# A simple ensemble approach for more robust process-based sensitivity analysis of case studies in convection-permitting models

Article

Accepted Version

the End User Agreement.

# www.reading.ac.uk/centaur

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online
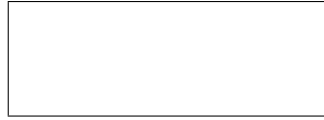
# A Simple Ensemble Approach for More Robust Process-Based Sensitivity Analysis of Case Studies in Convection-Permitting Models

David L. A. Flack*, Suzanne L. Gray and Robert S. Plant

*Department of Meteorology, University of Reading, UK*

*Correspondence to: D. L. A. Flack, LMD-ENS, 24 rue Lhomond, 75231 Paris Cedex 05, France. E-mail: dflack@lmd.ens.fr

Current Affiliation: LMD/IPSL, Département de Géosciences, ENS, PSL Research University, Ecole Polytechnique, Université Paris Saclay, Sorbonne Universités, UPMC Univ Paris 06, CNRS, Paris, France

**Case studies remain an important method for meteorological parameter sensitivity process studies. However, these types of study often use just a few case studies (typically up to three) and are not tested for statistical significance. This approach can be problematic at the convective scales, since uncertainty in the representation of an event increases, and the variability in the atmosphere arising from convective-scale noise is not routinely taken into account. Here we propose a simple ensemble method for performing more robust sensitivity analysis without the need for an operational-style ensemble prediction system and demonstrate it using a case study from the 2005 Convective Storm Initiation Project. Boundary layer stochastic potential temperature perturbations with Gaussian spatial structure are used to create small ensembles to examine the impact of increasing cloud droplet number concentration (CDNC) on precipitation. Whilst there is a systematic difference between the experiments, such that increasing the CDNC reduces the precipitation, there is also an overlap between the different ensembles implying that convective-scale variability should be taken into account in case study process-based sensitivity studies.**

## 1.   Introduction

Case studies remain an important method of understanding how models represent meteorological processes and the sensitivities of these processes to certain parameters (e.g. Thompson *et al.* 2004; Zampieri *et al.* 2005; Mahoney and Lackmann 2006; Adams-Selin *et al.* 2013; Barthlott *et al.* 2017). However, these types of sensitivity studies are often performed using single deterministic forecasts and consider just a few case studies [e.g. Mahoney and Lackmann (2006) and Gilmore *et al.* (2004) considered one and three cases, respectively]. The statistical significance of the differences between runs are not considered due to the small sample and furthermore such studies often use a small number of variants of the parameter being tested (e.g. Zhang and Anthes 1982; Gilmore *et al.* 2004; Oleson *et al.* 2008; Planche *et al.* 2015).

The robustness of such sensitivity experiments has been considered though the use of multiple resolutions in the vertical or horizontal (e.g. Sun and Bosilovich 1996; Planche *et al.* 2015); varying the domain size (e.g. Barthlott *et al.* 2017); or by using multiple models (e.g. Derbyshire *et al.* 2004). However, it may be that in these experiments sensitivities exhibited to model configurations are larger than the sensitivities exhibited to the parameter being examined. In particular, such model configuration changes can modify the timing and location of the weather event of interest (or even prevent it occurring). For example Done *et al.* (2006) demonstrate that a mesoscale convective system simulated with explicit convection can have a very different structure (and associated precipitation rates) to that simulated using parametrized convection, even at the same resolution, and Clarke *et al.* (2019) demonstrate that mis-representation of mesoscale convective system structure by convective parametrization schemes can impact downstream forecasts. Given the chaotic and indeterminate nature of the atmosphere, the possibility for numerical artefacts (e.g. Ancell *et al.* 2018) and its reduced predictability at convective scales relative to synoptic scales (e.g. Hartmann *et al.* 1995; Zhang *et al.* 2003; Hohenegger and Schär 2007), it is not obvious that such process-based case study sensitivity tests can yield robust results, particularly when using convection-permitting models.

Operationally, the variability in the atmosphere is addressed through the use of ensemble prediction systems (e.g. Buizza and Palmer 1995; Toth and Kalnay 1997; Bowler *et al.* 2008, 2009). Such ensembles are used to indicate the multiple possible outcomes of the forecasts by including representations of errors in the initial and lateral boundary conditions, as well as in the model itself. Ensemble systems designed for use at convective scales are described by Baldauf *et al.* (2011), Seity *et al.* (2011) and Hagelin *et al.* (2017) for example, and have been shown to add value over deterministic forecasts, particularly for events that have low predictability such as fog formation (Price *et al.* 2015; McCabe *et al.* 2016).

The inclusion of convective-scale noise in operational versions of such ensemble systems is becoming an increasingly common strategy: e.g. McCabe *et al.* (2016) describe the impact of the Met Office random parameter scheme adapted for use in the convection-permitting ensemble prediction system in which ten parameters, five from each of the microphysics and boundary layer parametrization schemes, are perturbed at regular intervals throughout the forecast to represent convective-scale uncertainty. However, in the context of parameter sensitivity tests, it may be impractical to run full operational ensembles due to availability of access and due to the number of members that would be required to be run to ensure that the conclusions of sensitivity to the parameter changes are robust given the other forms of uncertainty taken into account by the operational ensemble. We do not use an operational ensemble for these tests as, although the limited predictability of convective-scale motions is an important consideration, the uncertainties in the initial and boundary conditions are not of concern in order to establish whether there is sensitivity to the given parameter. Indeed, these uncertainties would often dominate in impact over that due to the changes in parameter both in magnitude and spatial scale. Rather, the purpose of using ensemble forecasts in parameter sensitivity studies is to facilitate the distinction between consistent tendencies driven by a varying a parameter from nonlinear effects introduced when varying model resolution or configuration. The aim of this study is to demonstrate a simple and computationally-cheap ensemble approach suitable for testing parameter sensitivity in process-based

case studies that yields improved robustness compared to small ensembles generated through configuration changes. Stochastic noise indicative of unresolved boundary-layer fluctuations is added on a spatial scale that is near the grid scale and of magnitude typical of unresolved fluctuations at this scale. These perturbations are used to create a small ensemble of possible realisations to distinguish whether variations in parameters yield systematic differences in the simulation outcome. It is anticipated that this ensemble technique could provide an accessible means for other research groups to improve the robustness of their sensitivity tests without the need to embed their tests within a full operational-style ensemble.

Specifically we consider simple ensemble simulations for sensitivity tests to cloud droplet number concentration (CDNC) for a case study from the Convective Storm Initiation project (CSIP; Browning *et al.* 2007). The case corresponds to Intensive Observation Period 8 (IOP 8) and was previously considered by Planche *et al.* (2015) who were also interested in the sensitivity of the convective showers to CDNC. Those authors found that increasing the CDNC by a factor of three led to a decrease in the average surface precipitation by around one sixth. Planche *et al.* (2015) used the Met Office Unified Model (MetUM) for their simulations and, to build confidence in their conclusion, ran the model with several different grid spacings. Here we revisit the case as a test of concept for our simple ensemble sensitivity method and further pursue the analysis to provide insights into the changed characteristics of the convection that may be associated with reduced precipitation.

The paper is organised as follows. The case study is described in Section 2, followed by details of the model and perturbation strategy in Section 3. Diagnostics are presented in Section 4 and the results are presented and discussed in Section 5. A summary is provided in Section 6.
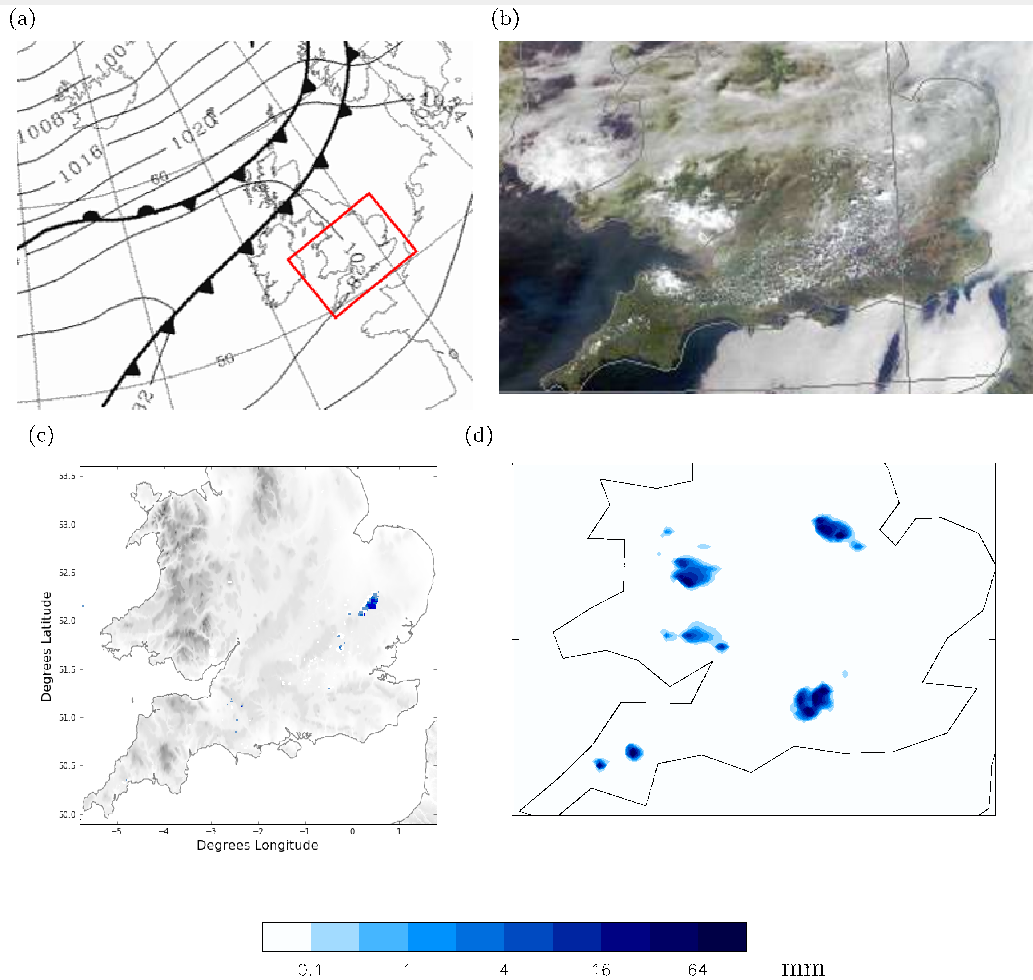
## 2. CSIP IOP 8

The case considered is IOP 8 of the CSIP campaign, which occurred on 13 July 2005. The synoptic situation (Fig. 1a) shows a retreating high pressure region over the United Kingdom (UK), indicating light winds. There were two trailing cold fronts on the northern fringes of the UK, which weakened as they progressed south and had dissipated by 0000 UTC on 14 July. Convection occurred in the mid-afternoon (Figs. 1c–d) in the southern half of the country, with the main focus to the north of London (Fig. 1c; see also Bennett 2007; Khodayar 2009) where temperatures reached 29°C. There was weak synoptic-scale forcing for convection, which was characterised by the diurnal cycle of boundary layer development later leading to cumulus. The convective development was limited by pre-existing mid-tropospheric capping inversions, and decayed in the evening partly due to the inland propagation of a sea breeze. The CSIP study region, over which the diagnostics presented are calculated, is indicated by the box in Fig. 1a.

The MetUM with a 4-km horizontal grid length produced more convective showers than were observed and they were also in the wrong location (compare Fig. 1d with Fig. 1c). These are not untypical errors for convection-permitting models (e.g. Clark *et al.* 2016). A possible source for the differences between the forecast and observed convection in this case may be hypothesised to be due to the locally increased CDNC observed over the southern half of the UK. For example, increased levels of CDNC (relative to typical levels) were observed by the LIDAR at the Chilbolton observatory (Clark and Lean 2006), and one might expect this to suppress the formation of convective precipitation. The control model simulation in Fig. 1d used the default (hereafter STANDARD) CDNC in the model and did not include boundary layer perturbations.

## 3. Model

The model used for this study is the MetUM. It is discussed in Section 3.1, along with the control experiments (without boundary layer perturbations), while the perturbation strategy used to create the ensemble is described in Section 3.2.

**Figure 1.** An overview of CSIP IOP 8: (a) the Met Office surface analysis at 0000 UTC (Courtesy of the Met Office, Crown Copyright, 2005, Met Office), where the box represents the area used in the study; (b) visible image from the MODIS satellite at 1330 UTC (NERC Satellite Receiving Station 2005); (c) the radar-derived hourly accumulation from 1500–1600 UTC (Met Office 2003); and, (d) the control model representation of the hourly accumulation for 1500–1600 UTC (using the default CDNC values and no boundary layer perturbations).

*3.1.   Control Experiments*

The MetUM at version 6.1 was used for this study. This is based on the "New Dynamics" dynamical core (Davies *et al.* 2005) which is semi-implicit, semi-Lagrangian and non-hydrostatic. The parametrizations used include the Edwards and Slingo (1996) radiation scheme, the Wilson and Ballard (1999) microphysics scheme, and the Lock *et al.* (2000) boundary-layer scheme with the Met Office Surface Exchange Scheme (MOSES: Essery *et al.* 2001).

More specifically, we use a 4-km horizontal grid-length configuration of the MetUM, with a limited area domain which covers the UK. The lateral boundary conditions were provided from a 12-km horizontal grid-length configuration and all of the runs considered were initiated at 0700 UTC on 13 July 2005 and run for 14 h, after which all of the convection in the case had dissipated (Clark and Lean 2006). The convection-permitting class of models begins at grid lengths of around 4 km, however at the larger grid lengths of this class of model it has been noted that shallow convection is not well represented (e.g. Lean *et al.* 2008; Clark *et al.* 2016). This poor representation means in practice that some 4-km models run convection schemes that are modified so that the convection is not over/under estimated and to reduce the formation of grid point storms. In the 4-km configuration of the MetUM the Gregory and Rowntree (1990) convection parametrization scheme is enabled with its behaviour being strongly modified by the Roberts (2003) CAPE dependent closure modification. The essential idea of the Roberts (2003) modification is to make the closure timescale an increasing function of CAPE: in this way deep convection can be represented explicitly wherever feasible, but nonetheless shallow overturning is parametrized and excessive accumulation of gridscale CAPE can be avoided. In the simulations analysed here, we checked that any precipitation originating from the convection parametrization was a small fraction of that originating through the explicit dynamics.

Table 1. CDNC over land and sea points for the four different experiments considered and values from the Planche *et al.* (2015) study for comparison.

| Experiment label | CDNC (cm$^{-3}$) | |
| --- | --- | --- |
| | land | sea |
| STANDARD | 300 | 100 |
| INCREASED | 600 | 150 |
| LAND | 600 | 600 |
| 2×LAND | 1200 | 1200 |
| Planche *et al.* (2015) control | 300 | 100 |
| Planche *et al.* (2015) "clean" | 100 | 100 |
| Planche *et al.* (2015) "polluted" | 900 | 100 |

The microphysics scheme used in our simulations is the "3D" version of the microphysics scheme described in detail by Wilson and Forbes (2004). The most important aspect of the microphysics scheme for this study is the autoconversion: the process of changing cloud water into precipitation. The autoconversion is based on Tripoli and Cotton (1980) and is given by

$$P_{RAUT} = A_1 E_{auto} \left(\rho q_{cl}\right)^{A_2-1} \frac{q_{cl}}{\left(n_d\right)^{A_3}},$$

for $P_{RAUT}$ the autoconversion rate, $A_1, A_2, A_3$ defined parameters, $E_{auto}$ a collision/collection coefficient, $q_{cl}$ the cloud liquid water, $\rho$ the density, and $n_d$ the CDNC. Thus sensitivity in the autoconversion mainly derives from uncertainties in the CDNC, which is raised to the power $A_3$=1/3. The "3D" scheme used here is an updated version of that used in Planche *et al.* (2015). The differences include the option for using up to two ice quantities, a consistent sub-gridscale model for all species of water and additional capabilities in the autoconversion to calculate CDNC from sulphate aerosol (although that option was not used within this study).

To test the sensitivity of the formation of convection to the CDNC, four different unperturbed control experiments were set up. The concentrations for each case are given in Table 1 and are constants in space and time. The STANDARD experiment uses the model's default values, as in the control run of Planche *et al.* (2015). The last two experiments use either the INCREASED land CDNC across the entire domain or else twice that value. Such concentrations for CDNC are typical of those considered in other sensitivity studies. For example Storer *et al.* (2010) used a linearly decreasing profile with height from the surface to a value of 100 cm$^{-3}$ at 4 km, with surface values ranging from 100 to 6400 cm$^{-3}$.

*3.2. Perturbation Strategy*

For each of the four CDNC experiments in Table 1, a small (seven member) simple ensemble is created consisting of a control plus six perturbed runs. This ensemble size was chosen due to computational expense limitations and because previous studies have found that it is for about seven members that results become statistically meaningful (e.g. Leoncini *et al.* 2013). The initial and boundary conditions are identical in all of the simulations to ensure that the large-scale dynamical situation remains indistinguishable from the control in all members and so focus analysis purely on the impact of the changing CDNC in the experiments. Ensemble members differ only through different realisations of boundary-layer fluctuations that are indicative of the variability in unresolved turbulence. Specifically, Gaussian potential temperature perturbations are centred on each horizontal grid point $(x_0,y_0)$ within the model domain with the form

$$\text{perturbation}(x,y) = A\exp\left[\frac{-\left(x-x_0\right)^2 + \left(y-y_0\right)^2}{2\sigma^2}\right],$$

for $A$ the amplitude, $(x,y)$ the position in the zonal and meridional directions and $\sigma$ the spatial scale of the perturbations. The overall perturbation field is formed as the superposition of these Gaussian perturbations centred on each grid point. The lengthscale $\sigma$ is set to 8 km and the amplitude is scaled such that the maximum value in the superposed field is 0.1 K. Perturbation fields are added to the

121 model state every 30 min throughout the entire run at a hybrid-model height of 720 m (chosen to be within the daytime convective
122 boundary layer). Perturbations of the same form have been used in studies by Leoncini *et al.* (2010), Done *et al.* (2012) and Flack *et al.*
123 (2018) in which more information can be found on the motivation for, and any sensitivities to, details of the implementation.

124     The members were created as paired ensemble members (i.e. the perturbations for even-numbered ensemble members are the same
125 in structure and amplitude to those for a corresponding odd-numbered member but opposite in sign) and the same random seeds were
126 used to create a given member in each of the four CDNC experiments. Some choices of seed caused the model to become unstable.
127 This was due to an occasional instability in this particular configuration of the model (Lean, pers. comm. (2018)). Such choices were
128 discarded and after several trials a suitable set of seeds was obtained that could be used in all four experiments. We do not anticipate
129 that the need to need to discard some seeds will have influenced the conclusions drawn here or that it affects the general applicability
130 of the methodology proposed. Moreover, in studies of some more recent convective-scale cases using this style of perturbation with
131 newer model configurations, Flack *et al.* (2018) found no problems with the specification of initial seeds.

## 132  4.  Diagnostics

133 There are a range of diagnostics that could be used to consider the impact of ensemble sensitivity experiments, including probabilistic-
134 based approaches such as the rank histogram and continuous rank probability score (e.g. Casati *et al.* 2008). However to fully
135 demonstrate the advantages of our method we need diagnostics that are simple and easily applied to both small ensemble and
136 deterministic forecasts. The complementary diagnostics we use are chosen to enable consideration of the variability within the
137 ensembles due to differences in both the magnitude (Sec. 4.1) and spatial structure (Sec. 4.2-4.3) of the precipitation. All diagnostics
138 presented are computed over the CSIP study region shown in Fig. 1c,d, which is also indicated by the box in Fig. 1a.

### 139  *4.1.  Mean Square Difference*

The Mean Square Difference (MSD) is a standard diagnostic used for considering errors in forecasts, or comparing ensemble members
(e.g. Hohenegger and Schär 2007; Clark *et al.* 2009; Leoncini *et al.* 2013). It evaluates differences in the local magnitude of a field and,
as such, may be subject to the "double penalty problem" (Roberts and Lean 2008) when applied to precipitation in which a forecast is
penalised twice for producing precipitation in the wrong location. This diagnostic thus requires careful interpretation unless either i) a
neighbourhood is applied to the MSD (for more on neighbourhood-based approaches see Ebert 2008) or ii) the MSD is considered in
components representing the common, control and perturbed parts (as in Flack *et al.* 2018). The MSD is computed here as

$$\text{MSD} = \frac{\Sigma \left(P_p - P_c\right)^2}{\Sigma P_c^2},$$

140 where $P_c$ is the hourly accumulation of precipitation in a control forecast and $P_p$ is that in a perturbed forecast. The summations
141 extend over all points in the study region for which either $P_c$ or $P_p$ exceeds a threshold of 0.125 mm with the threshold set to focus
142 on convective storms and ignore any drizzle from stratiform cloud regions. The threshold thus identifies, as convective, storms of size
143 $(4 \text{ km})^2$ or larger with hourly accumulations which would be equivalent to at least 2 mm if concentrated into boxes of $(1 \text{ km})^2$. The
144 MSD is also considered in component form (see Section 5.4).

### 145  *4.2.  Fraction of Common Points*

Common points are defined as those having an hourly precipitation accumulation in excess of 0.125 mm in both of the forecasts being
compared (e.g., the control and a perturbed member). The fraction of common points was introduced in Leoncini *et al.* (2010) and

modified by using a slightly different normalisation in Flack *et al.* (2018). The form used here is the latter, specifically

$$F_{\text{common}} = \frac{N_{1,2}}{N_1 + N_2 - N_{1,2}},$$

146 where $N_x$ is the number of precipitating points for forecast $x = 1$ or 2, and with subscript "$1, 2$" denoting the number in common in

147 the two forecasts. This is a simple diagnostic with values ranging between zero and unity that gives an indication of ensemble spread

148 across the study region in a spatial context.

149 *4.3. Fractions Skill Score*

The Fractions Skill Score (FSS; Roberts and Lean 2008) is a neighbourhood-based technique that can be used to compare ensembles

or as a verification diagnostic. It determines the fraction of points above a specified threshold within an area in two different products,

in our case the control and sensitivity forecasts. The FSS is given by:

$$FSS = 1 - \frac{\Sigma \left( f - o \right)^2}{\Sigma f^2 + \Sigma o^2},$$

150 where $f$ is the value of the output field within the control forecast and $o$ is the value of the same output field in the sensitivity forecast.

151 The FSS varies between zero and unity, with zero representing forecasts that are completely different spatially and unity representing

152 forecasts that are spatially identical. We apply the FSS to the precipitation field and generate binary fields of precipitating and non-

153 precipitating points using a threshold of hourly accumulations of 0.125 mm. The application of the FSS to a binary field means that

154 the diagnostic provides information only on the spatial displacement of events; no information is provided about differences in the

155 magnitude of precipitation (beyond that it meets the required threshold).

156     Here, the FSS has been calculated between the STANDARD CDNC members and the corresponding sensitivity experiment members

157 (similar to the member pairs comparisons in Dey *et al.* 2014). A "skilful" scale for these ensembles is then determined as the lowest

158 spatial scale for which the FSS exceeds 0.75, averaged across the pair comparisons (FSS values above this threshold represent forecast

159 pairs in good agreement for the given scale). This threshold is greater than the 0.5 FSS threshold usually used in other studies (e.g.

160 Roberts and Lean 2008; Dey *et al.* 2014) due to the short lead time for the present forecasts.

161 **5.    Results**

162 *5.1.    Comparisons of individual simulations*

163 As in many of the previously referenced studies (e.g. Sun and Bosilovich 1996; Adams-Selin *et al.* 2013; Planche *et al.* 2015;

164 Barthlott *et al.* 2017) we consider simple comparisons of single runs with the different CDNC settings, as presented for precipitation

165 accumulations in Figure 2. Comparing the control runs (i.e. without perturbations: Fig. 2a, b) corresponds to the approach of Planche

166 *et al.* (2015) and the STANDARD CDNC control is found to produce the largest values of the peak hourly-precipitation and cumulative

167 precipitation. For the peak hourly-precipitation we find that the 2×LAND experiment has the least precipitation whereas the LAND

168 and INCREASED are very similar (Fig. 2a). Furthermore, the cumulative precipitation across the forecast indicates that there is less

169 precipitation in the INCREASED experiment compared to the STANDARD run. By the end of the forecast the LAND and 2×LAND

170 experiments, whilst having less precipitation than the other experiments, are very similar to one another (Fig. 2b).

171     The implication of the existence of convective-scale noise is that a test of the sensitivity to CDNC values in an operational context

172 might easily have produced results corresponding to any member of the ensemble experiments. Moreover, given that the only difference
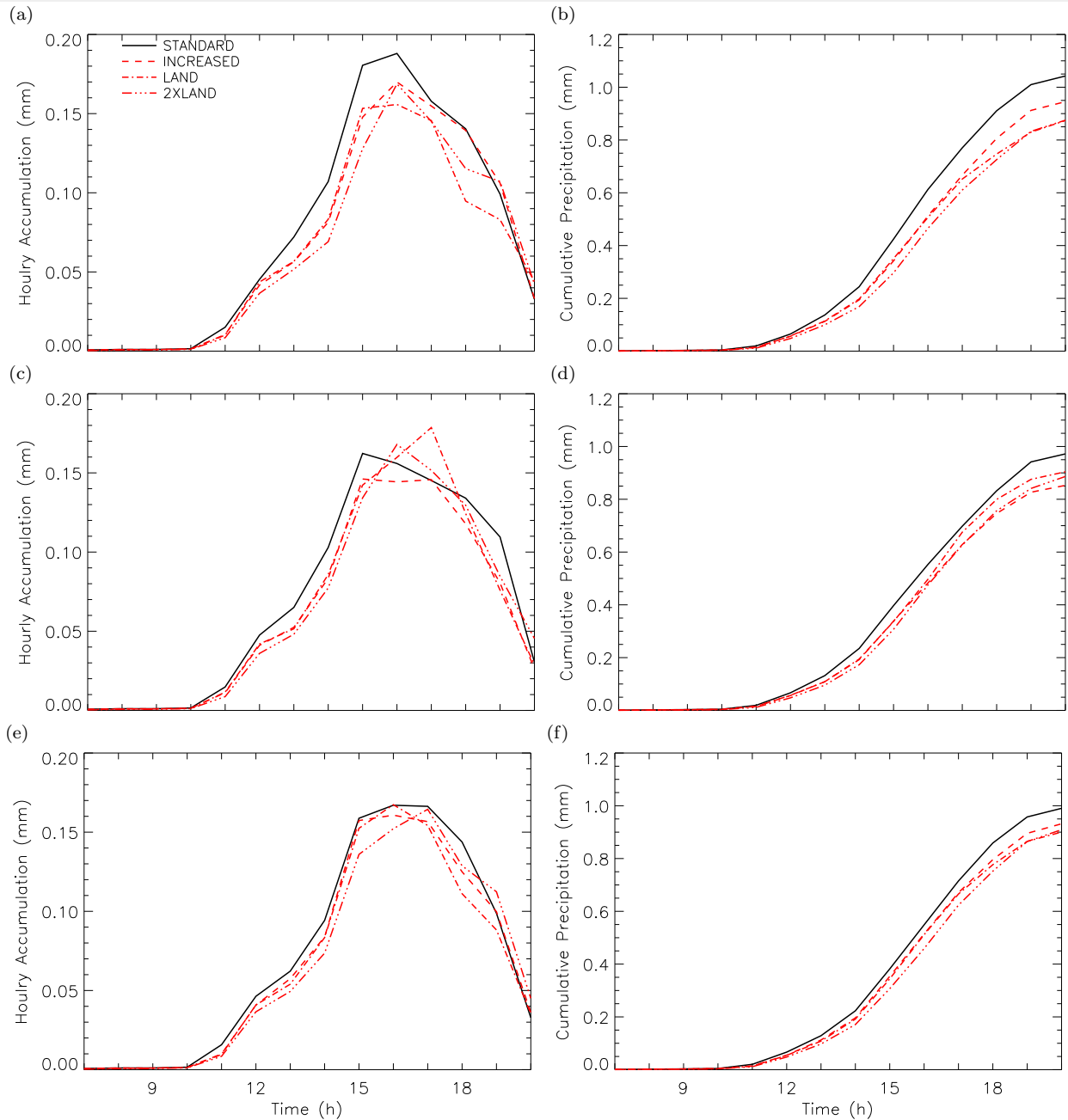
**Figure 2.** Domain-averaged hourly accumulations (a, c, e) and the associated cumulative precipitation (b, d, f) for (a, b) the control member, (c, d) member 1 and (e, f) member 3 of each ensemble. The solid line is for the STANDARD experiment, dashed for INCREASED CDNC, dot-dash for LAND CDNC dot-dot-dot-dash the 2×LAND CDNC. All values are plotted at the start of the hour. The legend in (a) refers to all panels.

between the members is the use of a random seed to introduce convective-scale noise, any one of the members presents an equally valid choice for CDNC comparisons. It is therefore important to consider whether the findings from the control simulations in Figs. 2a, b are robust for all members of the ensemble. For example, consider the results for members 1 and 3 of the ensembles shown in Figs. 2c, d and e, f, respectively, chosen as examples that are not consistent with our earlier findings.

Considering the peak hourly accumulations, for the control simulations the largest peak occurs with the STANDARD CDNC experiment (Fig. 2a), while for the member 1 simulations the LAND experiment has the highest peak (Fig. 2c) and for the member 3 simulations all experiments produce similar peaks, albeit at different times (Fig. 2e). In both perturbed ensemble members the cumulative precipitation is, as for the control members, highest for the STANDARD CDNC experiment throughout, but the ordering for the three experiments with increased CDNC is different for different ensemble members. For example, at the end of the simulations, the lowest cumulative precipitation occurs for the INCREASED CDNC experiment when comparing the member 1 results (Fig. 2d),

but for the LAND and 2×LAND CDNC experiments when comparing the member 3 results (Fig. 2f) or the control results (Fig. 2b) (with the values for the two experiments indistinguishable as plotted).

The largest differences in domain-average precipitation generally occur between the STANDARD and 2×LAND CDNC simulations, and "postage stamps" for both of these ensembles are presented for accumulations over 1500–1600 UTC in Figure 3. These plots serve to highlight our remarks above because on initial inspection there are very limited differences between the precipitation distributions in the different panels and it is not immediately clear whether the two sets have come from distinct underlying distributions. Closer inspection does suggest a possible reduction in the area of precipitating cells in the 2×LAND plots, but analysis of the full ensembles is desirable in order to test this hypothesis.
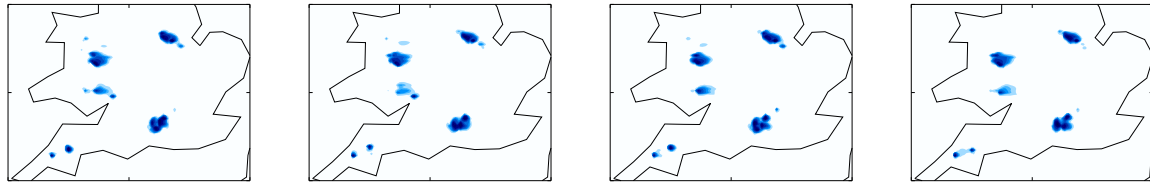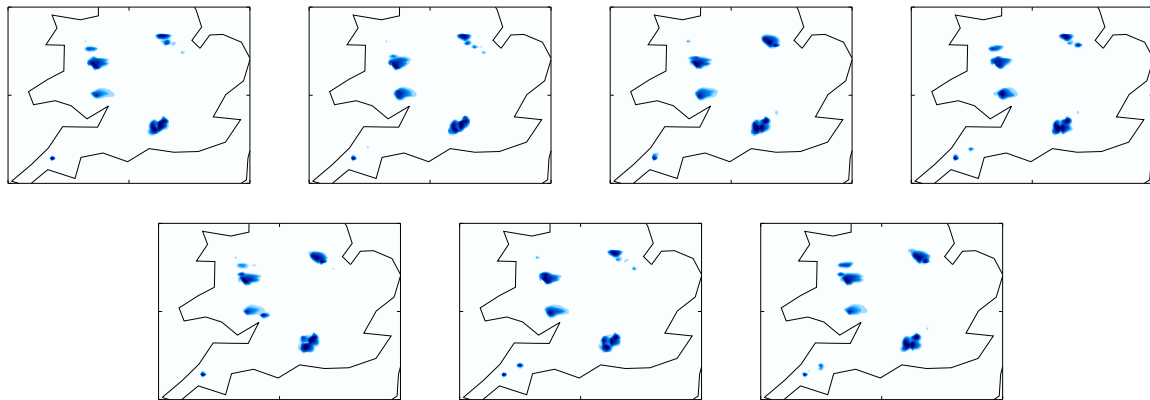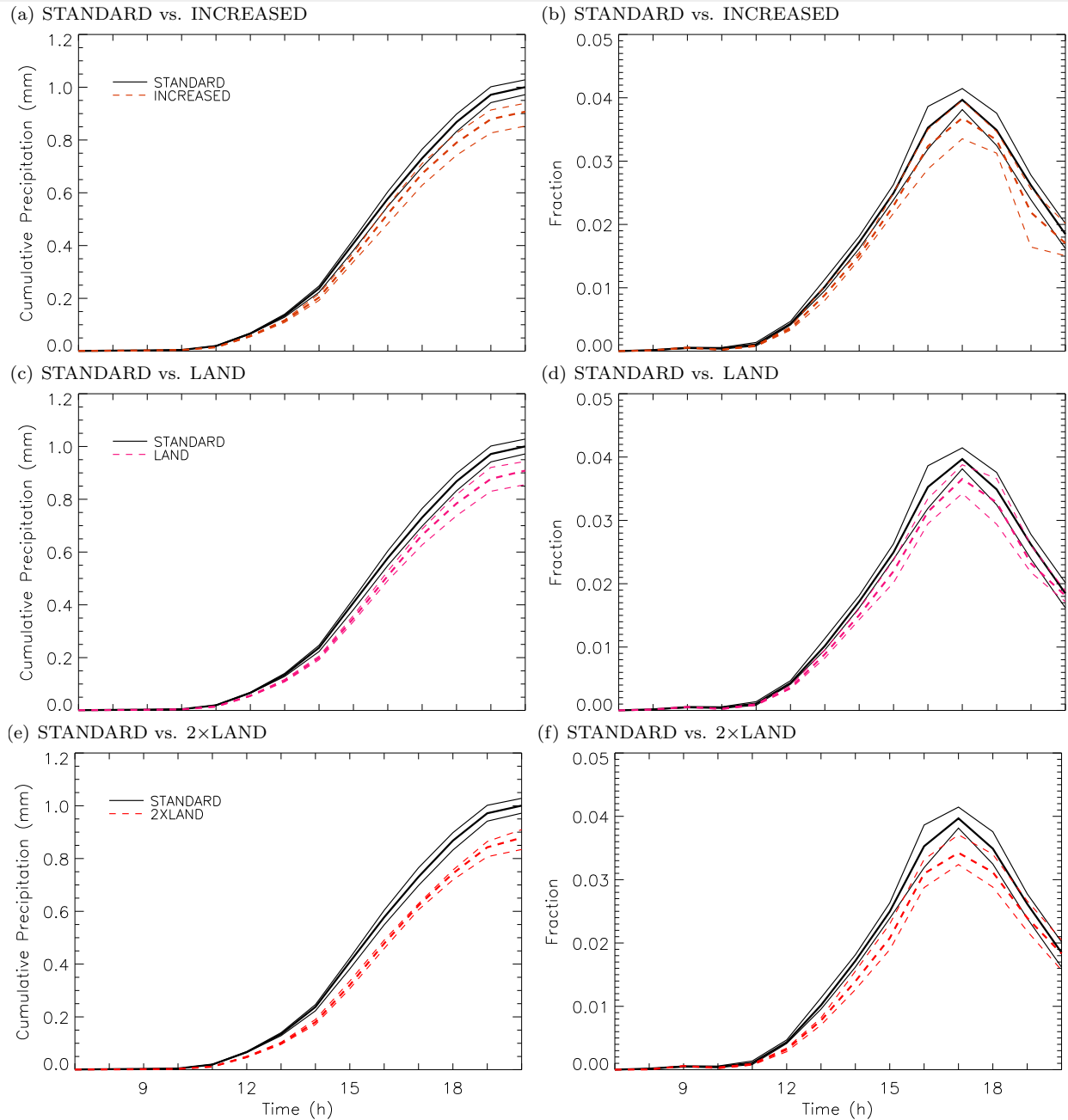


**Figure 3.** Postage stamps plots of the hourly precipitation accumulation from 1500–1600 UTC for a) the STANDARD CDNC and b) the 2×LAND CDNC ensemble. In each case, the control member is the first plot on the left and the corresponding members are shown in equivalent positions. The colour scale refers to all plots.

In summary, the small boundary-layer perturbations representative of turbulent fluctuations induce a level of variability in the precipitation results that cast doubt as to whether differences between single simulations can be considered reliable in revealing systematic impacts of different CDNC values. It is therefore clear that the use of ensemble diagnostics is required to be able to increase our confidence in the sensitivity analysis.

### 5.2. Comparisons of ensembles

In the left hand panels of Fig. 4 we consider the ensemble mean of the cumulative precipitation over the study area, along with the ensemble range, and in the right hand panels the fraction of precipitating points. It is clear from the figure that although there are reductions in precipitation in the experimental ensembles with increased CDNC, these impacts require a cautious assessment as the uncertainties within the simulations, represented by the ensemble range, can be of similar size.

(a) STANDARD vs. INCREASED

(b) STANDARD vs. INCREASED

(c) STANDARD vs. LAND

(d) STANDARD vs. LAND

(e) STANDARD vs. 2×LAND

(f) STANDARD vs. 2×LAND



**Figure 4.** (a, c, e) cumulative precipitation in the ensemble simulations, and (b, d, f) the fraction of points with precipitation exceeding an hourly accumulation of 0.125 mm. The solid lines on each panel present results for the STANDARD CDNC ensemble while the dashed lines are for (a, b) the INCREASED CDNC ensemble, (c, d) the LAND CDNC ensemble and (e, f) the 2×LAND CDNC ensemble. The thick lines represent the ensemble mean and the thinner lines the extremes of the ensemble. All values are plotted at the start of the hour. The legends in (a, c, e) refer to (b, d, f), respectively.

Figures 4a, c and e all show indications of a decrease in the cumulative precipitation for the increased CDNC ensembles in comparison with the STANDARD CDNC ensembles. The largest differences are seen between the STANDARD CDNC and 2×LAND CDNC ensembles, and these ensembles appear to be distinct. A Wilcoxon Signed Rank test has been used to test which ensembles are distinct (e.g. Hamill 1999). The test is an appropriate choice here because the distributions being compared are non-Gaussian. Such tests indicate that differences between the LAND or the 2×LAND ensembles and the STANDARD CDNC ensembles can be considered significant throughout the forecast period (i.e. at each hour) at the 5% significance interval. By contrast, the INCREASED CDNC ensemble shows significant differences from the STANDARD CDNC ensemble only in the final three hours of the simulations.

*5.3. Mechanism for reduced precipitation*

Thus far we have considered the effects of aerosol on the total precipitation within the study area, finding reduced precipitation from increased CDNC. Naturally we would like to understand something of the reasons for such reduction. One possible explanation could

*Prepared using qjrms4.cls*

be that increased CDNC acts to prevent convective rainfall from developing in some locations, and so that there is a reduction in the

fraction of precipitating points. The right hand panels of Fig. 4 do indeed indicate such a reduction, albeit with some overlap between

the ensembles which reiterates the need to perform the analysis in the context of the simple ensemble experiments. Considering the

time with peak accumulations (taken as 1500 UTC) and a time when the convection is beginning to dissipate (1800 UTC) we find

that there are reductions of 17% and 11%, respectively, in the ensemble mean number of points reaching the 0.125 mm hourly rainfall

accumulation threshold for the 2×LAND ensemble relative to the STANDARD ensemble.

Planche *et al.* (2015) demonstrated that ice microphysics does not play a strong role in this case and so argued that the reduction of

precipitation they found from increased CDNC was a result of the suppression of warm rain processes. If so, one might indeed expect to

see a reduction in the number of precipitating points and, more specifically, a reduction in the lightest precipitation accumulations with

limited differences for the heavier values. To consider this, Fig. 5 presents histograms of the hourly accumulations for each ensemble

experiment. It provides some support for the suggestion: the STANDARD ensemble has the most points in the lightest 0.125–0.5 mm

and 0.5–1.0 mm bins, but there are limited differences between ensembles in the number of points in the bins beyond 3.0 mm.
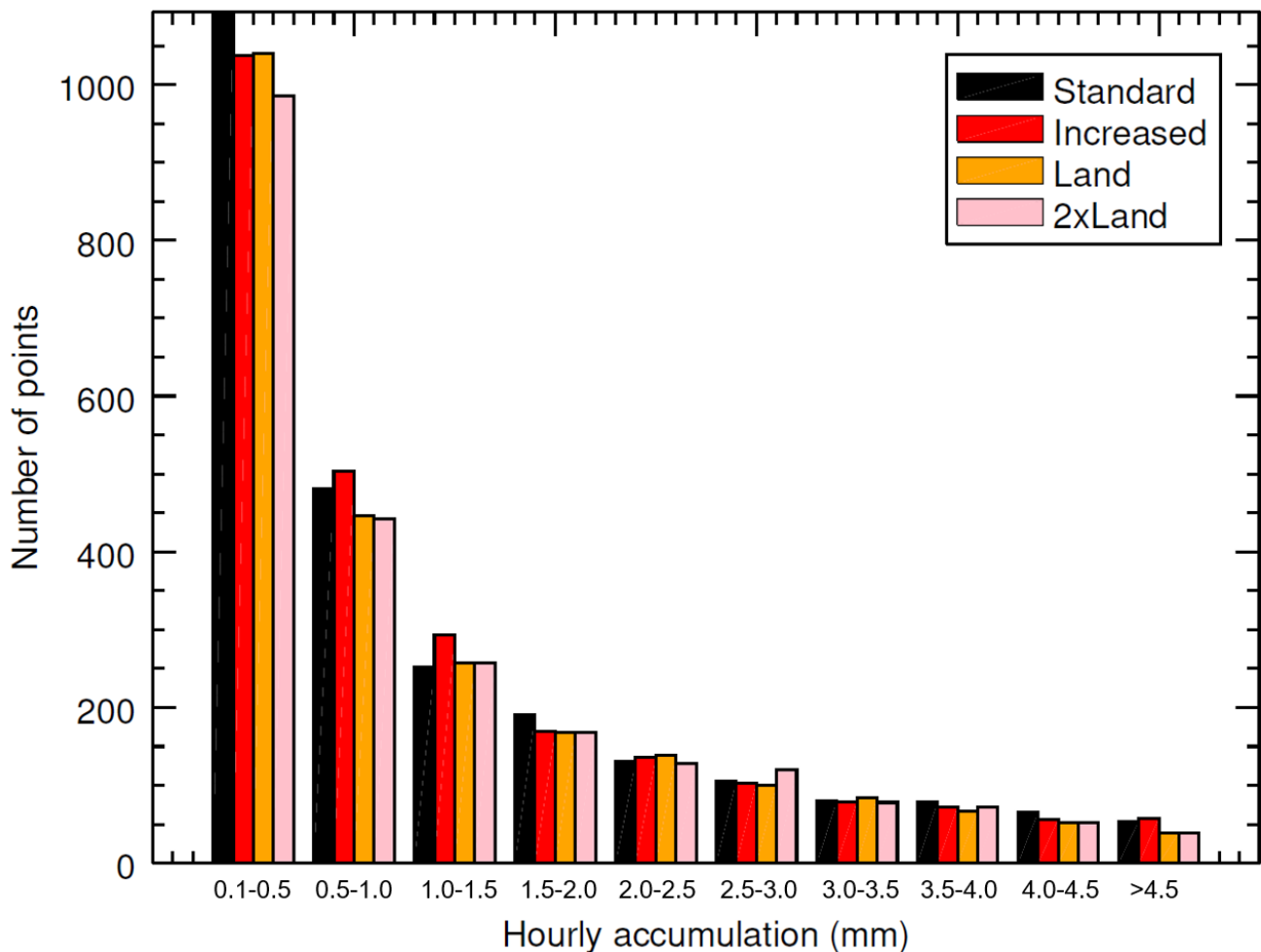


**Figure 5.** Distributions of hourly accumulations from each ensemble experiment. Black frequency bars represent the STANDARD CDNC ensemble, subsequent paler shades represent the INCREASED ensemble, the LAND ensemble and the 2×LAND ensemble. All bin widths are 0.5 mm except for the first, which ranges from 0.125−0.5 mm, and the last bin which represents all values > 4.5 mm.

*5.4. Causes of ensemble differences diagnosed using MSD*

In Fig. 6 we show the MSD between the 2×LAND CDNC and the STANDARD CDNC ensembles, focusing on times after 1200 UTC

once precipitation has developed. As expected, given the short length of the simulations, the MSD grows throughout the period and

does not saturate. Fairly steady, almost linear, growth occurs in each pair comparison until around 1700 UTC after which the MSD

becomes noisy as the precipitation starts to decay. However, interpretation of the MSD is complicated by the "double penalty" problem in convective-scale numerical weather prediction. Computing the MSD over the common points only (not shown) indicates that despite some reductions in warm rain (Sec. 5.3) there are only limited differences that can be attributed to changes at points that precipitate in both the paired ensemble members. Thus the increases in MSD must arise in large part due to different positioning of the convective precipitation.
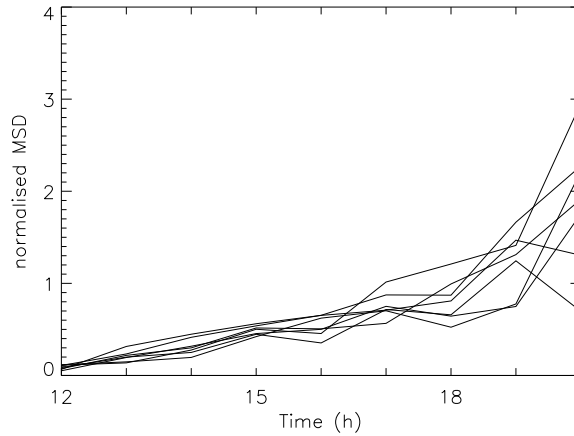


**Figure 6.** The Mean Square Difference between the STANDARD CDNC and 2×LAND CDNC ensemble members, with differences computed between the corresponding members.

The MSD being affected by the displacement of the cells is not unexpected as it is a known impact from boundary-layer perturbations such as those used here, particularly in cases of scattered showers (e.g. Done *et al.* 2012; Flack *et al.* 2018). The extent to which displacements were induced can be quantified using $F_{\text{common}}$, as in Fig. 7. The fraction decreases gradually with lead time. Around 70% of the first precipitating cells were co-located in corresponding members of the two ensembles, and by the time of maximum precipitation around half of the precipitating points remain in common. Further evidence of this displacement is found via the computation of the FSS, which indicates the amplitude of typical displacements in location between forecast events. For the ensemble comparisons here, it results in an average skilful scale of 7.7 km. This scale corresponds to approximately two grid boxes and is therefore on a similar spatial scale to that of the perturbations themselves. This small displacement implies that the MSD-related errors associated with displacement are likely to be more associated with a drop in the fraction of precipitating points, corroborated by $F_{\text{common}}$ and Figs. 4 and 5.
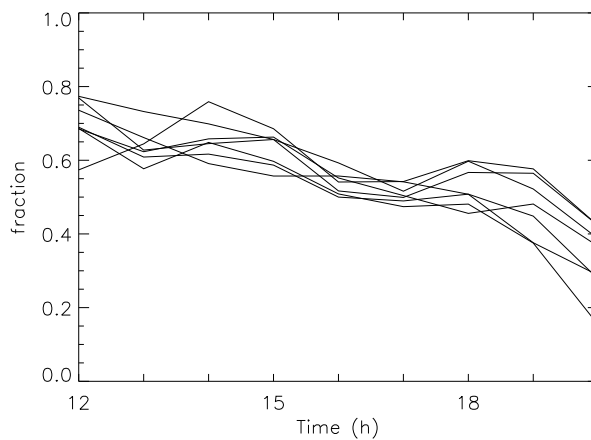


**Figure 7.** The fraction of common points in the STANDARD CDNC and 2×LAND CDNC ensembles, with fractions computed for the corresponding members.

## 6. Summary

While process-based sensitivity case studies remain an important tool for advancing our knowledge of how models perform, current sensitivity tests typically only consider one run of each variant and often for a limited number of cases (often between one and three

*Prepared using qjrms4.cls*

cases; e.g. Gilmore *et al.* 2004; Mahoney and Lackmann 2006). This small sample size means statistical tests cannot be produced. It is also a particular problem when resolution is increased because convective events studied with convection-permitting models are likely to have low predictability (e.g. Hohenegger and Schär 2007). Here we have studied an example that highlights the implications of this low predictability for case study process-based sensitivity tests of model behaviour. It is suggested that a simple ensemble technique can provide a clearer and cleaner view of the sensitivity to physical processes involved. Specifically, we have used boundary-layer potential temperature perturbations to create a small ensemble (7 members) with the 4-km grid-length configuration of the MetUM to determine the impact of increasing CDNC on a case of isolated convective showers.

When single pairs of control runs are considered, we find similar results to the previous study of this case by Planche *et al.* (2015) in that increasing CDNC leads to a suppression of the number of precipitating points, which in turn acts to reduce the average hourly-precipitation accumulations. This result was also found when considering the ensemble means. However, there are overlaps in the ensembles produced for different CDNC settings, so that the result would not necessarily be apparent were an arbitrary pair of runs to be compared. For example, if one were to consider only member 1 or member 3 of the ensembles then a misleading conclusion might have been reached.

Analysis of the full ensembles showed the LAND and 2×LAND CDNC ensembles to be statistically different from the STANDARD CDNC ensemble. Further analysis revealed that the reduced precipitation on increasing CDNC primarily occurs through convection being suppressed, with reduced cell numbers particularly for the lighter rain rates. The overlap shown in the ensembles casts doubt on single run experiments at the convective scale and the method suggested here provides a cheap alternative to operational ensembles to create more robust sensitivity test analysis to alleviate that doubt. This method provides a useful starting point for taking uncertainty into consideration, and may also provide a means for (cheap) initial testing of physics parameters in an operational context before determining the impact in a fully-operational ensemble; however, further testing would be required to show that this method is robust for all circumstances and types of cases.

## References

Adams-Selin RD, van den Heever SC, Johnson RH. 2013. Sensitivity of bow-echo simulation to microphysical parameterizations. *Wea. Forecasting* **28**: 1188–1209, doi:10.1175/WAF-D-12-00108.1.

Ancell BC, Bogusz A, Lauridsen MJ, Nauert CJ. 2018. Seeding chaos: The dire consequences of numerical noise in NWP perturbation experiments. *Bull. Amer. Meteorol. Soc.* **99**: 615–628, doi:10.1175/BAMS-D-17-0129.1.

Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T. 2011. Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Mon. Wea. Rev.* **139**: 3887–3905, doi:10.1175/MWR-D-10-05013.1.

283 Barthlott C, Mühr B, Hoose C. 2017. Sensitivity of the 2014 Pentecost storms over Germany to different model grids and microphysics schemes. *Q. J. R.*
284 *Meteorol. Soc.* **143**: 1485–1503, doi:10.1002/qj.3019.

285 Bennett LJ. 2007. Obersvations of boundary-layer development and initiation of precipitating convection. PhD thesis, University of Leeds.

286 Bowler NE, Arribas A, Beare SE, Mylne KR, Shutts GJ. 2009. The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system.
287 *Q. J. R. Meteorol. Soc.* **135**: 767–776, doi:10.1002/qj.394.

288 Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**:
289 703–722, doi:10.1002/qj.234.

290 Browning KA, Morcrette CJ, Nicol J, Blyth AM, Bennett LJ, Brooks BJ, Marsham J, Mobbs SD, Parker DJ, Perry F, Clark PA, Ballard SP, Dixon MA, Forbes
291 RM, Lean HW, Li Z, Roberts NM, Corsmeier U, Barthlott C, Deny B, Kalthoff N, Khodayar S, Kohler M, Kottmeier C, Kraut S, Kunz M, Lenfant J, Wieser
292 A, Agnew JL, Bamber D, McGregor J, Beswick KM, Gray MD, Norton E, Ricketts HMA, Russell A, Vaughan G, Webb AR, Bitter M, Feuerle T, Hankers
293 R, Schulz H, Bozier KE, Collier CG, Davies F, Gaffard C, Hewison TJ, Ladd DN, Slack EC, Waight J, Ramatschi M, Wareing DP, Watson RJ. 2007. The
294 Convective Storm Initiation Project. *Bull. Amer. Meteorol. Soc.* **88**: 1939–1955, doi:10.1175/BAMS-88-12-1939.

295 Buizza R, Palmer TN. 1995. The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.* **52**: 1434–1456, doi:10.1175/1520-0469(1995)
296 052$⟨$1434:TSVSOT$⟩$2.0.CO;2.

297 Casati B, Wilson LJ, Stephenson DB, Nurmi P, Ghelli A, Pocernich M, Damrath U, Ebert EE, Brown BG, Mason S. 2008. Forecast verification: current status
298 and future directions. *Meteorol. Appl.* **15**: 3–18, doi:10.1002/met.52.

299 Clark AJ, Gallus Jr WA, Xue M, Kong F. 2009. A comparison of precipitation forecast skill between small convection-allowing and large convection-
300 parameterizing ensembles. *Wea. Forecasting* **24**: 1121–1140, doi:10.1175/2009WAF2222318.1.

301 Clark P, Lean H. 2006. An overview of high resolution UM performance for CSIP cases. *Met Office, Joint Centre for Mesoscale Meteorology Report* **155**: 44.

302 Clark P, Roberts N, Lean H, Ballard SP, Charlton-Perez C. 2016. Convection-permitting models: A step-change in rainfall forecasting. *Meteorol. Appl.* **23**:
303 165–181, doi:10.1002/met.1538.

304 Clarke SJ, Gray SL, Roberts NM. 2019. Downstream influence of mesoscale convective systems:Part 1, influence on forecast evolution. *submitted to Q. J. R.*
305 *Meteorol. Soc.* .

306 Davies T, Cullen M, Malcolm A, Mawson M, Staniforth A, White A, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling
307 of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782, doi:10.1256/qj.04.101.

308 Derbyshire S, Beau I, Bechtold P, Grandpiex JY, Piriou JM, Redelsperger JL, Soares P. 2004. Sensitivity of moist convection to environmental humidity. *Q. J.*
309 *R. Meteorol. Soc.* **130**: 3055–3079, doi:10.1256/qj.03.130.

310 Dey SRA, Leoncini G, Roberts NM, Plant RS, Migliorini S. 2014. A spatial view of ensemble spread in convection permitting ensembles. *Mon. Wea. Rev.* **142**:
311 4091–4107, doi:10.1175/MWR-D-14-00172.1.

312 Done J, Craig G, Gray S, Clark P, Gray M. 2006. Mesoscale simulations of organized convection: Importance of convective equilibrium. *Q. J. R. Meteorol. Soc.*
313 **132**: 737–756, doi:10.1256/qj.04.84.

314 Done J, Craig G, Gray S, Clark PA. 2012. Case-to-case variability of predictability of deep convection in a mesoscale model. *Q. J. R. Meteorol. Soc.* **138**:
315 638–648, doi:10.1002/qj.943.

316 Ebert EE. 2008. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorol. Appl.* **15**: 51–64, doi:10.1002/met.25.

317 Edwards J, Slingo A. 1996. Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Q. J. R. Meteorol. Soc.* **122**: 689–719,
318 doi:10.1002/qj.49712253107.

319 Essery R, Best M, Cox P. 2001. MOSES 2.2 technical documentation. Technical Report 30, Hadley Centre. Available at http://www.metoffice.gov.
320 uk/research/hadleycentre/pubs/HCTN/index.html.

321 Flack DLA, Gray SL, Plant RS, Lean HW, Craig GC. 2018. Convective-scale perturbation growth across the spectrum of convective regimes. *Mon. Wea. Rev.*
322 **146**: 387–405, doi:10.1175/MWR-D-17-0024.1.

323 Gilmore MS, Straka JM, Rasmussen EN. 2004. Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics
324 scheme. *Mon. Wea. Rev.* **132**: 2610–2627, doi:10.1175/MWR2810.1.

325 Gregory D, Rowntree P. 1990. A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea.*
326 *Rev.* **118**: 1483–1506, doi:10.1175/1520-0493(1990)118$⟨$1483:AMFCSW$⟩$2.0.CO;2.

327 Hagelin S, Son J, Swinbank R, McCabe A, Roberts N, Tennant W. 2017. The Met Office convective-scale ensemble, MOGREPS-UK. *Q. J. R. Meteorol. Soc.*
328 **143**: 2846–2861, doi:10.1002/qj.3135.

Hamill TM. 1999. Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting* **14**: 155–167, doi:10.1175/1520-0434(1999)014⟨0155: HTFENP⟩2.0.CO;2.

Hartmann D, Buizza R, Plamer T. 1995. Singular vectors: The effect of spatial scale on linear growth of disturbances. *J. Atmos. Sci.* **52**: 3885–3894, doi: 10.1175/1520-0469(1995)052$⟨$3885:SVTEOS$⟩$2.0.CO;2.

Hohenegger C, Schär C. 2007. Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteorol. Soc.* **88**: 1783–1793, doi: 10.1175/BAMS-88-11-1783.

Khodayar S. 2009. High-resolution analysis of the initiation of the deep convection formed by boundary-layer processes. PhD thesis, Univeristy of Karlsruhe.

Lean HW, Clark PA, Dixon M, Roberts NM, Fitch A, Forbes R, Halliwell C. 2008. Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.* **136**: 3408–3424, doi:10.1175/2008MWR2332.1.

Leoncini G, Plant R, Gray S, Clark P. 2010. Perturbation growth at the convective scale for CSIP IOP18. *Q. J. R. Meteorol. Soc.* **136**: 653–670, doi: 10.1002/qj.587.

Leoncini G, Plant R, Gray S, Clark P. 2013. Ensemble forecasts of a flood-producing storm: Comparison of the influence of model-state perturbations and parameter modifications. *Q. J. R. Meteorol. Soc.* **139**: 198–211, doi:10.1002/qj.1951.

Lock A, Brown A, Bush M, Martin G, Smith R. 2000. A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Wea. Rev.* **128**: 3187–3199, doi:10.1175/1520-0493(2000)128$⟨$3187:ANBLMS$⟩$2.0.CO;2.

Mahoney KM, Lackmann GM. 2006. The sensitivity of numerical forecasts to convective parameterization: A case study of the 17 february 2004 east coast cyclone. *Wea. Forecasting* **21**: 465–488, doi:10.1175/WAF937.1.

McCabe A, R S, Tennant W, A L. 2016. Representing model uncertainty in the met office convection-permitting ensemble prediction system and its impact on fog forecasting. *Q. J. R. Meteorol. Soc.* **142**: 2897–2910, doi:10.1002/qj.2876.

Met Office. 2003. 1 km resolution UK composite rainfall data from the Met Office Nimrod System, NCAS British Atmospheric Data Centre. http://catalogue.ceda.ac.uk/uuid/27dd6ffba67f667a18c62de5c3456350. [accessed 02/2017].

NERC Satellite Receiving Station. 2005. NERC Satellite Receiving Station, Dundee Unviersity, Scotland. http://www.sat.dundee.ac.uk/. [accessed 21/02/17].

Oleson KW, Bonan GB, Feddema J, Vertenstein M. 2008. An urban parameterization for a global climate model. Part II: Sensitivity to input parameters and the simulated urban heat island in offline simulations. *J. Appl. Meteorol. Climatol.* **47**: 1061–1076, doi:10.1175/2007JAMC1598.1.

Planche C, Marsham JH, Field PR, Carslaw KS, Hill AA, Mann GW, Shipway BJ. 2015. Precipitation sensitivity to autoconversion rate in a numerical weather prediction model. *Q. J. R. Meteorol. Soc.* **141**: 2032–2044, doi:10.1002/qj.2497.

Price J, Porson A, Lock A. 2015. An observational case study of persistent fog and comparison with an ensemble forecast model. *Boundary Layer Meteorol.* **155**: 301–327, doi:10.1007/s10546-014-9995-2.

Roberts N. 2003. Stage 2 report from the storm-scale numerical modelling project. Technical Report 407, Met Office R&D. Available at: http://www.metoffice.gov.uk/research/nwp/publications/papers/technicalreports/index.html.

Roberts N, Lean H. 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.* **136**: 78–97, doi:10.1175/2007MWR2123.1.

Seity Y, Brousseau P, Malardel S, Hello G, Bénard P, Bouttier F, Lac C, Masson V. 2011. The AROME-France Convective-Scale Operational Model. *Mon. Wea. Rev.* **139**: 976–991, doi:10.1175/2010MWR3425.1.

Storer RL, van den Heever SC, Stephens GL. 2010. Modelling aerosol impacts on convective storms in different environments. *J. Atmos. Sci.* **67**: 3904–3915, doi:10.1175/2010JAS3363.1.

Sun WY, Bosilovich MG. 1996. Planetary boundary layer and surface layer sensitivity to land surface parameters. *Boundary Layer Meteorol.* **77**: 353–378, doi:10.1007/BF00123532.

Thompson G, Rasmussen RM, Manning K. 2004. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.* **132**: 519–542, doi:10.1175/1520-0493(2004)132$⟨$0519:EFOWPU$⟩$2.0.CO;2.

Toth Z, Kalnay E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**: 3297–3319, doi:10.1175/1520-0493(1997)125$⟨$3297: EFANAT$⟩$2.0.CO;2.

Tripoli GJ, Cotton WR. 1980. A numerical investigation of several factors contributing to the observed variable intensity of deep convection over south florida. *J. App. Meteor.* **19**: 1037–1063, doi:10.1175/1520-0450(1980)019⟨1037:ANIOSF⟩2.0.CO;2.

Wilson D, Forbes R. 2004. Unified Model documentation paper 26. The large-scale precipitation parametrization scheme. version 6.01. Technical report, Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB.

376 Wilson DR, Ballard SP. 1999. A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.* **125**:
377 1607–1636, doi:10.1002/qj.49712555707.

378 Zampieri M, Malguzzi P, Buzzi A. 2005. Sensitivity of quantitative precipitation forecasts to boundary layer parameterization: a flash flood case study in the
379 western Mediterranean. *Nat. Hazards and Earth Sys. Sci.* **5**: 603–612, doi:10.5194/nhess-5-603-2005.

380 Zhang D, Anthes RA. 1982. A high-resolution model of the planetary boundary layer – sensitivity tests and comparisons with SESAME-79 data. *J. Appl.*
381 *Meteorol.* **21**: 1594–1609, doi:10.1175/1520-0450(1982)021⟨1594:AHRMOT⟩2.0.CO;2.

382 Zhang F, Snyder C, Rotunno R. 2003. Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.* **60**: 1173–1185, doi:10.1175/1520-0469(2003)
383 060⟨1173:EOMCOM⟩2.0.CO;2.