 2000; Smith et al., 2006). This observational data set is only available over the UK (as shown in Figure 1), and so most of the validation in the following focuses on this region. The forecasts were assessed on the basis of 6-hourly rainfall accumulations, every 6 hours, for lead times from 0 to 54 hours.

Table 2. Start times of forecasts investigated in this study (all dates in July 2009).

10th 18UTC	16th 18UTC	21st 06UTC	27th 18UTC
11th 06UTC	17th 06UTC	21st 18UTC	28th 06UTC
11th 18UTC	17th 18UTC	22nd 06UTC	28th 18UTC
12th 06UTC	18th 06UTC	23rd 06UTC	29th 06UTC
12th 18UTC	18th 18UTC	23rd 18UTC	29th 18UTC
13th 06UTC	19th 06UTC	24th 18UTC	30th 06UTC
14th 06UTC	19th 18UTC	25th 06UTC	30th 18UTC
15th 18UTC	20th 06UTC	25th 18UTC	
16th 06UTC	20th 18UTC	26th 06UTC	

2.4.1 Fractions skill score

This score (denoted FSS) was developed by Roberts and Lean (2008), and was used by Kober et al. (2015) to assess the quality of deterministic forecasts produced using the Plant–Craig scheme for two case studies. Note that we use the term ‘deterministic’, in this manuscript, to refer to forecasts providing a single quantity (for example, a single-member forecast, or the ensemble mean), and ‘probabilistic’ to refer to forecasts providing a probabilistic distribution (or, at the very least, a deterministic forecast, with, in addition, an assessment of its uncertainty). The FSS is determined, at a given grid point X, by comparing the fractions of observed, O, and forecast, F, grid points exceeding a specific rainfall threshold, within a specific spatial window centred at X. Here we define:

$$FSS = 1 - \frac{\langle (F-O)^2 \rangle}{\langle F^2 \rangle + \langle O^2 \rangle} \tag{1}$$

where the angled brackets $\langle \ldots \rangle$ indicate averages over the grid point centres X for which observations are available, over the different forecast initialization times, and here over the different ensemble members (so that effectively a separate score is calculated for each ensemble member and these are averaged to produce the overall score denoted here by FSS). The spatial window (over which the fractions are evaluated) gives the scale at which the score is applied, so that the FSS can be used to assess the performance of forecasts both at the grid scale and at larger scales. The division by $\langle F^2 \rangle + \langle O^2 \rangle$ normalizes against the smoothing applied at the given scale, so that the score always ranges between 0 and 1. The FSS is positively oriented.

2.4.2 Brier scores

In order to determine whether or not the variability introduced by the Plant–Craig scheme is added where it is most needed, the Brier skill score (Wilks, 2006) was applied to both forecast sets, using the same observational data, to assess the respective quality of the probabilistic forecasts. The Brier score is a threshold-based probabilistic verification score, and is given by the mean difference between the forecast probability of exceeding a given threshold (this probability is here simply taken to be the fraction of ensemble members which forecast precipitation greater than the threshold) and the observed probability (i.e 1 if the observed precipitation is above the threshold and 0 if it is below). To obtain the Brier skill score, BSS, this is compared with a reference score; the reference score is here taken to be that calculated from always forecasting a probability taken from the observation data set (i.e. the proportion of times the observed precipitation is above the threshold). Thus,

$$BSS = 1 - \frac{\langle (f-o)^2 \rangle}{\langle (\langle o \rangle - o)^2 \rangle} \tag{2}$$

where f is the forecast probability, o is the observation (0 or 1) and $\langle o \rangle$ is the 'climatological' probability based on the observation set. The angle brackets denote an average over the entire forecast set. Although $\langle o \rangle$ is only available *a posteriori* to the event, it does provide a useful 'base' for comparison: if the forecast issued is no better than one given by simply always issuing a climatological average (i.e. if $BSS \leq 0$) then the forecast can be said to have no skill.

2.4.3 Ensemble added value

This measure aims to assess the benefit of using an ensemble, as against a single forecast randomly selected from the ensemble. It was recently developed and described in detail by Ben Bouallègue (2015) and a brief outline is given here. The score is of particular interest to the present study, as this measure should highlight the advantages and disadvantages of using the stochastic Plant–Craig methodology, and provides an assessment that is less affected by structural differences between the Plant–Craig scheme and the Gregory-Rowntree (GR) scheme.

The ensemble added value (EAV) is based on the quantile score (QS) (Koenker and Machado, 1999; Gneiting, 2011), which is used to assess probabilistic forecasts at a given probability level (equivalently, the Brier score assesses probabilistic forecasts at a given value threshold). If a quantile forecast ϕ_τ of the τth quantile of a meteorological variable is given, then the quantile score for that quantile is interpreted as

$$q_\tau = \langle (\omega - \phi_\tau)(\tau - I\{\omega < \phi_\tau\}) \rangle \tag{3}$$

where ω is the observed value, the function $I(x)$ is defined as 1 if x is true and 0 if x is false and the angle brackets denote an average over all forecasts, as for the Brier skill score. In this way, a forecast for a low quantile is penalized more heavily if it is above the observed value, than if it is below the observed value, and vice-versa for a forecast for a high quantile (note that the score is negatively oriented). The score for the 50% quantile is simply the mean absolute error.

The QS can, like the Brier score, be decomposed into a reliability and a resolution component (Bentzien and Friederichs, 2014). In order to calculate the EAV, a potential QS Q_τ is defined as the total QS minus its reliability component. The QS is here evaluated by first sorting the ensemble members, and interpreting the mth sorted ensemble member as the $(m-0.5)/24$ quantile forecast.

The EAV is then given by summing the potential QSs Q_m over the 24 members, and comparing with an equivalent sum over reference potential QSs:

$$EAV = 1 - \frac{\sum_m Q_m}{\sum_m Q_m^{\mathrm{ref}}}. \tag{4}$$

The reference forecast is created by defining the quantile as simply a randomly-selected member of the ensemble, so that the reference forecast represents the score which could have been obtained with only one forecast (a single member is randomly selected, with replacement, once for the entire period, but separately for each quantile). The EAV thus measures the quality of the ensemble forecast, relative to the quality of the individual members of the ensemble.

2.5 Separation into weakly- and strongly-forced cases

Groenemeijer and Craig (2012) applied the Plant–Craig scheme in an ensemble forecasting system for seven case studies, with various synoptic conditions, and showed that the proportion of ensemble variability arising from the use of the stochastic scheme (as against that arising from variations in the initial and boundary conditions) depends on the strength of the large-scale forcing, as measured by the large-scale vorticity maximum. In particular, the stronger the large-scale forcing, the lower the proportion of the variability that comes from the stochastic scheme.

Kober et al. (2015) investigated two of the case studies further, by verifying forecasts using the Plant–Craig scheme and using a non-stochastic convection scheme. They found that the improvement in forecast quality from using the Plant–Craig scheme was significantly higher for the more weakly-forced of the two cases, since the additional grid-scale variability introduced by the stochastic scheme is more important.

As part of the present study, we extend the work of Kober et al. (2015) by separating our validation period into dates for which the synoptic forcing is relatively weak or strong. We then compare any improvement in the forecasts using the Plant–Craig scheme, over those using the Gregory-Rowntree scheme, for the two sets of forecasts, to assess over an extended period whether the benefit of using a stochastic scheme is indeed greater when the synoptic forcing is weaker.

The separation into weakly- and strongly-forced cases was carried out *a posteriori* to the event based on surface analysis charts. The aim here is not to develop an adaptive forecasting system, but rather to develop understanding of the behaviour of the Plant–Craig scheme. Nonetheless, the results may also be interpreted as providing evidence that such a system may be feasible if the strength of the synoptic forcing could be predicted in advance (using, for example, the convective adjustment time scale as discussed by Keil et al. (2014)). The period was divided into 12-hour sections, centred on 00 or 12 UTC, and a surface analysis chart valid at the respective centre-time was used to determine whether to categorize the section as weakly- or strongly-forced. The 00 UTC analyses were taken from Wetterzentrale (2009) and the 12 UTC analyses from Eden (2009).

The separation was conducted by assigning periods with discernible cyclonic and/or frontal activity over or close to the UK as strongly-forced and the rest as weakly-forced, with some additional adjustment of the preliminary categorization based on the written reports by Eden (2009). The periods were categorized as in Table 3.

Table 3. Categorization of 12-hour periods (centred at the time given) investigated in this study, into weak and strong synoptic forcing (all dates in July 2009).

10th 00UTC Weak	17th 12UTC Strong	25th 00UTC Weak
10th 12UTC Strong	18th 00UTC Strong	25th 12UTC Weak
11th 00UTC Strong	18th 12UTC Weak	26th 00UTC Strong
11th 12UTC Strong	19th 00UTC Strong	26th 12UTC Strong
12th 00UTC Strong	19th 12UTC Weak	27th 00UTC Strong
12th 12UTC Strong	20th 00UTC Weak	27th 12UTC Weak
13th 00UTC Weak	20th 12UTC Weak	28th 00UTC Strong
13th 12UTC Weak	21st 00UTC Strong	28th 12UTC Strong
14th 00UTC Strong	21st 12UTC Strong	29th 00UTC Strong
14th 12UTC Strong	22nd 00UTC Strong	29th 12UTC Strong
15th 00UTC Weak	22nd 12UTC Strong	30th 00UTC Weak
15th 12UTC Weak	23rd 00UTC Weak	30th 12UTC Weak
16th 00UTC Weak	23rd 12UTC Weak	31st 00UTC Weak
16th 12UTC Weak	24th 00UTC Weak	31st 12UTC Strong
17th 00UTC Strong	24th 12UTC Weak	

3 Results

3.1 Fractions skill score

The quality of the respective deterministic forecasts (i.e. those produced by individual ensemble members, with no supplementary indication of the forecast uncertainty) using Gregory-Rowntree (GR) and Plant–Craig (PC) is assessed using Figures 2, 3 and 4. The performance of the schemes is overall similar, with PC being superior for low thresholds (in contrast to the findings of Kober et al. (2015)) and short lead times and GR for moderate thresholds. With upscaling (Figs. 3 and 4), the performance of both schemes improves for all thresholds and lead times. The PC scheme benefits particularly from the upscaling at higher thresholds and longer lead times, sometimes performing significantly better than the GR scheme where at the grid scale the performance was equal. In general, the difference in the scores between the two schemes does not reach such high values as those seen in Kober et al. (2015), although this could be due to the fact that they investigated individual case studies which were specifically selected to test the impact of the stochastic scheme, whereas our results are scores averaged over an extended period.

In general, then, the schemes perform similarly overall, and the impact of using a stochastic scheme on the FSS is modest. Indeed, the fact that there is no skill for the highest threshold, for either scheme, is more important. This lack of skill could be simply due to the fact that the case study period was too short to obtain a statistically significant sample of extreme rain events. However, it is also true that MOGREPS significantly overforecasts heavy rain over the UK for this period (see Figure 13).

3.1.1 Separation into weakly- and strongly-forced cases

Figure 5 shows the difference in FSS between PC and GR, for forecasts separated into weakly- and strongly-forced cases, as described in Section 2. It can be seen that, with no averaging, PC is better for the smallest thresholds but worse for the moderate thresholds, while with upscaling the relative performance for moderate and higher thresholds is improved, especially for the weakly-forced cases.

PC generally performs better than GR for weakly-forced cases, and worse for strongly-forced cases. While both schemes benefit from upscaling the score, this benefit is greater for PC. The results agree well with those of Kober et al. (2015) for two example cases, where the Plant–Craig scheme benefits more from the upscaling than the non-stochastic scheme, and performs relatively better for the weakly-forced than for the strongly-forced case.

Moreover, it is clear that the upscaling is more beneficial to the PC scheme (relative to the GR scheme) for the weakly-forced cases than for the strongly-forced cases. The interpretation is that the PC scheme provides a better statistical description of small-scale, weakly-forced convection than a non-stochastic scheme. This will not provide any improvement to the FSS evaluated at the grid scale, since the convection is placed randomly, but it does improve the FSS when it is evaluated over a neighbourhood of grid points, so that it becomes a more statistical evaluation of the quality of the scheme.

3.2 Brier score

The quality of the probabilistic forecasts, with respect to forecasts using the observed climatology, is assessed using Brier skill scores, plotted in Figure 6. While neither scheme has skill for high thresholds, PC performs substantially better for medium and low thresholds, for all lead times. In particular, PC has skill in predicting whether or not rain will occur (zero threshold), while GR does not. Further analysis shows that this is also the case for thresholds between 0 and 0.05 (not shown).

The decomposition of the Brier score into reliability (Figure 7) and resolution (Figure 8) is also shown (note that the difference is taken in the opposite direction for reliability so that the colour scale must not be reversed). The Plant–Craig scheme improves both components of this score; the improvement for reliability is rather higher than that for resolution. The scores for both reliability

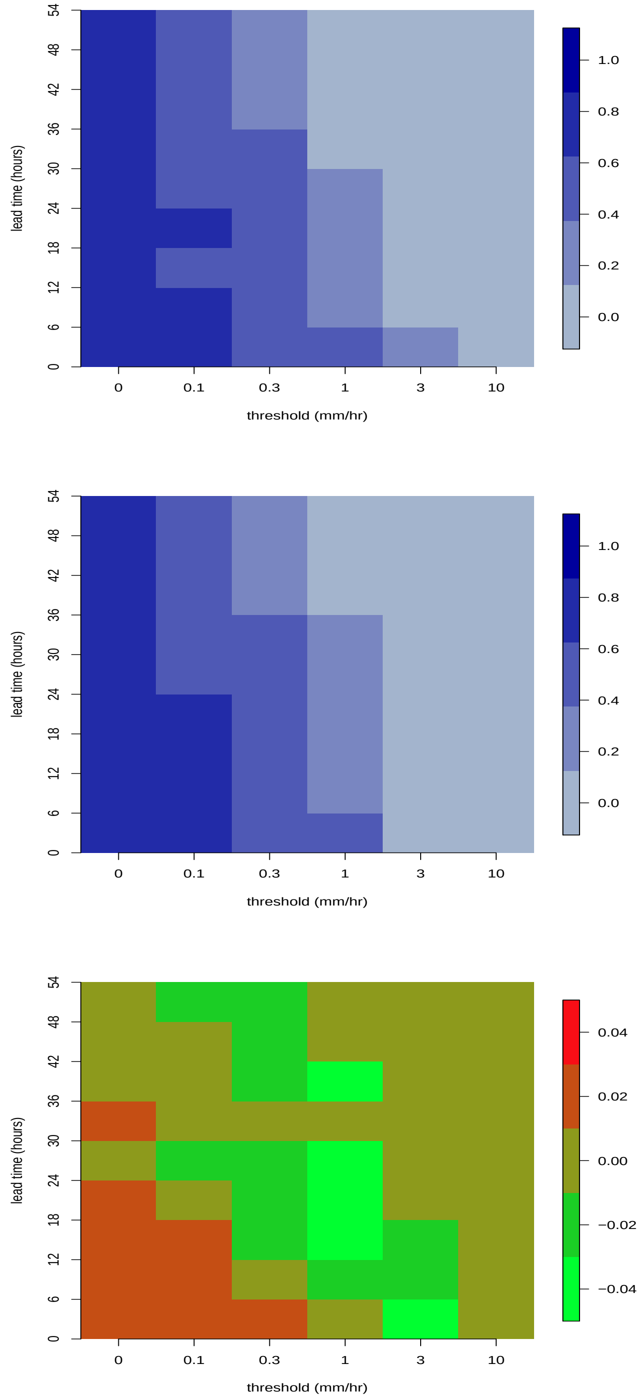

Figure 2. Fractions skill score computed for grid-scale data for the Gregory-Rowntree scheme (top), the Plant–Craig scheme (centre) and the difference between the two schemes (Plant–Craig minus Gregory Rowntree, bottom).

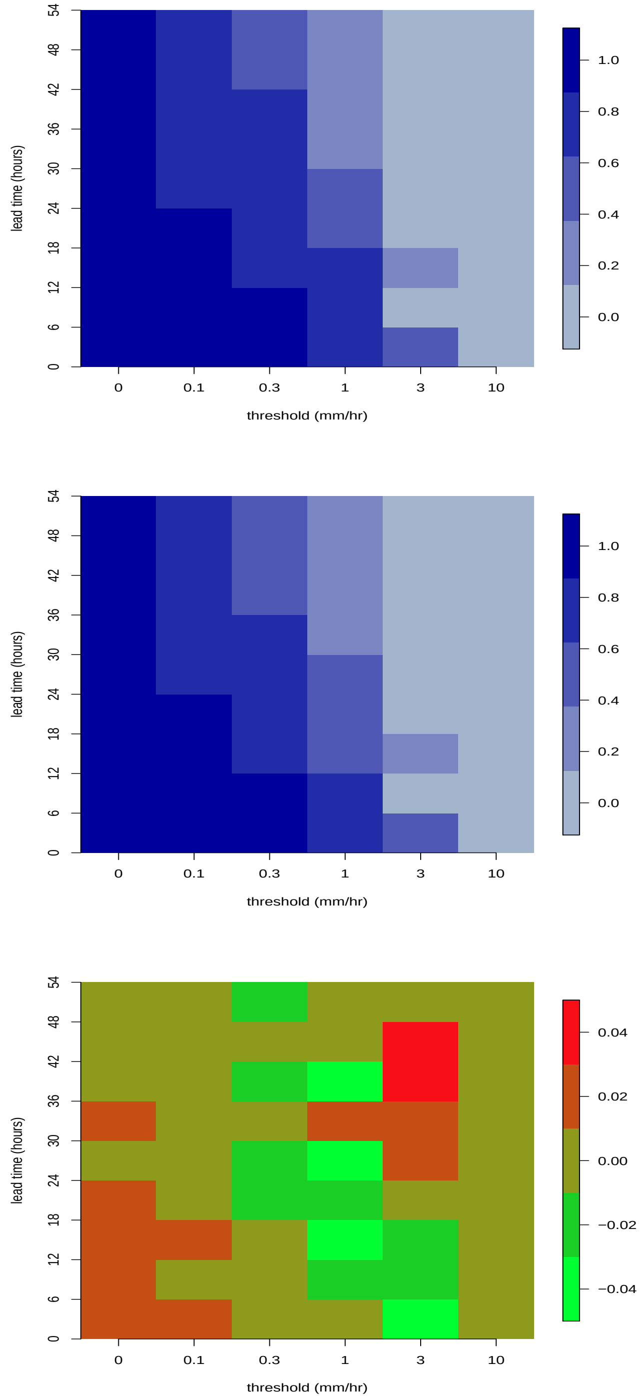

Figure 3. Fractions skill score for the Gregory-Rowntree scheme (top), the Plant–Craig scheme (centre) and the difference between the two schemes (Plant–Craig minus Gregory Rowntree, bottom). The neighbourhood area is $(120\text{km})^2$, corresponding to the central grid box and two grid boxes in each direction.

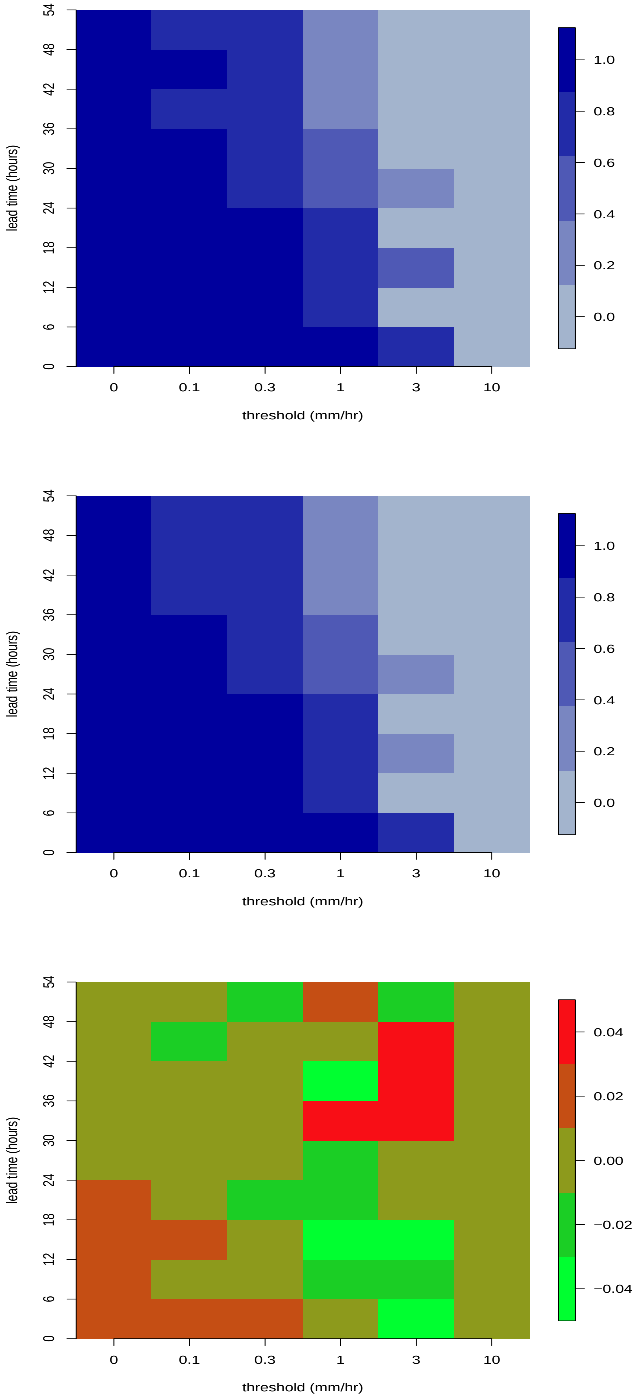

Figure 4. Fractions skill score for the Gregory-Rowntree scheme (top), the Plant–Craig scheme (centre) and the difference between the two schemes (Plant–Craig minus Gregory Rowntree, bottom). The neighbourhood area is $(216\mathrm{km})^2$, corresponding to the central grid box and four grid boxes in each direction.

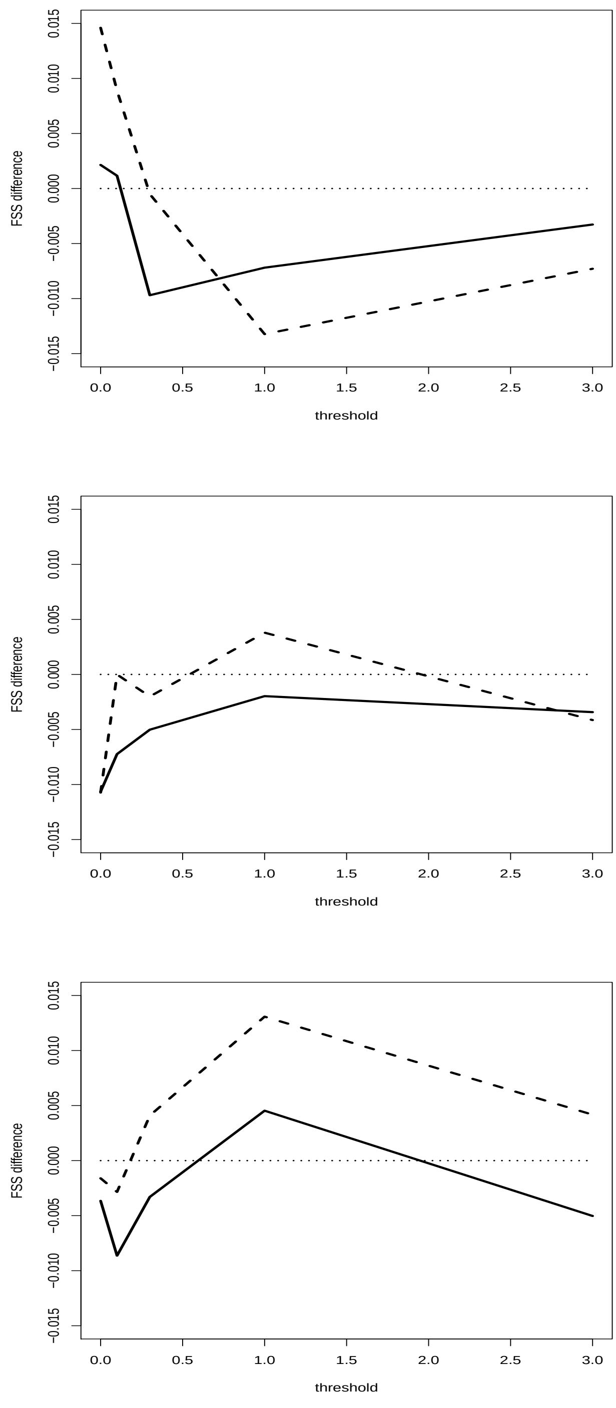

Figure 5. Fractions skill score for the Plant–Craig scheme, minus that for the Gregory-Rowntree scheme, for strongly forced cases (full lines) and weakly forced cases (dashed lines), with no averaging (top), with a neighbourhood area of two grid boxes in each direction (centre) and with a neighbourhood area of four grid boxes in each direction (bottom). The score shown is the average over all lead times.

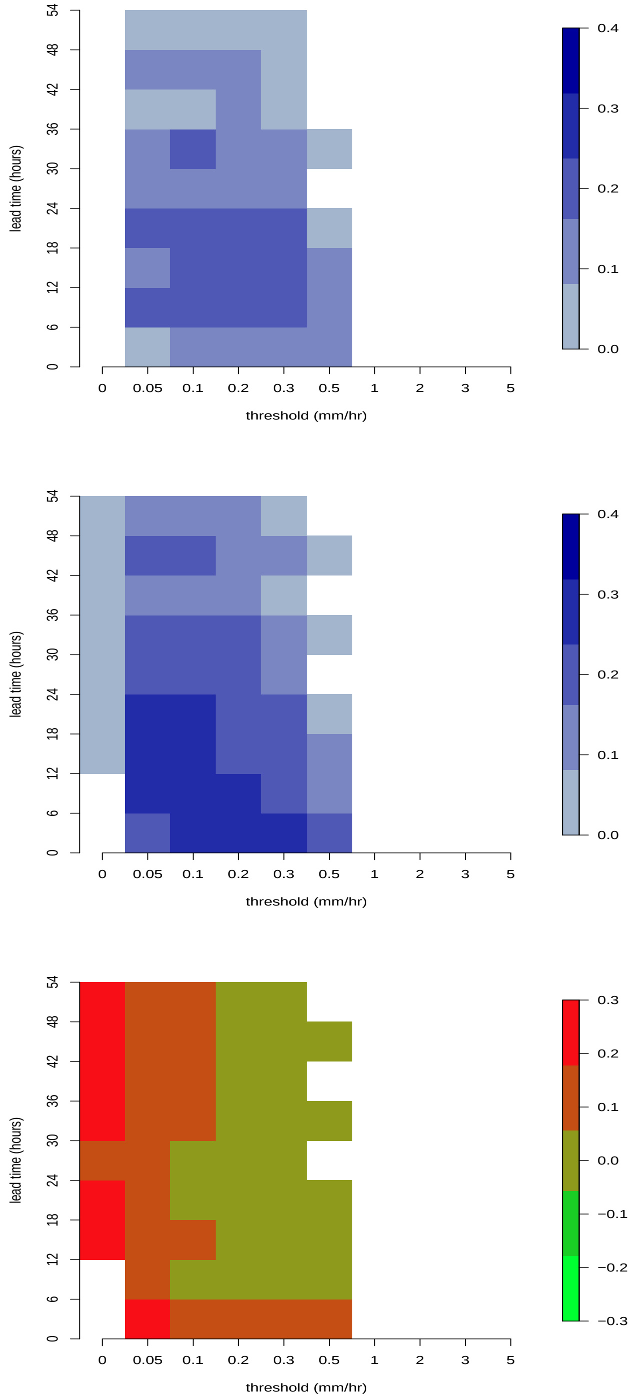

Figure 6. Brier skill score for the Gregory-Rowntree scheme (top), the Plant–Craig scheme (centre) and the difference between the two schemes (Plant–Craig minus Gregory Rowntree, bottom). For the difference plot, instances where both skill scores are lower than zero are not plotted.

and resolution are low for the higher thresholds, which is probably a consequence of the fact that there are insufficient data to assess such extreme values.

3.2.1 Separation into weakly- and strongly-forced cases

Figure 9 shows the Brier skill scores as a function of threshold, separated into strongly- and weakly-forced cases. The forecasts are improved using PC for both sets of cases, and the difference is considerably greater for weakly-forced cases, where GR has almost no skill. This can be interpreted in terms of the fact that small-scale variability is relatively more important for the weakly-forced cases, and ensemble members using the Plant–Craig scheme differ from each other more than for the strongly-forced cases, where initial and boundary condition variability is relatively more important (Groenemeijer and Craig, 2012). Our result is similar to what was found by Kober et al. (2015), where the Plant–Craig scheme was found to perform better than a non-stochastic scheme for a weakly-forced case, and at low thresholds, but worse than the non-stochastic Tiedtke, M. (1989) scheme for a strongly-forced case.

3.3 Ensemble added value (EAV)

The EAV is plotted in Figure 10. The PC scheme performs substantially better for this score across lead times, and the improvement is of a similar magnitude to that of the Brier score. This suggests that the improvement in the probabilistic forecast from using PC comes from the stochasticity of the scheme, since the EAV is measured against individual forecasts from the same ensemble: it should, therefore, be 'normalized' against differences in the underlying convection scheme which are not related to the stochasticity. The interpretation here is that while structural differences between two convection schemes will lead to differences in the quality of the ensemble forecasts, this will mainly be due to differences in the quality of individual members of the ensemble. The stochastic character of the PC scheme may or may not improve the quality of the individual members, but it is primarily designed to improve the quality of the ensemble as a whole.

Note that the ensemble forecasts using the GR scheme also have a positive EAV, representing the value added by the multiple initial and boundary conditions provided by the global model, and by the stochasticity coming from the random parameters scheme. Since these factors are also present in the ensemble forecasts using the PC scheme, it can be interpreted that the fractional difference between the two EAVs represents the value added by the stochastic character of the PC scheme as a fraction of the value added by all the ensemble generation techniques in MOGREPS.

3.4 General climatology

Although Nimrod radar observations were only available over a restricted part of the forecast domain, it is also of interest to compare the forecasts over the whole domain. Figure 11 shows the convective fraction: that is, the amount of rainfall which came from the convection scheme divided

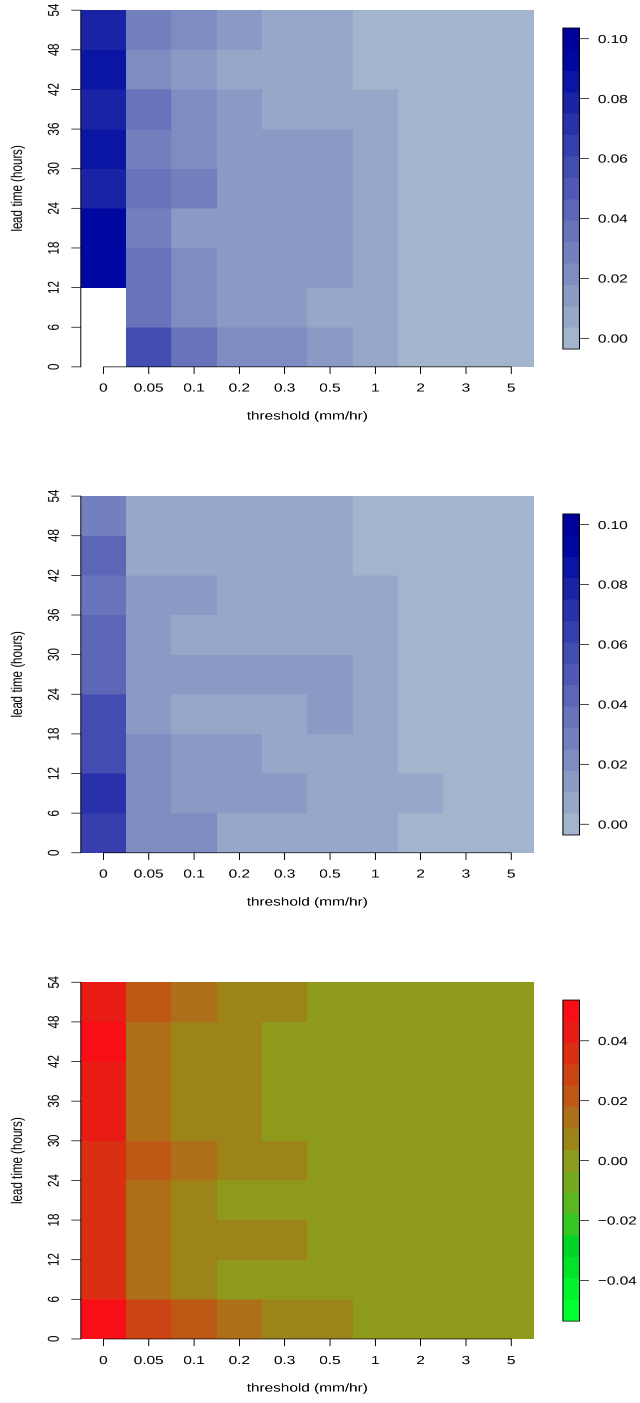

Figure 7. Brier score reliability for the Gregory-Rowntree scheme (top), the Plant–Craig scheme (centre) and the difference between the two schemes (Gregory Rowntree minus Plant–Craig, bottom).

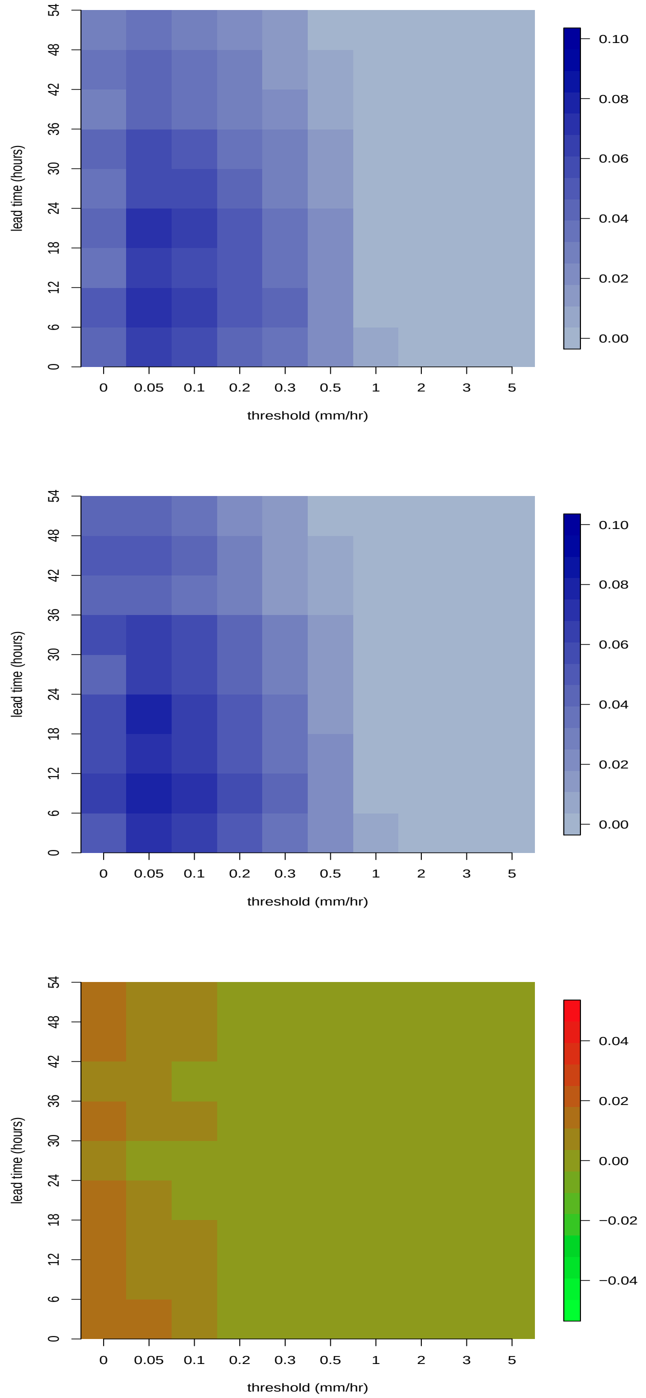

Figure 8. Brier score resolution for the Gregory-Rowntree scheme (top), the Plant–Craig scheme (centre) and the difference between the two schemes (Plant–Craig minus Gregory Rowntree, bottom).

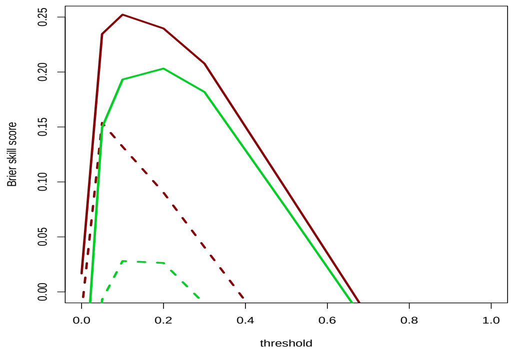

Figure 9. Brier skill score for the Gregory-Rowntree scheme (green lines) and the Plant–Craig scheme (red lines), averaged over all lead times, for cases with strong forcing (full lines) and weak forcing (dashed lines), as a function of threshold. The reference for the skill score is the observed climatology. The axes have been chosen to focus on where the skill score is above zero.

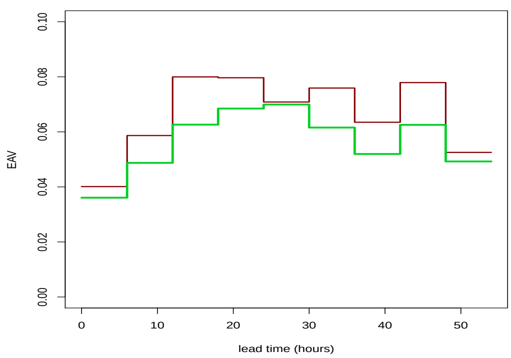

Figure 10. Ensemble added value (EAV) for the Gregory-Rowntree scheme (green line) and the Plant–Craig scheme (red line) as a function of forecast lead time.

by the total amount of rain from the convection scheme and grid-scale precipitation. Both schemes produce more convective rain over land, and the difference between the fractions over land and sea is in proportion to the fraction over the whole domain; the fractions are fairly constant with forecast lead time. As discussed in Sec. 2.1, the convective fraction is much lower for PC than for GR, suggesting that adjusting parameters to increase this fraction would further increase the PC influence on the forecast (for example, Groenemeijer and Craig (2012) used a reduced closure time scale to increase the activity of the PC scheme). It is important, however, to reiterate that the total amount of rainfall produced (including from the grid-scale dynamics) did not vary significantly, and that there is no correct value for the convective fraction.

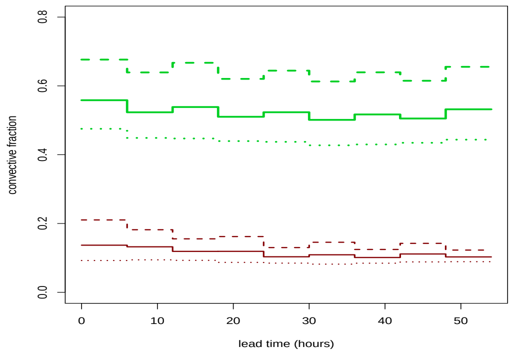

Figure 11. Convective fraction as a function of forecast lead time, for the Gregory-Rowntree scheme (green lines) and the Plant–Craig scheme (red lines), over land (dashed lines), over ocean (dotted lines) and in total (full lines), for the full NAE domain.

The ensemble spread is shown as a function of lead time in Figure 12, over the whole domain and separately over land and over ocean. Both schemes produce more spread over land, but the difference between PC and GR is also much greater over land. This is presumably due to the fact that PC has a higher convective fraction over land, and is therefore more able to influence the spread. The spread increases with forecast lead time, and does so more quickly with PC than with GR.

Figure 13 shows density plots of rainfall from the two schemes, and from the observations, over the UK part of the domain, for a lead time of 30 to 36 hours. It is clear that the model produces too many instances of heavy rainfall for this period, and that this is exacerbated by the extra variability introduced by the PC scheme. However, as shown earlier in this Section, neither scheme has any skill for large thresholds. It is clear from Figure 13 that this is partly due to over-production of heavy rain, although it is also the case that the case study was of insufficient length to fully assess such extreme values.

Figure 14 shows that the PC scheme also produces more heavy rainfall than the GR scheme over ocean (here for a lead time of 30 to 36 hours). This suggests that one possible approach to tuning the PC scheme could be to apply less input averaging over the ocean, since Keane et al. (2014) have shown that applying more input averaging increases the variability and, therefore, the tails of the distribution.

Although a lead time of 30 to 36 hours was chosen for Figures 13 and 14, similar conclusions could be drawn for the plots for other lead times (not shown). The exception to this statement is that for the first 6 hours, for which the forecasts had not developed sufficiently for the curves to lie significantly apart from each other.

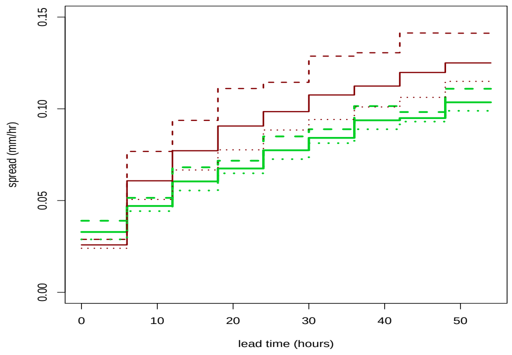

Figure 12. Ensemble spread as a function of forecast lead time, for the Gregory-Rowntree scheme (green lines) and the Plant–Craig scheme (red lines), over land (dashed lines), over ocean (dotted lines) and in total (full lines), for the full NAE domain.

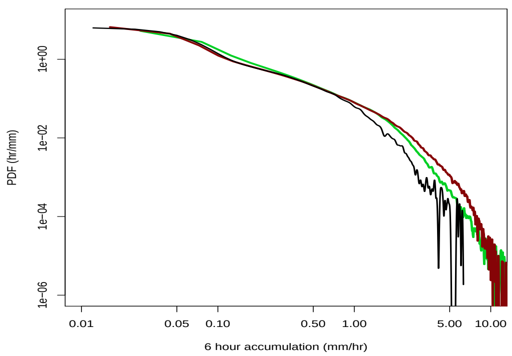

Figure 13. Density plots for accumulated rainfall for the period of 30 to 36 hours lead time, over the UK part of the domain, for forecasts with the Gregory-Rowntree scheme (green line), the Plant–Craig scheme (red line) and observations (black line).

3.4.1 Validation over the whole NAE domain

A validation using the routine verification system was also performed for the two setups, covering land areas over the whole forecast domain. This calculates various forecast skill scores, by comparing against SYNOP observations at the surface and at a height of 850 hPa, and yielded a mixed assessment of the performance of the PC scheme against the GR scheme. For example, the continuous ranked probability score, which assesses both the forecast error and how well the ensemble spread predicts the error (Hersbach, 2000), was improved by roughly 10% on using the PC scheme for rainfall, but degraded by about 10% for temperature and pressure. The impact on the wind forecast was broadly neutral.

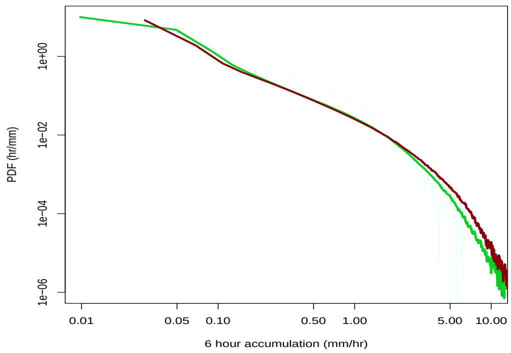

Figure 14. Density plots for accumulated rainfall for the period of 30 to 36 hours lead time, over the entire NAE domain, for forecasts with the Gregory-Rowntree scheme (green line) and the Plant–Craig scheme (red line) over ocean.

This shows that, while the improvements demonstrated in this Section hold for other areas outside the UK, this has come at a cost to the quality of the forecast for some of the other variables. An important advantage of using a stochastic convection scheme, over a statistical downscaling procedure, is its feedback on the rest of the model, and it is important that this feedback is of benefit. The recent analysis by Selz and Craig (2015a) is very encouraging in this regard, demonstrating the processes of upscale error growth from convective uncertainties can be well reproduced by the PC scheme, in good agreement with the behaviour of large-domain simulations in which the convection is simulated explicitly (Selz and Craig, 2015b).

4 Conclusions

A physically-based stochastic scheme for the parameterization of deep convection has been evaluated by comparing probabilistic rainfall forecasts produced using the scheme in an operational ensemble system with those from the same ensemble system with its standard deep convection parameterization. The impact of using a stochastic scheme on deterministic forecasts is broadly neutral, although there is some improvement when larger areas are assessed. This is relevant to applications such as hydrology, where rainfall over an area larger than a grid box can be more relevant than rainfall on the grid box scale.

The Plant–Craig scheme has been shown to have a positive impact on probabilistic forecasts for light and medium rainfall, while neither scheme is able to skillfully forecast heavy rainfall. The impact of the scheme is greater for weakly-forced cases, where subgrid-scale variability is more important. Keil et al. (2014) studied a convection-permitting ensemble without stochastic physics, and found that deterministic forecast skill was poorer during weak than during strong forcing conditions. They developed a convective adjustment time-scale to measure the strength of the forcing conditions. This quantity can be calculated from model variables and could therefore be used in advance to determine how predictable the convective response will be for a given forecast. This could potentially be useful in an adaptive ensemble system using two convection parameterizations (see, for example, Marsigli et al. (2005)), one of which is stochastic and is better suited to providing an estimate of the uncertainty in weaker forcing cases.

Although the Plant–Craig scheme clearly produces improved probabilistic forecasts, it is not certain whether this is due to its stochasticity, or to different underlying assumptions between it and the standard convection scheme. In order to make a clean distinction, further studies could be performed in which the performance of the Plant–Craig scheme is compared against its own non-stochastic counterpart, which can be constructed by using the full cloud distribution and appropriately normalizing, instead of sampling randomly from it (cf Keane et al., 2014). Nonetheless, the results from applying the recently-developed ensemble added value metric do provide some relevant information for this question. This metric aims to assess the quality of the ensemble in relation to the underlying member forecasts, and the Plant–Craig scheme has been shown to increase it. This indicates that the stochastic aspect of the scheme can increase the value added to a forecast by using an ensemble, since other aspects of the scheme would be expected (broadly) to affect the performance of the ensemble as a whole, and of the individual members, equally.

The results of this study justify further work to investigate the impact of the Plant–Craig scheme on ensemble forecasts. Since the version of MOGREPS used in this study has been superseded, it is not feasible to carry out more a more detailed investigation beyond the proof-of-concept carried out in the present study. Interestingly, the resolution used in this study is now becoming more widely used in global ensemble forecasting, and so future work could involve implementing the scheme in a global NWP system, for example the global version of MOGREPS. This would enable assessments to be made as to whether the scheme provides benefits for the representation of tropical convection, in addition to those aspects of mid-latitude convection that were demonstrated here.

5 Code and/or data availability

The source code for the Plant–Craig parameterization, as it was used in this study, can be made available on request, by contacting r.s.plant@reading.ac.uk.

Acknowledgements. We would like to thank Neill Bowler for helping to plan and set up the numerical experiments, and Rod Smyth for helping to set up preliminary experiments on MONSOON.

References

Abhilash, S., Sahai, A. K., Pattnaik, S., Goswami, B. N., and Kumar, A.: Extended range prediction of active-break spells of Indian summer monsoon rainfall using an ensemble prediction system in NCEP Climate Forecast System, International Journal of Climatology, pp. n/a–n/a, doi:10.1002/joc.3668, http://dx.doi.org/10.1002/joc.3668, 2013.

Ball, M. A. and Plant, R. S.: Comparison of stochastic parameterization approaches in a single-column model, Phil. Trans. Roy. Soc. A, 366, 2605–2623, 2008.

Bechtold, P.: Convection in global numerical weather prediction, in: Parameterization of Atmospheric Convection. Volume 2: Current Issues and New Theories, edited by Plant, R. S. and Yano, J.-I., chap. 15, pp. 5–45, World Scientific, Imperial College Press, 2015.

Ben Bouallègue, Z.: Assessment and added value estimation of an ensemble approach with a focus on global radiation, Mausam: Quarterly Journal of Meteorology, Hydrology & Geophysics, 66, 541–550, 2015.

Bengtsson, L., Steinheimer, M., Bechtold, P., and Geleyn, J.-F.: A stochastic parametrization for deep convection using cellular automata, Quarterly Journal of the Royal Meteorological Society, 139, 1533–1543, doi:10.1002/qj.2108, http://dx.doi.org/10.1002/qj.2108, 2013.

Bentzien, S. and Friederichs, P.: Decomposition and graphical portrayal of the quantile score, Quarterly Journal of the Royal Meteorological Society, 140, 1924–1934, doi:10.1002/qj.2284, http://dx.doi.org/10.1002/qj.2284, 2014.

Berner, J., Ha, S.-Y., Hacker, J. P., Fournier, A., and Snyder, C.: Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations, Monthly Weather Rev., 139, 1972–1995, doi:http://dx.doi.org/10.1175/2010MWR3595.1, 2011.

Bouttier, F., Vié, B., Nuissier, O., and Raynaud, L.: Impact of Stochastic Physics in a Convection-Permitting Ensemble, Monthly Weather Rev., 140, 3706–3721, doi:http://dx.doi.org/10.1175/MWR-D-12-00031.1, 2012.

Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B., and Beare, S. E.: The MOGREPS short-range ensemble prediction system, Q. J. R. Meteorol. Soc., 134, 703–722, 2008.

Buizza, R., Miller, M., and Palmer, T. N.: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, Q. J. R. Meteorol. Soc., 125, 2887–2908, 1999.

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y.: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems, Monthly Weather Rev., 133, 1076–1097, 2005.

Buizza, R., Bidlot, J.-R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., and Vitart, F.: The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System), Quarterly Journal of the Royal Meteorological Society, 133, 681–695, doi:10.1002/qj.75, http://dx.doi.org/10.1002/qj.75, 2007.

Christensen, H. M., Moroz, I. M., and Palmer, T. N.: Stochastic and Perturbed Parameter Representations of Model Uncertainty in Convection Parameterization, Journal of the Atmospheric Sciences, 72, 2525–2544, doi:10.1175/JAS-D-14-0250.1, http://dx.doi.org/10.1175/JAS-D-14-0250.1, 2015.

Clark, A. J., Kain, J. S., Stensrud, D. J., Xue, M., Kong, F., Coniglio, M. C., Thomas, K. W., Wang, Y., Brewster, K., Gao, J., Wang, X., Weiss, S. J., and Du, J.: Probabilistic Precipitation Forecast Skill as a Function of Ensemble Size and Spatial Scale in a Convection-Allowing Ensemble, Monthly Weather Rev., 139, 1410–1418, doi:http://dx.doi.org/10.1175/2010MWR3624.1, 2011.

Davies, L., Jakob, C., Cheung, K., Genio, A. D., Hill, A., Hume, T., Keane, R. J., Komori, T., Larson, V. E., Lin, Y., Liu, X., Nielsen, B. J., Petch, J., Plant, R. S., Singh, M. S., Shi, X., Song, X., Wang, W., Whitall, M. A., Wolf, A., Xie, S., and Zhang, G.: A single-column model ensemble approach applied to the TWP-ICE experiment, J. Geophys. Res., 118, 6544–6563, 2013.

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N.: A new dynamical core for the Met Office's global and regional modelling of the atmosphere, Q. J. R. Meteorol. Soc., 131, 1759–1782, 2005.

Ebert, E. E., ulrich Damrath, Wergen, W., and Baldwin, M. E.: The WGNE Assessment of Short-term Quantitative Precipitation Forecasts, Bull. Am. Meteorol. Soc., 84, 481–492, doi:http://dx.doi.org/10.1175/BAMS-84-4-481, 2003.

Eden, P.: July 2009 A hot start, then very unsettled with several heavy falls of rain, Weather, 64, i–iv, doi:10.1002/wea.496, http://dx.doi.org/10.1002/wea.496, 2009.

Gebhardt, C., Theis, S., Paulat, M., and Bouallègue, Z. B.: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries, Atmospheric Research, 100, 168 – 177, doi:http://dx.doi.org/10.1016/j.atmosres.2010.12.008, http://www.sciencedirect.com/science/article/pii/S0169809510003455, <ce:title>Uncertainty Propagation in Advanced Hydro-Meteorological Forecast Systems</ce:title>, 2011.

Gneiting, T.: Making and Evaluating Point Forecasts, Journal of the American Statistical Association, 106, 746–762, doi:10.1198/jasa.2011.r10138, http://dx.doi.org/10.1198/jasa.2011.r10138, 2011.

Gregory, D. and Rowntree, P. R.: A Mass Flux Convection Scheme with Representation of Cloud Ensemble Characteristics and Stability-Dependent Closure, Monthly Weather Rev., 118, 1483–1506, 1990.

Groenemeijer, P. and Craig, G. C.: Ensemble forecasting with a stochastic convective parametrization based on equilibrium statistics, Atmospheric Chemistry and Physics, 12, 4555–4565, doi:10.5194/acp-12-4555-2012, http://www.atmos-chem-phys.net/12/4555/2012/, 2012.

Harrison, D. L., Driscoll, S. J., and Kitchen, M.: Improving precipitation estimates from weather radar using quality control and correction techniques, Meteorological Applications, 7, 135–144, doi:10.1017/S1350482700001468, http://dx.doi.org/10.1017/S1350482700001468, 2000.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Wea. Forecasting, 15, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Kain, J. S.: The Kain-Fritsch convective parameterization: An update, J. Appl. Meteor., 43, 170–181, 2004.

Kain, J. S. and Fritsch, J. M.: A One-Dimensional Entraining / Detraining Plume Model and Its Application in Convective Parameterization, J. Atmos. Sci., 47, 2784–2802, 1990.

Keane, R. J. and Plant, R. S.: Large-scale length and time-scales for use with stochastic convective parametrization, Quarterly Journal of the Royal Meteorological Society, 138, 1150–1164, doi:10.1002/qj.992, http://dx.doi.org/10.1002/qj.992, 2012.

Keane, R. J., Craig, G. C., Zängl, G., and Keil, C.: The Plant-Craig stochastic convection scheme in ICON and its scale adaptivity, J. Atmos. Sci., doi:10.1175/JAS-D-13-0331.1, 2014.

Keil, C., Heinlein, F., and Craig, G. C.: The convective adjustment time-scale as indicator of predictability of convective precipitation, Quarterly Journal of the Royal Meteorological Society, 140, 480–490, doi:10.1002/qj.2143, http://dx.doi.org/10.1002/qj.2143, 2014.

Khouider, B., Biello, J., and Majda, A. J.: A stochastic multicloud model for tropical convection, Comm. Math. Sci., 8, 187–216, 2010.

Kober, K., Foerster, A. M., and Craig, G. C.: Evaluation of a stochastic and deterministic convection parameterization in the COSMO model, Monthly Weather Rev., doi:10.1175/MWR-D-15-0012.1, 2015.

Koenker, R. and Machado, J. A. F.: Goodness of Fit and Related Inference Processes for Quantile Regression, Journal of the American Statistical Association, 94, 1296–1310, doi:10.1080/01621459.1999.10473882, http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1999.10473882, 1999.

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., and Halliwell, C.: Characteristics of High-Resolution Versions of the Met Office Unified Model for Forecasting Convection over the United Kingdom, Monthly Weather Rev., 136, 3408–3424, doi:http://dx.doi.org/10.1175/2008MWR2332.1, 2008.

Lin, J. W.-B. and Neelin, J. D.: Toward stochastic deep convective parameterization in general circulation models, Geophysical Research Letters, 30, doi:10.1029/2002GL016203, 2003.

Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T.: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, Nonlinear Processes in Geophysics, 12, 527–536, doi:10.5194/npg-12-527-2005, http://www.nonlin-processes-geophys.net/12/527/2005/, 2005.

Martin, G. M., Ringer, M. A., Pope, V. D., Jones, A., Dearden, C., and Hinton, T. J.: The Physical Properties of the Atmosphere in the New Hadley Centre Global Environmental Model (HadGEM1). Part I: Model Description and Global Climatology, J. Climate, 19, 1274–1301, doi:http://dx.doi.org/10.1175/JCLI3636.1, 2006.

Mishra, A. and Krishnamurti, T.: Current status of multimodel superensemble and operational NWP forecast of the Indian summer monsoon, Journal of Earth System Science, 116, 369–384, doi:10.1007/s12040-007-0037-z, http://dx.doi.org/10.1007/s12040-007-0037-z, 2007.

Montani, A., Cesari, D., Marsigli, C., and Paccagnella, T.: Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges, Tellus A, 63, 605–624, doi:10.1111/j.1600-0870.2010.00499.x, http://dx.doi.org/10.1111/j.1600-0870.2010.00499.x, 2011.

Plant, R. S. and Craig, G. C.: A Stochastic Parameterization for Deep Convection Based on Equilibrium Statistics, J. Atmos. Sci., 65, 87–105, 2008.

Plant, R. S., Bengtsson, L., and Whitall, M. A.: Stochastic aspects of convective parameterization, in: Parameterization of Atmospheric Convection. Volume 2: Current Issues and New Theories, edited by Plant, R. S. and Yano, J.-I., chap. 20, pp. 135–172, World Scientific, Imperial College Press, 2015.

Ragone, F., Fraedrich, K., Borth, H., and Lunkeit, F.: Coupling a minimal stochastic lattice gas model of a cloud system to an atmospheric general circulation model, Quarterly Journal of the Royal Meteorological Society, doi:10.1002/qj.2331, 2014.

Roberts, N. M. and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, Monthly Weather Review, 136, 78–97, http://centaur.reading.ac.uk/31220/, 2008.

Roy Bhowmik, S. K. and Durai, V. R.: Multi-model ensemble forecasting of rainfall over Indian monsoon region, Atmósfera, 21, 225–239, 2008.

Selz, T. and Craig, G. C.: Simulation of upscale error growth with a stochastic convection scheme, Geophysical Research Letters, 42, 3056–3062, doi:10.1002/2015GL063525, 2015a.

Selz, T. and Craig, G. C.: Upscale Error Growth in a High-Resolution Simulation of a Summertime Weather Event over Europe, Mon. Wea. Rev., 143, 813–827, 2015b.

Smith, R. N. B., Blyth, E. M., Finch, J. W., Goodchild, S., Hall, R. L., and Madry, S.: Soil state and surface hydrology diagnosis based on MOSES in the Met Office Nimrod nowcasting system, Meteorological Applications, 13, 89–109, doi:10.1017/S1350482705002069, http://dx.doi.org/10.1017/S1350482705002069, 2006.

Teixeira, J. and Reynolds, C. A.: Stochastic Nature of Physical Parameterizations in Ensemble Prediction: A Stochastic Convection Approach, Monthly Weather Review, 136, 483–496, doi:10.1175/2007MWR1870.1, http://dx.doi.org/10.1175/2007MWR1870.1, 2008.

Tennant, W. and Beare, S.: New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS, Quarterly Journal of the Royal Meteorological Society, pp. n/a–n/a, doi:10.1002/qj.2202, http://dx.doi.org/10.1002/qj.2202, 2013.

Thirel, G., Regimbeau, F., Martin, E., Noilhan, J., and Habets, F.: Short- and medium-range hydrological ensemble forecasts over France, Atmospheric Science Letters, 11, 72–77, doi:10.1002/asl.254, http://dx.doi.org/10.1002/asl.254, 2010.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Monthly Weather Review, 117, 1779–1800, doi:http://dx.doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Wetterzentrale: UK Met Office Surface Analysis, http://www.wetterzentrale.de/topkarten/fsfaxbra.html, accessed: 12th November 2014, 2009.

Wilks, D.: Statistical Methods in the Atmospheric Sciences: An Introduction, International Geophysics Series, Elsevier Academic Press, http://books.google.de/books?id=_vSwyt8_OGEC, 2006.

Yang, C., Yan, Z., and Shao, Y.: Probabilistic precipitation forecasting based on ensemble output using generalized additive models and Bayesian model averaging, Acta Meteorologica Sinica, 26, 1–12, doi:10.1007/s13351-012-0101-8, http://dx.doi.org/10.1007/s13351-012-0101-8, 2012.

Zhu, Y.: Ensemble forecast: A new approach to uncertainty and predictability, Advances in Atmospheric Sciences, 22, 781–788, doi:10.1007/BF02918678, http://dx.doi.org/10.1007/BF02918678, 2005.