


Extreme rainfall sensitivity in convective-scale ensemble modelling over Singapore

Aurore N. Porson¹  | Susanna Hagelin² | Douglas F.A. Boyd³ | Nigel M. Roberts¹ | Rachel North³ | Stuart Webster³ | Jeff Chun-Fung Lo⁴

¹MetOffice@Reading, Department of Meteorology, University of Reading, Reading, UK

²Swedish Meteorological and Hydrological Institute, Norrköping, Sweden

³Met Office, Exeter, UK

⁴Centre for Climate Research Singapore, Singapore

Correspondence

Aurore N. Porson, MetOffice@Reading, Department of Meteorology, Meteorology Building, University of Reading, Reading, Berkshire, RG6 6BB, UK.
Email: aurore.porson@metoffice.gov.uk

A convective-scale ensemble system was developed to predict the occurrence of heavy convective rainfall around Singapore with a focus on the prediction of high-impact events. The new ensemble SINGV-EPS has been nested within two global ensembles, MOGREPS-G (UK Met Office) and EC-ENS (ECMWF). Predicting the occurrence of convective rainfall in an area such as Singapore is challenging and this article discusses the use of the convection-permitting ensemble to characterize the uncertainties in the prediction of such localized heavy rainfall. First, verification of wind, temperature, and precipitation is performed for a month-long period to assess the relative performance of each ensemble. This reveals differences, but no robust signal to say one is better than the other. The results are not statistically significant and not all variables are consistently better with one ensemble or the other. Secondly, the precipitation characteristics of SINGV-EPS are analysed from probabilities of precipitation and variability among the ensemble members. SINGV-EPS is sensitive to the choice of the global ensemble providing the initial conditions and boundaries. The results suggest there is benefit, in some cases, from combining the two ensembles. Thirdly, the spread of the ensemble precipitation is analysed using the dispersion Fractions Skill Score (dFSS). We compare the impact of the initial perturbations and the perturbations in lateral boundary conditions in both nesting options. The initial perturbations dominate in the beginning of the forecasts, with influence up to T+24 h, and are associated with an upscale growth of the uncertainties. The impact of the parent ensemble and lateral boundary conditions dominate at the end of the forecast and tend to influence larger scales more.

KEYWORDS

convective-scale ensembles, ensemble spread, tropical meteorology

1 | INTRODUCTION

Singapore has one the world's busiest airports, and the Straits of Malacca is one of the world's most important shipping lanes linking the Indian Ocean and Pacific Ocean. Together with a high population density, any damage caused by extreme

weather could have a significant impact on the infrastructure of the region.

At Singapore, intense rainfall events can occur throughout the year. These events can happen at any time of the day, either following the onset of daytime convection or as a result of specific organization, such as squall lines, at the end of the

night or early morning. The ability to forecast the onset, intensity, and evolution of deep convective storms is the subject of ongoing research and is of primary importance to operational meteorologists.

Lo and Orton (2016) described the features of Sumatra Squalls. These squalls usually initiate late at night over Sumatra, move eastwards through the Malacca Straits and make landfall in Singapore in the early hours of the morning. Strong wind gusts accompany the heavy rain produced by these squalls. Many factors can affect the onset, development, and propagation of these squalls. For example, Sakurai *et al.* (2005) described the seasonal influence of the Intertropical Convergence Zone on the propagation of the storms developing over Sumatra and found that eastward movement of the storms (i.e. affecting Singapore) is more uncertain than the westward movement. Wu *et al.* (2009) showed the importance of the topography over Sumatra for the onset of the heavy rainfall over the sea to the west, while more generally, Letkewicz and Parker (2011) showed how sensitive squall-line development is when interacting with terrain. Mesoscale flows over Sumatra are also associated with gravity waves resulting from different heating profiles, as described in Love *et al.* (2011).

With so many forcing mechanisms for initiating convection in this area (i.e. temperatures of the shallow seas and large mountain ranges), it is clear that we need a model that can capture the uncertainties related to the nonlinear nature of the convection that initiates and develops over Sumatra and the Malaysian Peninsula in order to predict the landfall and intensity of the rainfall at Singapore. Additionally, the nonlinear evolution of convection related to variations in background shear and instability, and perhaps to a lesser extent for the Tropics, larger-scale dynamical forcing, add to the difficulty of predicting high-impact rainfall.

Convection-permitting ensembles are required to estimate the uncertainty in the prediction of precipitation events and provide additional details in the timing and spatial development of these events (Marsigli *et al.*, 2001). Increased computer power has allowed more of a focus on convection with the development of convection-permitting ensembles in several meteorological centres (Gebhardt *et al.*, 2010; Schwartz *et al.*, 2010; Beck *et al.*, 2016; Hagelin *et al.*, 2017). The use of convective-scale models improves both the physical realism and the skill of rainfall forecasts when analysed on appropriate scales (Roberts and Lean, 2008; Clark *et al.*, 2011; 2016; Schwartz and Sobash, 2017). Convective-scale models have also been shown to give an improved representation of diurnally forced convection compared to parametrized-convection models (Kendon *et al.*, 2012).

Despite the potential benefits, little research has been done with convection-permitting ensembles to understand and predict the occurrence of the squalls in tropical regions, compared to the midlatitude severe convective systems (Hanley *et al.*, 2013; Bednarczyk and Ancell, 2015).

The type and scale of the perturbations used in setting up a convective-scale Ensemble Prediction System (EPS) is an ongoing area of research. Durran and Weyn (2016) studied the sensitivity of perturbation growth to the scale of those perturbations for a squall-line simulation in horizontally homogeneous conditions. They found that errors at the large scale are perhaps more important than errors at the small scale. Nielsen and Schumacher (2016) also demonstrated the importance of the large-scale errors by computing the evolution of the spread of the ensembles using multiple variables and an amplitude factor on the initial and lateral boundary perturbations of the ensemble. Gebhardt *et al.* (2010) showed that the dispersion of an ensemble is affected by the type of perturbations: initial conditions (IC) control the dispersion at the beginning of the forecasts, while lateral boundary conditions (LBC) control the following forecast hours. The sensitivity of these may depend upon the size of the domain (Warner *et al.*, 1997). Hanley *et al.* (2013) also identified the importance of the large-scale forcing on a squall-line simulation as a part of the Convective and Orographically-induced Precipitation Study (COPS) campaign. Most of these studies, however, relate to midlatitude meteorology.

In collaboration with Meteorological Service Singapore (MSS), the Met Office have set up an EPS at the convective scale, namely SINGV-EPS, to capture the possible outcomes in forecasts of heavy (and extreme) rainfall around Singapore and analyse the benefits of running an ensemble at such a resolution. SINGV-EPS has the option of being nested within MOGREPS-G (Unified Model UM-SINGV) or nested within EC-ENS (EC-SINGV). Since the Met Office Global and Regional EPS (MOGREPS-G) and ECMWF Ensemble EC-ENS have more uncertainty in the Tropics (Park *et al.*, 2008), compared to midlatitudes, the convective-scale ensembles nested within these global ensembles are also likely to exhibit noticeable differences.

The aims of the article are to determine the benefits of running a convective-scale ensemble for capturing heavy or extreme rainfall over Singapore. The first objective is to determine whether there are differences in skill and variability between UM-SINGV and EC-SINGV. The second objective is to determine the relative impact of the perturbations from the ICs or from the LBCs on the spread of the ensembles. We examine the differences in spread between the two ensembles and ask how that compares with the spread characteristics regarding the impact of the ICs and LBCs. The article is divided into five main sections. In section 2, we describe UM-SINGV and EC-SINGV ensembles and their parent ensembles as well as the metrics used in this study. A spatial representation, given by the Fractions Skill Score, is used here, following Roberts and Lean (2008), Roberts (2008), Mittermaier *et al.* (2013), Schwartz *et al.* (2010), Dey *et al.* (2014) and Flack *et al.* (2018). In section 3, we use objective verification to evaluate a month-long trial to determine

the differences between the two ensembles and assess their diurnal cycles. In section 4, we study the meteorology of four high-impact cases to understand the differences between UM-SINGV and EC-SINGV, the sensitivity of SINGV-EPS to the parent ensemble and the relative importance of the ICs and LBCs. The relative importance of these perturbations is studied by quantifying the ensemble spread at different scales for precipitation and compared against the differences in the spread between the two ensembles.

2 | DESCRIPTION OF THE ENSEMBLE MODELS AND METRICS

2.1 | The ensemble models EC-SINGV and UM-SINGV

The SINGV ensemble, SINGV-EPS, is run as a downscaler of the global ensemble similar to the set-up used by the UK ensemble MOGREPS-UK up to March 2016 (Hagelin *et al.*, 2017). The initial and boundary conditions for SINGV-EPS are taken from the global ensemble running at the European Centre for Medium-range Weather Forecasts (ECMWF) at 0.2° resolution with 51 members (Buizza and Palmer, 1995; Marsigli *et al.*, 2001; Molteni *et al.*, 2001; Buizza *et al.*, 2008; Leutbecher and Palmer, 2008) and the global ensemble running at the Met Office, MOGREPS-G at $\sim 0.28^\circ$ resolution with 18 members (Bowler *et al.*, 2008; Bowler and Mylne, 2009; Tennant and Beare, 2014). Comparing the global ensembles from ECMWF and MOGREPS-G, Zsoter *et al.* (2009) showed that, by using the ensemble mean and the control forecasts, both global ensembles tend to follow their own ensemble forecasts more closely than the forecasts from the other EPS. The two ensembles differ in their way of creating the member perturbations, with the EC-ENS model based on singular vector perturbations (Buizza and Palmer, 1995) and MOGREPS-G based on the ensemble transform Kalman filter (ETKF) data assimilation scheme (Bowler and Mylne, 2009). Park *et al.* (2008) highlighted different characteristics of the ensembles, particularly in the Tropics; Fig. 7 in their work shows that EC-ENS is less dispersive than MOGREPS-G over the first 3 days.

For ease of reference, SINGV-EPS within MOGREPS-G is named UM-SINGV and SINGV-EPS within the ECMWF ENS is named EC-SINGV.

The formulation of SINGV-EPS is based on MOGREPS-UK (Hagelin *et al.*, 2017). However, some of the parametrizations used in SINGV-EPS are slightly different from MOGREPS-UK. SINGV-EPS uses a tropical configuration while MOGREPS-UK uses a midlatitude configuration. The main differences are explained in Bush *et al.* (2019) and include: a different set of vertical levels (more levels in the upper troposphere to allow for a higher tropopause), the presence of boundary-layer stochastic perturbations in

the midlatitude configuration (useful to initiate convection earlier) and not in the tropical configuration, as well as the use of a prognostic cloud scheme (PC2) in the tropical configuration (Morcrette, 2012a; 2012b). Details about other physics parametrizations can be found in Dipankar *et al.* (personal communication, 2019) and Hagelin *et al.* (2017).

An important difference in the initial perturbations in the UM-SINGV and EC-SINGV ensembles is that the sea-surface temperatures (SSTs) are perturbed in UM-SINGV, but are nearly identical in the EC perturbed members. This is because the EC-ENS model relies on an evolving temperature parametrization to create the spread of the ensemble. This, together with differences in the parent ensembles (Park *et al.*, 2008) may have some implications for the spread of the ensemble as described in section 4.

SINGV-EPS is running with a 4.5 km (0.0405°) grid spacing with 364 points in longitude, 342 in latitude, and 80 vertical levels. Sensitivity tests with the UM deterministic model show that the changes introduced with the tropical convective-scale science contribute more to the performance of the model than an increase in horizontal resolution. The domains are illustrated in Figure 1. Singapore is offset from the domain centre so that the source of the Sumatran squalls, namely the Sumatran mountains, are wholly within the domain. Within EC-ENS, SINGV-EPS is initialized twice a day, at 0300 and 1500 UTC, with 51 members. Within MOGREPS-G, mirroring the settings of MOGREPS-UK, SINGV-EPS is initialized four times a day, at 0300, 0900, 1500 and 2100 UTC, with 18 members. We will only use the 1500 UTC forecasts in this article as this is a time at which both ensembles are initialized and is in the period of transition between the daytime and night-time convection systems (for example, squall lines), as local time is UTC +8 h. Each forecast is run for 36 h.

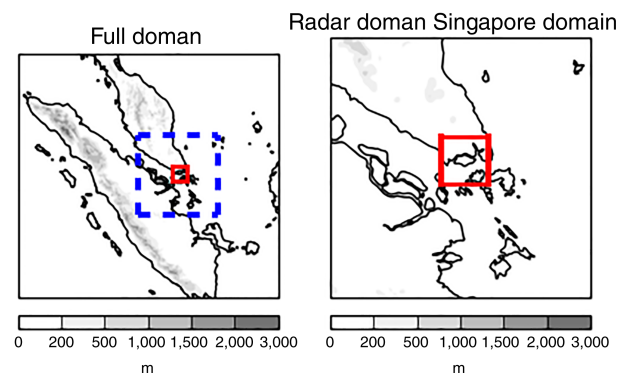


FIGURE 1 Illustrations of the different domains used here. (a) “Full domain” with “Radar domain” (box in dotted lines) and “Singapore domain” (small box in solid lines). (b) “Radar domain” and “Singapore domain” (small box in solid lines). The “Singapore domain” contains 20 by 20 grid points centred around Singapore [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | Neighbourhood metrics

The probabilities generated from the ensembles make use of neighbourhood post-processing. Neighbourhood ensemble probabilities (*NEP*) are computed by averaging the neighbourhood probability *NP* defined here over the ensemble members (as described in Eqn 3 in Schwartz and Sobash (2017)). Neighbourhood processing is used to increase the size of the ensemble (Roberts and Lean, 2008; Clark *et al.*, 2010, 2016; Schwartz *et al.*, 2010, 2015; Golding *et al.*, 2016). Probabilities are given by fractions of occurrence in the “neighbourhood” square surrounding each grid point and this is applied for each ensemble member to produce a final ensemble probability. The purpose is to produce more smoothly-varying probabilities by accounting for the small-scale uncertainty that is not captured by a small ensemble. A neighbourhood *S* around the *i*th point, *S_i*, is here a box of 15×15 points centred on the point itself, with *N_b* the number of grid points included in *S_i*. The following calculation is repeated for every grid point of the “radar domain” in Figure 1, except for the grid points which are too near the boundaries. For each member *j*, this follows Eqn 2 from Schwartz and Sobash (2017) as described here:

$$NP(q)_{ij} = \frac{1}{N_b} \sum_{k=1}^{N_b} BP_{kj}, \quad k \in S_i,$$

with *BP* the binary probability for threshold *q* of event occurrence at the *i*th point for the *j*th member.

The dispersion metric dFSS (Dey *et al.*, 2014) is also used in this article to study the spatial characteristics of the ensemble spread for precipitation. This dispersion metric is based on the Fractions Skill Score metric (FSS) as in Roberts and Lean (2008) and Roberts (2008) and can be used to compare the similarity between ensemble members and a field of observations/truth.

First we define a square neighbourhood of length *L* over which probabilities (area fractions) are computed at each grid square for a specified rainfall threshold (e.g. 2 mm/h). Let *P_{XL}(i,j)* and *P_{YL}(i,j)* be the neighbourhood probabilities of an event from two ensemble members *X* and *Y* respectively at grid point (*i,j*) on a domain of dimension *m* × *n* grid points. Then the FSS is given by:

$$FSS_{X,Y,L} = 1 - A_{X,Y,L} / B_{X,Y,L};$$

$$A_{X,Y,L} = \frac{\sum_i^m \sum_j^n [P_{X,L}(i,j) - P_{Y,L}(i,j)]^2}{m \times n};$$

$$B_{X,Y,L} = \frac{[\sum_i^m \sum_j^n [P_{X,L}(i,j)]^2 + \sum_i^m \sum_j^n [P_{Y,L}(i,j)]^2]}{m \times n},$$

with *A_{X,Y,L}* representing the mean squared differences between the neighbourhood probabilities, and *B_{X,Y,L}* the maximum possible mean squared differences (Roberts, 2008). The FSS

ranges between 0 and 1. If the two probability fields are identical at every grid square, the FSS = 1. In that case the fields are perfectly spatially matched. If the two fields have non-zero probabilities at different locations everywhere in the domain, the FSS = 0. In that case the two fields are completely mismatched. It is therefore possible to determine the spatial agreement between two fields at different scales by computing the FSS for a range of neighbourhood lengths *L*. The bigger the FSS the better the spatial agreement at the scale (neighbourhood length) of interest.

We can extend this measure of spatial differences between two fields to obtain a measure of the spatial ensemble spread by following the approach of Dey *et al.* (2014). If the metric FSS is computed for a pair of ensemble members, it becomes known as dFSS to relate to the dispersion metric of the ensemble. For an ensemble forecast, the dFSS is computed for each member compared with every other member. Then the mean value over all the comparisons, dFSSmean, is found. The closer dFSSmean is to 1 the better the mean spatial agreement between the ensemble members and the lower the spatial ensemble spread; conversely a low value of dFSSmean indicates a large spatial spread.

dFSSmean is computed over a range of neighbourhood sizes (*L*) varying in length from *L* = 3 to 31 points, and we obtain one value of dFSSmean for each forecast time and neighbourhood size. We can then use dFSSmean to estimate how the spatial ensemble spread differs between UM-SINGV and the EC-SINGV as the forecasts evolve.

3 | MONTHLY ANALYSIS FOR OCTOBER 2017

3.1 | Objective verification

Initial objective verification is performed here against land-surface observations of rain accumulations (frequency every 6 h), wind speed and temperature over the month of October 2017. The following verification scores: Continuous Rank Probability Score (CRPS), Rank Probability Score (RPS) and Relative Operating Characteristic (ROC) area, are used here to analyse the performance of EC-SINGV and UM-SINGV against the land-surface observations as illustrated in Figures 2–6. A “spin-up period” comprising the first 6 h is not included in the wind and temperature scores. To collect as many data as possible, the verification for the full ensemble comparison is done over the 0300 and 1500 UTC cycles, which are the two cycles in common between UM-SINGV and EC-SINGV. A small neighbourhood length of 3×3 points (13.5×13.5 km) is used, which was found to be beneficial in the objective verification of MOGREPS-UK (Mittermaier, 2014; Hagelin *et al.*, 2017; Mittermaier and Csimas, 2017).

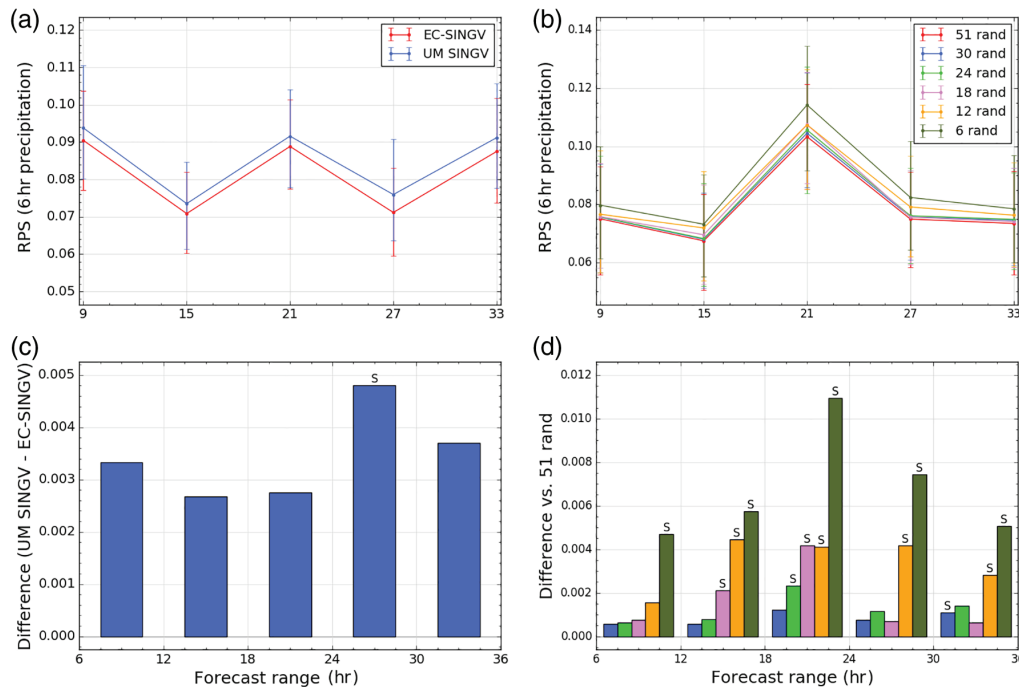


FIGURE 2 Ranked probability scores for 6-hourly precipitation for October 2017. (a) Comparison UM-SINGV (18 members) with EC-SINGV (51 members), (b) Difference plots between UM-SINGV and EC-SINGV for simulations starting at 0300 and 1500 UTC only, (c) Ensemble sizes of 6, 12, 18, 24 and 30 are here compared for the ranked probability scores of the full ensemble EC-SINGV for simulations starting at 1500 UTC only and the differences to the full-size EC-SINGV ensemble are depicted in (d). A neighbourhood size of 3 grid points is used here. Ninety per cent confidence intervals are calculated using Monte-Carlo method [Colour figure can be viewed at wileyonlinelibrary.com]

Figures 2, 3 and 4a,b compare the skill of both EC-SINGV and UM-SINGV for the 6-hourly precipitation, temperature and wind speed respectively. In these plots, 90% confidence intervals were generated by bootstrapping, using 500 bootstrap resamples and the percentile method (Wilks, 2011). Statistical significance at the 0.05 level was determined using the Wilcoxon signed-rank test (Wilks, 2011). Figures 2 and 4 show that EC-SINGV has better (lower) ranked probability scores for precipitation and wind, while UM-SINGV has better continuous ranked probability scores for the 1.5 m temperature (Figure 3), although most differences are within the error bars and therefore there is little real discernible difference. This behaviour matches the performance of the ensembles regarding the mean biases (not shown): while both ensembles have a positive bias in wind speed, EC-SINGV has a smaller bias than UM-SINGV whereas UM-SINGV has a smaller bias in temperature than to EC-SINGV. In Figure 5, the ROC areas, representing the resolution of the ensemble, for the precipitation are higher (i.e. better) for EC-SINGV, but Figure 6, for the wind speed, shows a better performance for EC-SINGV at lower thresholds (Figure 6a) and for UM-SINGV at higher thresholds (Figure 6b). Neither of the ensembles has a better spread consistently across all lead times (not shown): UM-SINGV starts with a better spread in temperature and wind speed up to about T + 18 h, in agreement with the results of Park *et al.* (2008). The spread in precipitation is the subject of section 4.

To complete this analysis, we also take into account the difference in ensemble size between the two ensembles: EC-SINGV has 51 members while UM-SINGV has 12 members. Sensitivity to ensemble size has been the subject of many studies (Buizza and Palmer, 1998; Clark *et al.*, 2011; Schwartz *et al.*, 2014; Hagelin *et al.*, 2017; Raynaud and Bouttier, 2017). However, there has been little study of the ensemble size of convective-scale ensembles over the Tropics. Here, for each day of the trial period, a new random set of members is used by the verification, creating a random selection of members throughout the month (each draw automatically includes the control member of the ensemble). Figures 2, 3 and 4c,d display the impact of ensemble size on the EC-SINGV ensemble on the ranked probability scores for precipitation, temperature and wind speed respectively for an ensemble of 6, 12, 18, 24 and 30 members against the full-size EC-SINGV ensemble for the 1500 UTC forecasts. A lead time of T + 15 h corresponds to the initiation of the daytime convection and a lead time of T + 30 corresponds to the night-time convection. Using the 1500 UTC forecasts only has the advantage of comparing the verification results directly against the local time and the diurnal cycle, and so to understand the growth of errors corresponding to the diurnal cycle.

The largest change in ensemble skill occurs between an ensemble size of 6 to 12 members (this is in agreement with for example Buizza and Palmer (1998) and Raynaud and

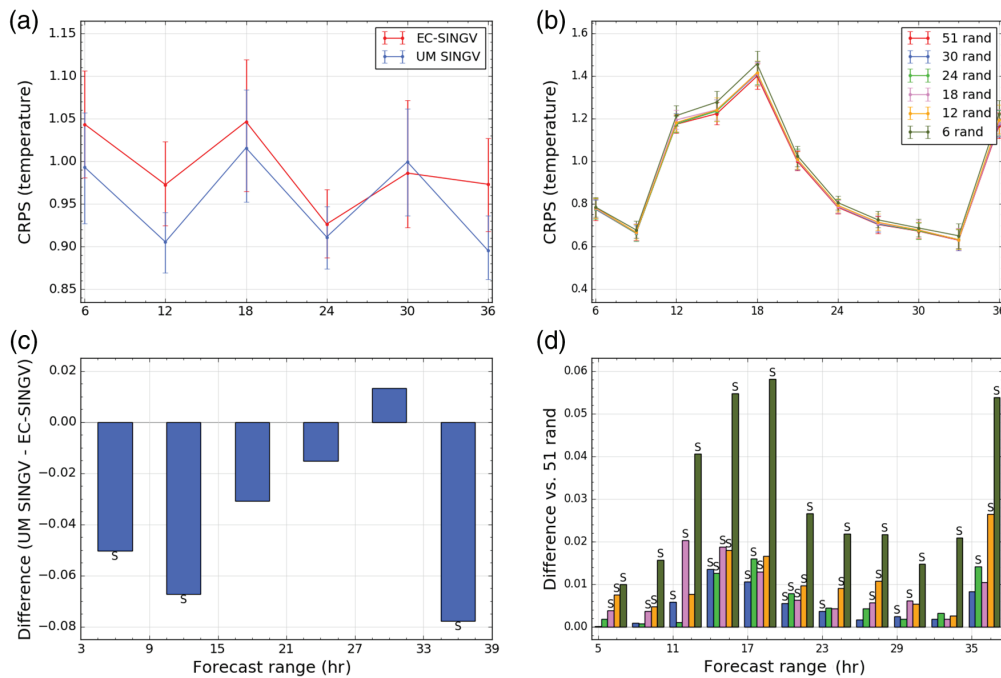


FIGURE 3 Continuous ranked probability scores for temperature at 1.5 m for October 2017. (a) Comparison UM-SINGV (18 members) with EC-SINGV (51 members) for simulations starting at 0300 and at 1500 UTC, (b) Difference plots between UM-SINGV and EC-SINGV, (c) Ensemble sizes of 6, 12, 18, 24 and 30 are here compared for the continuous ranked probability scores of the full ensemble EC-SINGV for simulations starting at 1500 UTC only, and (d) Difference plots against the full-ensemble EC-SINGV. A neighbourhood size of 3 grid points is used here. Ninety per cent confidence intervals are calculated using Monte-Carlo method [Colour figure can be viewed at wileyonlinelibrary.com]

Bouttier (2017)). However, the impact of further increasing the size of the ensemble to 18 members is more important for the wind speed than for the temperature (so a size of 18 members starts to be more comparable with the full-size ensemble as in Schwartz *et al.* (2014)). Significant, albeit small, differences can still hold even for an ensemble size of 30 members (again in agreement with studies over midlatitudes as in Raynaud and Bouttier (2017) and Hagelin *et al.* (2017)). The growth in errors happens during the daytime and night-time convection and the impact of ensemble size is largest at these times. There is also some evidence that for the temperature, the 6-member ensemble is most dissimilar to the 51-member ensemble when the error growth peaks. Similar results are found for precipitation: an ensemble size of 12 members is essentially enough to capture the skill of the 51-member ensemble, except over the T + 15 to T + 21 period (i.e. the daytime convection), during which a larger ensemble size, as 24 members, may be needed.

The resolution of the ensemble for different sizes is also studied using areas under ROC. For the wind speed, there are some lead times at which a smaller size ensemble is as skilful as the full-size ensemble (not shown). For the 6-hourly precipitation, the ROC area improves with increasing ensemble size at low thresholds (Figure 7a). At higher thresholds however (i.e. 32 mm in Figure 7b, as opposed to lower thresholds as in Figure 7a), the smaller sized ensembles are as skilful as the full-size ensembles at some lead times. Although the skill of

the ensemble does depend on ensemble size (Figures 2–4), the spread (i.e. standard deviation between the members and the ensemble mean) does not vary with ensemble size. This is in agreement with the results for the precipitation in section 4.

Reliability diagrams for wind speed and precipitation were also compared, revealing similar behaviours for the two ensembles (not shown).

An important note to make here is that given UM-SINGV uses only 12 members, the differences in ranked probability scores in Figures 2, 3 and 4 between the full-size EC-SINGV and a 12-member EC-SINGV ensemble are as large as the differences between the UM-SINGV and the full-size EC-SINGV ensemble from the left panels of the same figures.

Finally, we did not decide to reduce the ensemble size of EC-SINGV based on these results for later sections. To exploit the full capability of the EC-SINGV ensemble, the 51 members of EC-SINGV are used for the analysis of the case-studies in section (this is so we get a better chance to assess the behaviour of the extreme members as explained in Clark *et al.* (2011)). The impact of ensemble size will be revisited in section 4.

3.2 | Monthly mean characteristics of the diurnal cycle

To examine the spin-up and diurnal behaviour of the characteristics of our UM-SINGV and EC-SINGV ensembles,

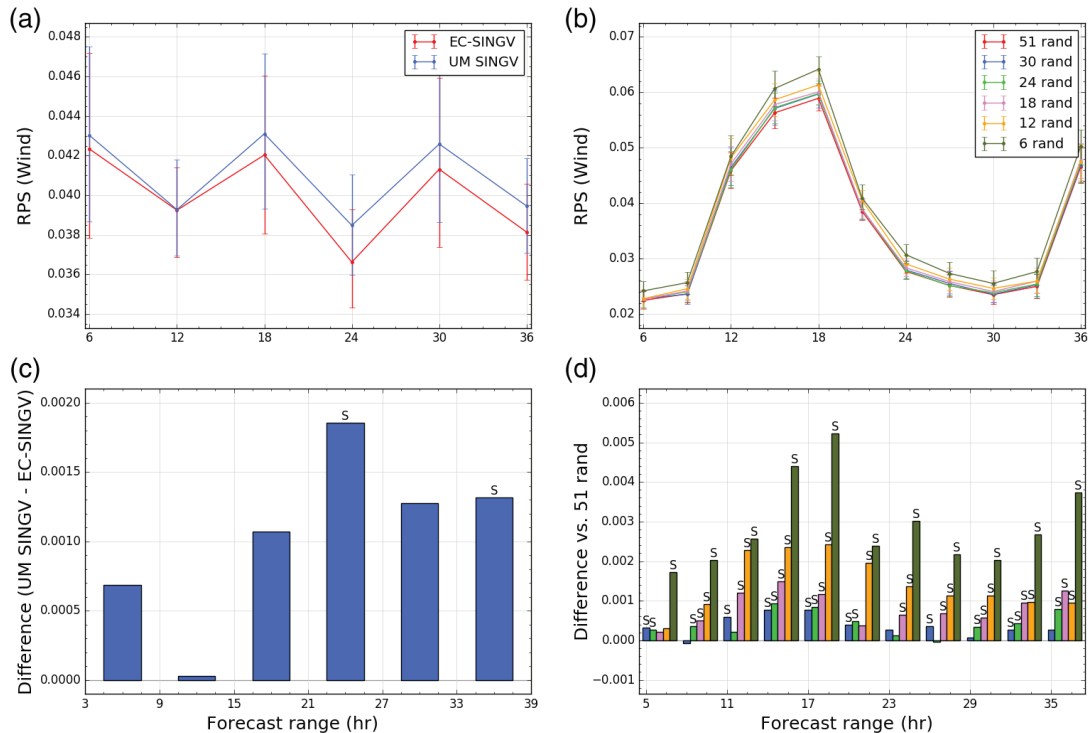


FIGURE 4 Ranked probability scores for 10 m wind speed for October 2017. (a) Comparison UM-SINGV (18 members) with EC-SINGV (51 members) for simulations starting at 0300 and 1500 UTC, (b) Difference plots between UM-SINGV and EC-SINGV, (c) Ensemble sizes of 6, 12, 18, 24 and 30 are here compared for the continuous ranked probability scores of the full ensemble EC-SINGV for simulations starting at 1500 UTC only, and (d) Difference plots against the full-ensemble EC-SINGV. A neighbourhood size of three grid points is used here. Ninety per cent confidence intervals are calculated using Monte-Carlo method [Colour figure can be viewed at wileyonlinelibrary.com]

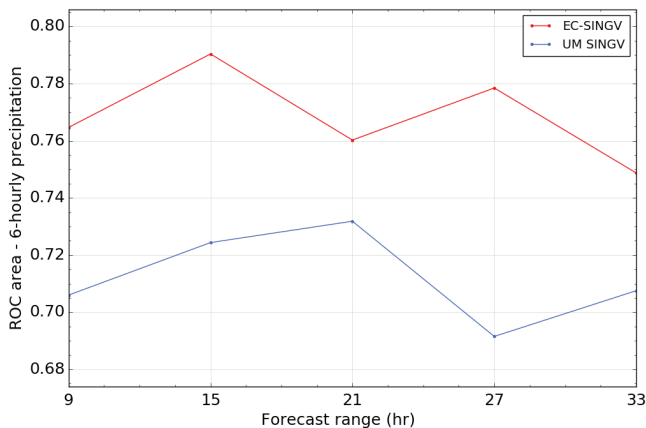


FIGURE 5 ROC area for UM-SINGV and EC-SINGV (full size) for 6-hourly precipitation for a threshold of 4 mm. A neighbourhood size of three grid points is used here [Colour figure can be viewed at wileyonlinelibrary.com]

we have plotted the mean rainfall over the “Full domain” and “Singapore domain” (Figure 1), for the whole month of October for both EC-SINGV and UM-SINGV (Figure 8). In this region, rain occurs frequently, so while a month-long period is not a particularly long time for comparison, rain occurs for about 20 days in total during October 2017, hence we believe the results are meaningful. From Figure 8, we see four main characteristics. First, UM-SINGV (red) has an

earlier development in the spin-up (see for example T + 5 or 2000 UTC or 0400 local time), associated with night-time convection. Secondly, EC-SINGV has more rain over land (Figure 8c,d) between T + 15 and T + 20 h, which corresponds to the afternoon convection (i.e. between 0600 and 1100 UTC, or between 1400 and 1900 local time). Thirdly, UM-SINGV is drying out more quickly than EC-SINGV following the afternoon convection (Figure 8c–d, around T + 20 h). Finally, we see that UM-SINGV (red) again shows an earlier development in the night-time convection for the second day of the forecast, see T + 27 h (Figure 8b,d). So, UM-SINGV has stronger night-time convection than EC-SINGV at both T + 3 h and T + 27 h (i.e. 1800 UTC or 0200 local time).

When comparing against the GPM (Global Precipitation Measurements) satellite data, a spin-up time of about 6 h is clearly visible in Figure 8a–c. Figure 8c shows that the peak in the diurnal cycle (T + 15 to T + 20 h) is too strong in both ensembles, but there is less bias in UM-SINGV. The dissipation of the convection is however better captured by EC-SINGV. In Figure 8a, EC-SINGV is also better at capturing the amount of rain over both land and sea. So again, from this set of observations, both ensembles show different characteristics and neither of them clearly outperforms the other. Note also that a satellite-based product such as the GPM is naturally likely to detect deeper and more mature convection

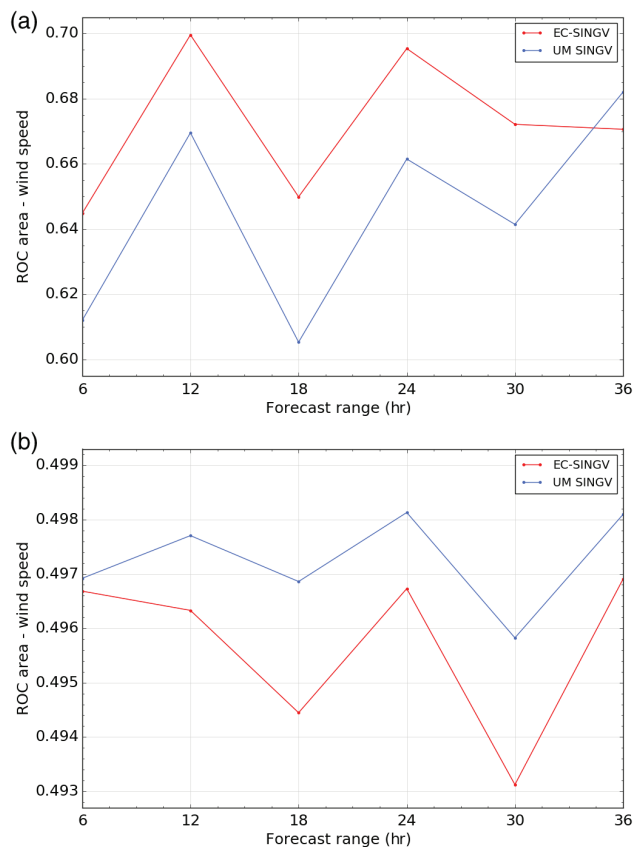


FIGURE 6 ROC areas for UM-SINGV and EC-SINGV (full size) for 10 m wind speed. A neighbourhood size of three grid points is used here. (a) Threshold of 5.65 m s⁻¹, (b) threshold of 8.74 m s⁻¹ [Colour figure can be viewed at wileyonlinelibrary.com]

(with anvils) more easily than younger developing storms, and may be biased towards a later afternoon peak than might actually occur in reality.

In the rest of this article, we have used radar imagery to examine the local spatial development and organization of the showers, since the GPM was only used to identify domain-wide patterns of the diurnal cycle. A radar scan of 240 km around Singapore is used. Within Singapore, the radar data was well calibrated using a dense network of rain-gauges. Further evaluation of the impact of the convective-scale model nested within the global EC and global UM models will be done in future work in the context of the deterministic model.

4 | SPATIAL ENSEMBLE SPREAD

We will now examine the spatial differences between the ensemble members, and hence the spatial ensemble spread, for the four case-studies. The objective is to determine the contribution to the spatial ensemble spread in both UM-SINGV and EC-SINGV from the ICs and LBCs. We do this by using the dispersion metric for the Fractions Skill Score (dFSS) as in Dey *et al.* (2014) as defined in section

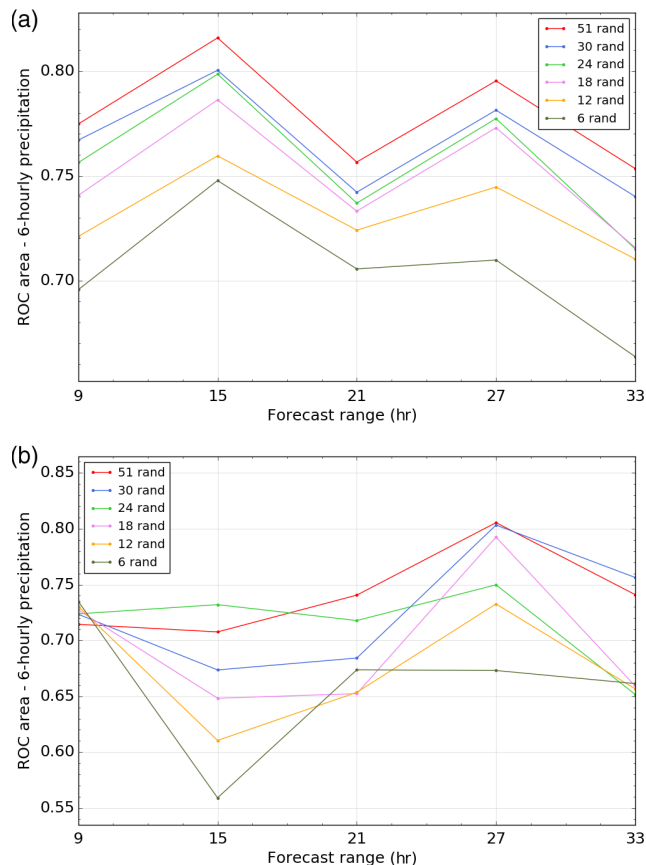


FIGURE 7 Impact of ensemble size on EC-SINGV on the ROC areas for the 6-hourly precipitation. Ensemble size of 6, 12, 18, 24 and 30 are here compared against the ROC areas of the full ensemble EC-SINGV for simulations starting at 1500 UTC. A neighbourhood size of three grid points is used here. (a) Threshold of 0.5 mm, (b) threshold of 32 mm [Colour figure can be viewed at wileyonlinelibrary.com]

2.2. In order to do so, we have chosen four case-studies representative of high-impact rainfall. We will show the variability for each day as well as the day-to-day variability of the dFSS metric over the four case-studies and assess the importance of this variability compared to the mean over the four days.

4.1 | Analysis of case-studies

October is in the inter-monsoon season and can have varied wind directions. Lo and Orton (2016) show that the number of squall-line events peaks in October and November. As illustrated in their paper, squall lines tend to initiate over Sumatra and the Straits of Malacca. Convergence is often found along the Straits of Malacca late in the evening. Given suitable wind conditions, the squalls move eastwards towards Singapore, as shown in the Hovmöller diagrams in Fig. 2 of Lo and Orton (2016). Additional mesoscale circulations come from the sea-surface temperature and the sea-breeze/land-breeze systems. Sea breezes are dependent on the combination of multiple parameters such as atmospheric stability, time-averaged integrated surface heat flux near the coastline, offshore wind,

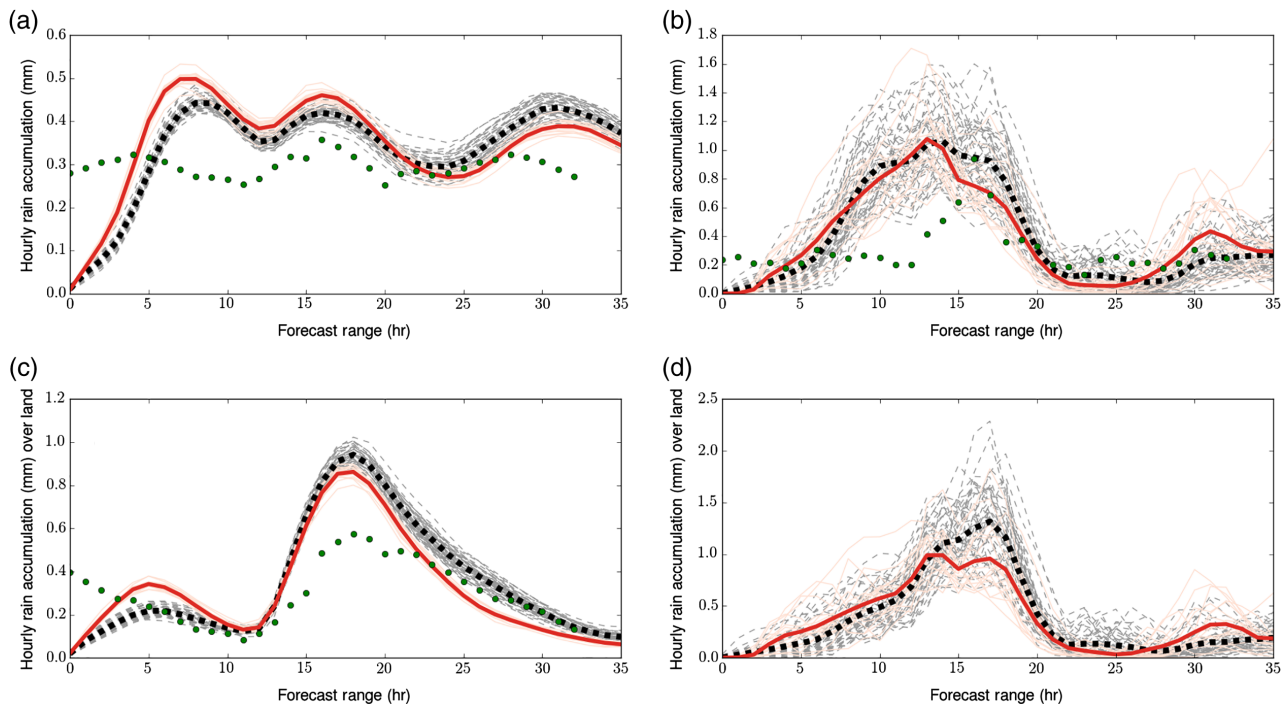


FIGURE 8 Monthly analysis of hourly rainfall accumulations as a function of lead times for October 2017 for UM-SINGV in red and EC-SINGV in black. This mean analysis is calculated over land and sea for the “Full domain” in (a), over land only over the “Full domain” in (b), over land and sea over the “Singapore domain” in (c) and over land only over the “Singapore domain” in (d). For information about the domains, please see Fig 1. GPM data (see text for details) are illustrated with small circles [Colour figure can be viewed at wileyonlinelibrary.com]

as well as local topography (Porson *et al.*, 2007a, 2007b). The uncertainty in the initial development of the squalls is a result of these complex interactions in the local winds as well as secondary development related to cold pools. An ensemble approach is needed to capture this uncertainty, which cannot be specified in a deterministic forecast.

The four case-studies were chosen on the basis of evidence for squall lines or convective storms from local rain-gauge data and radar animations. The performance of the two ensemble systems is examined for two squall-line cases (making landfall at Singapore in the early mornings of 9 and 12 October 2017) and two afternoon rain cases, with heavy rain hitting Singapore in the late afternoons of 7 and 30 October 2017.

Southwest Monsoon conditions with low-level winds blowing from the southwest prevailed over Singapore and the surrounding region on 9 and 12 October 2017. A tropical depression formed over the central part of the South China Sea on the morning of 9 October while tropical cyclone *Khanun* formed as a tropical depression over the western North Pacific on the morning of 12 October. Formation of tropical cyclones over the South China Sea and the western North Pacific can draw low- to mid-level strong winds (925 to 700 hPa) over the Singapore and Sumatra region which is favourable for the formation of Sumatra Squalls. The occurrence of Sumatra squalls on 9 and 12 October also brought moderate to heavy thundery showers to many parts of Singapore in the morning. The highest daily rainfall recorded

during the period was 129.8 mm around Tuas (westernmost part of Singapore) on 12 October 2017.

Since squall lines usually happen in the early morning from about 2200 UTC (0600 local time) to approximately 0300 UTC (1100 local time) and the forecasts are initialized at 1500 UTC, we focus on the early morning event on the second day of the forecast (lead times of $T + 30$ to $T + 36$ h), to avoid the spin-up period and allow for more differences to develop between forecasts.

In the first case (9 October) in Figure 9, both ensembles capture the location of the developing squall line along the Straits of Malacca, with higher probabilities for UM-SINGV, although EC-SINGV better captures the formation of the arc. In the second case (12 October) in Figure 10, the probabilities are, again, higher for UM-SINGV than for EC-SINGV for the threshold shown. For 12 October, in both ensembles, the control members of the ensembles (i.e. without the perturbation in ICs from the perturbed members of the respective parent ensembles) miss the squall line (not shown). This highlights the benefit of running convective-scale ensembles for extreme cases, compared to deterministic models, in order to obtain a bigger range of scenarios.

We have also repeated this analysis for the two cases of afternoon convection, 30 October and 7 October 2017. On both days, the Southwest Monsoon gave way to inter-monsoon conditions where winds became lighter and more variable. During the day, thundery showers developed in the afternoon and persisted into the evening, due to strong

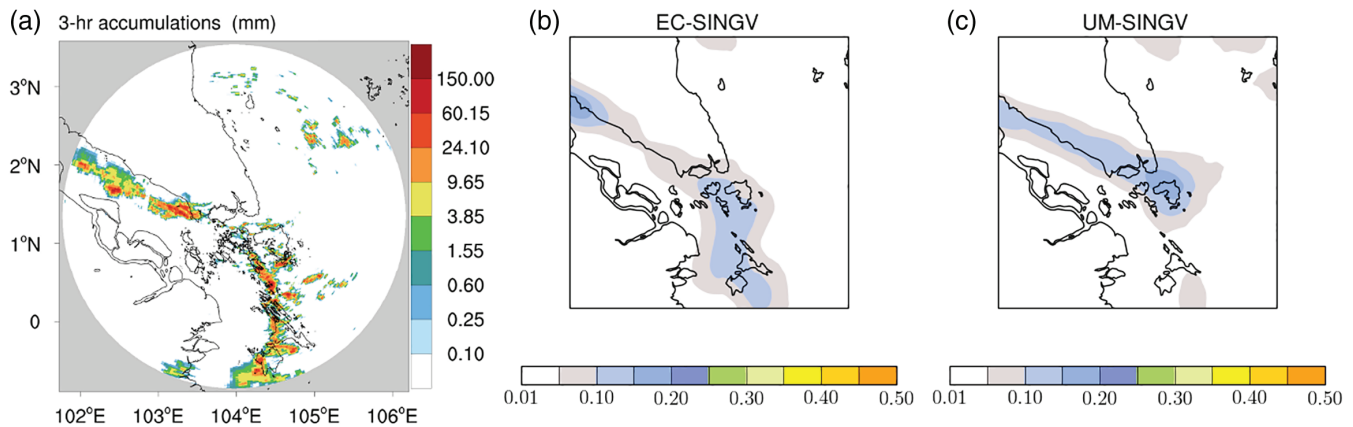


FIGURE 9 Date 9 October. (a) Radar at [2300 UTC 8 October–0200 UTC 9 October], (b) Neighbourhood ensemble probability (NEP) of accumulated rainfall 2300–0200 UTC to exceed rain amounts of 12 mm over the 3 h for EC-SINGV, and (c) UM-SINGV. The simulations are initialized at 1500 UTC 7 October. Note that the probability scale ranges from 0 to 0.5 [Colour figure can be viewed at wileyonlinelibrary.com]

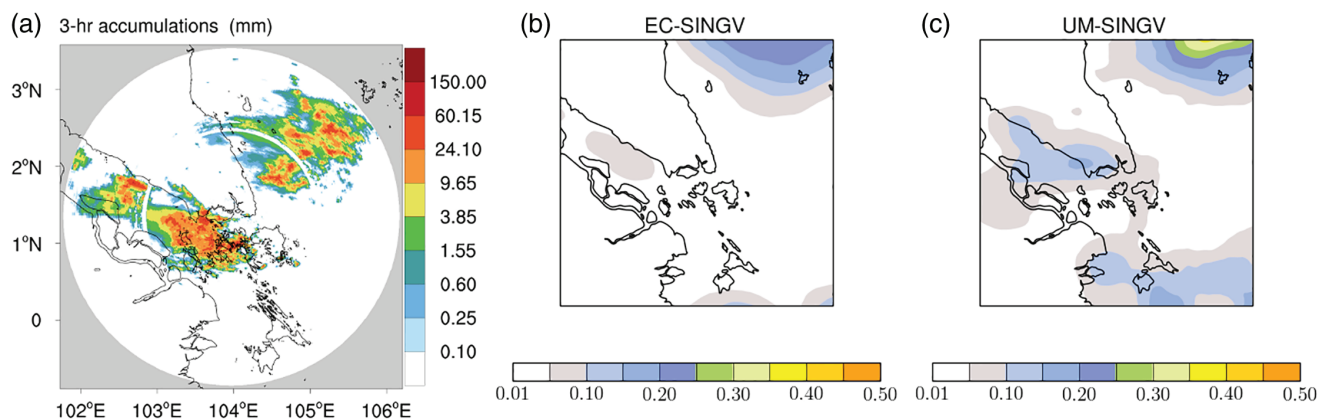


FIGURE 10 Date 12 October. (a) Radar at [2300 UTC 11 October–0200 UTC 12 October], (b) neighbourhood ensemble probability (NEP) of accumulated rainfall 2300–0200 UTC to exceed rain amounts of 3 mm over the 3 h for EC-SINGV, and (c) UM-SINGV. The simulations are initialized at 1500 UTC 10 October [Colour figure can be viewed at wileyonlinelibrary.com]

solar heating of land areas coupled with the convergence of winds over Singapore and the surrounding area.

On 30 October, storms developed around 0900 UTC, following the northern shoreline of the Straits of Malacca. By 1100 UTC, the rain band had moved to Singapore. Figure 11 shows that both ensembles can represent the risk of heavy precipitation at the right location and time, with higher probabilities for UM-SINGV. UM-SINGV also shows high values of probability over Peninsular Malaysia, within the limits and outside of the radar coverage area.

For 7 October, Figure 12 shows similar spatial variability between the two ensembles over this 3-hour window. But the details of the precipitation over the northeast of Sumatra Island and over the Malaysian Peninsula are different between the two ensembles. UM-SINGV is wetter generally over Sumatra Island, with some of these areas lying outside the radar coverage area.

Again, based on this subjective evaluation of these rain events, there is little evidence here to suggest that one

ensemble is much better than the other, which matches the inconclusive results from the objective verification.

The ensembles were also compared at a smaller scale over the 20×20 grid-point Singapore domain. Figures 13 and 14 compare the ensembles for the squall-line cases and afternoon rain cases respectively.

In Figure 13, we use meteograms or box plots to represent the mean hourly rain amounts (kg/m^2) over the “Singapore domain”, covering 20 by 20 grid points (i.e. 90 by 90 km) in each direction, centred on Singapore (Figure 1). The size of the domain is chosen to be big enough so that the ensembles have some skill at predicting the rainfall at this length-scale (see the monthly analysis of Figure 8 for more detail), but still small enough to focus on the Singapore area. There is a notable difference in the predictive skill of these two cases: in the first case, both ensemble means show a marked increase in the domain mean precipitation around the period of interest (2100–0300 UTC), while in the second case, only the mean of UM-SINGV has a small increase. There is a small difference in the timing of the squall line for the first

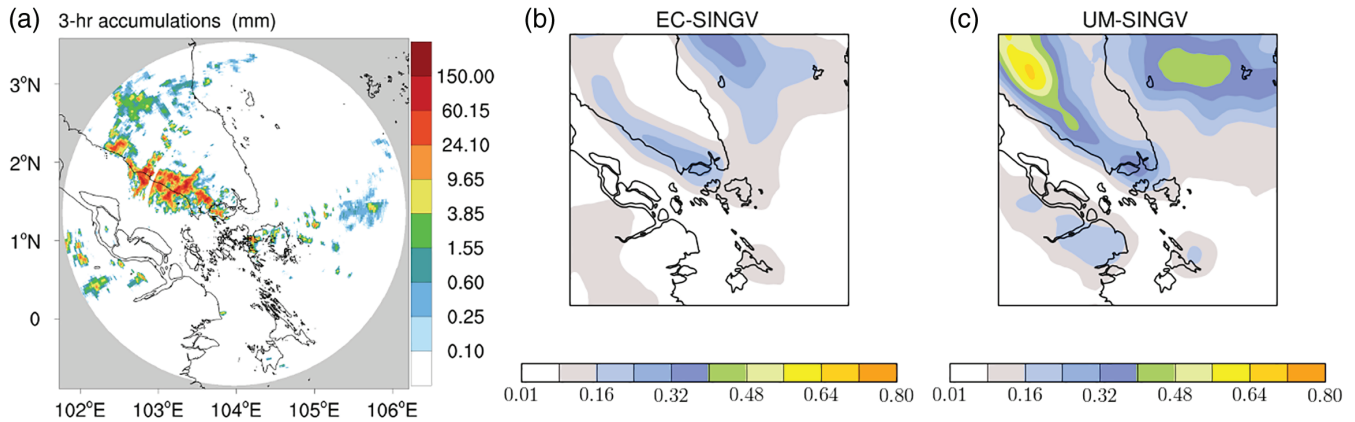


FIGURE 11 (a) Radar at [0800 UTC–1100 UTC] on 30 October, (b) neighbourhood ensemble probability (NEP) of accumulated rainfall 0800–1100 UTC on 30 October to exceed rain amounts of 6 mm over the 3 h for EC-SINGV, and (c) UM-SINGV. The simulations are initialized at 1500 UTC 29 October [Colour figure can be viewed at wileyonlinelibrary.com]

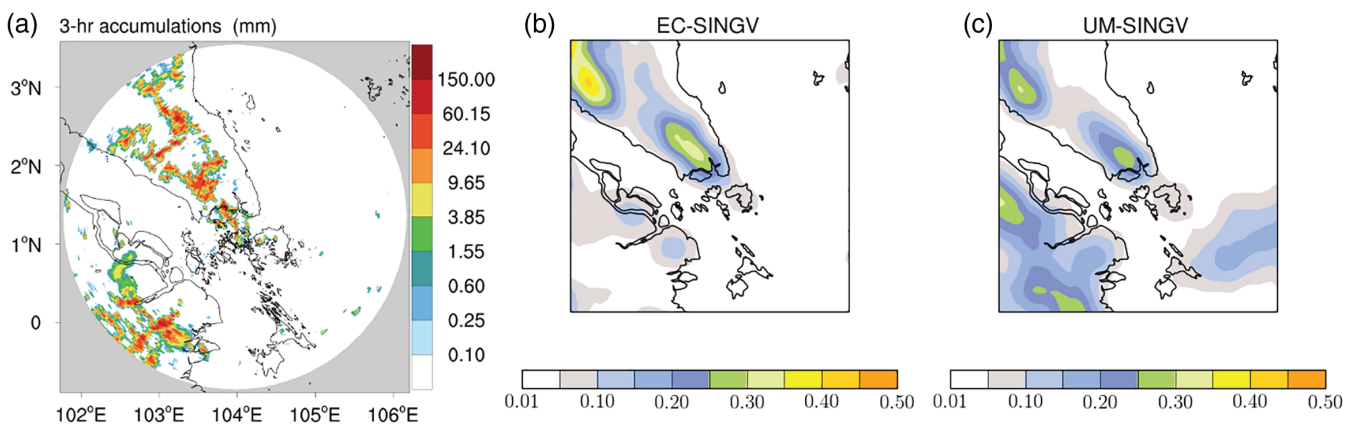


FIGURE 12 (a) Radar at [0800 UTC–1100 UTC] on 7 October, (b) neighbourhood ensemble probability (NEP) of accumulated rainfall 0800–1100 UTC on 7 October to exceed rain amounts of 6 mm over the 3 h for EC-SINGV, and (c) UM-SINGV. Simulations are initialized at 1500 UTC 6 October [Colour figure can be viewed at wileyonlinelibrary.com]

case, with UM-SINGV 2–3 h earlier. The rainfall amount is larger for UM-SINGV in both cases. In terms of the extremes (outliers in circles), the range seen in both ensembles from these meteograms is fairly similar. Despite the smaller number of members in UM-SINGV, the extremes in UM-SINGV cover the same range as the extremes in EC-SINGV. In the second case, it is particularly noticeable that a lot of the members in both ensembles are just simply “dry” (i.e. little-to-no precipitation in the region of interest) and only a few of them show heavy rain, indicating less predictability.

The same characteristics regarding the spread and the extremes of the ensembles apply to the afternoon rain cases. In Figure 14, we use bin diagrams depicting the number of hourly events for each member, associated with each bin, over the same box of 20 by 20 grid points. The bin ranges used here are in mm per hour: 0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128 and 256. These diagrams for both afternoon rain cases give similar mean amounts of rain for the afternoon events from both ensembles, with more extreme values (i.e. extreme members) in EC-SINGV for 30 October. On 7 October, for

hourly accumulations of over 40 mm per hour, UM-SINGV has, however, more spread than EC-SINGV.

The comparison of the two ensembles so far has shown that EC-SINGV and UM-SINGV have a different spatial organization of convection for these heavy rain events (i.e. squall line and afternoon rain) in some cases, with some areas wetter than others and different squall-line developments. However, they also show similar distribution in their extremes, which particularly indicates that UM-SINGV is able to capture a range of outcomes despite having fewer members.

4.2 | Dispersion metric dFSS applied to each case-study

We have chosen to compute the dFSS using a threshold of 2 mm/h. All these calculations are done over a domain slightly smaller than the full domain, which excludes 14 points (i.e. corresponding to the maximum neighbourhood size used here) from the boundaries of the full domain, Figure 1. We did not examine the first 6 h to avoid model spin-up and because the FSS is sensitive to the frequency

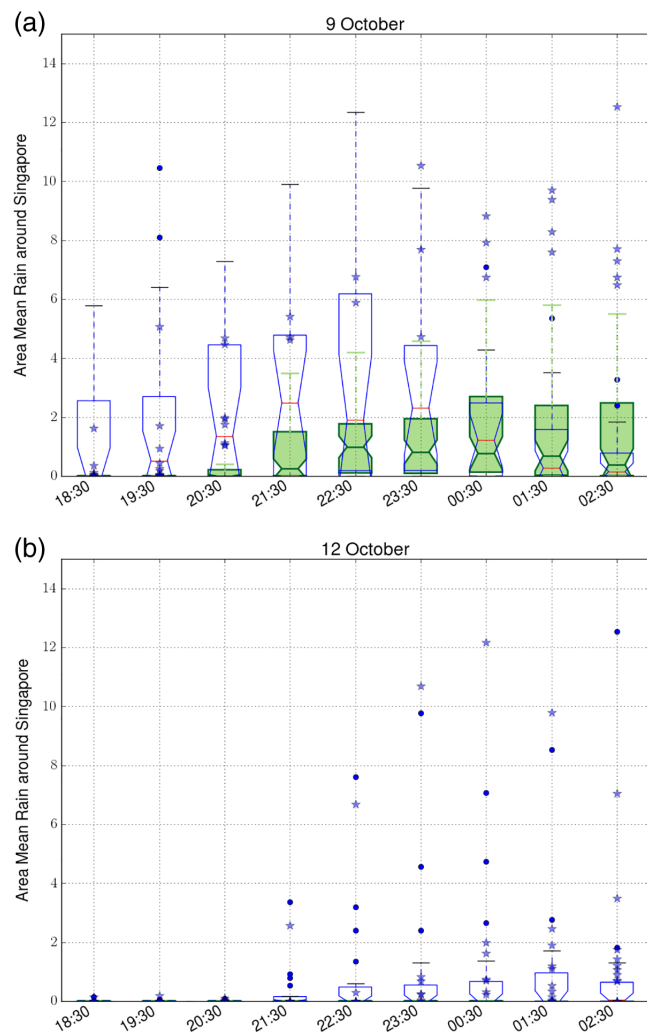


FIGURE 13 Meteorograms representing hourly rain accumulations over the “Singapore domain” (see Fig 1 for reference) for the 9 October in (a) and 12 October in (b). The UM-SINGV ensemble is represented with white boxes, red lines for medians, and blue circles for outliers; the EC-SINGV ensemble is represented with green boxes, green lines for medians and blue stars for outliers. The boxes represent the first and third quartiles, while the whiskers are set at the $1.5 \times \text{IQR}$ (interquartile range) from below and above the first and third quartiles respectively. Outlier circles here are thus found above the 75th centile $+1.5 \text{ IQR}$ [Colour figure can be viewed at wileyonlinelibrary.com]

bias which can tend to become larger when rainfall coverage is small.

Spatial comparison may depend upon ensemble size. Leoncini *et al.* (2013) found that an ensemble size of 8 was large enough to capture the time evolution of the anomalies relative to a 50-member ensemble for a flash-flood event for the United Kingdom; however, there was not much spread between members in that study. Hagelin *et al.* (2017) found that ensemble size matters mostly for the spatial details in the precipitation areas, so perhaps influencing the smallest spatial scales. To check if the impact on the spread of the precipitation is dependent upon the size of the ensemble, we have recalculated dFSSmean for three sets of ensemble members

from EC-SINGV (Figure 15), approximately equal in size to UM-SINGV. Set1 uses members 0 to 17. Set2 uses members 0 and then 18 to 34; set3 uses members 0 and then 35 to 50. These three sets sample all the ensemble members. Figure 15 shows that the smaller ensemble samples from EC-SINGV have dFSSmean values very much like the full ensemble with the dots lying almost exactly along the diagonal. Therefore, for these cases, we have reasonable confidence that the results from the comparison of EC-SINGV and UM-SINGV should not be significantly affected by the bigger size of EC-SINGV (51 members), compared to UM-SINGV (18 members). We do not claim that this is a statistically significant result for all situations; for that a larger sample would be needed.

As in Dey *et al.* (2014), Figure 16 shows dFSSmean as a function of neighbourhood size (i.e. spatial scale) and forecast time for EC-SINGV for all four cases (a–d) and for the mean over these four cases in Figure 16e. Similar behaviour can be seen across all the cases. It is most apparent that the spatial agreement between members increases with increasing spatial scale (dFSS is larger for larger neighbourhoods) showing that the spatial ensemble spread is greatest at small scales. The dFSS is larger around 0900 UTC, which is in the afternoon local time when convection is most widespread and reflects the greater spatial predictability at that time. Three of the cases (9 October being the exception) show an increase in the spatial agreement at early times up to 0100 UTC, which probably indicates that the ensemble is still adjusting to the perturbations and increasing precipitation coverage (the impact of the initial perturbations will be studied later on).

Roberts and Lean (2008) and Roberts (2008), when comparing forecasts with radar, defined a smallest skilful scale as the neighbourhood length at which $\text{FSS} = 0.5 + f/2$, where f is the fractional coverage of pixels exceeding the threshold on the domain. Skok (2015) and Skok and Roberts (2018) show that a value of $\text{FSS} = 0.5$ can be used instead because the neighbourhood size at which $\text{FSS} = 0.5$ is effectively a direct measure of the mean spatial agreement between two fields (as long as the frequency bias is small). In an ensemble context, values of dFSSmean of at least 0.5 give the scales (neighbourhood sizes) over which the forecasts have sufficient agreement (sufficiently low spread) to have confidence in the prediction of the outcomes provided by the ensemble forecast. This, of course, assumes that the ensemble has a good representation of the true spatial uncertainty. We see there is little agreement between members at small scales: dFSSmean is smaller than 0.5 for small neighbourhood lengths and dFSSmean is larger than 0.5 for neighbourhood lengths longer than 10 to 15 grid points (~ 40 to 70 [$15 \text{ times } 4.5$] km). The 12 October has even less agreement at small scales than the other cases, with good agreement only at scales larger than $\sim 120 \text{ km}$ other than around 0900 UTC when agreement is at scales larger than $\sim 70 \text{ km}$ at the time

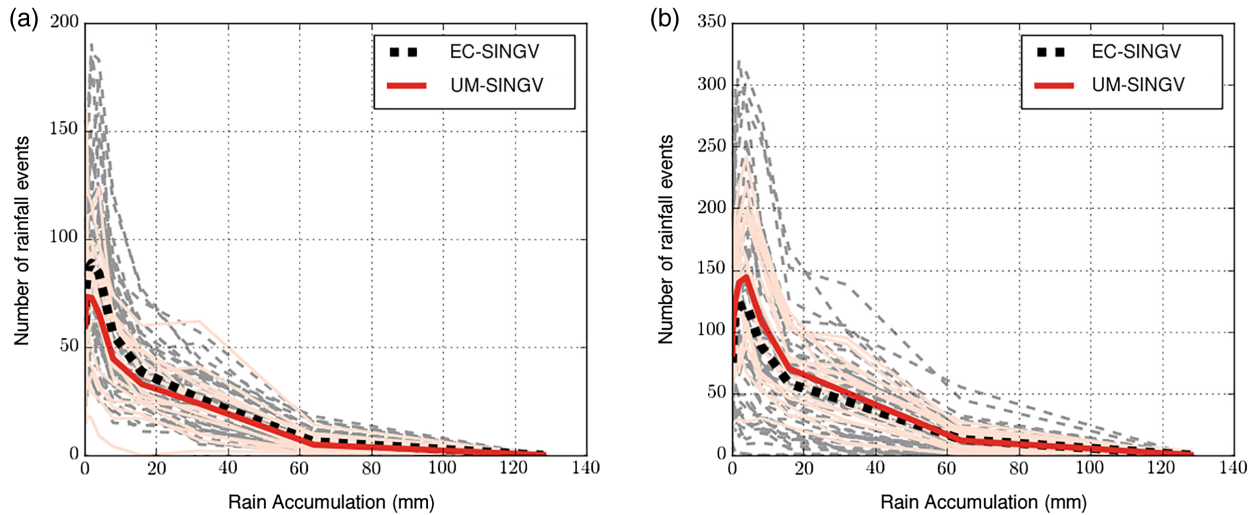


FIGURE 14 Bin diagrams of hourly accumulated rainfall over the “Singapore domain”, 0400 UTC–1500 UTC, for simulations starting at 1500 UTC the day before. (a) 7 October, (b) 30 October 2017. Ensemble members are depicted in thin lines and the mean over the members in thick lines [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

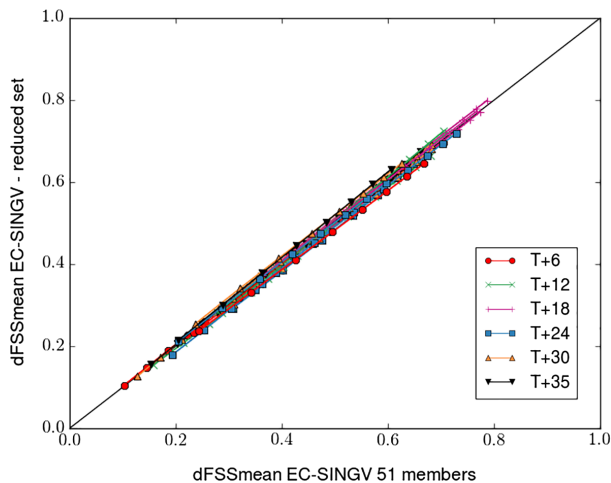


FIGURE 15 Comparison of dFFS for all the case studies between different sets of ensemble members defined as set1, set2 and set3 (see text for more explanation) and EC-SINGV for different lead times (as in legend) and different neighbourhood scales (3 by 3, 7 by 7, 11 by 11, 15 by 15, 19 by 19, 23 by 23, 27 by 27 and 31 by 31 points) [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

of maximum convection. This reflects the large disagreement between members for this case, as seen in Figure 13.

The main characteristics of the diurnal cycle and the dependence on neighbourhood size are still present in the mean in Figure 16e.

These results can be compared to those of Leoncini *et al.* (2013). They found that the least spatial agreement between forecasts occurred before and after the development of storms because the uncertainty in the initiation and decay dominated the convection at those times. The best spatial agreement occurred during the most intense part of the storm. These results are consistent with the findings in Figure 16. In all the diagrams, an increase in dFFSmean happens around 0900 UTC (i.e. 1700 local time, which corresponds to the daytime

convection). Although all these plots show an increase in FSS during the afternoon, they do not reveal any second increase in FSS related to the development of the squall line later in the simulations. This is probably an indication of much lower predictability for that feature. An analysis over a smaller domain focused around Singapore (not done here) may be helpful to determine the impact of these squall lines on the spread, although care is needed in the choice of domain size because the use of too small a domain might introduce significant errors associated with proximity to the boundaries (Skok and Roberts, 2018) and make the scores more erratic for such a small sample.

4.3 | Spatial spread in UM-SINGV and EC-SINGV

To illustrate the differences in the ensemble spread between UM-SINGV and EC-SINGV, we plot the differences in dFFSmean between UM-SINGV and EC-SINGV for the four cases in Figure 17a–d and for the mean over the four cases in Figure 17e. Figure 17 shows that at the start of the forecasts, dFFSmean is higher in UM-SINGV (red colours) than it is for the EC-SINGV at all scales. Therefore the spread in UM-SINGV is smaller at the start of the forecasts. However, for later forecast hours, dFFSmean from UM-SINGV is smaller than from EC-SINGV (blue colours), particularly at larger scales, indicating that the spatial spread in UM-SINGV is larger. Averaging over the four cases (Figure 17e) for this type of metric reveals a much smoother response than on a day-to-day basis, but still retains the important signal that the EC-SINGV ensemble has more spread at the start and the UM-SINGV ensemble later. However, this suggests that individual case-study analysis may reveal higher differences in this distribution of spread and so averaging this type of product over a long period may sometimes be misleading.

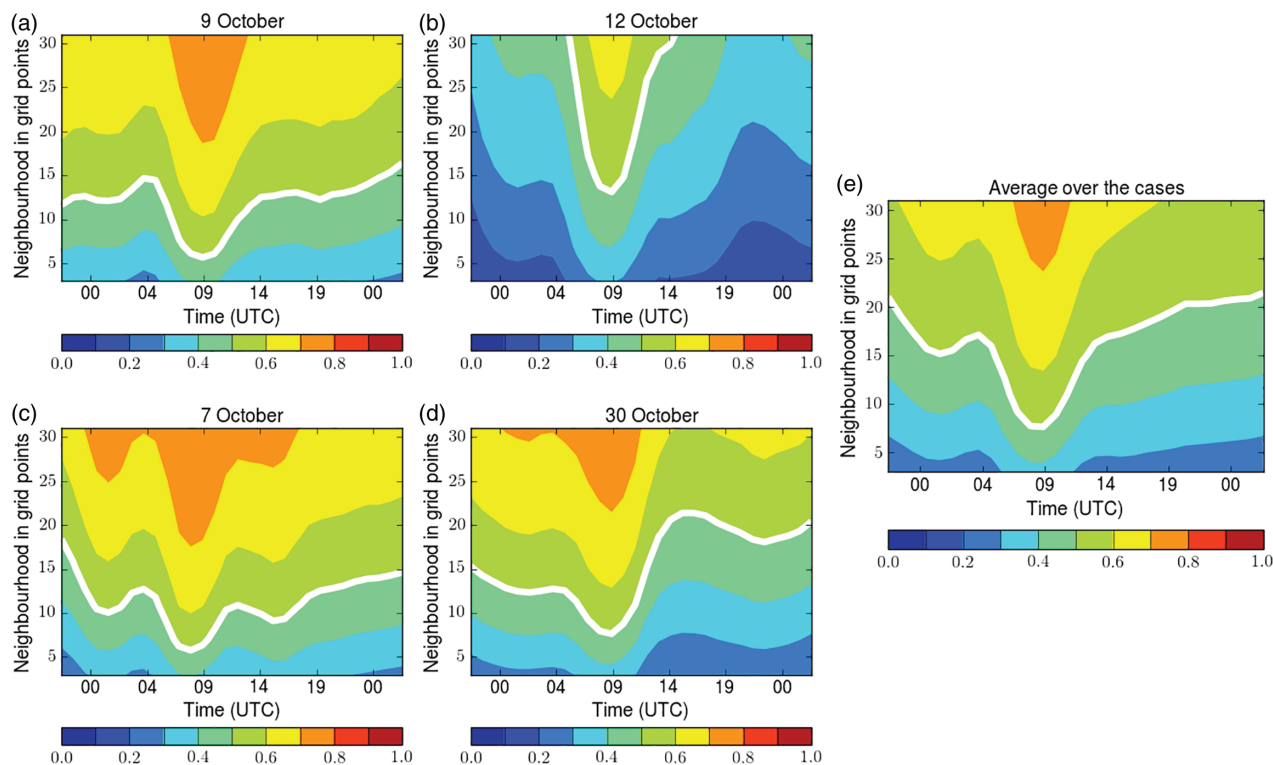


FIGURE 16 Diagram representing dFSSmean (see text) as a function of neighbourhood size and time of day for EC-SINGV and for the four case-studies. Values of dFSSmean of 0.5 are depicted in white. Different neighbourhood scales (3 by 3, 7 by 7, 11 by 11, 15 by 15, 19 by 19, 23 by 23, 27 by 27 and 31 by 31 points) are used here. (a) 9 October, (b) 12 October, (c) 7 October, (d) 30 October 2017, (e) average over all cases [Colour figure can be viewed at wileyonlinelibrary.com]

Note that these results do not indicate whether larger or smaller spread is better, only that the two ensembles show different behaviour.

As illustrated in sections 3.2 and 4.1, the two SINGV-EPS ensembles produce different organization in the development of the convection, so it is not surprising to see differences in the spread of the precipitation from these ensembles.

The particularly large differences from 1800 UTC onwards on 9 October occur at the time of the squall line. The smaller spread in EC-SINGV may be associated with the presence of more developed convergence lines in more members in EC-SINGV than in UM-SINGV, which would make the EC-SINGV ensemble more confident than UM-SINGV about rainfall occurrence. Note, though, that the dFSS analysis is carried out for the whole domain (“Full domain”, Figure 1) and therefore not all the signal is associated with the squall line. As mentioned previously, looking at a smaller domain might be helpful to isolate the signal, but also brings additional difficulties, particularly if one ensemble has more rain than the other, increasing the frequency bias. While it is beyond the scope of this article to assess which characteristics of the ICs in the parent ensembles are responsible for these differences in the spread, the differences in spread may be related to the differences in the perturbations in SSTs given to EC-SINGV and UM-SINGV, as well as the characteristics of the parent ensembles, as described in section 2.1.

A common problem with convective-scale EPS is that they are often found to be under-spread. Combining different convective-scale ensembles has been used in Beck *et al.* (2016). They show that combining EPS models gives a better performance against observations and increases ensemble spread, when using grid-scale metrics. Since we are examining two ensembles, the next step is to compare the dFSSmean metric for a combination of the two ensembles (i.e. 69 members including the 51 EC-SINGV members as well as the 18 UM-SINGV members) to the original UM-SINGV and EC-SINGV to see how this affects the spatial spread.

Figure 18a shows that, as in Figure 17, the UM dFSSmean are larger (less spread) at earlier times in the forecasts and smaller (more spread) at later times. Figure 18b,c show how the dFSSmean from EC-SINGV and the dFSSmean from UM-SINGV compare against the two SINGV-EPS ensembles combined (here EC + UM). This shows that the EC + UM ensemble has smaller dFSSmean (higher spread) than EC-SINGV alone, but the results are different when it comes to comparing EC + UM against UM-SINGV. Indeed, after T + 6 h, the dFSSmean from EC + UM are larger (smaller spread) than the 18-member UM-SINGV ensemble. Thus, grouping the ensembles does not necessarily lead to an increase in the ensemble spread in spatial terms! The combination of the two ensembles has not produced larger spread than that of the larger-spread ensemble on its own,

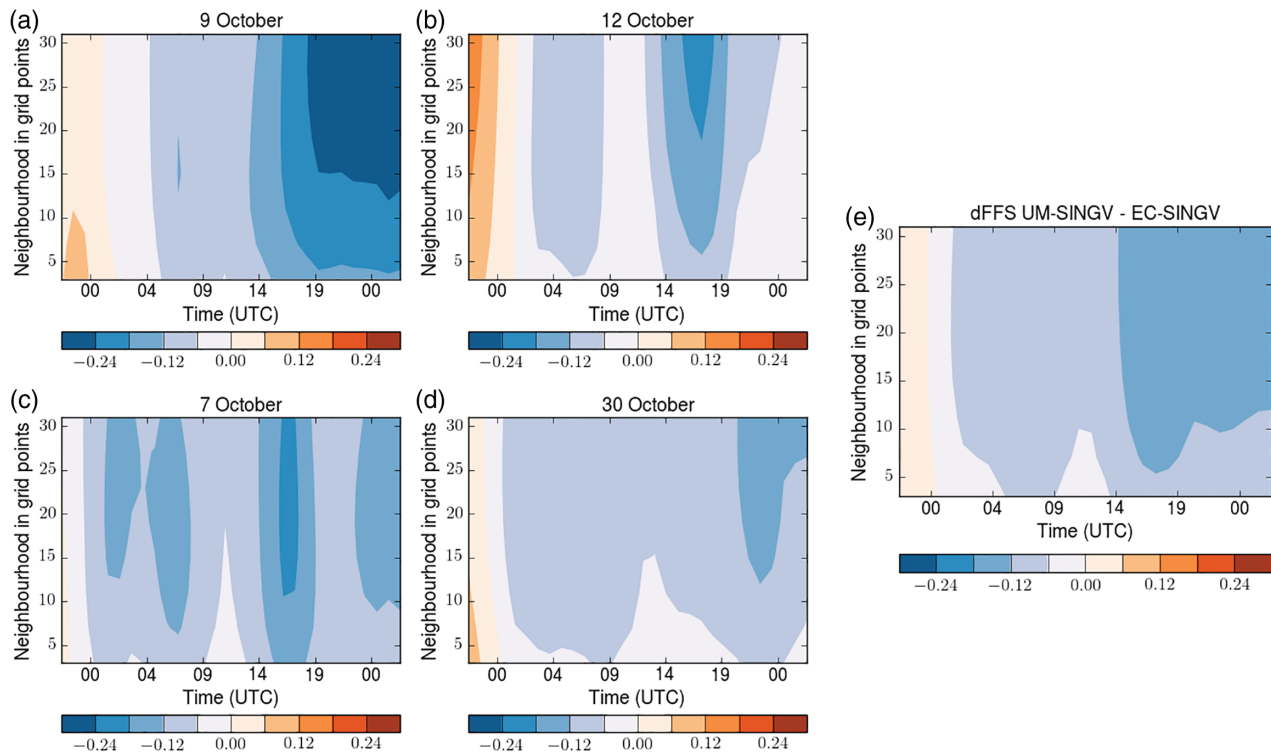


FIGURE 17 Diagram representing the difference in dFSSmean (see text) between UM-SINGV and EC-SINGV for the four case-studies. Blue (red) colours mean dFSSmean smaller (larger) for UM-SINGV. Different neighbourhood scales (3 by 3, 7 by 7, 11 by 11, 15 by 15, 19 by 19, 23 by 23, 27 by 27 and 31 by 31 points) are used here. (a) 9 October, (b) 12 October, (c) 7 October, (d) 30 October 2017, (e) average over all cases [Colour figure can be viewed at wileyonlinelibrary.com]

although it has created larger spread than that of the lower-spread ensemble on its own. This again suggests that the EC-SINGV ensembles produced rain in a more focused region than the UM-SINGV, and relates to the differences in the spread in the ICs.

4.4 | Spatial spread and the initial and lateral boundary conditions

We now investigate the ICs and LBCs on the evolution of spread as measured by the dFSS. First, we rerun each ensemble by setting the ICs to be the same for every forecast (so defining two new forecasts UM-IC for UM-SINGV and EC-IC for EC-SINGV) whilst continuing to use different LBCs. Then we run the forecasts using the different ICs whilst making the LBCs identical (from the control), thus defining two new forecasts UM-LBC for UM-SINGV and EC-LBC for EC-SINGV. Next we examine the relative influence of the ICs and LBCs by subtracting the dFSS-mean values in the ensembles with fixed ICs or LBCs from the standard ensembles (Figure 19). So, the impact of the ICs is illustrated for EC-SINGV by subtracting EC-IC from EC-SINGV (similarly for UM-SINGV). The impact of the LBCs is illustrated by subtracting EC-LBC from EC-SINGV (similarly for UM-SINGV).

In Figure 19, the results are averaged over the four case-studies directly. The day-to-day variability is filtered out,

but robust signals remain. Averaging over the different cases lead to similar magnitudes between the four cases for the ICs. The magnitude of the impact from the LBCs is however more sensitive to the day-to-day variability.

As expected, the impact of the ICs (Figure 19a,b) is the largest at the beginning of the forecast, up to approximately T + 18 h (0900 UTC) and the impact of the LBCs (Figure 19c,d) is the largest towards the end of the forecast from approximately T + 18 h onwards.

The impact of the ICs tends to be greatest at the smaller to intermediate scales at the earliest times and then extends to all scales equally, perhaps even upscaling. The impact of the LBCs is greatest at the later times (mostly after 0900 UTC) and tends to influence the larger scales more. Smaller scales appear to be affected more by LBCs in EC-SINGV than UM-SINGV.

For both configurations, over all lead times, the maximum impact of the ICs is a bit larger (i.e. bigger dFSS differences) than the impact of the LBCs. The impact of the LBCs on the ensemble spread at the end of the forecasts is in agreement with the role of these perturbations on the ensemble dispersion in Gebhardt *et al.* (2010). These results are also in agreement with Peralta *et al.* (2012) who show that the impact of the ICs on the ensemble dispersion can extend up to T + 14 h for higher thresholds (i.e. higher than 1 mm/h). The lead time at which the ICs or LBCs become most influential

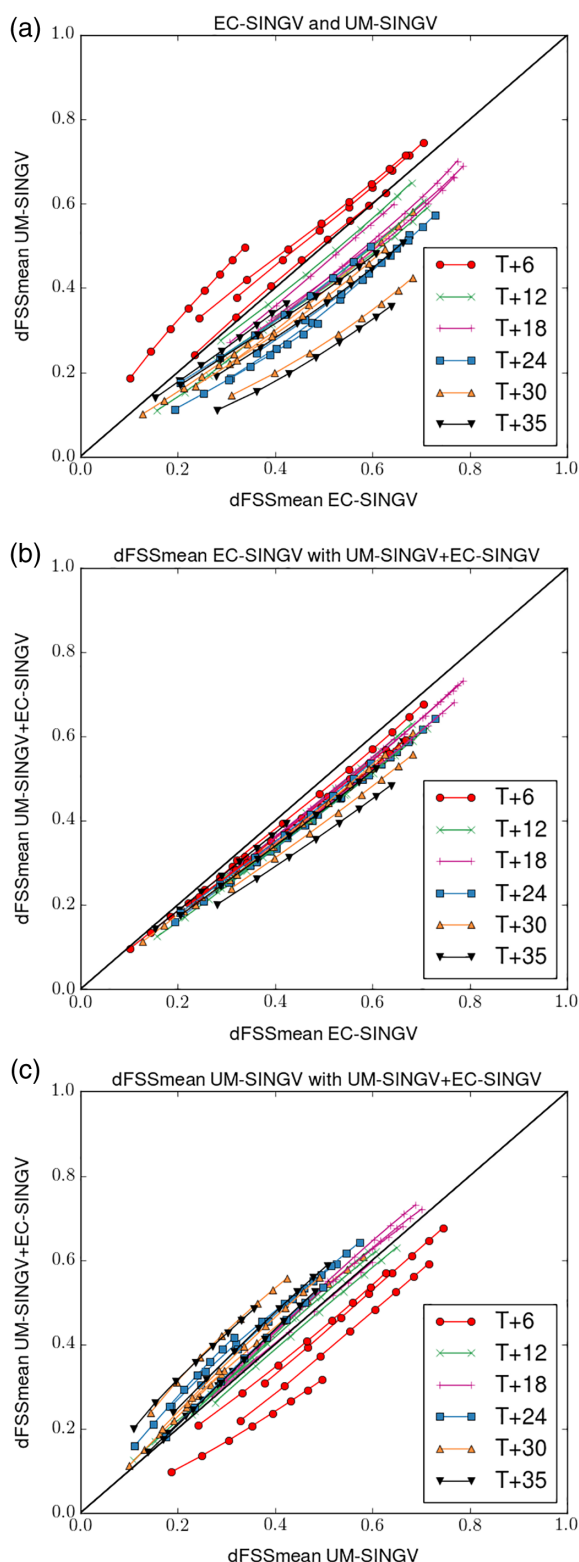


FIGURE 18 Comparison of dFFS for all the case studies and as a function of lead times (see legend) between (a) UM-SINGV and EC-SINGV (b) UM-SINGV combined with EC-SINGV against EC-SINGV (c) UM-SINGV combined with EC-SINGV against UM-SINGV. Different neighbourhood scales (3 by 3, 7 by 7, 11 by 11, 15 by 15, 19 by 19, 23 by 23, 27 by 27 and 31 by 31 points) are used here [Colour figure can be viewed at wileyonlinelibrary.com]

depends on the domain size of these simulations (Warner *et al.*, 1997), but also on the meteorological conditions and therefore the geographical region in which the system is run.

A comparison of Figure 19 and Figure 17 shows that, on average, the dFFSmean differences associated with the ICs or LBCs are more important than the differences related to the choice of the ensemble (UM-SINGV or EC-SINGV).

5 | CONCLUSIONS

The main objective of this article is to study the predictability of convective rainfall over Singapore using convection-permitting ensembles. To this end, we set up an ensemble, named SINGV-EPS, with a grid spacing of 4.5 km. SINGV-EPS is nested either within the ECMWF global ensemble (EC-SINGV with 51 members) or within the Met Office global ensemble MOGREPS-G (UM-SINGV with 18 members).

Both ensembles were compared using objective verification against land-surface data and satellite GPM, but it is difficult to draw robust conclusions about relative performance. For example, ranked probability scores show a better performance in wind speed and precipitation for EC-SINGV, while UM-SINGV shows better performance for temperature. When taking into account the difference in size of the ensemble between the two ensembles however, the impact of the ensemble size can be of similar magnitude to the differences between the two full-size ensembles. This may suggest perhaps a better performance of UM-SINGV if ensemble size is taken into account. An analysis of the monthly-mean rainfall with forecast lead time shows systematic differences between the two SINGV-EPS ensembles. The UM-SINGV ensemble tends to initiate convection earlier and have a larger peak around the time of squall-line occurrence, whereas the EC-ensemble tends to produce more rainfall over the period of most convective activity. A comparison against the GPM shows that UM-SINGV captures the daytime peaks of convection better, but EC-SINGV is better at capturing the dissipation of convection. So, unless one of these ensembles is shown to be clearly systematically better in general (which it is not), it would seem to be beneficial to utilize a combined ensemble that has the capability to give a bigger range of plausible outcomes.

Using case-study analysis, we illustrate the benefits of running such an ensemble for providing plausible alternative higher-impact scenarios. Because of the inherent low predictability associated with a single forecast, running a convective-scale ensemble is thus needed to capture the possible occurrence of high rainfall amounts (perhaps extreme) over Singapore and the use of both ensembles is useful for capturing plausible events. The results show that the convective-scale ensemble is sensitive to the parent ensemble model, which tells us something about the nature of the

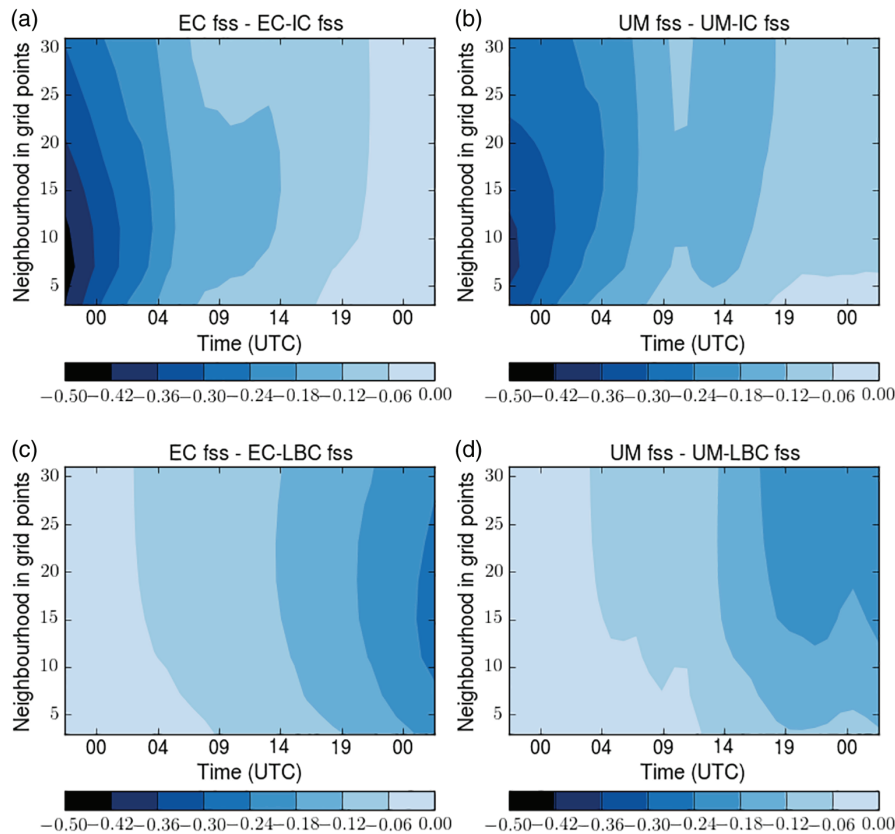


FIGURE 19 Mean impact of initial conditions (ICs) and lateral boundary conditions (LBCs) on the forecast over the four cases (9, 7, 12 and 30 October 2017). (a,b) Differences in dFSSmean metrics (see text) between the standard ensemble and the ensemble with the same ICs. (a) EC; (b) UM. (c,d) Differences in dFSSmean metrics between the standard ensemble and the ensemble with the same LBCs. (c) EC; (d) UM. Different neighbourhood scales (3 by 3, 7 by 7, 11 by 11, 15 by 15, 19 by 19, 23 by 23, 27 by 27 and 31 by 31 points) are used here [Colour figure can be viewed at wileyonlinelibrary.com]

predictability of the convective events. It appears that the differences in the spread in the ICs between UM-SINGV and EC-SINGV, as well as differences in the parent ensembles, perhaps in the dynamical configuration or parametrizations, can determine the nature of the convection in the convection-permitting forecasts. It is possible the one ensemble may provide systematically more convective instability or a larger convective inhibition or a warmer sea-surface temperature than the other and this matters when convection tries to develop explicitly in the higher-resolution forecasts.

Our second research question is to understand the nature of the spatial spread of the ensemble regarding the precipitation because it is important to be able to forecast where heavy rain is likely to occur. For the four case-studies examined here, we have studied the impact of the ensemble perturbations coming from the initial and lateral boundary conditions in both UM-SINGV and EC-SINGV.

The impact of the different perturbations on the spread of the ensemble rainfall (i.e. UM-SINGV or EC-SINGV, initial perturbations or lateral boundary conditions) vary with lead time and spatial scales. The initial perturbations dominate more at the beginning of the forecasts (up to T + 18 h) and are more associated with smaller to intermediate scales. The

lateral boundary conditions tend to influence the spread more after around T + 18 h and are more influential at intermediate to larger scales. The size of the impact associated with each of the different types of perturbations is similar, although slightly larger for the initial-condition perturbations. The results were compared against the mean over the four cases. In particular we find a lot of day-to-day variability in the spread between EC-SINGV and UM-SINGV, with the mean resulting in a much smoother response, but yet still a robust signal. Future work could evaluate how different this impact would be for different weather regimes throughout the year.

These results highlight the importance of the ICs. The differences related to the initial perturbations can persist up to 24 h into the forecast. Hence, any work on the ICs through the use of data assimilation techniques may be of important consideration for the ensemble spread. Note as well that the relative behaviour of the ICs and BCs may differ in midlatitude conditions, or even other tropical regions. Further research should also evaluate the sensitivity of convective-scale ensembles nested in different global ensembles and the potential of using combined information from these different ensembles.

ACKNOWLEDGEMENTS

The authors would like to thank Humphrey Lean, Alison Stirling and Adrian Lock for some useful scientific insights regarding the scientific configuration and the characteristics of the convection around Singapore. We would also like to thank David Walters for his constructive comments on the main ideas of the article. Many thanks to Warren Tennant for advice on the differences between the global ensembles used in this work. Finally, many thanks to the anonymous reviewers for their insightful comments.

ORCID

Aurore N. Porson  <https://orcid.org/0000-0002-5023-8522>

REFERENCES

- Beck, J., Bouttier, F., Wiegand, L., Gebhardt, C., Eagle, C. and Roberts, N.M. (2016) Development and verification of two convection-allowing multi-model ensembles over Western Europe. *Quarterly Journal of the Royal Meteorological Society*, 142, 2808–2826. <https://doi.org/10.1002/qj.2870>.
- Bednarczyk, C.N. and Ancell, B.C. (2015) Ensemble sensitivity analysis applied to a Southern Plains convective event. *Monthly Weather Review*, 143, 230–249. <https://doi.org/10.1175/MWR-D-13-00321.1>.
- Bowler, N.E. and Mylne, K.R. (2009) Ensemble transform Kalman filter perturbations for a regional ensemble prediction model. *Quarterly Journal of the Royal Meteorological Society*, 135, 757–766. <https://doi.org/10.1002/qj.404>.
- Bowler, N.E., Arribas, A., Mylne, K.R., Robertson, K.B. and Beare, S.E. (2008) The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134, 703–722. <https://doi.org/10.1002/qj.234>.
- Buizza, R. and Palmer, T.N. (1995) Singular vector structure of the atmospheric global circulation. *Journal of the Atmospheric Sciences*, 52, 1434–1456. [https://doi.org/10.1175/1520-0469\(1995\)052<1434:TSVSOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2).
- Buizza, R. and Palmer, T.N. (1998) Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, 126, 2503–2518. [https://doi.org/10.1175/1520-0493\(1998\)126<2503:IOESOE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2).
- Buizza, R., Leutbecher, M. and Isaksen, L. (2008) Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134, 2051–2066. <https://doi.org/10.1002/qj.346>.
- Bush M, Allen T, Boutle IA, Edwards JM, Finnenkoetter A, Franklin C, Hanley K, Lean H, Lock A, Manners J, Mittermaier M, Morcrette C, North R, Petch J, Short C, Vosper S, Walters D, Webster S, Weeks M, Wilkinson J, Wood N and Zerroukat M. (2019) The Met Office Regional Atmosphere and Land configurations (RAL) – Version 1. *Geoscientific Model Development*. Discuss.
- Clark, A.J., Gallus, W.A., Jr. and Weisman, M.L. (2010) Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Weather and Forecasting*, 25, 1495–1509. <https://doi.org/10.1175/2010WAF2222404.1>.
- 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. (2011) Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Monthly Weather Review*, 139, 1410–1418. <https://doi.org/10.1175/2010MWR3624.1>.
- Clark, P., Roberts, N.M., Lean, H., Ballard, S.P. and Charlton-Perez, C. (2016) Review convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications*, 23, 165–181. <https://doi.org/10.1002/met.1538>.
- Dey, S.R.A., Leoncini, G., Roberts, N.M., Plant, R.S. and Migliorini, S. (2014) A spatial view of ensemble spread in convection-permitting ensembles. *Monthly Weather Review*, 142, 4091–4107. <https://doi.org/10.1175/MWR-D-14-00172.1>.
- Dipankar A, Webster S, Furtado K, Wilkinson J, Sanchez C, Lock A, North R, Sun X, Vosper S, Huang X-Y, Barker D. (2019) SINGV: a convective-scale weather-forecast model for Singapore. To be submitted.
- Durran, D.R. and Weyn, J.A. (2016) Thunderstorms do not get butterflies. *Bulletin of the American Meteorological Society*, 97, 237–244. <https://doi.org/10.1175/BAMS-D-15-00070.1>.
- Flack, D.L.A., Gray, S.L., Plant, R.S., Lean, H.W. and Craig, G.C. (2018) Convective-scale perturbation growth across the spectrum of convective regimes. *Monthly Weather Review*, 146, 387–405. <https://doi.org/10.1175/MWR-D-17-0024.1>.
- Gebhardt, C., Theis, S.E., Paulat, M. and Ben Bouallegue Z. (2010) Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation in lateral boundaries. *Atmospheric Research*, 100, 168–177. <https://doi.org/10.1016/j.atmosres.2010.12.008>.
- Golding, B., Roberts, N.M., Leoncini, G., Mylne, K.R. and Swinbank, R. (2016) MOGREPS-UK convection-permitting ensemble products for surface water flood forecasting: rationale and first results. *Journal of Hydrometeorology*, 17, 1384–1406. <https://doi.org/10.1175/JHM-D-15-0083.1>.
- Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N.M. and Tennant, W. (2017) The Met-Office convective-scale ensemble, MOGREPS-UK. *Quarterly Journal of the Royal Meteorological Society*, 143, 2846–2861. <https://doi.org/10.1002/qj.3135>.
- Hanley, K.E., Kirshbaum, D.J., Roberts, N.M. and Leoncini, G. (2013) Sensitivities of a squall line over Central Europe in a convective-scale ensemble. *Monthly Weather Review*, 141, 112–132. <https://doi.org/10.1002/qj.877>.
- Kendon, E.J., Roberts, N.M., Senior, C.A. and Roberts, M.J. (2012) Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate*, 25, 5791–5806. <https://doi.org/10.1175/JCLI-D-11-00562.1>.
- Leoncini, G., Plant, R.S., Gray, S.L. and Clark, P.A. (2013) Ensemble forecasts of a flood-producing storm: comparison of the influence of model-state perturbations and parameter modifications. *Quarterly Journal of the Royal Meteorological Society*, 139, 198–211. <https://doi.org/10.1002/qj.1951>.
- Letekewicz, C.E. and Parker, M.D. (2011) Impact of environmental variations on simulated squall lines interacting with terrain. *Monthly Weather Review*, 139, 3163–3183. <https://doi.org/10.1175/2011MWR3635.1>.
- Leutbecher, M. and Palmer, T.N. (2008) Ensemble forecasting. *Journal of Computational Physics*, 227, 3515–3539. <https://doi.org/10.1016/j.jcp.2007.02.014>.
- Lo, J.C.-F. and Orton, T. (2016) The general features of tropical Sumatra squalls. *Weather*, 71(7), 175–178. <https://doi.org/10.1002/wea.2748>.
- Love, B., Matthews, A.J. and Lister, G.M.S. (2011) The diurnal cycle of precipitation over the Maritime Continent in a high-resolution

- atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, 137, 934–947. <https://doi.org/10.1002/qj.809>.
- Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., Tibaldi, S., Molteni, F. and Buizza, R. (2001) A strategy for high-resolution ensemble prediction. II: Limited-area experiment in four Alpine flood events. *Quarterly Journal of the Royal Meteorological Society*, 127, 2095–2115. <https://doi.org/10.1002/qj.49712757613>.
- Mittermaier, M.P. (2014) A strategy for verifying near-convection-resolving forecasts at observing sites. *Weather and Forecasting*, 29, 185–204. <https://doi.org/10.1175/WAF-D-12-00075.1>.
- Mittermaier, M.P. and Csima, G. (2017) Ensemble versus deterministic performance at the kilometer scale. *Weather and Forecasting*, 32, 1697–1709. <https://doi.org/10.1175/WAF-D-16-0164.1>.
- Mittermaier, M., Roberts, N.M. and Thompson, S.A. (2013) A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorological Applications*, 20, 176–186. <https://doi.org/10.1002/met.296>.
- Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F. and Paccagnella, T. (2001) A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. *Quarterly Journal of the Royal Meteorological Society*, 127, 2069–2094. <https://doi.org/10.1002/qj.49712757612>.
- Morcrette, C.J. (2012a) Improvements to a prognostic cloud scheme through changes to its cloud erosion parametrization. *Atmospheric Science Letters*, 13, 95–102. <https://doi.org/10.1002/asl.374>.
- Morcrette, C.J. (2012b) Prognostic-cloud-scheme increment diagnostics: a novel addition to the case-study tool kit. *Atmospheric Science Letters*, 13, 200–207. <https://doi.org/10.1002/asl.380>.
- Nielsen, E.R. and Schumacher, R.S. (2016) Using convection-allowing ensembles to understand the predictability of an extreme rainfall event. *Monthly Weather Review*, 144, 3651–3676. <https://doi.org/10.1175/MWR-D-16-0083.1>.
- Park, Y.-Y., Buizza, R. and Leutbecher, M. (2008) TIGGE: preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, 134, 2029–2050. <https://doi.org/10.1002/qj.334>.
- Peralta, C., Ben Bouallegue, Z., Theis, S.E., Gebhardt, C. and Buchhold, M.J. (2012) Accounting for initial conditions uncertainties in COSMO-DE-EPS. *Journal of Geophysical Research*, 117, D07108. <https://doi.org/10.1029/2011JD016581>.
- Porson, A.N., Steyn, D.G. and Schayes, G. (2007a) Sea-breeze scaling from numerical model simulations. Part I: Pure sea breezes. *Boundary-Layer Meteorology*, 122, 17–29. <https://doi.org/10.1007/s10546-006-9090-4>.
- Porson, A.N., Steyn, D.G. and Schayes, G. (2007b) Sea-breeze scaling from numerical model simulations. Part II: Interaction between the sea breeze and slope flows. *Boundary-Layer Meteorology*, 122, 31–41. <https://doi.org/10.1007/s10546-006-9092-2>.
- Raynaud, L. and Bouttier, F. (2017) The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts. *Quarterly Journal of the Royal Meteorological Society*, 143, 3037–3047. <https://doi.org/10.1002/qj.3159>.
- Roberts, N.M. (2008) Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, 15, 163–169. <https://doi.org/10.1002/met.57>.
- Roberts, N.M. and Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136, 78–97. <https://doi.org/10.1175/2007MWR2123.1>.
- Sakurai, N., Murata, F., Yamanaka, M.D., Mori, S., Hamada, J.-I., Hashiguchi, H., Tauhid, Y.I., Sribimawati, T. and Suhardi, B. (2005) Diurnal cycle of cloud system migration over Sumatera Island [Sumatra]. *Journal of the Meteorological Society of Japan*, 83, 835–850. <https://doi.org/10.2151/jmsj.83.835>.
- Schwartz, C.S. and Sobash, R.A. (2017) Generating probabilistic forecasts from convection-allowing ensembles using neighbourhood approaches: a review and recommendations. *Monthly Weather Review*, 145, 3397–3418. <https://doi.org/10.1175/MWR-D-16-0400.1>.
- Schwartz, C.S., Kain, J.S., Weiss, S.J., Xue, M., Bright, D.R., Kong, F., Thomas, K.W., Levit, J.J., Coniglio, M.C. and Wandishin, M.S. (2010) Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, 25, 263–280. <https://doi.org/10.1175/2009WAF2222267.1>.
- Schwartz, C.S., Romine, G.S., Smith, K.R. and Weisman, M.L. (2014) Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Weather and Forecasting*, 29, 1295–1318. <https://doi.org/10.1175/WAF-D-13-00145.1>.
- Schwartz, C.S., Romine, G.S., Sobash, R.A., Fossell, K.R. and Weisman, M.R. (2015) NCAR's experimental real-time convection-allowing ensemble prediction system. *Weather and Forecasting*, 30, 1645–1654. <https://doi.org/10.1175/WAF-D-15-0103.1>.
- Skok, G. (2015) Analysis of fraction skill score properties for a displaced rainband in a rectangular domain. *Meteorological Applications*, 22, 477–484. <https://doi.org/10.1002/met.1478>.
- Skok, G. and Roberts, N.M. (2018) Estimating the displacement in precipitation forecasts using the Fractions Skill Score. *Quarterly Journal of the Royal Meteorological Society*, 144, 414–425. <https://doi.org/10.1002/qj.3212>.
- Tennant, W. and Beare, S. (2014) New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. *Quarterly Journal of the Royal Meteorological Society*, 140, 1150–1160. <https://doi.org/10.1002/qj.2202>.
- Warner, T.T., Peterson, R.A. and Treadon, R.E. (1997) A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society*, 78, 2599–2617. [https://doi.org/10.1175/1520-0477\(1997\)078<2599:ATOLBC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2).
- Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*, 3rd edition. Amsterdam: Elsevier.
- Wu, P., Hara, M., Hamada, J.-I. and Yamanaka, M.D. and Kimura, M.D. (2009) Why a large amount of rain falls over the sea in the vicinity of western Sumatra Island during nighttime. *Journal of Applied Meteorology and Climatology*, 48, 1345–1361. <https://doi.org/10.1175/2009JAMC2052.1>.
- Zsoter, E., Buizza, R. and Richardson, D. (2009) “Jumpiness” of the ECMWF and Met Office EPS control and ensemble-mean forecasts. *Monthly Weather Review*, 137, 3823–3836. <https://doi.org/10.1175/2009MWR2960.1>.

How to cite this article: Porson AN, Hagelin S, Boyd DFA, *et al.* Extreme rainfall sensitivity in convective-scale ensemble modelling over Singapore. *Q J R Meteorol Soc.* 2019;1–19. <https://doi.org/10.1002/qj.3601>