UNIVERSITY OF READING

Department of Meteorology

A spatial approach to the analysis of convective scale ensemble systems

Seonaid Rosemary Austin Dey

A thesis submitted for the degree of Doctor of Philosophy

January 2016

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Seonaid Dey

Abstract

The use of kilometre-scale ensembles in operational forecasting provides challenges for forecast evaluation and interpretation. New spatial methods for characterising and verifying convection permitting ensembles are developed, and tested on the 12 member Met Office 2.2km resolution UK ensemble. Each ensemble member is regarded as an equally plausible realisation of the true atmospheric state.

A novel methodology is presented for spatial ensemble characterisation based on the Fractions Skill Score. Characterising the domain-wide ensemble behaviour, these methods identify useful spatial scales and spin-up times for the model, and demonstrate the upscale growth of errors and forecast differences. The ensemble spread is shown to be highly dependent on the spatial scales considered and the threshold applied to the field. Comparing differently-generated ensemble systems shows the utility of spatial ensemble evaluation techniques for assessing different ensemble perturbation strategies.

It is also important to consider location-dependent ensemble behaviour. A new method for calculating the location-dependent spatial agreement of ensemble members is presented. Through comparing with radar observations, the location-dependent spatial skill of the ensemble is also quantified. These methods are verified using an idealised experiment. Six convective cases, and a summer season, are used to demonstrate the methods in an operational context, with links made to physical processes. Overall, the ensemble system is reasonably well-spread spatially. Poorer spread-skill is associated with a low fractional coverage of rain, and low synoptic-scale rain rates. Higher confidence in the location of precipitation is found to the northwest of the UK.

To investigate coherent physical structures in the ensemble, the spatial approach was used to inform the calculation of multivariate correlations. Using the spatial approach, physically-meaningful correlations which demonstrate inter-variable relationships are obtained. Overall, the spatial approach is found to give useful information for forecasting, and for the interpretation and evaluation of convection-permitting ensembles.

Dedication

This thesis is dedicated to the memory of my father Neil A. Dey (1946 – 2015) whose love, support, and encouragement made this work possible.

Acknowledgements

I would first and foremost like to acknowledge the help and support of my three supervisors Bob Plant, Nigel Roberts and Stefano Migliorini. Throughout my PhD, all three have been actively involved and interested in my project. In particular, I would like to thank Bob for being enthusiastic and positive even when things go wrong, Nigel for the many ”5 minute” discussions that have gone on for 50 minutes, and Stefano for his input into the project, even when far away at ECMWF, or the Met Office. I would also like to thank the many other people who have helped during the last four years including: Humphrey Lean, Giovanni Leoncini, Carol Halliwell, Chloe Eagle, Khalid Mahmood, A J Watling and Catherine Turner, and my monitoring committee, Sue Gray and Sarah Dance. For proof reading my thesis I would like to thank Laura Davison and Ian Dey.

The Department of Meteorology is a fantastic place to do a PhD: fun, sociable and inclusive. I would like to thank my office mates in 1U07, particularly those who have been my neighbours and put up with my mutterings (Brodie Pearson, Mike Glinton, Hannah Parker, Leo Saffin and Duick Young). Thanks also to everyone from MetOffice@Reading for many enjoyable and distracting coffee breaks. I would like to thank those people (you know who you are) who have looked after me during the last six months.

Last but by no means least, I would like to thank my family for their love, patience and belief which has allowed me to reach this point. Special mention goes to my Uncle, Dr. Ian Dey, for being a fantastic role model, and my boyfriend Duncan Anderson for helping me be the person I want to be.

– Seonaid Dey, January 2016

Contents

Chapter 1

Introduction

1.1 Motivation

Recent increases in computing power have allowed a shift towards higher resolution numerical weather prediction (NWP) where physical processes such as convection can be explicitly simulated. However, although these high resolution simulations produce realistic features (Mass et al., 2002; Lean et al., 2008; Weisman et al., 2008; Schwartz et al., 2009), errors grow rapidly (Hohenegger and Schär, 2007; Melhauser and Zhang, 2012; Radhakrishna et al., 2012), and small-scale predictability is maintained for only a few hours. Hence, to benefit fully from convection permitting NWP it is necessary to understand and quantify the forecast uncertainty. Ensembles, where several forecasts are run with different starting conditions, have been successfully used for this purpose in larger scale NWP (e.g. Palmer, 2000, and references therein), and are now being run at convection permitting resolutions. In particular, convection permitting ensembles have been investigated for a range of case studies (Hanley et al., 2011; Leoncini et al., 2013; Clark et al., 2013; Hanley et al., 2013), nowcasting applications (Migliorini et al., 2011), and are now run, or about to be run, operationally at several forecasting centres (Baldauf et al., 2011; Gebhardt et al., 2011; Bouttier et al., 2012; Golding et al., 2014). This opens up the possibility of forecasting both high-impact convective events, often associated with flooding and sometimes even loss of life, and also the less extreme convective events and local weather which, although less costly, affect large numbers of people on a daily basis.

Despite the anticipated benefits of running convection-permitting ensembles, questions remain about the best methods for their interpretation and evaluation. In particular the ensemble mean, successfully used for smoothly varying, large scale fields, may not be physically appropriate at the convective scale (e.g. Ancell, 2013; Hollan and Ancell, 2015). This is particularly true for forecast quantities with high spatial variability, such as precipitation; for these fields the ensemble mean field does not retain the physical structures present in the individual member forecasts. Other open questions relate to the interpretation of forecast uncertainty, given the tiny fraction of realisations covered by the ensemble members, and methods of forecast verification. Standard forecast verification and comparison measures, such as the Root Mean Squared Error (RMSE e.g. Wilks, 2011) are unsuited to the convective scale as they overly penalise spatial differences. Several more suitable methods have been proposed for verifying convection-permitting deterministic forecasts (e.g. Ebert, 2008; Gilleland et al., 2009; Johnson and Wang, 2012) that can now be developed for convection permitting ensembles.

It has been shown that the skill of convective scale forecasts is scale dependent, with skill increasing as a function of spatial scale (Roberts and Lean, 2008; Roberts, 2008; Ben Bouallègue and Theis, 2014; Mittermaier et al., 2013b; Mittermaier, 2014). Clark et al. (2011) showed this was also true for ensemble forecasts, with ensemble skill increasing with both spatial scale and ensemble size. Given this dependence on spatial scale, methods have also been developed to evaluate the differences between ensemble member forecasts, a measure of the ensemble spread, at different spatial scales. In particular, Johnson et al. (2014) used wavelet decomposition to investigate perturbation growth and Surcel et al. (2014) used spectral decomposition to investigate the filtering properties of the ensemble mean. However, a general methodology for the evaluation of convective scale ensembles still needs to be developed.

This PhD aims to develop new approaches for the evaluation and characterisation of convective scale ensembles by considering the forecasts not only at individual grid points, but also over the regions or neighbourhoods surrounding these points. In particular, the aim is to develop neighbourhood methods for evaluating forecasts of convective events from an operational convective scale ensemble. There are several key requirements for these methods:

- Results from different spatial locations in the forecast, or at different forecast times, must relate to each other in a physically meaningful manner. For example, results across a range of forecast times (say an hour) should give a consistent message, and results at neighbouring grid points should be related.
- The method must not be overly sensitive to changes in ensemble membership. This is particularly important when operational forecasts with a small number of members are evaluated: the results should reflect the ensemble as a whole and

not be disproportionally influenced by the behaviour of each individual ensemble member.

- Metrics for the evaluation of convective scale ensembles often consider only precipitation, due to its high spatial variability and importance for forecasting (e.g. Ebert, 2008). However, other variables, and the relationships between them (inter-variable relationships), are also important for understanding and characterising the performance of convection permitting ensembles. Hence, the methods developed in this thesis should also consider different variables, and the inter-variable relationships.

In this thesis methods will be developed in line with these requirements. A variety of uses of these methods will be investigated, for example for forecasting and in data assimilation, with a focus on understanding convective events over the UK.

1.2 Thesis aims and structure

To aid the development of the methods outlined in Section 1.1, six questions have been formulated: the aim of the thesis will be to answer these questions.

1. How can the spatial variability of ensemble member forecasts be meaningfully summarised? What are suitable summary measures, and what do these tell us about ensemble performance and spatial predictability?
2. Over what spatial scales should the ensemble be interpreted and evaluated? How can these spatial scales be defined?
3. What information does this spatial approach provide for forecasting?
4. How can correlations in the vertical be used to allow physical structures and multivariate relationships from the ensemble to be usefully summarised? To what extent do these correlations reveal properties of the convection and the convective environment?
5. What is the dependence of vertical correlation structure on the horizontal scales used? How does this relate to the scales obtained from the spatial methods?
6. In what ways might these methods be useful in other areas such as data assimilation?

To address these questions, it is necessary to combine knowledge from three key research areas: predictability and ensembles, neighbourhood verification methods and the use of multi-variate methods. To provide a context and background for the work of later chapters, **Chapter 2** will introduce these research areas in turn and, for each, give an overview of relevant literature.

Chapter 3 introduces the NWP model and convection permitting ensemble system used for the work in this thesis. Other data sources including radar and satellite data are also introduced, and the overall methodology of this thesis discussed.

In **Chapter 4**, with the aim of answering Question 1, a new methodology, based on the Fractions Skill Score (FSS Roberts and Lean, 2008; Roberts, 2008), is introduced for evaluating the spatial differences between members of a convection permitting ensemble. The spatial spread of the ensemble members is determined and the realism of the ensemble spread is tested by comparing with the skill against radar derived precipitation accumulations. These techniques are presented through the use of two convective cases studies. Discussion focuses on the forecast evolution through different lead times, the effect of considering different threshold values for the fields used to calculate the FSS, the comparison of different forecast variables and the consideration of different model formulations within the ensemble. The work of Chapter 4 has been published in Dey et al. (2014).

Chapter 5 will be a short introduction to the six convective case studies used in Chapters 6 and 7.

The methods presented in Chapter 4 provide a compact summary of the ensemble performance over the whole domain at a given scale. However, by considering only one value to represent the domain, different meteorological phenomena, such as convective and frontal precipitation, are considered together, when each individually may have an inherently different predictability and ensemble spread. In **Chapter 6** a new, location and scale dependent measure of ensemble spread that does not suffer from this drawback is introduced with the aim of answering Question 2. Through the application of these techniques to an idealised ensemble system, and to the six case studies introduced in Chapter 5, we will start to address Question 3. The work in Chapter 6 is published in Dey et al. (2016a).

Chapter 7 applies the spatial approach, developed in Chapters 4 and 6, to the calculation of multivariate correlations with the aim of answering Questions 4 and 5. The six convective cases introduced in Chapter 5 are used as examples to investigate physical structures and meteorological processes in the ensemble forecasts. This leads to a discussion of Question 6.

Going beyond the case study approach, **Chapter 8** applies the location-dependent techniques for ensemble evaluation, developed in Chapter 6, to a summer season of rain rate forecasts. To quantify the spatial ensemble spread the ensemble member forecasts are compared; to quantify the spatial ensemble skill, the ensemble forecasts are compared with radar observations. With the aim of answering Question 3, the spatial predictability of precipitation over the UK for a summer season is quantified, and some of the influential factors investigated. Question 2 will also be further discussed. The work in Chapter 8 is published in Dey et al. (2016b, under review).

Finally, in **Chapter 9**, the overall results of this thesis are presented and the key conclusions summarised. Areas of future investigation are also discussed.

Chapter 2

Background and literature review

To address the aims of this PhD, outlined in Section 1.2, it is necessary to combine knowledge from three key research areas: predictability and ensembles, forecast verification methods (in particular neighbourhood methods and the quantification of ensemble spread) and the use of multivariate methods, particularly in data assimilation. To provide a context and background for the work of later chapters, this chapter will introduce these research areas in turn and, for each, give a review of relevant literature (Sections 2.1 to 2.3). Following on from this, in Section 2.4 a brief introduction is given to the current understanding of convection over the UK.

2.1 Predictability and ensembles

2.1.1 Global and synoptic scales

Since the pioneering work of Lorenz it has been known that there is a finite limit to atmospheric predictability: "certain turbulent systems, possibly including the earth's atmosphere, possess for practical purposes a finite range of predictability" (Lorenz, 1969). Errors in the initial state and model formulations, along with the non-linearity of the governing atmospheric equations, mean that two model trajectories, almost identical initially, diverge to completely different solutions that are statistically uncorrelated (Ehrendorfer, 1997; Palmer, 2000). This is what is meant here by the limit of predictability, a limit that, at the synoptic scale, is estimated to be of order days (Ehrendorfer, 1997; Palmer, 2000; Hohenegger and Schär, 2007).

To quantify the nature of predictability it is useful to consider the problem probabilistically. This is done by focusing on the temporal evolution of the probability density function (PDF) of possible forecast outcomes, $p(t)$. At the start of the forecast, the PDF (hereafter p_0) represents errors in the starting conditions. In operational forecasting systems a Monte Carlo method is used as a basis for describing the time evolution of the PDF (e.g. Ehrendorfer, 1997). The Monte Carlo forecasting method was proposed in Leith (1974) and involves the generation of an ensemble of initial model states by randomly sampling from p_0. These states are then independently evolved using the NWP model. An overview of Monte Carlo forecasting is given in (Kalnay, 2003, chapter 6).

In operational ensemble forecasting systems a number of methods are used to select initial states from which to run ensemble member forecasts. For example, an ensemble can be created from forecasts starting from different initial conditions (e.g. Rabier et al., 1996), from different models (Multimodel ensembles, e. g. Marsigli et al., 2014), different model start times (e.g. Raynaud et al., 2015) or using different parametrisations, physics schemes and perturbations. These latter perturbation methods are discussed further in Section 2.1.2 with respect to high resolution ensembles.

Given the large number of degrees of freedom in atmospheric models it is not computationally possible to run a sufficient number of ensemble members to span completely the PDF of possible forecast outcomes. Hence it is necessary to select starting conditions for the ensemble member forecasts which span as much of the PDF as possible. Several different methods have been developed for this purpose. For example, singular vectors mathematically select the fastest growing modes (Buizza, 1997; Buizza et al., 2005; Palmer, 2000) and bred vectors are produced by considering how perturbations grow with forecast lead time (Toth and Kalnay, 1997) (i.e. forecasts with different lead times, say 2 hours and 6 hours, are compared at same verification time, say 12Z). Other methods of initialising ensemble perturbations include methods based on the Kalman Filter (Bishop et al., 2001; Majumdar et al., 2002; Evensen, 2009, Chapter 4) which derives ensemble perturbations as part of the Data Assimilation scheme. This method is used at the Met Office to generate boundary and initial conditions for the Met Office Global and Regional Ensemble Prediction System (MOGREPS Bowler et al., 2008, 2009). Mathematical details of these methods can be found in Palmer (2000); Kalnay (2003); Leutbecher and Palmer (2008); Evensen (2009). It is important to note that these methods focus only on the quantification of initial condition uncertainties. Work into methods to quantify uncertainties incurred through approximations in the modelling process (model errors) is ongoing and will be discussed in Section 2.1.2. Note also that the methods described above are applied to global data assimilation systems and rely on assumptions such as linearity.

2.1.2 Meso and convective scales

At the convective scale (the subject of this thesis) the assumptions used in global data assimilation are not necessarily valid and the methods need adapting (e.g. Caron, 2013; Craig and Würsch, 2013; Wang et al., 2014; Tennant, 2015). Additionally, at small spatial scales, forecast errors grow more rapidly than at the global and synoptic scales (Lorenz, 1969; Ehrendorfer, 1997; Palmer, 2000, and references therein). Convective scale ensemble generation is further complicated by the influences of boundary conditions due to the use of a limited area domain. These questions are increasingly important as advances in computer power allow models to be run at ever higher resolutions (Mass et al., 2002; Lean et al., 2008, and references therein).

Several forecasting centres now operationally run convection permitting ensembles. For example, the 12 member, 2.2 km grid length, convection permitting ensemble of the UK Met Office (MOGREPS-UK Mylne, 2013; Golding et al., 2014) has been running operationally since May 2013 and the 20 member, 2.8 km grid length, COSMO-DE (Consortium for Small-scale Modeling, regional model centred on Germany) ensemble prediction system of Deutscher Wetterdienst (DWD) became operational in 2012 (Baldauf et al., 2011; Gebhardt et al., 2011, 2012). Other convection permitting ensembles are under development such as an ensemble based on the Applications of Research to Operations at Mesoscale (AROME) model (Bouttier et al., 2012) and as part of the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Spring Experiment (e.g. Clark et al., 2012; Johnson and Wang, 2013, and references therein). It is important to make the best use of these ensemble systems. Therefore, recent research has also focused on the applications of these models, for example for calculation of tornado path-lengths (Clark et al., 2013) and nowcasting (e.g. Migliorini et al., 2011; Sun et al., 2014; Golding et al., 2014). Further research has focused on the effects of ensemble grid length (Kong et al., 2006, 2007), and the use of convection permitting ensembles to investigate individual case studies (e.g. Hanley et al., 2011; Leoncini et al., 2013; Hanley et al., 2013).

The work of this thesis is only possible because of the availability of high resolution ensemble systems (such as those detailed above) that can produce convection explicitly without the need for a convection scheme. This allows the physical processes associated with the convection to be investigated in an ensemble framework. Note that these ensembles are 'convection permitting', not 'convection resolving': to fully resolve convection, grid lengths of less than 1 km are needed (Weisman et al., 1997; Bryan et al., 2003; Craig and Dörnbrack, 2008; Birch et al., 2014; Hanley et al., 2015; Verrelle et al., 2015).

Hohenegger and Schär (2007) compared the predictability for low resolution (around 80 km grid spacing) and convection-permitting (2.2 km grid spacing) simulations and found error doubling times around ten times faster for the higher resolution simulations. Further work has investigated the links between mesoscale processes and predictability with faster error growth (lower predictability) linked to moist dynamics (Zhang et al., 2002, 2003; Hohenegger et al., 2006), and higher predictability linked to high orography (Walser et al., 2004) and large catchments (Walser and Schär, 2004). Recent studies have sought to separate equilibrium convection (where precipitation amounts can be accurately predicted if a large enough area is considered) and triggered convection (unpredictable in the sense that local triggers are needed to overcome convective inhibition) in order to distinguish different modes of predictability in convective events (Keil and Craig, 2011; Zimmer et al., 2011; Craig et al., 2012; Keil et al., 2014).

It has also been argued that instability indices such as absolute and convective instability are a poor indicator of the forecast predictability (Hohenegger et al., 2006), suggesting that new methods of evaluating predictability are needed. The predictability of precipitation is of particular interest due to its high impact and spatial variability. This is particularly true at the current time, when NWP models are starting to be used to run flood forecast models (Cloke and Pappenberger, 2009; Gaborit et al., 2014; Caseri et al., 2015). In this thesis, considering the precipitation fields is central both to the analysis of the ensemble spatial characteristics (Chapters 4, 6 and 8) and also inter-variable relationships (Chapter 7).

In addition to using high resolution ensembles to investigate predictability, recent investigations have been conducted into different perturbation strategies for high resolution ensembles. To obtain an ensemble distribution as close to the true atmospheric PDF as possible it is crucial that the ensemble perturbations are an accurate representation of errors in initial conditions and in the modelling process. Recent studies have looked into the effects of different perturbation strategies including downscaled initial condition perturbations (Kühnlein et al., 2014), boundary condition perturbations (Gebhardt et al., 2011), boundary layer parameters (Martin and Xue, 2006; Leoncini et al., 2010; Done et al., 2012) and fixed parameter perturbations (e.g. Stensrud et al., 2000; Hacker et al., 2011; Hally et al., 2014; Vié et al., 2012; Wheatley et al., 2014; Schumacher and Clark, 2014; Muhlbauer et al., 2014). Interestingly, when only considering physics perturbations and perturbing 4 model physics schemes of the Weather Research and Forecasting (WRF) model, Hacker et al. (2011) found a lack of domain wide systematic response to the parameters. This conclusion is also supported by the work of Vié et al. (2012) who found that microphysical perturbations (autoconversion, accretion and evaporation) had little effect on probabilistic scores. In contrast, Stensrud et al. (2000); Keil et al. (2014); Hally et al. (2014) found that microphysical perturbations could have an impact on forecast skill. The effect of different model physics is considered in Chapter 4 in the context of ensemble spatial spread.

Instead of considering fixed parameters, Bouttier et al. (2012); Baker et al. (2014); Romine et al. (2014) consider stochastic physics perturbations. Bouttier et al. (2012) conclude that using an adaption of the ECMWF Perturbation of Physics tendencies scheme improves reliability and spread-skill consistency for the AROME convection permitting ensemble. This contrasts with the results from Baker et al. (2014) who find little effect from adding a random parameter scheme to a 1.5km configuration of the Met Office ensemble system. However, there are differences in the studies that could account for these differing results. In particular, Baker et al. (2014) considered a single case study whereas Bouttier et al. (2012) apply a statistical approach using two weeks of data.

Given the recent research effort in developing these different ensembles, they should be evaluated appropriately. In particular the approach needs to be multivariate and appropriate to the convective scale. Recent work in forecast verification has sought to address the latter point and is discussed in the next section.

2.2 Forecast verification methods

2.2.1 Neighbourhood methods

Traditional metrics for evaluating forecasts against observations, such as the Root Mean Square Error, standard deviation and rank histograms (e.g. Barlow, 1989; Wilks, 1995) operate on the grid scale. For smoothly varying fields, or large spatial scales (as found on the grid scale for low resolution simulations), these metrics are acceptable: their ranking of forecasts agrees with qualitative estimates of the forecast accuracy (Stratman et al., 2013). However, for high resolution forecasts, this is no longer the case and spatial errors are penalised twice, once for being a miss, and again for being a false positive. This is known as the double penalty problem and is a particular issue in the verification of precipitation fields. Of course, these issues apply equally to the comparison of ensemble member forecasts as they do to the comparison of forecasts with observations.

In response to these issues a number of different verification methods have been developed. The Spatial Forecast Verification Inter Comparison Project (Gilleland et al., 2009; Ebert, 2008) provided a framework for comparing these methods and assessing their applicability to different situations. Gilleland et al. (2009) separate the methods into four main categories: feature based, field deformation, scale-separation and neighbourhood. The first two categories are "displacement" methods which categorise forecast skill by the amount the forecast field must be manipulated to give results within an acceptable error of the observations. The second two categories are "filtering" methods that first apply a spatial filter to the forecast before calculating traditional verification statistics. By applying the filter at increasingly large spatial scales information can be obtained about the scales at which the forecast is skillful. A second intercomparison project "Mesoscale Verification Intercomparison in Complex Terrain" (MesoVICT) commenced in 2014 (Mittermaier et al., 2013a; Dorninger et al., 2013). One neighbourhood verification metric considered in the inter-comparison project, the Fractions Skill Score (FSS Roberts and Lean, 2008) will be applied in Chapter 4 to the evaluation of convection permitting ensembles, and is discussed there in detail.

2.2.2 The spread-skill relationship

To be of most use, an ensemble system needs to be calibrated such that the range of possibilities derived from the ensemble are representative of the true uncertainty in the forecast (Buizza, 1997; Leutbecher and Palmer, 2008). This relationship between the *spread* in ensemble forecasts and the *skill* of the ensemble when compared to observations is known as the spread-skill relationship: for a perfect ensemble, spread should equal skill. This spread-skill relationship is often (e.g. Buizza et al., 2005; Kong et al., 2007; Bouttier et al., 2012; Baker et al., 2014) used as a tool to analyse ensemble performance where the spread is given by the standard deviation of ensemble members and the skill given by the RMSE in the mean for comparison against observations. Thus, the spread is an estimate of the variance of the ensemble member forecasts, and the skill a measure of the average differences between ensemble members and observations.

Both the Standard Deviation and RMSE are point metrics and suffer from the problems discussed above. New spatial methods of characterising and evaluating the ensemble spread and skill (referred to as the "spatial spread-skill relationship") will be presented in Chapters 4 and 6.

2.3 Multivariate methods

Standard ensemble evaluation techniques calculate the ensemble mean and standard deviation. Along with the problems associated with applying these metrics at the grid scale there are also issues relating to their physical meaning. In particular, Ancell (2013); Hollan and Ancell (2015) show how the ensemble mean is not on the same attractor as the ensemble members and consequently is unphysical. Hence, new evaluation methods that are physically representative of the ensemble are needed. This section provides the background theory of multivariate correlations, a tool used in the Data Assimilation community to represent inter-member relationships. Correlations will be used in Chapter 7 of this thesis to investigate spatial and inter-variable relationships within a convective scale ensemble.

2.3.1 Basic formulation of correlations

The variance of a population x measures how the population is distributed around the mean and is given by

$$V(x) = \langle (x - \langle x \rangle)^2 \rangle, \tag{2.1}$$

where $\langle\rangle$ denotes the expectation value. The standard deviation of x is then given by

$$\sigma = \sqrt{V(x)}. \tag{2.2}$$

The covariance is calculated between two populations x and y:

$$cov(x,y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle. \tag{2.3}$$

Although Equations 2.1- 2.3 are true for an infinite sample, when we are considering a limited sample of size N the equations above are biased for Gaussian errors. To remove this bias the equations above are altered to the following form:

$$stdev = \sigma_x = \sqrt{\frac{\sum_{i=1,N}(x^i - \overline{x})^2}{N-1}}, \tag{2.4}$$

$$cov(x,y) = \frac{\sum_{i=1,N}(x^i - \overline{x})(y^i - \overline{y})}{N-1}, \tag{2.5}$$

where x^i and y^i are ensemble member forecasts of quantity x and y respectively and $\overline{x}, \overline{y}$ are the ensemble mean values.

The covariance is extended for n such quantities (represented by n-dimensional vectors $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)^T$), to give the covariance matrix:

$$cov(\mathbf{x},\mathbf{y}) = \begin{pmatrix} cov(x_1,y_1) & cov(x_1,y_2) & \cdots & cov(x_1,y_n) \\ cov(x_2,y_1) & cov(x_2,y_2) & \cdots & cov(x_2,y_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(x_n,y_1) & cov(x_n,y_2) & \cdots & cov(x_n,y_n) \end{pmatrix}. \tag{2.6}$$

Thus, each element *p,q* in the covariance matrix is given by the covariance between the N ensemble forecasts of quantities x_p and y_q. When $\mathbf{x} \neq \mathbf{y}$ we have a multivariate covariance matrix and for $\mathbf{x} = \mathbf{y}$ a univariate covariance matrix. For the univariate case the elements along the leading diagonal are the variances. For both cases the matrix is symmetric.

The correlation coefficient is given by the covariance normalised by the standard deviation as given in Equation (2.7). This gives a dimensionless measure $-1 \leq$

correlation$(x, y) \leq 1$:

$$\text{correlation}(x, y) = \rho(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y}. \tag{2.7}$$

The sampling error in the correlation coefficient σ_ρ can also be estimated (Barlow, 1989):

$$\sigma_\rho = \frac{1 - \rho^2}{\sqrt{N - 1}}. \tag{2.8}$$

2.3.2 Correlations in Data Assimilation: overview

In data assimilation the covariance information for all state variables, and across all spatial points, is contained in a matrix known as the **B** matrix. This matrix contains error information about the background (a priori) state needed to provide an initial guess for the assimilation scheme. The structure of **B** is similar to that described in Equation 2.6 but instead of vectors **x** and **y** representing one quantity, they represent all the forecast variables at each point in three dimensional space. For example, if our forecast only contained variables T, p and u then we would have $\mathbf{x} = (T_1, T_2, \cdots, T_n, p_1, p_2, \cdots, p_n, u_1, u_2, \cdots, u_n)^T$ and $\mathbf{y} = \mathbf{x}$, where the indices $1, \cdots, n$ represent all spatial points (Bannister, 2008a).

An accurate representation of **B** is crucial for a successful data assimilation scheme as **B** spreads information both in three dimensional space and between variables. This can be seen from the covariances in Equation 2.6 that relate different variables and spatial points. **B** also gives a weighting to the background state giving a measure of the importance that should be allocated to it in the assimilation. It is impossible to work with **B** explicitly as it is prohibitively large, containing of order 10^{14} elements. Hence, a number of approximations are necessary to use **B** in a variational data assimilation framework. Common simplifications include the assumption that **B** is static (no flow dependency), hydrostatic, in near geostrophic balance, and that control variable structures do not vary in the horizontal (Bannister, 2008b).

The selection of control variables varies between forecasting centres; common choices include vorticity or horizontal wind speeds, temperature and surface pressure (together referred to as unbalanced mass), stream function, velocity potential and specific humidity (Bannister, 2008b). Some of these variables, of particular interest to the data assimilation community, will be considered in the work of this thesis (Chapters 5 to 7). Recent research has focused on specific humidity, as it has been shown to have a significant impact on forecast skill (Andersson et al., 2005, 2007; Ingleby et al., 2013). Due to its importance in error growth processes specific humidity is also of interest for future work on convective scale assimilation. At the convective scale another important variable to assimilate is rainfall through the use of radar data; this work is ongoing (Sobash and Stensrud, 2013; Simonin et al., 2014) .

2.3.3 Correlations in Data Assimilation: estimating **B**

Initially the background error covariances must be measured in what is known as the 'calibration step'. Methods of computing this include the analysis of innovations (differences between observations and the background), differences between forecasts of the same event at different lead times and the ensemble method (Bannister, 2008a). Results from these approaches can then be compared with the assumed background error structures to evaluate the validity of the assumptions made in modelling **B**.

B at synoptic scales

Empirically it is found that covariances depend on the latitude, altitude and the variables considered. Longer horizontal length scales are found in the tropics and stratosphere, and short vertical length scales are found in the tropics (Ingleby, 2001). Ingleby (2001) found that results obtained from the UK Met Office global 3D-var assimilation scheme were found to capture the main features of the empirical results, with the exception of the lengthening of horizontal length scales in the stratosphere. It was also found that, in general, longer horizontal length scales resulted in longer vertical length scales and, for balanced variables, that there was more variation in vertical correlation than for those not in balance.

B at the mesoscale

On the mesoscale, correlation length scales are found to be smaller than those found using global models, with more variability in the vertical (Berre, 2000). This is attributed to the inclusion of smaller scales in mesoscale models, and the absence of global length scales in limited area models. Interestingly, in order for them to be better represented, mesoscale correlations must be calculated on the full high-resolution grid: correlations downscaled from a lower resolution model are noisy below the lower resolution model's grid scale (Bölöni et al., 2015). At the mesoscale, the couplings between the horizontal and vertical are important with the degree of geostrophy decreasing with decreasing scale for mass and vorticity fields (Berre, 2000). As expected, large scale balance may break down at small scales. Zhang (2005) focused on the error growth in mesoscale forecast errors, and found that the error growth scales also depended on the background dynamics. Although comparison between the full mesoscale model run and a dry run with no latent heating or cooling showed large differences in the magnitude of the variance, the covariance structures were similar between the runs. This suggests that, even for biased forecasts, the covariance structures may contain useful information.

B at the convective scale

Results from Vetra-Carvalho et al. (2012) and Bannister et al. (2011) suggest that at the convective scale (a grid spacing of 1.5km was taken for these studies) traditional balance relations no longer hold. For example, hydrostatic balance was found to be a good approximation in non-convective regions but not when convection was present (Vetra-Carvalho et al., 2012). In particular, in the lower part of the free atmosphere where convection is strongest (as measured by the relative strength of unbalanced to balanced ensemble member perturbations), the correlations were higher for the model than would be found through hydrostatic equilibrium: greater convective vertical mixing results in stronger vertical correlations. A departure from spatial isotropy is found at the convective scale for some variables such as potential temperature and specific humidity, but not for other variables such as the velocity potential (Bannister et al., 2011). Agreeing with Vetra-Carvalho et al. (2012), Bannister et al. (2011) found that the degree of balance is reduced inside regions of convection, suggesting that to accurately represent errors on the convective scale, the balance constraints on **B** should be modified. This highlights the need for research into the covariance structures seen at the convective scale.

2.3.4 Correlations from ensembles: compensating for sampling uncertainty

In the preceding discussion the effects of sampling uncertainty have been mentioned in several contexts: in the selection of ensemble initial conditions, in the verification of high resolution forecasts, and in the formulation of correlations from a limited sample (Equation 7.5). In Data Assimilation the problem of sampling uncertainty occurs when the **B** matrix is calculated from an ensemble with limited membership. Several methods have been developed to help address this problem including the localisation of correlations in spatial (Hamill et al., 2001), spectral (Buehner, 2005) and wavelet (Pannekoucke et al., 2007) space, and combinations of these methods (Buehner, 2012). Adaptive localisation methods, that contain information about "errors of the day" are also being tested (Anderson, 2007; Bishop and Hodyss, 2007, 2009a,b). Localisation methods will be discussed in the proceeding subsections, and links with the spatial methods of the verification community highlighted. In addition to covariance localisation, to ensure stability of the Data Assimilation scheme, it is necessary to inflate the ensemble estimate of **B** through multiplication with an inflation factor (greater than one). The optimal inflation factor is dependent on the ensemble size (e.g. Hamill et al., 2001), with smaller ensembles requiring larger optimal inflation factors.

Localisation in physical space

Location in physical space damps correlations at large spatial separations, making use of the fact that correlation magnitudes tend to decrease with increasing spatial separations, and that low value correlations tend to be noisier. The theoretical basis for this spatial localisation was first discussed in Gaspari and Cohn (1999), and it has since been shown to dramatically improve results from the Ensemble Kalman Filter (Hamill et al., 2001). The optimum filter length scale depends on the ensemble size (Hamill et al., 2001; Kirchgessner et al., 2014); even without considering variability in location and meteorological situation there is not a fixed scale for all situations.

Localisation in spectral space

Spectral localisation weights correlations according to their separation in wave number, again making use of the reduction in correlation magnitude seen at increasing separations. Localisation in spectral space can be related to that in physical space. In particular, assumptions of homogeneity and isotropy in physical space give a diagonal correlation matrix in spectral space (Buehner, 2005). Localisation of covariances, variances and correlations in spectral space also relates to a spatial averaging in physical space as reviewed by Berre and Desroziers (2010); Buehner (2010). Focusing on correlations, Buehner and Charron (2007) show how localisation in spectral space can be obtained by spatially shifting and weighting ensemble members in physical space. This method is closely linked to the neighbourhood methods used to increase sample size in forecast verification; the basis of the methods used in this thesis (Chapter 7).

Adaptive localisation methods

Localisation methods that operate in both spectral and spatial space have been developed and provide similar results to localisation in wavelet space (Buehner, 2012). As a wavelet diagonal approach amounts to locally averaging correlations (Pannekoucke et al., 2007) variations in correlation structure across the domain can be accounted for. Note that the amount of spatial variation still has to be specified in advance so further methods would be needed to allow adaption to a given meteorological situation. One possible adaptive method takes an element-wise power (for powers greater than one) of the correlation matrix. This damps small correlations, that are likely to be noise, more than large correlations that are more likely to be real (Bishop and Hodyss, 2007). This method is adaptive as it varies in time according to the correlation matrix (the smoothing does not have to be pre-specified) and was found to be superior to non adaptive methods (Bishop and Hodyss, 2009a,b). Another possible adaptive method where the sampling error is estimated by taking regression components from multiple groups of ensemble members is proposed in Anderson (2007). However, this method has the drawback of being very expensive: it may be better to simply enlarge the ensemble.

2.3.5 Correlations and sensitivity analysis

Sensitivity analysis seeks to determine how a particular forecast component impacts on the other forecast variables: a similar interpretation can be considered for interpreting multivariate correlations. Traditionally, methods based on the adjoint (reverse) model have been used to determine the sensitivity of different model parameters to changes in initial conditions (e.g. Errico and Vukicevic, 1992; Rabier et al., 1996). This method is effective but also expensive, requiring an adjoint model to be run (and to exist). An alternative method of sensitivity analysis based on an ensemble is described by Torn and Hakim (2008). This method is attractive as it requires no further model iterations. The ensemble sensitivity is given by

$$\frac{\partial J}{\partial x} = \frac{cov(\mathbf{J}, \mathbf{x})}{var(\mathbf{x})}, \tag{2.9}$$

where **J** is the forecast metric and **x** the analysis state variable. Both **J** and **x** are vectors of length equal to the number of ensemble members, with the ensemble mean removed from all elements. Equation 2.9 is derived by Ancell and Hakim (2007) who show this methodology to be equivalent to adjoint sensitivity analysis in the limit of a diagonal initial-time error covariance matrix. It is informative to compare the ensemble sensitivity equation (Equation 2.9) with that for calculating correlations (Equation 2.7). Both are based on a covariance with the difference lying in the normalisation: correlations are normalised using both variables (giving a dimensionless quantity) whereas, when calculating the sensitivity, the covariance is only divided by the variance of one variable retaining inverse dimensions of the quantity with respect to which sensitivity is calculated. The ensemble-sensitivity method is used in several recent studies (e.g. Garcies and Homar, 2009; Hanley et al., 2013; Bednarczyk and Ancell, 2015).

2.4 Convection over the UK

As this thesis is focused on evaluating the performance of high resolution ensembles for UK convective cases, this section provides a background to convection over the UK. The literature relating to the climatology of UK convection is reviewed, followed by sections focusing on locally forced convection, Mesoscale Convective Systems (MCSs) and the upper level forcing of convection. To maintain brevity and focus, the extensive research into convection over other parts of the world, in particular the United States, has not been reviewed in detail. References to articles reviewing convection outside of the UK, and to convective processes, are included throughout this section.

2.4.1 Overview and climatology

Overall, the north and west of the UK are wetter than areas to the southeast (e.g. Malia, 2012; Met Office, 2015; Fairman et al., 2015). Hand et al. (2004) investigated extreme UK twentieth century rainfall events. Fifty cases were identified and classified into convective (30 cases), frontal (15 cases) and orographic (5 cases). A strong seasonal variation was observed with the majority of cases (and all convective cases) occurring in the months of June to September suggesting that storms are forced by high surface temperatures. This was consistent with the findings of Warren (2014, Chapter 4) for a five year climatology of convection and quasi-stationary convective systems (QSCSs), over the UK. Convective initiation was investigated as part of the Convective Storm Initiation Project (CSIP Browning et al., 2007), and reviewed by Bennett et al. (2006) with convective initiation over the UK classified as either boundary layer forced, upper level forced. These mechanisms of convective initiation are discussed further in Sections 2.4.2 to 2.4.4.

2.4.2 Locally forced convection

Local forcing of convection is often closely linked to the surface, for example through topographically induced convergence, land-sea contrasts, and moisture variability (Bennett et al., 2006). Many of these processes are linked to stationary features (such as topography) possibly increasing the predictability for the initiation of precipitation by these mechanisms (similar to the increased predictability over orography found by Walser et al. (2004); Walser and Schär (2004) and discussed in Section 2.1.2).

One particular example of locally forced convection, convection associated with convergence over the English Southwest (hereafter SW) peninsula, was the focus of the COnvective Precipitation Experiment in Summer 2013 (COPE Blyth et al., 2015; Leon et al., 2015), and has been studied in detail for a case study of flash flooding over Boscastle and Crackington Haven in Cornwall. On the 16^{th} August 2004 a convectively unstable environment with weak uplift and highly moist low levels combined with the local topography (greater heating and roughness over land), and onshore winds, to produce a convergence line over the SW peninsula. This convergence line produced the triggering mechanism for a line of small intense cells that deposited over 200mm of rain in 4 hours (Burt, 2005; Golding et al., 2005). Golding et al. (2005) analysed the meteorology of this event through radar and satellite observations and a 1km configuration of the Met Office Unified Model (MetUM, to be discussed in Chapter 3) which was able to capture the narrow line of precipitation. They concluded that the depth reached by the convection, 6.5km, was important: convection was strong enough to generate heavy precipitation but also shallow enough to allow closely packed cells to form, with downdrafts sufficiently weak not to disturb the convergence line (and hence allow for the formation of more cells). Leoncini et al. (2013) also used the MetUM to investigate the 16^{th} August case study using an ensemble framework and introducing temperature perturbations. They concluded that the model was robust to the perturbations with all members having high levels of skill. This is consistent with the interpretation that the case had increased predictability linked to topography.

Flooding events such as that on 16^{th} August 2004 are not uncommon over Boscastle: Burt (2005) comments that similar flooding events were recorded in 1963, 1958 and 1957. The local topography over this region, with convergence lines possible from both coasts of the peninsula, and narrow valleys with small river catchments, is conducive to this type of summer flash flooding event (Burt, 2005). The combination of factors needed to produce the Boscastle flooding event are highlighted by Warren et al. (2014) who investigate a case study from 21^{st} July 2010, very similar to that of 16^{th} August 2004. Like the Boscastle case a narrow quasi-stationary convective line formed over the SW peninsula on 21^{st} July 2010. However, due to lower rain rates (resulting from a dryer environment with less Convective Available Potential Energy), a shorter duration (due to later initiation) and less stationarity (due to faster large scale evolution) these storms did not produce any flooding so did not have the impact, and hence did not generate the same interest as, the 16^{th} August 2004 case. This last point is important for the analysis of convective events: many instances of convection, though interesting physically and meteorologically, go unnoticed because they do not have newsworthy societal impact. However, analysis of these cases is still valuable and important for understanding those cases that do have impact, as demonstrated by Warren et al. (2014).

2.4.3 Mesoscale convective systems

Mesoscale convective systems form when individual convective clouds organise into a larger scale structure with horizontal extent greater than 100km. The typical structure of an MCS is a region of heavy convective precipitation, followed by a broad stratiform region (Houze and Betts, 1981). A comprehensive review of MCSs is given in Houze (2004): here we focus on MCSs over the UK. Gray and Marshall (1998) construct a climatology of MCSs over the UK from 1981-1997 which was extended to events from 1998-2008 by Lewis and Gray (2010). These studies show that MCSs occur on average twice a year (giving 57 cases from the combined climatologies), in the months May-September where temperatures are high enough to produce the required Convective Available Potential Energy (CAPE Gray and Marshall, 1998). The majority of MCSs initiated over continental Europe before being advected over the UK (44/57) resulting in the highest concentration of MCSs occurring over the South East of England.

Most MCSs over the UK occur in what is known as a "Spanish plume" (Lewis and Gray, 2010). This synoptic situation for a Spanish plume is shown in Figure 2.1. An upper level trough advects south-west across the UK with an associated surface cold front extending south through Spain, and a strong temperature gradient. Pressure drops

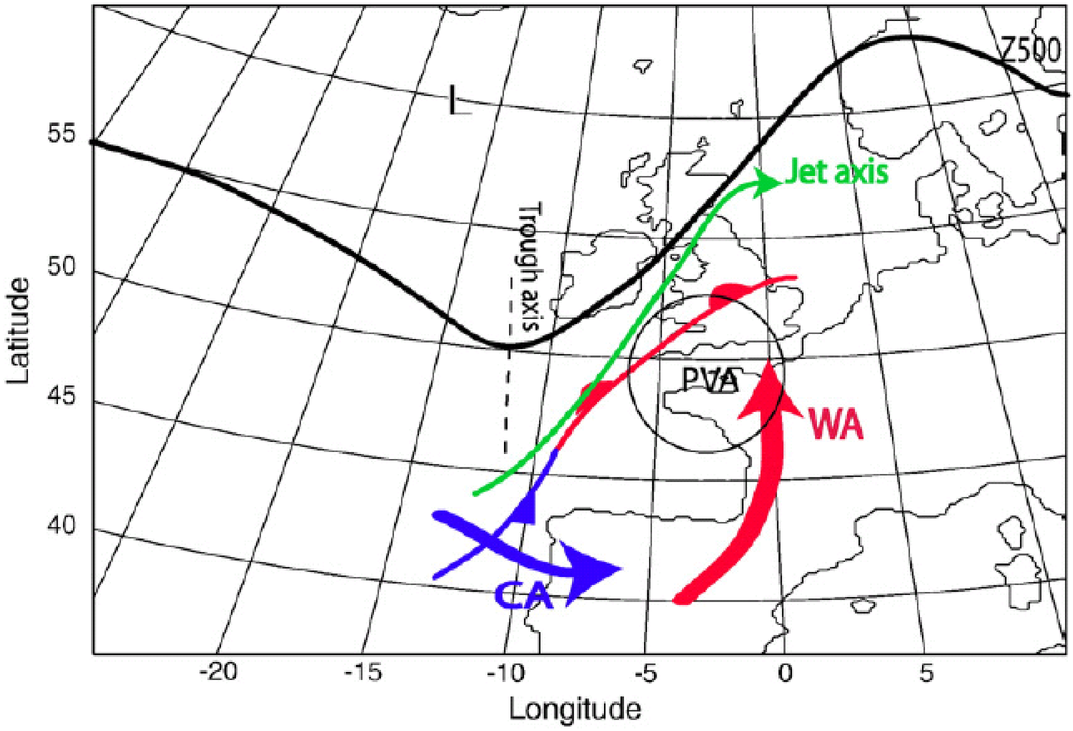


Figure 2.1: Schematic depicting the synoptic situation for a Spanish plume. Areas of cold advection (CA, blue), warm advection (WA, red), and positive vorticity advection (PVA, black circle) are shown along with the jet axis (green) and trough axis (black dashed). Taken from Lewis and Gray (2010) (their Figure 12a). Reproduced with permission from permission from Elsevier.

over Iberia ahead of the cold front due to positive vorticity advection, leading to warm advection over Iberia and a further drop in pressure and further ascent. Cold advection is seen over the Iberian west coast. As the warm dry area of Iberian air moves north it overrides warm moist air near the surface creating a strong lid which allows CAPE to build up.

Due to their high impact, several detailed case studies have been conducted of MCSs over the UK. Browning and Hill (1984) investigate an MCS using radar data and satellite imagery. Through this they develop a conceptual model for the system, splitting its evolution into four stages: transition from individual convective showers to a cluster of convection; rapid increase in system size, continued generation of cells and expanding stratiform region; decrease in convection strength but continued stratiform rain and decrease in size and rain intensity. Clark et al. (2014b,a) investigate an MCS occurring over southern England on 25th August 2005 as part of the CSIP project. Both in-situ observations and small grid length model simulations (1km grid length versions of the MetUM) were considered and it was concluded that the model results were remarkably similar to the observations. This is important for the work in Chapters 5, 6 and 7 where the ensemble behaviour is studied for an MCS case: the model (albeit at smaller grid spacing than we consider) is capable of representing the detail of MCS systems.

2.4.4 Upper level forcing of convection

As highlighted by Hand et al. (2004), convection over the UK is often associated with synoptic forcings. Potential instability can be generated though a positive potential vorticity (PV) anomaly associated with a lowering of the tropopause. Air below the PV anomaly is destabilised and the associated ascent can trigger convection. For an overview of PV the reader is referred to Martin (2006, Chapter 9) and Hoskins et al. (1985). Roberts (2000a) links the lowering of the tropopause to dry regions in satellite water vapour (WV) imagery. At wavelengths of around 6.7μm radiation is strongly absorbed by water in the atmosphere so that dry regions are depicted dark, and cloudy or moist regions depicted as white, on the WV imagery. As a lowering of the tropopause results in dry tropospheric air reaching lower levels, tropopause depressions appear as dark regions on the WV imagery. A tropopause fold is often part of the dry intrusion of a extratropical cyclone as described by Browning (1997).

The relationship between WV imagery and thunderstorms was investigated statistically by Roberts (2000b) who considered thunderstorm clusters over western Europe and the Atlantic from summer and winter periods. It was found that 60% of storm clusters showed a relationship with WV imagery falling into one of six categories based on the location of convection with respect to the dry region: dry edge, dry eye, inner crescent rim, dry tip, dry plume and orographic. Out of these categories dry edge storms were the most common (over 50% of cases) followed by dry arc (over 15% of cases). More detailed case studies linking thunderstorms to WV imagery are found in Browning and Roberts (1994, 1995); Browning et al. (1996). Given these results WV imagery will be used as a tool in the analysis of the case studies presented in Chapter 5 of this thesis.

Although this review has treated upper level and locally forced convection separately it should be emphasised that, in the real world, these occur rarely in isolation: for example, the Boscastle flooding case described in Section 2.4.2 occurred due to local effects in an environment that was already unstable. Further examples of cases that are both synoptically and locally forced are found in Morcrette et al. (2007), Russell et al. (2008) and Russell et al. (2009).

Chapter 3

Data, models, and overall methodology

3.1 Introduction

This chapter provides an introduction to the models, data, and overall methods used in Chapters 4 to 8. First, details of the numerical weather prediction (NWP) model used in this thesis, the Met Office Unified model (MetUM), will be presented in Section 3.2. Next the sources of observations will be discussed: radar data (Section 3.3.1) and satellite imagery (Section 3.3.2). Finally, Section 3.4 gives an overview of the analysis methodology used throught this thesis.

3.2 The Met Office Unified Model

The MetUM is a numerical weather and climate prediction model, developed and run by the Met Office. The model is “unified” because the same core code is used for all resolutions and timescales. The first version of the MetUM became operational in 1991, made possible by the developement of a general non-hydrostatic dynamical core (prior to this the Met Office ran non-unified models). Since 1991, the MetUM has been regularly improved and updated.

The MetUM is used for all the NWP forecasts discussed in this thesis; version 7.7 for the cases in Chapter 4 and version 8.2 for the cases in Chapters 5 to 8. The dynamical core and physics parametrisations of the MetUM will be discussed in Sections 3.2.1 and 3.2.2 respectively. This is followed by an introduction to the deterministic and ensemble configurations considered in this thesis.

3.2.1 Dynamical core

Versions 7.7 and 8.2 of the MetUM use the 'New Dynamics' dynamical core presented in Davies et al. (2005) with further details in Staniforth et al. (2006). The main features of this dynamical core are summarised here.

Governing equations

The MetUM dynamical core features a non-hydrostatic deep atmosphere formulation where the full Navier-Stokes equations are solved numerically. This gives the momentum and continuity equations

$$\frac{D\mathbf{u}}{Dt} = -2\mathbf{\Omega} \times \mathbf{u} - \frac{\nabla p}{\rho} - g\mathbf{k} + \mathbf{S^u} \tag{3.1}$$

and

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{3.2}$$

where $\mathbf{u} = (u, v, w)$ is the velocity vector, $\mathbf{\Omega}$ is the angular velocity vector of the Earth's rotation, ρ is the density, p is the pressure, t is time, $\mathbf{S^u}$ is the frictional force per unit mass, g is the gravitational field strength, $\mathbf{k}$ is the unit vector in vertical direction and $\frac{D}{Dt}$ is the Lagrangian derivative which, for a vector $\mathbf{A}$, is given by $\frac{D\mathbf{A}}{Dt} = \frac{\partial \mathbf{A}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{A}$.

Equations 3.1 and 3.2, along with the equation of state

$$\kappa \Pi \Theta_v \rho = \frac{p}{c_p}, \tag{3.3}$$

and thermodynamic equations of the form

$$\frac{Dm}{Dt} = P^m, \tag{3.4}$$

are solved for the prognostic variables velocity $\mathbf{u}$, virtual potential temperature Θ_v, Exner pressure $\Pi = \left(\frac{p}{p_o}\right)^{\kappa}$, dry density ρ_{dry} and mixing ratios of moist quantities. Here $\kappa \equiv \frac{R_{dry}}{c_p}$ where R_{dry} is the gas constant for dry air, c_p is the specific heat at constant pressure for dry air, p_0 is a reference pressure of 10^5 Pa and P^m are the physics tendencies for quantities m representing any of Θ_v, mixing ratios, or other included variables such as aerosols.

Equations 3.1–3.4 are solved in spherical polar coordinates for all prognostic variables except density, using a two time level semi-implicit semi-Lagrangian discretisation. To estimate the Lagrangian time derivative, the semi-Lagrangian scheme (Staniforth and Côté, 1991) considers air parcels at the end of each time step (the arrival point) and calculates where they came from (the departure point) (Kalnay, 2003, Ch.3). The two-time-level time integration scheme is semi-implicit and uses a predictor-corrector approach. In this scheme, the time-level $n+1$ terms needed for the semi-Lagrangian scheme are first approximated by their values at time-level n and then updated to give better estimates (Davies et al., 2005). To ensure the conservation of mass, a semi-implicit Eulerian flux form of the continuity equation is used. The discretisation leads to a coupled linear set of implicit equations, which are written as a single equation with one unknown Π', where $\Pi' = \Pi^{n+1} - \Pi^{n}$ at time-level n. The resulting elliptic equation is then solved using a generalised conjugate residual iterative solver (Eisenstat et al., 1983; Smolarkiewicz and Margolin, 1994).

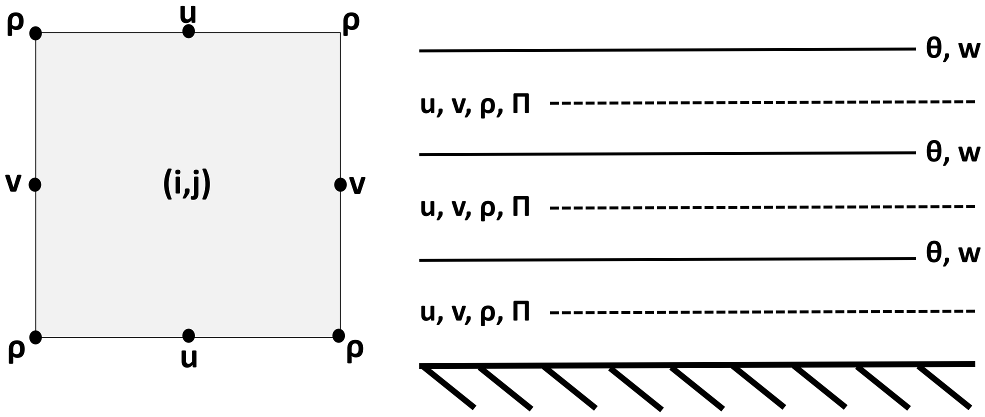
Model grid

The MetUM uses a horizontal latitude-longitude (ϕ,λ) staggered Arakawa C-grid (Arakawa and Lamb, 1977) where the horizontal wind components u and v are held at grid points displaced by half a grid spacing, in the λ and ϕ directions respectively, from other variables. Scalar variables are held at the poles. In the vertical a Charney-Phillips staggering (Charney and Phillips, 1953) is used with θ and w held on 'θ-levels' including the surface and model top, and u, v, ρ and Π held on 'ρ-levels' located half way between the θ-levels. Example Arakawa C and Charney-Phillips grids are given in Figure 3.1. These grids were chosen as they give the best geostrophic adjustment properties, and the best representation of the dissipation of gravity waves (Davies et al., 2005).

Terrain-following hybrid height coordinates η are used in the vertical, related to the true height coordinates z by

$$z = \begin{cases} \eta z_T + h\left(1 - \frac{\eta}{\eta_f}\right)^2, & 0 \le \eta < \eta_f \\ \eta z_T, & \eta_f \le \eta \le 1 \end{cases} \tag{3.5}$$

where h is the height of the orography above the mean earth radius (i.e. $h = 0$ over the



(a) Arakawa C-grid (b) Charney-Phillips grid

Figure 3.1: Examples of the horizontal (a) and vertical (b) grids used in the MetUM. Example variables west-east velocity component (u), north-south velocity component (v), vertical velocity component (w), density (ρ), potential temperature (θ) and Exner pressure Π, are shown.

ocean) and z_T is height of the model lid (fixed). Hence, model levels are terrain-following up to a height of η_f, above which the model levels become flat. Near the surface η is chosen to be quadratic to give a higher resolution at low levels: $\eta_k = \left(\frac{k}{N}\right)^2$ for model level k and total number of vertical levels N. Higher resolution near to the surface allows the larger vertical gradients and fluxes associated with the boundary layer to be better resolved. Above a certain height the vertical grid spacing is increased (i.e. it is no longer quadratic). This helps to reduce the reflection of vertically propagating waves from the upper boundary (Davies et al., 2005).

The equations of the MetUM dynamical core (Section 3.2.1) are written for a general grid, where the coordinate poles need not line up with the geographic poles. This is useful for limited area models (LAMs) as the coordinate grid can be rotated so that the LAM domain is centred upon the equator of the coordinate grid (instead of at its geographical location). Centring on the equator gives a more uniform grid.

3.2.2 Parametrisation schemes

In the MetUM, processes that cannot be resolved on the grid are modelled using physics parametrisations. This section gives a brief overview of the physics parametrisations active in the forecasts considered in Chapters 4 to 8.

Cloud

At versions 7.7 and 8.2, the MetUM has two cloud schemes available, a diagnostic cloud scheme based on Smith (1990) and the prognostic scheme of Wilson et al. (2008b,a). At the time of the forecasts considered in this thesis, the prognostic scheme (PC2) was operational in regional and global configurations of the MetUM, and the diagnostic scheme was operational in the UK model. The main features of the cloud schemes are summarised below; for more details of both schemes the reader is referred to Wilson (2011).

The diagnostic cloud scheme uses the grid-box mean total water content (liquid+ vapour) and liquid temperature, to calculate the condensation and evaporation of cloud liquid water. Fluctuations within the grid box are parametrised using a symmetric triangular distribution, defined using a critical value of relative humidity (RH_{crit}). The cloud fraction ranges from zero when $RH < RH_{\text{crit}}$, to 0.5 when the relative humidity is given by $RH < (5 + RH_{\text{crit}})/6$. A cloud fraction of 1.0 is obtained when RH=1 (Wilson, 2011). The value of RH_{crit} is pre-defined, and gradually decreases from 0.91 (in the lowest model levels) to 0.8 (Hanley et al., 2015).

In PC2 the cloud volume fraction and condensate (total, liquid, and ice) are prognostic variables which are, at each time step, modified by the other physical processes in the model: large scale precipitation, radiation, orographic drag, semi-Lagrangian advection, convection, boundary layer effects, and pressure changes. Cloud erosion processes due to the entrainment of dry air into cloudy regions, and mixed phase microphysics (Wilson and Ballard, 1999) are also included. More details of the PC2 scheme, and its implementation in the MetUM, can be found in Wilson et al. (2011). As the PC2 scheme takes into account of the effect of different physical processes at different times and locations it is considered to be more physical than the Smith (1990) scheme which only diagnoses cloud from the temperature and humidity at a given point and time.

Convection

The purpose of a convection scheme is to represent the effects of unresolved moist convection on the model variables. This is needed to prevent instability from building up in the model (possibly resulting in the build up of unrealistically strong, explicitly resolved storms), and to reduce biases in model variables, such as moisture, which are modulated by convective processes.

The convection scheme is active in the regional and global configurations of the MetUM used to provide boundary and initial conditions to the UK forecasts presented in this thesis. As resolution increases to within the so-called “grey-zone” (grid lengths of the order of 1-10 km), deep convection may be permitted by the explicit model dynamics, although it may not be adequately resolved. The treatment of convection at these resolutions is a subject of ongoing research (e.g. Holloway et al., 2014). If the convection scheme is enabled for simulations within the grey-zone, it may have the effect of removing instability from the model before the convection can be represented by the model dynamics. This can result in an underestimation of rain, and a poor representation of the diurnal cycle (e.g. Lean et al., 2008). The alternative is to run without a convection scheme at resolutions that are sufficiently high to permit convection. This is done for operational UK configurations of the MetUM (1.5km - 2.2km grid spacing). It has been shown that these models perform better without a convection scheme, with a more realistic organisation of showers (Roberts, 2003; Lean et al., 2008). However this solution is not perfect, with problems including the late initiation of precipitation, and too few, too heavy showers (e.g. Tang et al., 2013).

When convection is parametrised in the MetUM, this is done using a mass flux scheme based on Gregory and Rowntree (1990), but with many modifications (Stratton et al., 2012). This is a quasi-equilibrium scheme which assumes that convection is in near-balance with environmental forcing. Deep, shallow and mid-level convection are treated separately within the scheme and dry convection in the boundary layer is considered by the boundary layer scheme. First, to determine which form of the scheme to use, the stability of the surface layer is assessed and, if it is found to be unstable, an undilute parcel ascent is calculated from the top of the surface layer (0.1× boundary layer depth). Precipitation forms when the cloud condensate within the parcel exceeds a predetermined value. The top of the ascent is defined as the level at which the parcel is no longer positively buoyant. If this level is lower than 2.5 km, or lower than the freezing level, the shallow scheme is employed, if not then the deep scheme is used. Instability from points above the boundary layer is considered using the mid-level scheme.

A bulk cloud model, which considers the combined properties of an ensemble of clouds, is used to model the effects of the convective plumes on the resolved scales. This includes the consideration of updrafts, downdrafts and entrainment/detrainment into the cloud. Different entrainment/detrainment profiles are used for deep and shallow convection. The deep and mid-level convective schemes use a CAPE closure based on Fritsch and Chappell (1980), where the cloud base mass flux is calculated as that which would be required to reduce the CAPE to zero over a prescribed timescale, in the absence of convective forcing. The shallow scheme directly links the cloud base mass flux to the sub-cloud convective vertical velocity scale (Grant, 2001).

Precipitation

The large-scale precipitation ('microphysics') scheme is based on Wilson and Ballard (1999) but updated since (Wilkinson, 2012). The scheme considers the variables vapour, liquid, ice aggregates and rain. The microphysical processes between these variables are modelled, and the transfer of moisture and associated latent heating calculated. This includes processes relating to the nucleation, deposition, sublimation, aggregation and rimming of ice, the capture of raindrops by ice, evaporation of rain, accretion, autoconversion and fall of rain and ice under gravity. For the high resolution simulations considered in this thesis, precipitation and ice are both treated as prognostic variables, allowing horizontal advection of rain across grid boxes and allowing a delayed surface precipitation response to the dynamics. To allow the cloud field heterogeneity to be modelled, the grid boxes are divided into 8 sub-regions for the calculation of microphysical processes. A uniform distribution of liquid rain, ice cloud, and liquid water is assumed across each grid box.

Boundary layer and sub-grid turbulence

Turbulent processes occur on scales smaller than the model grid, and must hence be parametrised. This is particularly true in the boundary layer, where surface effects such as heating/cooling and friction result in significant turbulent fluxes of heat and moisture (e.g. Garratt, 1994). In the MetUM these processes are modelled using a boundary layer scheme based on Lock et al. (2000). This scheme uses a first order closure to parametrise the turbulent fluxes with the addition of non local terms. Several different boundary layer types are defined including stable, well-mixed and cumulus-capped. The boundary layer type is determined using a parcel ascent from the surface to locate unstable layers, and a parcel descent from the cloud top to identify the depth of turbulent mixing. The depths of the unstable and mixed layers are then used to determine the eddy-diffusivity surface and stratocumulus profiles, K^{surf} and K^{Sc}. Mixing coefficients based on the local Richardson number $K(R_i)$ are also calculated. The vertical eddy flux for a variable χ is then given by

$$\overline{w'\chi'} = -\max[(K_\chi^{surf} + K_\chi^{Sc}), K_\chi(R_i)]\frac{\partial\overline{\chi}}{\partial z} + K_\chi^{surf}\gamma_\chi, \tag{3.6}$$

where $K_\chi^{surf}\gamma_\chi$ is the non-local flux, and the mean and deviation from the mean of a quantity are represented by an over-bar and prime respectively. More details of the MetUM boundary layer scheme can be found in Lock and Edwards (2012).

At resolutions $O(10\text{km})$ or more, the one-dimensional (vertical) treatment of turbulent mixing used in the MetUM boundary layer scheme is appropriate. However, at higher resolutions ($O(1\text{km})$ or lower) large eddies are partially resolved, and it is also necessary to consider turbulent mixing in the horizontal direction. Finding the most appropriate method of including this mixing in NWP models is a topic of ongoing research (e.g. Hanley et al., 2015). In the operational high-resolution configurations of the MetUM considered in this thesis, a two-dimensional subgrid-turbulence scheme based on Smagorinsky (1963) was used for horizontal mixing, with vertical mixing left to the boundary layer scheme. As detailed in Halliwell (2014), this scheme calculates the eddy viscosity as

$$\nu = (c_s\Delta)^2 S f_h(R_i) = \lambda^2 S f_h(R_i), \tag{3.7}$$

where c_s is a predefined constant, Δ the maximum horizontal grid spacing, $f_h(R_i)$ a stability function defined by the boundary layer scheme, and

$$S = \frac{1}{\sqrt{2}}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right). \tag{3.8}$$

The sub-grid mixing is implemented in the MetUM as part of the explicit horizontal diffusion scheme.

Surface exchange

The MetUM boundary layer scheme diagnoses fluxes that are strictly above the surface: the surface fluxes are calculated by the surface exchange scheme (MOSES; Essery et al., 2003; Best, 2005). Version 2.2 of MOSES was operational in MetUM versions 7.7 and 8.2 and features a tiled formulation, where nine surface types are used to model the heterogeneity within each grid box. These surface types are broad-leaf trees, needle-leaf trees, temperate grass, tropical grass, shrubs, urban, inland water, soil and ice. For each tile (where appropriate) the leaf area index, canopy height , albedo and roughness length are specified. Precipitation and sub-surface properties are considered uniformly across the tiles within a grid box.

Radiation

Radiation processes occur on a molecular scale and must be parametrised in all resolutions of NWP models. In the MetUM this is done using a scheme based on Edwards and Slingo (1996), and described in detail in Edwards et al. (2014). The radiative transfer equations are simplified using a two-stream approximation where the angular variation of radiation is represented by upward, downward diffuse and direct solar fluxes. Long wave radiation (emitted from the surface) and short wave radiation (solar radiation reaching the top of the atmosphere) are treated separately. The spectrum is further decomposed into spectral bands and quasi-monochromatic regions over which the fluxes are calculated. The total flux is the sum of the quasi-monochromatic fluxes. The absorption and scattering by aerosols, water drops and ice crystals are taken into account along with Rayleigh scattering and gaseous and continuum absorption. A correction is also made for the orographic impact on short wave radiation, with short wave radiation reduced in shady regions and increased on slopes facing the sun. This is particularly important for high resolution simulations (Manners et al., 2012).

Gravity wave drag

In the MetUM two distinct parametrisations are used for gravity wave drag: a scheme to deal with drag induced by the sub-grid scale orography and another for the non-orographic drag. These are described in Webster et al. (2003); Webster and Wells (2011) and Scaife et al. (2002); Bushell et al. (2013) respectively.

The sub-grid scale orography (SSO) scheme diagnoses a surface layer h of height 2.5σ where σ is the standard deviation of the SSO. The average values of wind u, density ρ and buoyancy N^2 over the surface layer are then used to estimate the blocked layer depth b;

$$b = h - t = h - \frac{u\tau_s}{F_c N}, \tag{3.9}$$

where F_c is a pre-determined critical Froude number. The total surface stress τ_s is then partitioned into a gravity wave component (τ_{gwd}) and a blocked component (τ_{bk}) given by

$$\tau_{gwd} = \left(\frac{t}{h}\right)^2 \tau_s \tag{3.10}$$

and

$$\tau_{bk} = \tau_s - \tau_{gwd} \tag{3.11}$$

respectively. Wind increments to the large scale flow are determined by distributing τ_{bk} uniformly in the surface layer, and τ_{gwd} with magnitude decreasing in height above the surface layer.

The non-orographic drag scheme represents the spectrum of non-orographic gravity waves that propagate into the stratosphere and cannot be resolved by the model. The breaking of these non-stationary waves has a significant impact on the energy and momentum budgets (Bushell et al., 2013) and is important for simulating the large-scale stratospheric circulation, jet strengths, and the quasi-biennial oscillation. In high resolution MetUM configurations, run over limited area domains, the representation of these effects is not necessary (due to the large-scale information coming from lateral boundary conditions), and the non-orographic drag scheme is switched off.

3.2.3 Model configurations used in this thesis

Using the MetUM, the Met Office operationally runs both deterministic and ensemble forecasts, using global and limited area models. As only one model run is required for the deterministic forecasts, they are run at higher resolution than the corresponding ensemble forecasts. However, the ensemble forecasts have the advantage of forecasting the uncertainty, as well as the best-guess atmospheric state. The work in this thesis primarily uses the high resolution ensemble configuration with the high resolution deterministic configuration also briefly considered in Chapter 5. These configurations are discussed in detail below. The global and regional models are only considered in passing, and provide boundary and initial conditions for the limited area models. All models are run with 70 vertical levels with a model lid of around 40km for the UK models, and 80km for the global and regional models.

High resolution deterministic forecast

The Met Office operationally runs a high-resolution convection-permitting deterministic forecast, the UKV, on a grid with varying resolution. The inner part of this grid has a constant spacing of 1.5km covering the UK and Ireland as shown in Figure 3.2. To achieve this constant resolution grid, a rotated pole, located at the geographic location of 37.5°N, -177.5°E is used. Towards the edges of the 1.5km grid-spacing domain, the grid is stretched up to 4 km to reduce the jump in resolution when downscaling from a lower resolution model. The variable resolution grid is beneficial as it reduces artifacts near the boundaries and gives similar results to a nested model away from the boundaries (Tang et al., 2013). Thus, the cost of running the high resolution model is reduced. The current configuration of the UKV (and that used in this thesis) takes its boundary conditions directly from the global model. The initial conditions are given by a high resolution UK analysis. This analysis is produced with three-hour cycling incremental 3D-variational data assimilation, conducted on a 3km grid. Additionally, to allow the analysis to benefit from convective-scale radar observations, latent heat nudging and cloud nudging are applied to improve the initial representation of rain and cloud (Dixon et al., 2009; Dow and Macpherson, 2013; Simonin et al., 2014).

Ensemble prediction system

The Met Office Global and Regional Ensemble Prediction System (MOGREPS) consists of a global ensemble (MOGREPS-G, current grid-spacing 33km in the mid-latitudes) within which a regional ensemble (MOGREPS-R, retired in 2013, 18km grid spacing), and UK ensemble (MOGREPS-UK, 2.2km grid spacing), are one way nested. MOGREPS-G perturbations are generated using an ensemble transform Kalman filter, and then added to the Met Office 4D-Var analysis as described by Bowler et al. (2008, 2009). This perturbation strategy includes a stochastic kinetic energy backscatter scheme and localisation. Model error is addressed in MOGREPS-G using the random parameters scheme to account for sub-grid process uncertainty. The current configuration of MOGREPS-G runs eleven perturbed members and the unperturbed control every 6 hours; before version 8.2 MOGREPS-G and MOGREPS-R ran 23 members and the unperturbed control every 12 hours.

The UK ensemble, MOGREPS-UK, has been run routinely since July 2012 and operationally since June 2013 (Mylne, 2013; Golding et al., 2014). MOGREPS-UK runs with 12 members and a constant 2.2 km grid spacing over the UK and Ireland. MOGREPS-UK is one way nested directly inside the global ensemble MOGREPS-G and, to reduce the jump in resolution between the two models, the edges of the MOGREPS-UK grid are stretched up to 4km (i.e a variable resolution grid, similar to that of the UKV, with the same rotated pole, is used). The constant-resolution part of the MOGREPS-UK domain is shown in light grey in Fig. 3.2, and is the same as the constant resolution region of the UKV domain. Note that no further data assimilation is included when downscaling to the UK domain.

The cases presented in Chapters 5, 6 and 7 use this operational configuration of MOGREPS-UK. The cases discussed in Chapter 4 were run with a pre-operational configuration of the UK ensemble. In this configuration the setup of the UK domain was consistent with that of MOGREPS-UK. There were however differences in the downscaling: the pre-operational configuration UK domain was nested inside MOGREPS-R (which was nested inside MOGREPS-G) to reduce the jump in resolution from the global model, which at the time had a grid spacing of around 60km in the mid-latitudes. The horizontal grid spacings for the different ensemble configurations are given in Table 3.1.

	Version 7.7			**Version 8.2**	
Ensemble	**Global**	**Regional**	**UK**	**Global**	**UK**
Grid spacing	60km	18km	2.2km	33km	2.2km

Table 3.1: Horizontal grid spacings for the ensemble configurations used in this thesis. Note that the values quoted for global model configurations are appropriate in the mid-latitudes.

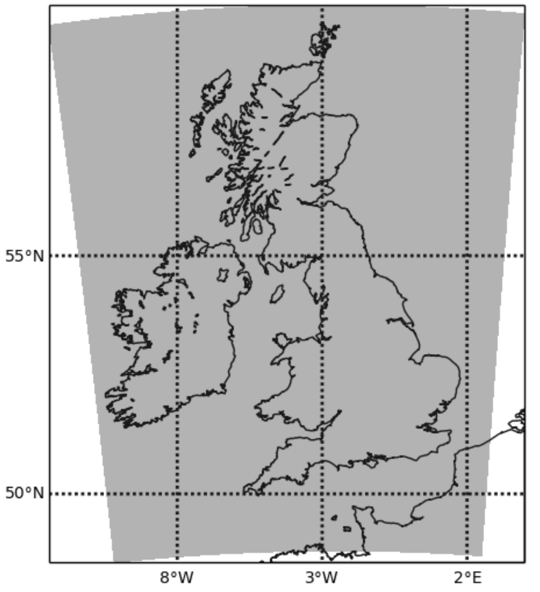


Figure 3.2: Constant resolution part of the MOGREPS-UK and UKV domains (grey).

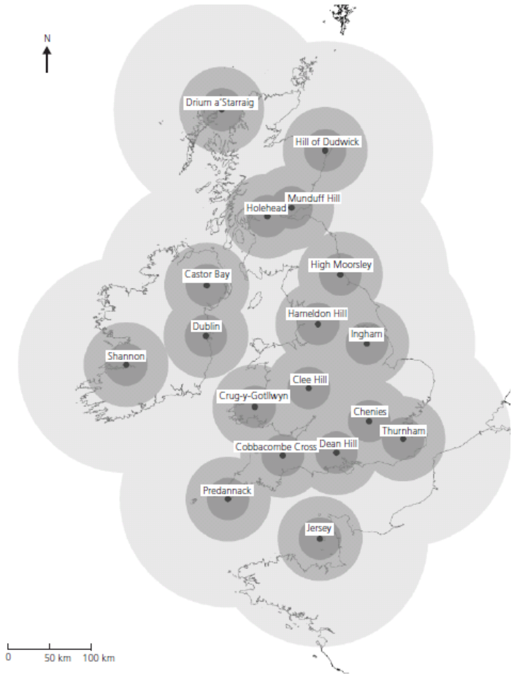


Figure 3.3: Locations and coverage of the operational UK weather radars. Ranges of 50km (dark grey), 100km (middle grey) and 250km (light grey) are shown. Taken from Harrison et al. (2012) (their Figure 1). Reproduced with permission from ICE publishing.

3.3 Other data sources

3.3.1 Radar data

Radar data are used to observe precipitation. By measuring the intensity of back-scattered radiation, meteorological radars calculate the radar reflectivity, Z, proportional to the sixth power of the hydrometeor diameter. Once the radar reflectivity is known, the rain rate R can be estimated using a Z-R relationship, $Z = aR^b$, where a and b are predetermined empirical constants (Collier, 1989, Chapters 2 and 3). In this thesis radar derived precipitation data are used both to understand the meteorology of different case studies, and to verify the forecasts. Radar data were chosen as a verification source for this work because of their high spatial coverage.

Eighteen C-band (5cm wavelength) radars are operated over the UK: 15 by the Met Office, 2 by Met Eirran and one by the state of Jersey. Each radar has a resolution in polar coordinates of $600m \times 1^\circ$ and scans 4 elevations every 5 minutes (Harrison et al., 2012). The radar locations are shown in Figure 3.3. Combining the data from all these radars, the Met Office creates, every 5 minutes, rain rate and hourly precipitation accumulation composites at 1km resolution over the UK, known as the Radarnet composite. The key features of the Radarnet system are outlined below; more details can be found in Harrison et al. (2009, 2012).

The Radarnet system grew from the Met Office Nimrod system (Golding, 1998; Harrison et al., 2000), and is often still referred to as 'Nimrod radar data'. There are six key steps in producing the Radarnet composite as presented in Harrison et al. (2012):

1. The data are quality controlled with a subtraction of the mean noise, application of a speckle filter, identification of spurious echoes, and application of a fuzzy logic approach to filter the spatial characteristics of the reflectivity. The effects of topography are also assessed and a correction made for the attenuation of the radar beam by the precipitation.

2. Variations in the vertical profile of reflectivity (the beam height increases with distance from the radar) are accounted for by using simple idealised vertical reflectivity profiles to estimate the near-surface reflectivity.

3. The radar reflectivity is converted to a precipitation rate R using the relationship $Z = 200R^{1.6}$.

4. Gauge data are used to estimate and remove any systematic bias in the radar rain rate estimates. A moving time window is used, dependant on the number of gauge-radar matches available.

5. Data from the individual radars are combined to form the rain rate composite. For points where data from several radars overlap, values are taken from the radar with the highest quality index (Harrison et al., 2009).

6. For the hourly precipitation composite, accumulations are computed using 1 minute pseudo-rain rates, calculated from the 5 minute rain rates using an advection method.

Note that, despite the measures described above to reduce errors in the Radarnet composites, there are likely to be unaccounted for errors (such as clutter that is not removed). Hence, for the work presented in this thesis, additional checks were made on the radar composites. In particular, the times when rain rates were apparently unphysical ($> 300mm\,hr^{-1}$), and times when several radars were offline, were not analysed. Occasionally, there were single points in the radar composite with no rain rate data. As these points usually occurred within dry regions, their rain rates were set to zero.

In this thesis, to make a fair comparison with the model, all Radarnet data are interpolated (bi-linearly) onto the 2.2 km resolution MOGREPS-UK grid. This interpolation was done before any comparisons were carried out.

3.3.2 Satellite imagery

Meteorological satellites provide vast amounts of information about the earth system including cloud cover and temperature, moisture and radiation. To understand the background meteorology of the case studies presented in this thesis, data were used from the Meteosat Second Generation (MSG) geostationary satellite operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). Visible and water vapour (WV) images from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) were used, obtained through the Dundee Satellite Receiving Station (NERC Satellite Receiving Station, 2015). A summary of the SEVIRI channels used for visible and water vapour images are given in Table 3.2, more details can be found in Schmetz et al. (2002); NERC Satellite Receiving Station (2015); EUMETSAT (2015). Two channels (5 and 6) were used for water vapour images: both give information about clouds in the upper troposphere with channel 5 focusing on a region around 10km, and channel 6 a slightly broader region around 8km. For the work in Chapter 5 the channel which showed the highest definition of cloud features was chosen for the WV imagery.

Channel	**Spectral range**	**Description**
1	0.56 - 0.71 μm	Visible green to visible red
5	5.35 - 7.15 μm	Mid infra-red / water vapour
6	6.85 - 7.85 μm	Mid infra-red / water vapour

Table 3.2: Description of the visible and water vapour SEVIRI channels used in this thesis (Schmetz et al., 2002).

3.4 Overall analysis methodology

In this thesis (Chapters 4 to 8) new methods are developed for the analysis of convective scale ensemble systems. The aim (as detailed in Chapter 1, Section 1.2) is to develop methods to summarise the spatial characteristics of the ensemble. Thus, it is hoped that these methods will reduce the need to inspect a large number of plots of individual ensemble member behaviour (stamp plots). However, to ensure that the new methods presented in this thesis behave as expected and give a consistent message, it is necessary to inspect the individual stamp plots. Hence, for the case studies presented in Chapters 4 and 6, stamp plots at each forecast lead time, and for the forecast variables precipitation, wind speed, horizontal divergence, temperature, cloud fraction and specific humidity, were visually inspected. Note that, as discussed in Chapters 1 and 2, the ensemble mean does not provide a suitable summary of the stamp plot information as physical characteristics of the individual fields are not maintained.

An example stamp plot, for rain rate forecasts for the 12 members of MOGREPS-UK at 17Z on 17^{th} July 2013 is given in Figure 3.4. At this time, a line of heavy precipitation is seen to the southeast of England, forecast (at slightly different locations) by 10 out of 12 of the ensemble members. The 17^{th} July 2013 is one of the case studies considered in Chapters 5 to 7.

In addition to stamp plots, Met Office surface analysis, Radarnet data, and satellite imagery were also used to ensure an accurate interpretation of the new spatial methods. To maintain brevity and focus, the use of stamp plots and observations to visually verify the results of spatial methods is not discussed further in this thesis.

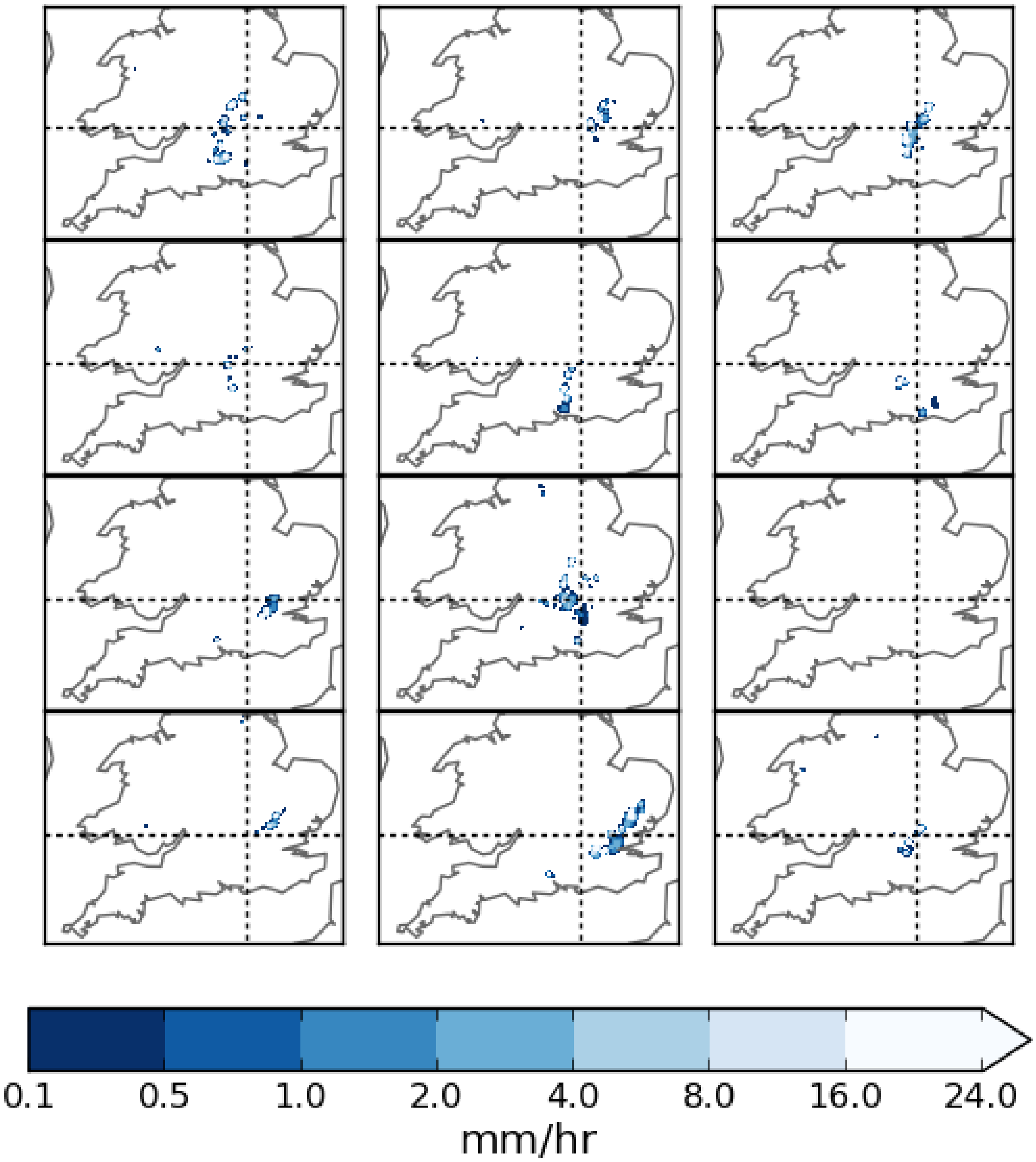


Figure 3.4: Example stamp plot of MOGREPS-UK ensemble member surface rain rates at 17Z on 17th July 2013.

Chapter 4

A spatial view of ensemble spread using the Fractions Skill Score

4.1 Introduction

As discussed in Chapter 2, to benefit fully from the information provided by convective scale ensembles, new methods are needed for their characterisation and evaluation. This chapter focuses on developing methods which give an overview of the spatial characteristics of the ensemble across the model domain. In particular, this chapter aims to answer the questions (thesis question 1; Section 1.2):

1. How can the spatial variability of ensemble member forecasts be meaningfully summarised?
2. What are suitable summary measures, and what do these tell us about ensemble performance and spatial predictability?

To answer Question 1, a new methodology is introduced for evaluating, thoroughly, the differences between members of a convection permitting ensemble and the dependence of these differences on spatial scale. These methods are based on the Fractions Skill Score (FSS, Roberts and Lean, 2008; Roberts, 2008). To answer Question 2, various considerations are discussed including the forecast evolution through different lead times, the effect of considering different threshold values for the fields used to calculate the FSS, and the comparison of different forecast variables. To demonstrate the techniques presented in this chapter two convective cases are considered using a pre-operational configuration of MOGREPS-UK (discussed in Chapter 3). The spatial spread of the ensemble members is characterised and the realism of the ensemble spread is tested by comparing with the skill against radar derived precipitation accumulations.

The technique used to determine spatial differences between members can also be used for the comparison of different model formulations within the ensemble. To demonstrate this, different model physics configurations were considered in addition to the MOGREPS ensemble members for the 08/07/2011 case study. This specific example is provided to demonstrate the utility of spatial evaluation techniques in the comparison of different ensemble formulations. Note, however, that a complete systematic evaluation comparing different types of physics configuration is outside the scope of this work. To do this it would be necessary to consider a large number of cases with different convective forcing as detailed by, for example, Stensrud et al. (2000); Keil et al. (2014). The spatial ensemble spread produced by different physics configuration strategies is evaluated and compared to that of the MOGREPS ensemble. In operational frameworks, different physics configurations are often considered in addition to initial and boundary condition perturbations and so the spatial spread produced by an ensemble with different MOGREPS members combined with different physics configurations is also investigated.

This chapter is structured as follows: in Section 4.2 the two case studies that provide examples throughout the chapter are introduced and the model configuration discussed. The FSS is then introduced in Section 4.3, followed by a description of our method of using the FSS to characterise convection permitting ensembles in Section 4.4. Section 4.5 provides example results for ensembles with different initial condition and lateral boundary condition perturbations. Results for different physics configurations are then discussed in Section 4.6. Finally, in Section 4.7 the conclusions from this work are summarised. The overall conclusions from the work of this Chapter (and those from Chapters 6, 7, and 8) are presented in Chapter 9, where areas of future investigation are discussed.

4.2 Method

4.2.1 Cases

Two convective cases were chosen to demonstrate the methods discussed in this chapter. In these cases convection occurred in different synoptic situations. The first case, 23/04/2011, was chosen as an example of organised spring convection over England

and will be referred to as the ‘organised Spring’ case. This case had a low pressure system centred to the northwest of the UK and a high pressure system centred over Scandinavia. A frontal structure stretched down across the western UK. As the front moved eastward a convergence line formed across eastern England ahead of the front. This convergence line is shown in the UK Met Office analysis at 18Z on the 23/04 (Figure 4.1a). Convective storms developed in the vicinity of this convergence line with precipitation first seen at 14Z on 23/04, and continuing until 03Z on 24/04. At 18Z a band of frontal precipitation entered the model domain from the northwest preceding an occluded front which entered the domain at 00Z on the 24/04.

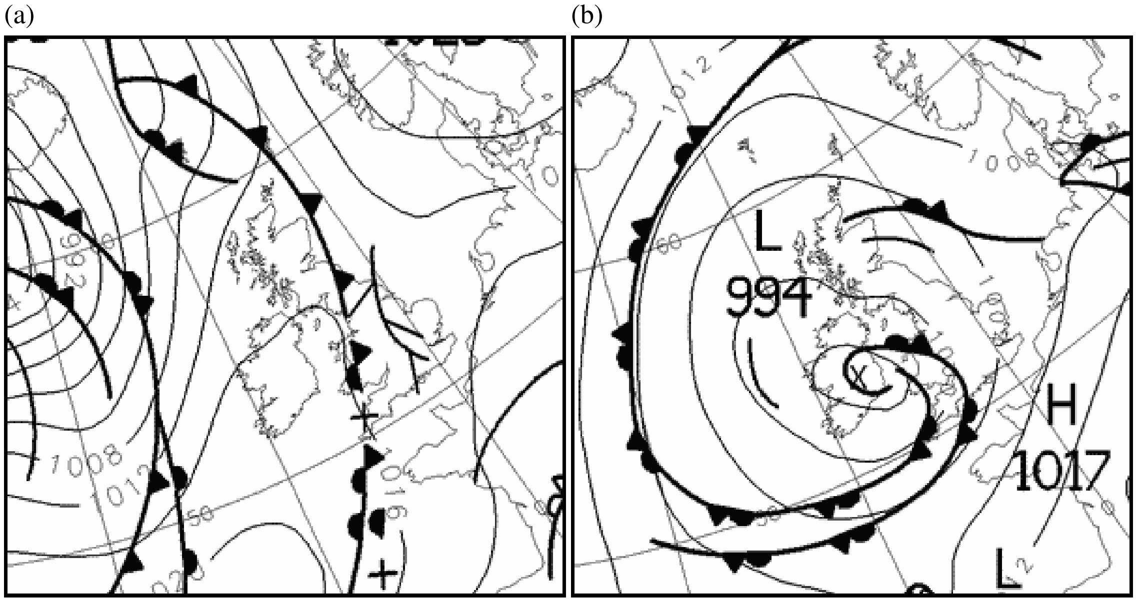


Figure 4.1: UK Met Office surface analysis valid at (a) 18Z on 23/04/2011 and (b) 06Z on 08/07/2011. Courtesy of the Met Office. Crown copyright.

The second case, 08/07/2011, featured a number of convective storms that formed over the UK in an area of instability within the circulation of a decaying low pressure system. At 06Z the low centre was situated over Ireland as shown in Figure 4.1b. Throughout the day the low centre then moved towards the northeast reaching the northeast of England by 18Z. By 14Z there were many heavy showers over Scotland as indicated by the Radarnet radar system (not shown). Convective clouds associated with these showers were also seen from visible satellite observations from the Meteosat geostationary satellite. For this case study we focus on one particular storm that formed over the

Edinburgh area of eastern Scotland and remained stationary for around four hours producing large rainfall totals (09Z – 21Z radar derived precipitation totals of over 64 mm) and flooding. In future discussion this will be referred to as the ‘flooding’ case. Previous analysis of this case by Leoncini et al. (2011) showed that the Met Office 2.2 km ensemble on this occasion gave a 30% to 40% chance of a flood-producing storm within 25 km of Edinburgh; a level of significant risk.

4.2.2 Model setup

For the work in this chapter a pre-operational version of MOGREPS-UK was used as described in Chapter 3. Analysis was completed over the 2.2km grid length region of the MOGREPS-UK domain, and additionally for the flooding case, over a small 100 km domain surrounding Edinburgh. This subdomain allows us to focus on the spatial characteristics of the flood-producing storm over Edinburgh: its location is highlighted in Figure 4.2 in dark grey.

For the organised Spring case an ensemble of 8 MOGREPS simulations were run (seven perturbed members plus a control). This reduction in ensemble size allowed 5 different physics configurations to be considered for each MOGREPS simulation (giving a total of 40 simulations). The different model configurations were:

1. A control ensemble with the standard model settings labelled “standard”.
2. An ensemble with a restricted version of the convection scheme (Roberts, 2003) as would be applied to the Met Office 4km deterministic model (labelled “conv”).
3. An ensemble with the time step increased from 25 s to 50 s labelled “time”. It is interesting to investigate the effects of a longer time step as increasing the time step reduces the computational cost of the simulation but may increase model error.
4. An ensemble with increased time step and restricted convection scheme labelled “conv+time”.
5. An ensemble with modifications to the graupel labelled “grp”. The graupel modification allows the production of graupel through the capture of rain by snow and results in an increased graupel mass. This modification has become a standard option in MetUM from version 8.0 onwards (Wilkinson, 2011).

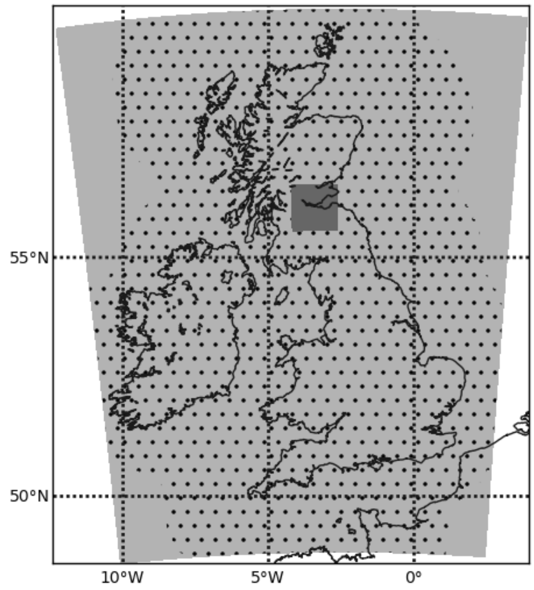


Figure 4.2: Domains of the UK 2.2 km model (light grey), 100 km subdomain for the summer flooding case (dark grey) and areas of radar coverage (dotted).

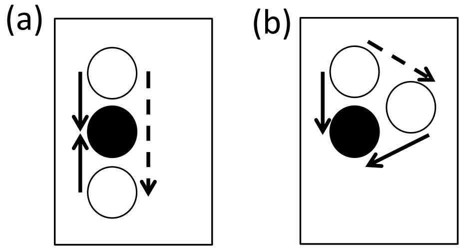


Figure 4.3: Two different idealised spatial distributions of precipitation. Individual ensemble members (shown in white) position the precipitation in different spatial locations. The control simulation (shown in filled black) may produce precipitation in the centre of that produced by individual ensemble members as shown in (a) or at the edge of the ensemble as shown in (b). Considering only the spatial separation of member-member pairs (solid arrows) indicates that (a) and (b) have the same spatial ensemble spread. Including both member-control and member-member pairs allows the differences in spread between (a) and (b) to be detected.

It must be emphasised that these model configurations were chosen to demonstrate the methodology presented in this chapter, and not as possible implementations to the UK Met Office ensemble prediction system. Note also that the model variations are neither stochastic nor designed to represent the model error, although they do, nevertheless, represent plausible alternative formulations.

The UK model for the organised Spring case was started at 06Z on 23/04/2011, and for the flooding case at 18Z on 07/07/2011. MOGREPS-G and MOGREPS-R were initiated 6 hrs and 3 hrs respectively before the UK model. For both cases the UK model was run up to lead times of 36 hours.

4.3 The Fractions Skill Score

To evaluate convection permitting ensembles in a sensible way it is necessary to choose a verification approach that considers multiple spatial scales and does not suffer from the double penalty problem where spatial errors are penalised twice: once for being a near miss, and again for being a false positive. Many possible spatial verification approaches have been proposed in recent years as discussed in Chapter 2. Here we have chosen to focus on the Fractions Skill Score (FSS) of Roberts and Lean (2008); Roberts (2008). The FSS is a fuzzy verification measure used to compare two fields within a given square neighbourhood.

Since its original formulation the FSS has been used for different applications and several further developments have been proposed. Schwartz et al. (2010) consider circular neighbourhoods to calculate the field of fractions at each grid point and then produce probabilistic guidance using the field of fractions as a neighbourhood probability. Duda and Gallus Jr (2013) also use the circular neighbourhood approach, verifying the precipitation of mesoscale convective systems. In this chapter the FSS is considered over a square neighbourhood as detailed in Roberts and Lean (2008); Roberts (2008).

Duc et al. (2013) extend the FSS neighbourhood to include temporal and ensemble dimensions to give a single FSS value representative of the ensemble. For example, the fields could be compared over a neighbourhood defined as having spatial horizontal size 10 by 10 grid points, including data from a one hour period, and from 3 ensemble members. Thus a single field of fractions including spatial, temporal and ensemble information is then compared with observations. This is useful for providing an overview of model performance but does not provide information regarding the spread - skill relationship of the ensemble or the spatial differences between individual pairs of ensemble members.

Rezacova et al. (2009) use the FSS to calculate the ensemble spread-skill relationship with the ensemble skill calculated from the FSS between ensemble member- radar comparisons and the ensemble spread from the FSS between perturbed ensemble members and the ensemble control. Following on from this Zacharov and Rezacova (2009) determine a relationship between the FSS estimates of ensemble spread and skill and use this to predict the ensemble skill given the spread. Zacharov and Rezacova (2009) consider together FSS results from differently sized neighbourhoods. This method was chosen because there is no fixed scale that can give a FSS skill value over different cases. However, as different physical behaviour is apparent at different spatial scales (e.g as shown in Roberts (2008)) it is informative also to investigate how the ensemble spread varies with spatial scale which is the subject of the work presented here. Whereas Rezacova et al. (2009); Zacharov and Rezacova (2009) only consider comparisons between perturbed ensemble members and the control, in this thesis the FSS between all independent member-member pairs is considered. Considering all members in this manner is the best representation of total spread as it includes fully the inter-member variability. Leoncini et al. (2016) discuss other possible methods of member comparison.

The FSS is described in Roberts and Lean (2008) and summarised here for ease of reading. To calculate the FSS a threshold is first selected, say for precipitation, either as a fixed value (e.g 4 mm hr^{-1}) or as a percentile (e.g top 1% of precipitation field). The field is converted to binary form with grid points set to 1 for values above the threshold and 0 otherwise. A neighbourhood size is then selected and, for each neighbourhood centred upon each grid point, the fraction of grid points with the value '1' within this square is computed. Two fields of fractions (denoted A and B), say from a model and observations, are then compared using the mean squared error (MSE). For a neighbourhood size n and domain size N_x by N_y grid points this is given by:

$$MSE_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [A_{(n)i,j} - B_{(n)i,j}]^2. \tag{4.1}$$

The fractions skill score is computed by comparing $MSE_{(n)}$ with a reference MSE, $MSE_{(n)ref}$,

$$FSS_{(n)} = 1 - \frac{MSE_{(n)}}{MSE_{(n)ref}}, \tag{4.2}$$

where $MSE_{(n)ref}$ is the largest possible MSE that can be obtained from fraction fields A and B:

$$MSE_{(n)ref} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [A^2_{(n)i,j} + B^2_{(n)i,j}]. \tag{4.3}$$

The FSS varies from 0 (complete mismatch between the fields) to one (perfect match between the fields).

Different neighbourhood sizes are considered in order to evaluate the FSS at different spatial scales. Here the neighbourhood size is defined to be the total width of the square neighbourhood in km. The smallest possible neighbourhood is 2.2 km, set by the grid scale.

No bias exists between the binary fields created using percentile thresholds as, by definition, the same number of points exceed the threshold for both fields. Hence, for percentile thresholds, the maximum possible spatial disagreement is found for two fields which place the points of interest at opposite edges of the domain. A perfect match is only obtained between fields with this maximum disagreement when they are compared over a neighbourhood of twice the smallest dimension of the domain. In other words, the FSS will only equal 1 when the neighbourhood size is equal to twice the smallest dimension of the domain. This sets the maximum neighbourhood size for percentile thresholds. For value thresholds the fields may be biased and this argument does not hold. For the examples presented here only percentile thresholds are considered and the maximum neighbourhood size is 1848 km for the UK domain and 200 km for the 100km subdomain.

4.4 How the Fractions Skill Score is used

The FSS can be calculated at a particular time between two different forecasts, or between a forecast and observations, the former giving a measure of spatial spread, the latter of spatial skill. The ensemble spread is characterised by calculating the FSS for all independent member – member pairs ($N_p(N)$, for an ensemble of N members) resulting in

$$N_p(N) = N \times (N-1)/2 \tag{4.4}$$

comparisons. Here, and for the remainder of this thesis, the control is treated as an additional ensemble member. This assumption was tested for the cases presented in this chapter, and for the six cases discussed in Chapter 5, by visual examination of the perturbed ensemble member forecasts and the control. As it was not possible to separate the control forecast from those of the perturbed members, the assumption is acceptable. Thus, for the flooding case we have 12 MOGREPS members (the 11 perturbed members and unperturbed control) and for the organised Spring case we have 8 MOGREPS members for each physics configuration (the 7 perturbed members and unperturbed control). Justification for this method comes from our interest in the total spatial ensemble spread. In this situation the spatial location of a feature in the control forecast is not necessarily at the centre of corresponding features in the perturbed members and therefore we do not wish to assign any special status to the control forecast. Figure 4.3 demonstrates the advantages of this method: when considering the control as an additional ensemble member one can distinguish the different spatial spread in Figures 4.3a and 4.3b, whereas when only comparing against the control the spread in Figures 4.3a and 4.3b is indistinguishable.

The ensemble skill is assessed by comparing the model hourly precipitation accumulations with those derived from the Met Office Radarnet radar system (described in Chapter 2). For the flooding case 1 km Radarnet radar-derived hourly precipitation accumulations were interpolated onto the 2.2 km model grid. Radarnet data at 1 km resolution were not available for analysis of the organised Spring case so 5 km data were instead used. The area of Radarnet coverage differs slightly from the UK 2.2 km domain over which the model is run and is indicated by the dotted region in Figure 4.2. All analysis involving radar data, or the comparison of model and radar data, only considers the area with radar coverage. We assume the radar data is representative of the precipitation that occurred and ignore observational errors, which would have to be considered within a routine verification framework. Visual examination of the radar fields found no obvious errors.

To assess ensemble skill each model simulation is separately compared with radar observations, whilst to assess ensemble spread we compare all possible pairings of the model runs. Again consider Figure 4.3, but this time take the filled black circles to represent the location of precipitation in the radar data. As a measure of ensemble skill we are only considering the spatial differences associated with the solid arrows. These measures of ‘spread’ and ‘skill’ consider different numbers of member-member or member-radar pairs, raising questions about a direct comparison of these metrics. Additional questions relate to the mathematical interpretation of these measures given that, for scalar Gaussian random variables, the mean euclidean distance between points is equal to twice the variance. Hence, we would expect the mean of the euclidean distances calculated between pairs of variables to equal twice the mean of the euclidean distances calculated by comparing those variables against their mean. However note that, as the ensemble member and radar forecasts are not Gaussian random variables, and the FSS does not calculate euclidean distances, this relationship does not necessarily apply to the comparison of dFSSmean and eFSSmean.

This chapter focuses on the characterisation of spatial ensemble spread, with spatial ensemble skill, as measured by eFSSmean, considered only to put the spread into context. Hence, answering questions about the direct quantitative comparison of dFSSmean and eFSSmean, discussed above, is not the subject of this chapter. Consequently, when the terms "spread"/"skill" or "over spread"/"under spread" are used in this chapter, they should always be considered qualitatively. In Chapter 6, Section 6.4, these comparison methods are considered further (for the location-dependent characterisation of ensemble spread and skill) using an idealised experiment.

Three different comparison strategies were used for the organised Spring case to characterise the differences between spatial spread in the MOGREPS ensemble and that produced through considering different physics configurations. 8 MOGREPS ensemble members ($N = 8$), and 5 different physics configurations ($N = 5$), were considered. Additionally results were produced using a subset of two physics configurations ($N = 2$) to allow spatial differences resulting from individual configurations to be investigated.

1. All independent comparisons were made between the MOGREPS members for a given physics configuration, with each physics configuration treated separately. Considering all 5 physics configurations in this manner gives $N_p(8) \times 5 = 140$ comparisons, a strategy denoted as MOGREPS5. Considering 2 physics configurations in this manner gives $N_p(8) \times 2 = 56$ comparisons, denoted as MOGREPS2.

2. All independent comparisons between the different physics configurations for a given MOGREPS member, with each MOGREPS member treated separately. Considering all 5 physics configurations gives $8 \times N_p(5) = 80$ comparisons for this strategy denoted as Physics5. Considering 2 physics configurations gives $8 \times N_p(2) = 16$ comparisons (Physics2).

3. Comparisons between different MOGREPS members which additionally have different physics configurations. For example, MOGREPS member 2 with the standard physics configuration might be compared with MOGREPS members 1,3,4,...,12 with the physics configurations conv, conv+time, time and grp. Considering all 5 physics configurations with this comparison strategy, referred to as MOGREPS5+Physics5, gives $N_p(8) \times N_p(5) = 280$ comparisons. Considering 2 physics configurations (MOGREPS2+Physics2), gives $N_p(8) \times N_p(2) = 28$ comparisons.

4.4.1 Defining the believable scale

In order to make a spatial comparison between different ensembles it is necessary to find scales which are believable and have a reasonable level of spatial agreement. For the purposes of this thesis, 'believable' scales for the intercomparison of ensemble members are derived in an equivalent manner to those scales that would be considered skillful if the comparison was instead against observations (assuming that the ensemble is well spread). This scale is quantified using the methodology of Roberts and Lean (2008) where a neighbourhood size is considered believable ('skillful') if a FSS value of

$$FSS \geq 0.5 + \frac{f_0}{2} \tag{4.5}$$

is obtained for that neighbourhood. f_0 is equal to fraction of the field considered in the FSS calculation (for example, considering the top 99^{th} percentile threshold would give $f_0 = 0.01$) and Equation 4.5 simplifies to an equality when the neighbourhood is twice the spatial difference between two binary fields (Roberts and Lean, 2008; Roberts, 2008). As f_0 is normally small Equation 4.5 can be approximated as $FSS \geq 0.5$.

4.4.2 Summarising the distribution of FSS pairs

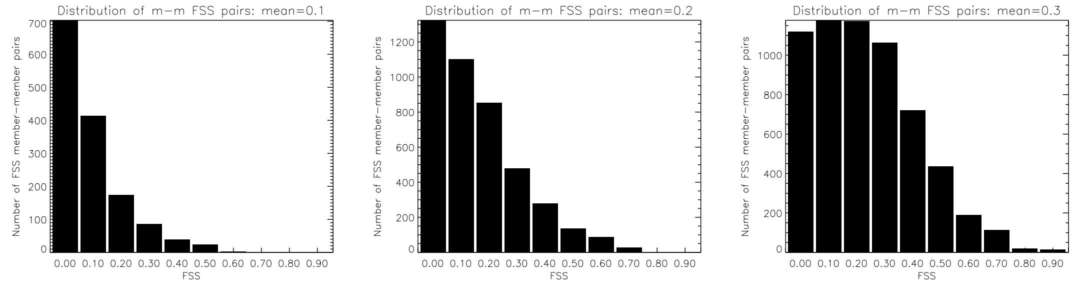
Given the large number of FSS values, FSS_i, (one calculated for each comparison) it is necessary to consolidate this information to provide an overview of spatial ensemble behaviour. Here, the mean is taken over the relevant set of FSS_i. When calculated over member-member pairs this is referred to as dFSSmean where 'd' indicates that this is a measure of ensemble dispersion. When calculated over member-radar pairs this is referred to as eFSSmean where 'e' indicates that this is a measure of ensemble error. dFSSmean gives an indication of the average spatial agreement within the ensemble for a given neighbourhood size. In other words, we can select a level of spatial agreement for the ensemble, represented by the value of dFSSmean, and ask at what neighbourhood size (or sizes) this agreement is obtained.

Note that, as FSS_i are each calculated separately using Equation 4.2, a different normalisation (i.e. a different value of $MSE_{(n)ref}$; Equation 4.3) is used for each member-member, or each member-radar comparison. Thus, each FSS_i is separately constrained to be between zero and one, irrespective of the difference between the fields A and B.

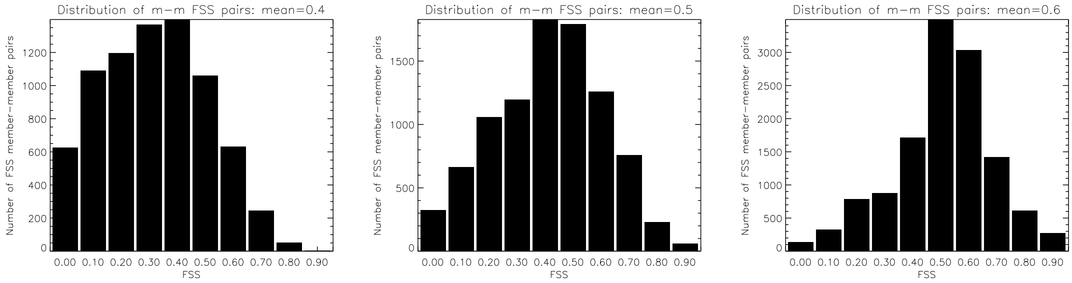
This method was used to maintain consistency with the FSS: by calculating FSS_i in this manner we create an ensemble of FSS values. Additionally, separately normalising each FSS_i allows us to focus on the spatial differences between the fields A and B. However, this method does have the disadvantage of complicating the mathematical meaning of dFSSmean and eFSSmean, resulting in a less direct link between the field differences (A and B) and the values of dFSSmean or eFSSmean.

Of course, dFSSmean and eFSSmean are only meaningful for the characterisation of the total distribution if the relevant distribution of FSS_i is approximately Gaussian. Figure 4.4 shows the distribution of FSS_i for different ranges of dFSSmean, using the standard physics ensemble for the organised Spring case. Similar distributions are found for eFSSmean, and for the flooding case. The distributions shown in Figure 4.4 are the sum of all distributions for different neighbourhood sizes, thresholds and forecast lead times, whose mean value falls within the specified range. From Figure 4.4 we can see that, for dFSSmean values from 0.45 to 0.75, the distribution is approximately symmetric, and hence the mean is a useful indicator of the distribution of FSS_i in these instances. This range is particularly important as it covers the believable scale (Section 4.4.1), which is used to quantify ensemble behaviour, and the differences between ensembles. Distributions with dFSSmean values less than 0.45 or greater than 0.75 are skewed to the right and left respectively, due to the constraint of $0 \leq FSS \leq 1$. Hence very small and very large values of dFSSmean are less well defined than those around 0.5. However, it is argued that this is not of consequence: very small dFSSmean values indicate that the members are so different that the ensemble is not believable, and should not be quantitatively evaluated at these scales, while very large dFSSmean values indicate that the members are almost identical. Additionally it should be noted that, even for very low/high vales of dFSSmean, the distributions of FSS_i are uni-modal, and the mean is still a sensible choice of a single value for characterising the distribution. Any correction to the mean due to the distributions being non-Gaussian would make the small/large dFSSmean values slightly less small/large and would not change the conclusions to the work presented in this chapter.

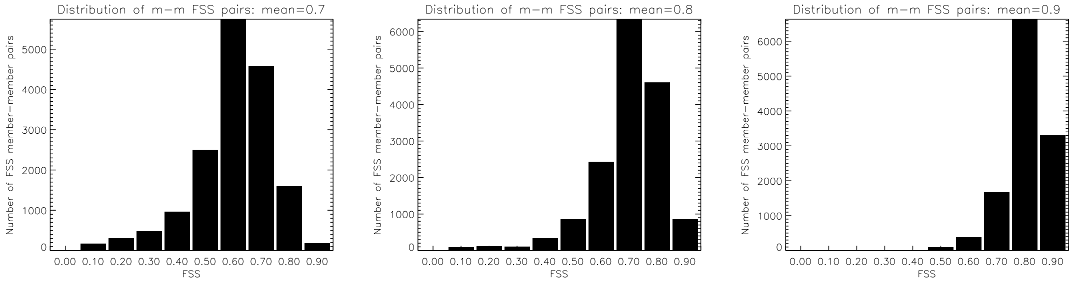
As shown in Figure 4.4, a range of FSS_i will be obtained from the different ensemble member-member pairs. For example, if the majority of ensemble members place rain at the same spatial location but a small number of members place the rain far away this



(a) $0.05 \leqslant mean < 0.15$ (b) $0.15 \leqslant mean < 0.25$ (c) $0.25 \leqslant mean < 0.35$



(d) $0.35 \leqslant mean < 0.45$ (e) $0.45 \leqslant mean < 0.55$ (f) $0.55 \leqslant mean < 0.65$



(g) $0.65 \leqslant mean < 0.75$ (h) $0.75 \leqslant mean < 0.85$ (i) $0.85 \leqslant mean < 0.95$

Figure 4.4: Distributions of all member-member FSS_i for the organised Spring case (standard physics only). The plots are separated by the mean of the distribution with data from all forecast lead times, neighbourhood sizes and the 99^{th},95^{th},90^{th} and 85^{th} percentile thresholds included.

may produce a similar value of dFSSmean as a situation in which all ensemble members place the rain at slightly different spatial locations. Hence it may also be important to summarise the range of FSS_i values surrounding dFSSmean. To do this the standard deviation of FSS values, dFSSstdev, is used. dFSSstdev is closely linked to the standard error in dFSSmean, $\frac{dFSSstdev}{\sqrt{N_{FSS}}}$ where N_{FSS} is the number of FSS_i samples used to calculate dFSSmean. As the purpose of this work is to focus on the spatial distribution of ensemble members, we consider dFSSstdev and avoid the $\frac{1}{\sqrt{N_{FSS}}}$ dependence on ensemble size. This allows the spatial distribution of differently sized ensembles to be compared.

4.4.3 Thresholding

The FSS can be calculated using either fixed value or percentile thresholds. Following on from the work of Roberts (2008); Mittermaier and Roberts (2010) we focus here on the use of percentile thresholds to allow the spatial distribution of phenomena to be investigated. Higher percentile thresholds are associated with smaller, more extreme forecast features, and lower percentile thresholds are associated with larger-scale smoother features (Roberts, 2008). Note that here, and in all future discussion, the percentile threshold is applied over the whole domain, including areas both with and without precipitation.

To understand the effect of applying percentile thresholds it is informative to investigate the values corresponding to each threshold. Examples for hourly precipitation values corresponding to the 90^{th} and 99^{th} percentile thresholds are given in Figure 4.5. These percentile thresholds are used as examples throughout this chapter. All ensemble members (grey solid lines) and radar (black lines) are shown for the organised Spring case (top) and Summer flooding case (bottom). From both cases and thresholds it can be seen that the radar percentile thresholds generally correspond to lower precipitation values than the model. This bias in the model compared to radar is an important consideration for model evaluation. However, it is also important to investigate the spatial distribution of precipitation: using percentile thresholds allows us to focus on this despite the model bias.

For the Spring case at the 90^{th} percentile threshold (Figure 4.5a) the radar values drop to zero after 16 hours. After this time radar derived precipitation covers less than 10% of the domain. This demonstrates that the 90^{th} percentile, and other percentile thresholds below the 90^{th}, are not a suitable threshold for radar precipitation accumulations for this case. For all cases (apart from the unlikely event of 100% coverage) there will be a limited area covered by precipitation in both the model and observations, and a corresponding minimum suitable percentile threshold. In an operational situation this minimum threshold could easily be calculated from the fraction of precipitation coverage. All FSS results presented in this chapter have been calculated using percentile thresholds above this minimum value.

For the Spring case the 8 MOGREPS members from the standard physics configuration are shown in dark grey in Figure 4.5a and Figure 4.5b and, although differing by up to 2.5 mm in accumulation values (for the 99^{th} percentile threshold), follow the

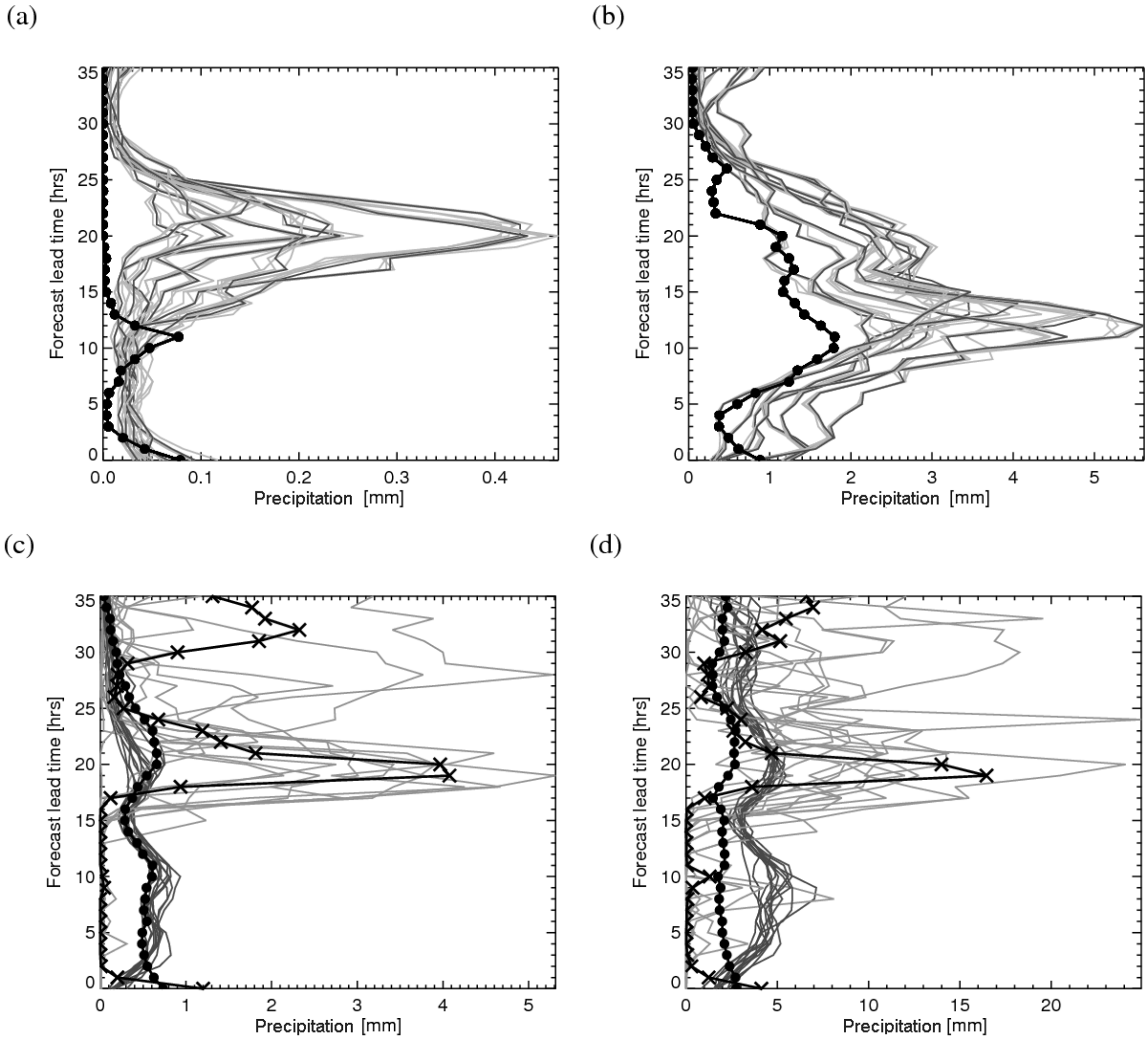


Figure 4.5: Hourly precipitation accumulation values corresponding to the 90^{th} (a,c) and 99^{th} (b,d) percentile thresholds. For the organised Spring case (top) results from all the simulations are shown. To highlight the grouping of members those with the standard physics configuration are shown in dark grey and those from other physics configurations in light grey. For the flooding case (bottom) percentile thresholds calculated using data for the full UK domain are shown in dark grey, and those for the limited area domain are shown in light grey. Radar data are shown from the area of the UK domain with radar coverage (black with circles) and, in (c,d) over the limited area domain (black with crosses).

same overall trend throughout the day. This suggests that the ensemble members produce precipitation features, such as that associated with frontal passage, at similar times. The simulations for all MOGREPS members and the other 4 physics configurations are shown in light grey with the different physics configurations clustering around the corresponding MOGREPS member. In these experiments the different physics configurations have little effect on the precipitation value corresponding to a given percentile threshold. Interestingly, Figure 4.5a and Figure 4.5b show peaks in precipitation values at different times: Figure 4.5a (90^{th} percentile) at a lead time of 20 hrs and Figure 4.5b (99^{th} percentile) at a lead time of 12 hrs. The higher threshold considers only the areas of convective precipitation, giving a corresponding value that peaks when these storms are strongest whereas the lower threshold includes frontal precipitation and peaks where this is heaviest.

The 12 members for the Summer flooding case are shown for thresholds calculated over the full UK domain (dark grey) and limited area domain (light grey). Beyond a lead time of 15 hours, when convection occurred over Edinburgh, values for the limited domain are up to 5 times larger than those over the UK domain. Considering this area separately using percentile thresholds allows the flood producing storm to be investigated. It should be noted that using high value thresholds over the UK domain would also select the Edinburgh area. However, for this highly-variable case some ensemble members missed the convection over Edinburgh, and do not produce sufficiently high precipitation values. It is not possible to choose a value threshold that is high enough to select only the area of convection, and yet low enough to include all the ensemble members. Again, this demonstrates the utility of percentile thresholds.

4.5 Results for LBC and IC perturbations

4.5.1 dFSSmean and eFSSmean

First we consider the realism of the spatial ensemble spread by comparing dFSSmean and eFSSmean for both cases. Both dFSSmean and eFSSmean were calculated over the section of the 2.2 km UK domain with radar coverage (highlighted by the dotted region in Figure 4.2). Figure 4.6 shows dFSSmean (left) and eFSSmean (right) for the organised Spring case (top) and flooding case (bottom) calculated for the 99^{th} percentile threshold over the whole UK domain. These results were computed for the 12 members of the flooding case and 8 MOGREPS members with standard physics for the organised Spring case. To check the validity of comparing these differently sized ensembles, results were also produced for the flooding case when only considering the first 8 ensemble members (not shown). These 8 member results differed only in small details from those calculated from 12 members, and lead to the same conclusions, so it was decided to show the results from the full 12 member comparisons.

Comparison of the dispersion measures (dFSSmean) for the two cases (Figures 4.6a and 4.6c) shows that, although these cases are synoptically different, with different convective forcing, the overall behaviour is broadly similar. At small scales ensemble members are very different resulting in low values of FSS. FSS values increase as the members become more similar when considered at larger scales. The temporal variability present in the ensemble spread, as measured by dFSSmean, is also clear at this threshold with the scale at which $FSS = 0.5$ varying from 150-500 km for the organised Spring case and 100-250 km for the flooding case. These scales are large because in both cases there is considerable uncertainty in the locations of the showers and showery areas. The temporal variability can be related to the evolution of physical processes. For example, in Figure 4.6a the area of larger ensemble spread (smaller dFSSmean) at lead times 13-20 hrs can be linked to greater convective activity and the highest rainfall instances (compare with Figure 4.5b) and the increase in dFSSmean (decrease in spread) from 20-25 hrs can be related to a area of spatially predictable frontal precipitation moving into the domain.

Overall there is less temporal variability in the FSS for the flooding case. This can again be related to the meteorology of the cases: precipitation in the flooding case was the result of one mechanism, instability associated with a decaying low pressure system, whereas precipitation in the Spring case was associated with both convective showers and frontal passage. Coincidentally, for both cases, the spatial ensemble spread increases with forecast lead time after 20 hours. This up-scaling of forecast spatial differences should be expected from a statistical evaluation of a large number of cases, but not necessarily from individual case studies where the physical processes of the day dominate. Using dFSSmean for individual case studies allows these processes, and their effect on the spatial ensemble spread and upscale growth of forecast differences, to be examined.

The error measures (eFSSmean, Figures 4.6b and 4.6d) show a similar structure to the dispersion measures with a similar magnitude for ensemble spread and skill. There are times, such as for the Spring case at a lead time of 20 hrs (Figure 4.6b), or the flooding case at lead times 0-5 hrs (Figure 4.6d) when the ensemble is clearly under-spread. For the Spring case a timing error results from a front passing into the domain in all members earlier than seen in the radar; for the flooding case convective showers present in the radar have yet to spin up in the model. In both cases there is little evidence that the

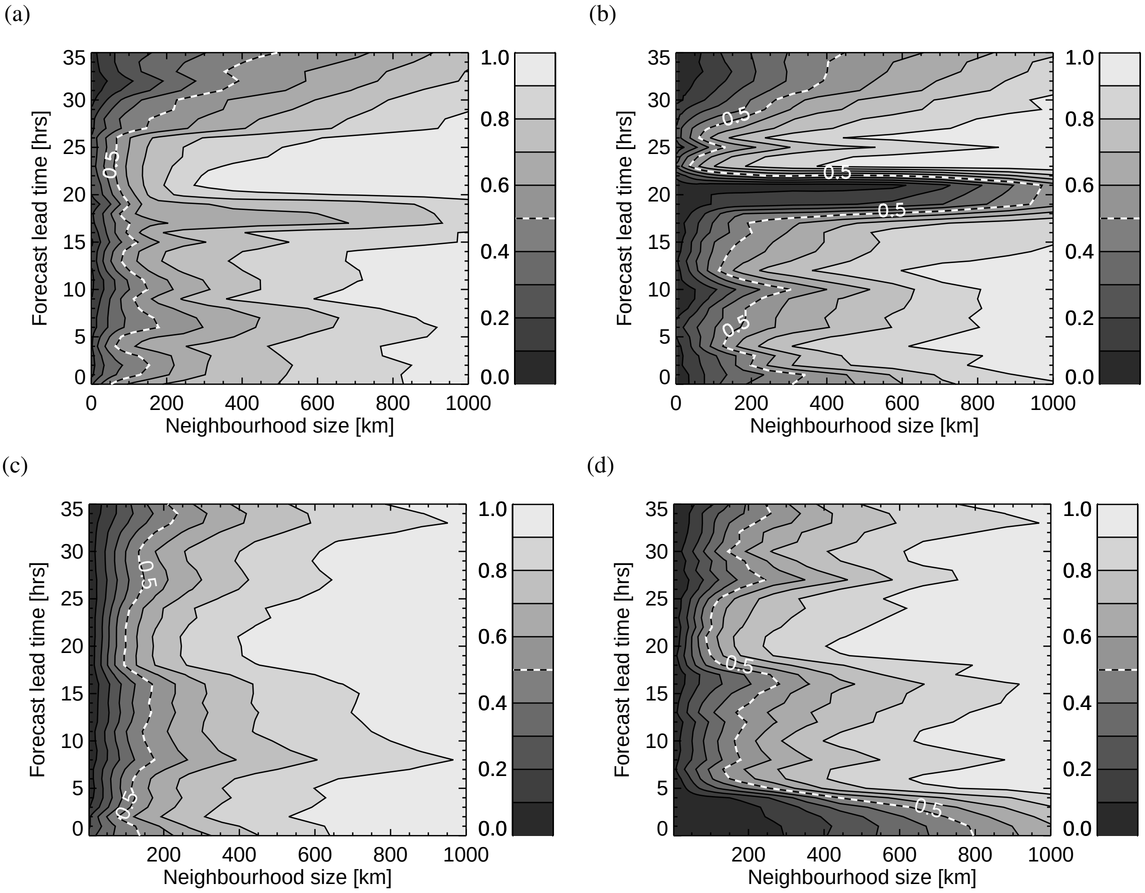


Figure 4.6: dFSSmean (a,c) and eFSSmean (b,d) for the organised Spring case (top) and the Summer flooding case (bottom). The standard physics configuration and the 99^{th} percentile threshold are considered. The white dashed line at 0.5 represents the believable scale. Results were calculated over the area of the UK domain with radar coverage.

ensemble is over-spread.

For the flooding case dFSSmean and eFSSmean have also been calculated over the 100 km limited area domain containing the flooding event. Selecting a subdomain in this manner allows us to focus on the spatial predictability of a specific event which can be very different from the UK domain averaged results. Differences between the domains can also be seen in the values corresponding to each percentile threshold as discussed in Section 4.4.3. dFSSmean and eFSSmean, calculated over the 100km domain are shown in Figures 4.7a and 4.7b respectively at forecast lead times 17 hrs - 26 hrs when convection was seen over Edinburgh. Comparison of Figure 4.7a and Figure 4.7b suggests that the ensemble spread and skill are similar and that, over this area, the ensemble is capturing the spatial variability of the precipitation well. This gives confidence in the ensemble for a spatially unpredictable flooding event. There are some differences between dFSSmean and eFSSmean, in particular that eFSSmean is more variable with time. This may be partly due to both the smaller number of comparisons in the error calculation, and also reflects differences between the model and observations in the temporal evolution of the storm. Note that, as the 99^{th} percentile threshold corresponds to different precipitation values over the UK and Edinburgh domains, we cannot do a direct comparison between Figures 4.6 and 4.7. This also suggests that we are indeed looking at different processes or phenomena within the different domains and confirms the need to use a suitable domain size to examine the spatial variability of particular features. The domain must be large enough to give representative results, but small enough to focus on the phenomena of interest. Of course, the same remarks will be true of any spatial measure.

4.5.2 dFSSstdev in addition to dFSSmean

In this section we discuss the benefits of considering dFSSstdev in addition to dFSSmean. Figure 4.8 shows dFSSmean and dFSSstdev calculated for the organised Spring case (top) and flooding case (bottom) when considering the 99^{th} percentile threshold for hourly precipitation accumulations. The FSS was calculated over the whole UK domain. dFSSstdev is shown in Figure 4.8c and Figure 4.8d and presents results consistent with those from dFSSmean. For example, the largest values of dFSSstdev occur in areas where low dFSSmean values extend to large scales. The greater spatial spread associated with low values of dFSSmean results in a wider range of possible values for FSS_i and larger dFSSstdev.

However, there is also some further information given by the standard deviation. In particular, for the flooding case (Figure 4.8d) there is an area of higher standard deviation seen in the first two hours of the forecast at neighbourhood sizes up to 500km which is associated with the spin-up of the model. This effect is even more apparent in results for the 99.9^{th} percentile threshold (not shown) and is the result of the convection-permitting model having to spin up showers during the first few hours of the forecast. Because the ensemble members spin-up showers at different locations, lower values of dFSSmean and a large range of values of FSS_i (resulting in a large dFSSstdev) are obtained. For the Spring case (Figures 4.8a and 4.8b) convective showers are not present at the forecast

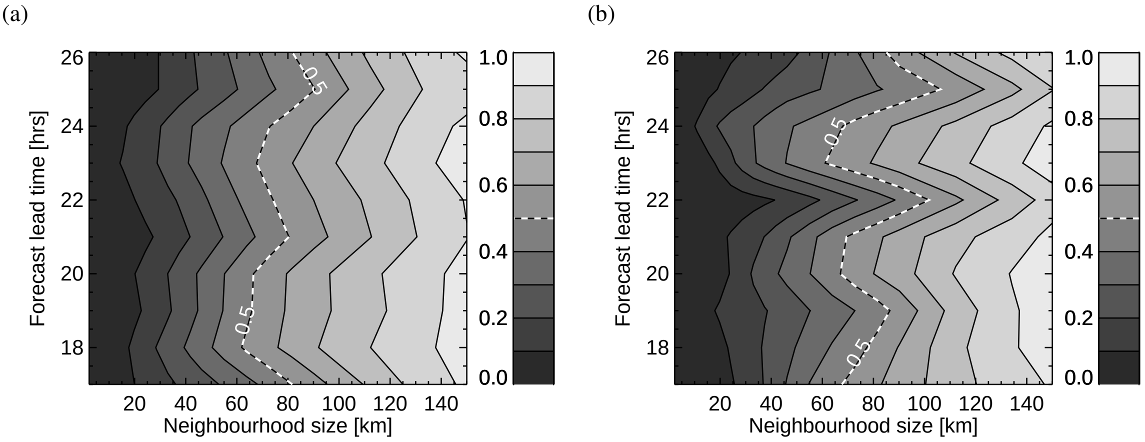


Figure 4.7: FSS calculations over the Edinburgh subdomain: (a) dFSSmean and (b) eFSSmean. The 99^{th} percentile threshold is considered. The white dashed line at 0.5 represents the believable scale.

start time and do not need to be spun-up from initial conditions. Hence spin-up effects are not seen in the precipitation diagnostics. It is useful to examine how the standard deviation behaves at different scales. The smallest values are found at both the grid scale, where differences are so large that similarly low values of the FSS are expected for all member pairs, and also at the largest scales, where all members are effectively the same.

4.5.3 Other fields and thresholds

The use of different percentile thresholds allows more information to be gained about the ensemble spread for different ranges of forecast values; for example a higher threshold will select more extreme values compared to a lower threshold which will select values that are more widespread. An example is given in Figure 4.9 for the organised Spring case where results for the top 99^{th} (LHS) and 85^{th} (RHS) percentiles are compared. This time we show a different diagnostic field, the 10 m horizontal wind speed. Like the hourly precipitation accumulations this field was selected as a suitable candidate for calculation of the FSS because of its high spatial variability. 10 m wind speeds are also used by the Met Office for routine forecast verification.

The 99^{th} percentile threshold selects only the highest wind speeds in the domain. At lead times 0-10 hrs these are found in areas to the north of the UK near the low pressure

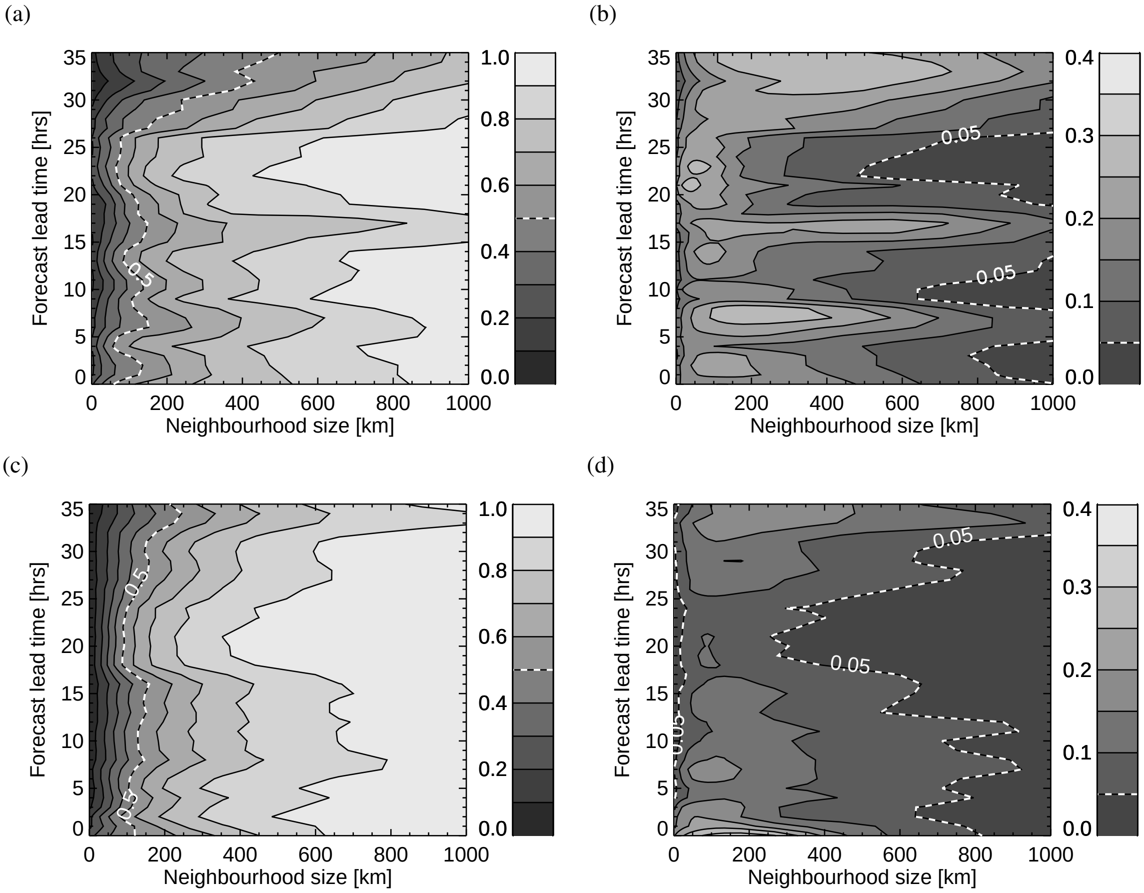


Figure 4.8: dFSSmean (a,c) and dFSSstdev (b,d) for the organised Spring case (top) and the flooding case (bottom). The white dashed line in (a,c) at 0.5 represents the believable scale. To guide the eye, in (b,d) the white dashed line at 0.05 represents the neighbourhood at which dFSSstdev is an order of magnitude smaller than the believable scale. The 99^{th} percentile threshold is considered and results are calculated over the whole UK domain.

centre. The exact placement of the highest winds varied considerably between the ensemble members, with some placing them to the northwest and others to the northeast of the UK. Hence there were large spatial differences between the members resulting in low dFSSmean values extending to large neighbourhoods at a lead time of 10 hrs as shown in Figure 4.9a. At lead times greater than 10 hours there is high spatial agreement amongst the ensemble members resulting in high values of dFSSmean. All members place the highest winds to the northwest of the UK associated with the frontal feature that enters the domain at this time.

Comparing Figure 4.9a and Figure 4.9b we see the unusual result that for a lead time of 12 hours, and after 28 hours, there is more agreement (larger FSS values) for

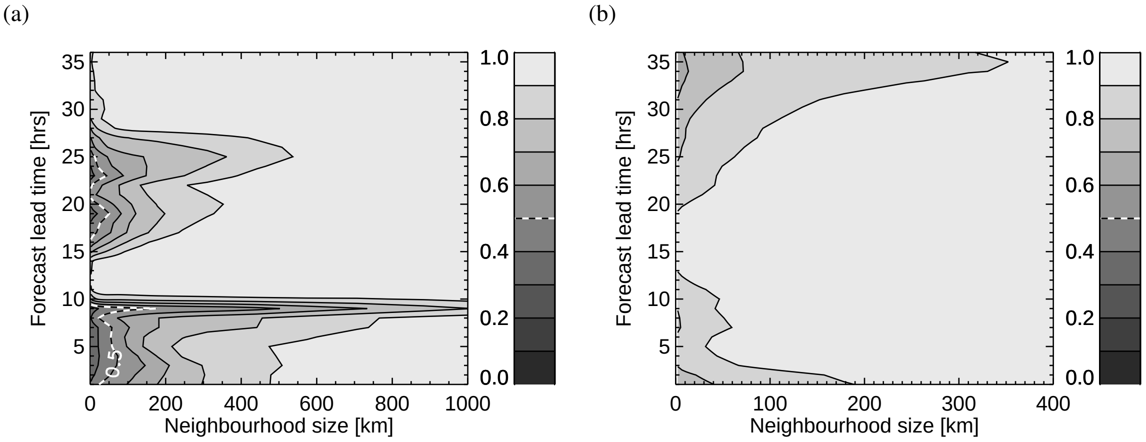


Figure 4.9: Comparison of dFSSmean calculated for the (a) 99th and (b) 85th percentile thresholds for the 10 m horizontal wind speed field in the organised Spring case. Results are calculated over the whole of the UK domain and only the standard physics configuration is considered. The white dashed line at 0.5 represents the believable scale.

the 99th than for the 85th percentile for a given neighbourhood size. This behaviour suggests that care must be taken in the interpretation of the 99th percentile threshold for the wind speed field. For the wind speed, local variability is superimposed upon a background gradient from the large scale situation. The 99th percentile is likely to include both local variability from points where the background field is moderate and also larger scale variability where the background field is high. Consequently, unlike for precipitation, we cannot cleanly examine local features in the wind speed field simply by selecting a high threshold value. It is necessary to also consider a lower threshold that includes features of the larger scale flow such as, for this case, the 85th percentile threshold. Figure 4.9b shows that, at lead times 12-20 hrs, the FSS values for the 85th percentile are particularly high. These areas of small spatial spread can be related to the synoptic situation: at a lead time of 12 hrs a highly predictable frontal feature entered the domain from the northwest and the top 15 % of wind speeds in the domain were closely associated with the flow in the vicinity of this front. Hence, there was very high spatial agreement between the members at these times. Before the front entered the domain the highest winds were associated with a less predictable decaying cold front. Moreover, after the front had progressed further into the domain greater differences between the members emerged at larger scales for the winds to the south of the occluded front.

The effect of different thresholds on the FSS for hourly precipitation accumulations can be seen by comparing Figures 4.6a and 4.6c with Figures 4.10a and 4.10b respectively. The latter show dFSSmean calculated for the 90^{th} percentile threshold. In particular, it can be seen that the large temporal variability seen in Figures 4.6a and 4.6c for the 99^{th} threshold has been replaced in the 90^{th} percentile results by a trend for ensemble spread to increase systematically with time. This trend is expected climatologically as forecast differences grow from small to larger scales with increasing forecast lead time. The rate of increase is different for the two cases. For the flooding case (Figure 4.10b) scales at which dFSSmean=0.5 increase gradually from 5 km to 100 km over 36 hours as forecast differences grow from small to larger scales. For the organised Spring case, dFSSmean values greater than 0.5 are seen even at the grid scale for lead times up to 25 hrs. After this time the scale at which dFSSmean=0.5 increases rapidly to 225km. This pattern is in agreement with the behaviour seen for the 99^{th} threshold and has the same interpretation: after 25 hrs an area of precipitation moves out of the domain but with timing differences between the members. Overall, there is better spatial agreement between the ensemble members at the 90^{th} percentile threshold than at the 99^{th}: the broader-scale features selected by the lower threshold are more predictable. When considering a range of different thresholds from the 99^{th} to 80^{th} percentile (not shown) the transition from large temporal variability to a trend of upscale growth of forecast differences with increasing lead time was found to be smooth: there is no sudden transition. It is likely that the range of thresholds over which such a transition occurs will be highly case dependent as the relative importance of local and large scale features changes. The FSS allows such behaviour to be investigated.

4.6 Results assessing different physics configurations

In this section an application of dFSSmean to the comparison of the multiphysics and MOGREPS ensembles for the organised Spring case is presented. Thus the spatial ensemble spread associated with LBC and IC perturbations is compared to that generated through different physics configurations as described in Section 4.4. The examples presented are for the 99^{th} percentile threshold of precipitation accumulation: lower thresholds showed smaller spatial differences (larger dFSSmean values) but lead to the same general conclusions. Note that the purpose is not to evaluate the merits of particular physics configurations but to show a method that can be used to examine the behaviour

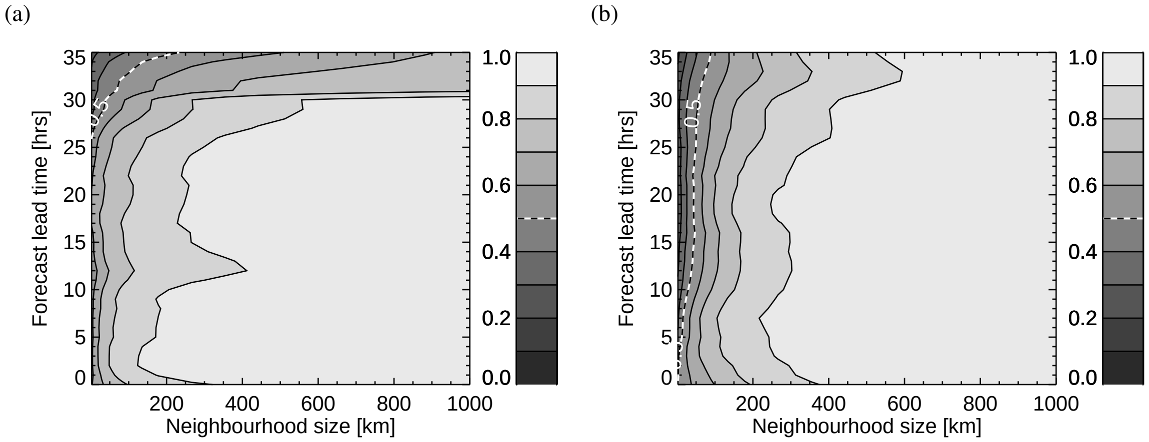


Figure 4.10: dFSSmean calculated using the 90^{th} percentile threshold of hourly precipitation accumulations for (a) the organised Spring and (b) the Summer flooding case. Results are calculated over the whole of the UK domain and only the standard physics configuration is considered. The white dashed line at 0.5 represents the believable scale.

of stochastic processes or physics changes in ensembles.

Figure 4.11b shows dFSSmean comparing the configuration with restricted convection scheme and increased time step (conv+time) to that with the modified treatment of graupel (grp) using the Physics2 comparison strategy (comparison strategy 2 in Section 4.4). This particular comparison is shown because it gives larger spatial differences than those found when comparing any other physics configuration pairs, or considering all physics configurations (the Physics5 comparison strategy). In Figure 4.11b FSS values of 0.5 are reached by a neighbourhood size of 5 km, and no spatial differences are seen for neighbourhoods greater than 100 km (where FSS$\approx$1). The lowest values of dFSSmean occur between lead times of 12 hrs and 16 hrs when the heaviest convective showers were present: it is during these events that modifications to the treatment of graupel are most noticeable.

Results from comparing only the MOGREPS members from conv+time and grp (comparison strategy MOGREPS2, 1 in Section 4.4) are shown in Figure 4.11a. These differ only in minor details from those shown in Figure 4.8a (dFSSmean calculated for the MOGREPS ensemble with the standard physics configuration). The MOGREPS2 results show that FSS values of 0.5 are reached on scales greater than 60 km, scales at which the Physics2 members are almost identical. In other words, the spatial variation introduced through different physics configurations is only seen close to the grid scale. If we consider FSS values lower than FSS $= 0.5$ to represent fields so different that the forecast is no longer useful, then the different physics configurations applied here, for this particular case, are simply moving around features that are known to be unpredictable from the MOGREPS ensemble. Of course, this is not to say that physics changes in general are unimportant for improving model performance, or that using different physics configurations is not sometimes a valuable component of an ensemble system, or that adding small scale perturbations is undesirable. For another case or for other physics perturbations the effects might be very different. Our purpose is simply to demonstrate a methodology that allows the spatial effects of different ensemble configurations to be thoroughly investigated and set into the context of other aspects of forecast uncertainty.

It is possible that, although the evaluation of Physics2 only showed forecast differences at small spatial scales, combining the different physics configurations with those from the MOGREPS2 ensemble may lead to large changes in the growth of spatial differences. To assess this, the comparison strategy MOGREPS2+Physics2 (comparison strategy 3 in Section 4.4) is employed. Again, examples are shown for the physics configurations conv+time and grp which show the largest spatial differences. The results of MOGREPS2+Physics2 are shown in Figure 4.11c. Differences between Figure 4.11c and Figure 4.11a are very small and hence, to aid interpretation, Figure 4.11d shows the difference between the MOGREPS2 and the MOGREPS2+Physics2 results. The differences are over an order of magnitude smaller than the dFSSmean values in Figures 4.11a and 4.11c. It is interesting that both positive and negative differences are seen: modifying the different physics configuration both adds and removes spatial spread. From Figure 4.11d it can also be seen that differences between MOGREPS2 and the MOGREPS2+Physics2 extend, with similar magnitude, across all spatial scales. However, in terms of the fractional difference relative to dFSSmean the differences at small neighbourhoods have more importance. At a lead time of 15 hrs the fractional difference in dFSSmean varies from 7% at 50 km to 3% at 250 km. It should be noted that these differences are still very small, especially at the larger more predictable scales (as indicated by the point where FSS$\geq$0.5 in the MOGREPS ensemble).

Analysis of the combined MOGREPS+Physics comparisons supports the conclusions drawn previously that the introduction of these differences in the physics only influences

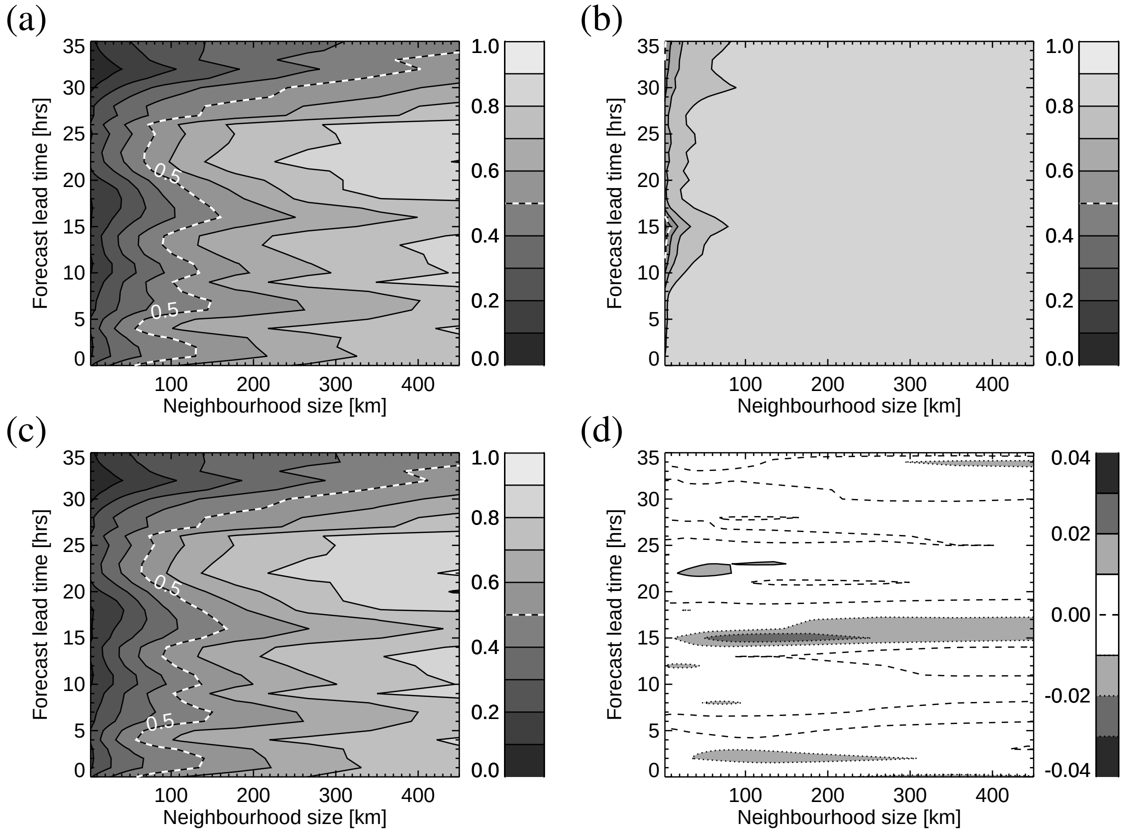


Figure 4.11: dFSSmean comparisons of the 'restricted convection with increased time step' and the 'graupel' physics configurations for the 99th percentile threshold of hourly precipitation accumulations. Results from different comparison strategies are shown: (a) MOGREPS2, (b)Physics2 and (c) MOGREPS2+Physics2. (d) shows the difference between sub-figures (c) and (a). Results are calculated over the whole of the UK domain. The white dashed line at 0.5 represents the believable scale.

scales much smaller than the predictable scales of the system (in this particular experiment). In practical terms, the variability of those scales could be addressed with spatial post processing and without the need for additional ensemble members. On the other hand, if the scales of the physics changes were to upscale to scales greater then the system's predictable scales then the performance of the ensemble might benefit from more perturbed-physics members. Systematic application of the methods shown here would provide a sound basis for making these decisions.

4.7 Chapter discussion and conclusions

In this chapter a new methodology for the detailed analysis of ensemble spread for high resolution forecasts, focusing on spatial variability, has been presented with the aim of answering the questions:

1. How can the spatial variability of ensemble member forecasts be meaningfully summarised?
2. What are suitable summary measures, and what do these tell us about ensemble performance and spatial predictability?

This method focuses on two different measures of ensemble spread: dFSSmean and dFSSstdev, the mean and standard deviation of the FSS calculated over all ensemble member-member pairs. dFSSmean gives a measure of the FSS value for the whole ensemble indicating the average spatial agreement within the ensemble over a particular size of neighbourhood i.e at a given spatial scale. dFSSstdev provides some further useful information about the range of FSS values used in the calculation of dFSSmean. A large range of FSS values, corresponding to a large value of dFSSstdev, indicates that the ensemble members are unevenly distributed.

To demonstrate the utility of these measures, and answer question 2, results were presented from two case studies. It was shown that dFSSmean and dFSSstdev allowed investigation of, for example, the temporal evolution of ensemble spread, model spin up, and saturation of forecast differences. Considering different percentile thresholds allowed information to be gained about the spatial spread of the ensemble for different physical regimes. In particular it was found that, for hourly precipitation accumulations, the dFSSmean for the 99^{th} percentile threshold had high temporal variability. This contrasted with the dFSSmean for the 90^{th} percentile threshold for which spatial differences between the ensemble members increased with time.

The realism of the ensemble spatial distribution was also tested by comparison with another metric, the mean FSS calculated over all member-radar pairs, denoted eFSSmean. This error measure can be compared with dFSSmean to investigate the spread-skill relationship of the ensemble at different times and spatial scales. For the two cases considered here these measures suggested that ensemble spread was reasonable. On occasion the ensemble was under-spread and this was linked to timing errors between the simulations and the observations and to the need for spin up of showers in a convection permitting model.

For one case study, results were presented for a comparison of spread between differently generated ensembles, including multiple physics configurations. This application illustrates a methodology for identifying the spatial scales that are influenced by modifications to physical processes. Examining the FSS for different spatial scales and over a range of times allowed a quantification of the effects of using different physics configurations compared to LBC and IC perturbations. For the case described here it was concluded that modifying the physics in the ways done here did not influence the ensemble evolution at scales where the forecast has skill. These results are not to be interpreted as general: well chosen physics modifications can and do improve forecasts as demonstrated by, for example by Stensrud et al. (2000); Keil et al. (2014). The key point is that evaluation techniques presented here allow clear statements about the impacts of physics modifications to be made since different ensemble configurations can be thoroughly investigated and the spatial impact of the changes quantified.

The work presented in this chapter provides a framework through which spatial ensemble spread can be analysed. There are some limitations to this study: in particular the consideration of two cases only and the limited consideration of physics perturbations. Another limiting factor is the methodology of calculating a single value of the FSS that is representative of a comparison across a whole domain. As discussed above this can mean that different meteorological phenomena, such as convective and frontal precipitation, are considered together, when each individually may have an inherently different predictability and ensemble spread. It is possible to select a smaller domain to consider events of interest, as highlighted with respect to Figure 4.7, although this is only useful in hindsight once the event has occurred. A spatially varying and scale dependent measure of ensemble spread that does not suffer from this drawback is presented and discussed in Chapters 5 and 8.

Despite these limitations there are some important conclusions from this chapter. In particular, we have stressed how the ensemble spread is highly dependent on the scales considered for evaluation. Consequently, to investigate the ensemble behaviour thoroughly it is necessary to consider multiple scales, and be mindful of the different expectations for skill at these scales. Forecasts should be verified, and the benefits of forecast model changes assessed, at scales that are believable.

Chapter 5

Analysis of six summer convective cases: introduction to the case studies

5.1 Introduction

This chapter introduces the six case studies used as examples in the proceeding two chapters. The choice of cases is discussed in Section 5.2. Section 5.3 then details the synoptic background to each case. The model representation of the meteorology for each case (hereafter denoted 'model meteorology') is discussed in Section 5.4 for both the Met Office Global and Regional Ensemble Prediction System UK ensemble (MOGREPS-UK) and the Met Office variable resolution UK deterministic (UKV) operational forecasts.

5.2 Selection of cases

The six cases were selected from summer 2013 during the period of the COnvective Orographic Precipitation experiment (COPE Blyth et al., 2015; Leon et al., 2015). The COPE field campaign concentrated on the SW Peninsula to investigate the processes controlling precipitation intensity in convection. Five of the cases discussed here are from COPE IOPs (intensive observing periods); the exception being the 17/07. The forecast start times were selected to be early enough for the high resolution model to have spun-up before the event of interest, but also close to the event to increase forecast accuracy. The cases were selected to represent a range of convective situations and differing levels of predictability. The first three cases (17/07, 23/07 and 27/07), and the morning of 02/08, featured deep convection with large scale forcing and will be discussed in Section 5.3.1. In the remaining three cases (29/07, 02/08 (afternoon) and 03/08), peninsula convergence, a topographical effect of the SW Peninsula which acts to align showers along the peninsula, played a role in the initiation and maintenance of convection. These peninsula convergence cases are discussed in Section 5.3.2. The dates and forecast start times for each case are summarised in Table 5.1 along with a short description of the meteorology for each case.

Date [dd/mm]	**Forecast start [dd/hh]**	**Meteorological situation**
17/07	17/03Z	Line of thunderstorms
23/07	22/21Z	Organised bands of thunderstorms
27/07	27/09Z	Mesoscale Convective System (MCS)
29/07	29/03Z	Peninsula convergence, convective showers
02/08	02/03Z	Peninsula convergence, convection moving north from France
03/08	03/03Z	Peninsula convergence

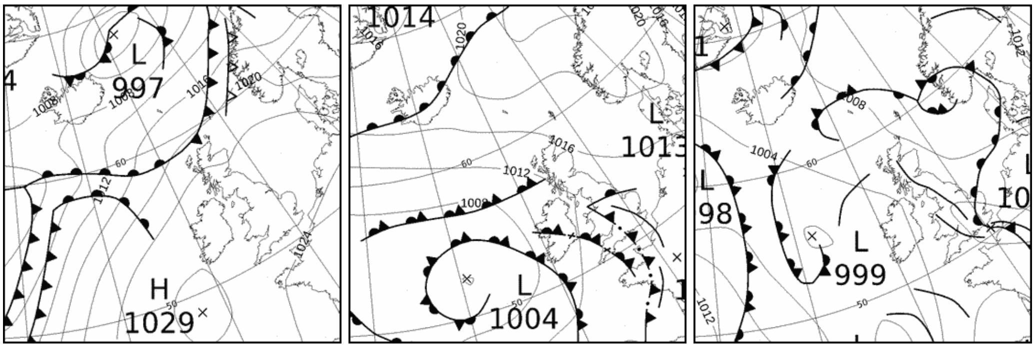
Table 5.1: Overview of the cases considered.

5.3 Synoptic overview

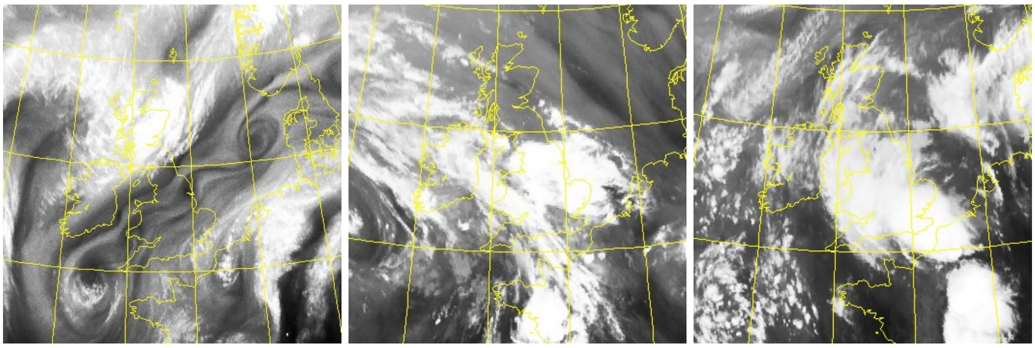
5.3.1 Deep convective cases

17/07

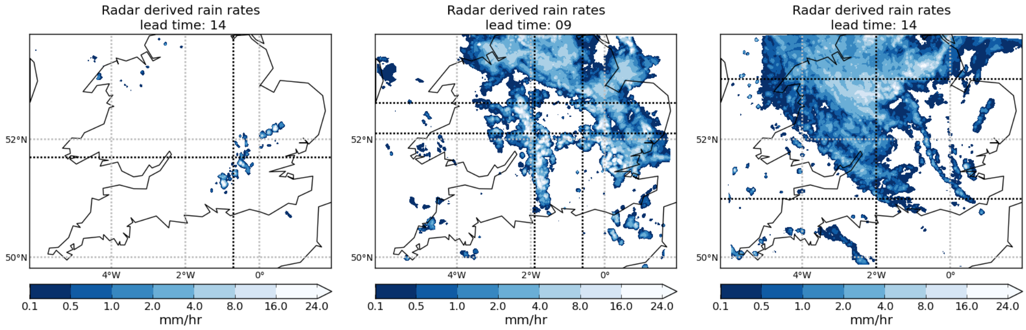
On the 17/07 localised thunderstorms developed within a generally settled anticyclonic period. Two low pressure systems (one over Scandinavia, the other over the Atlantic) are seen tracking to the north of the UK in the Met Office analysis at 12Z (Figure 5.1a). At this time, a layer of dry of upper-tropospheric air, originally associated with the Atlantic cyclone, has descended to lower levels. This dry air can be clearly seen as a dark band in the water vapour imagery at 12Z as shown in Figure 5.1d. Throughout the day this dry layer propagated towards the southeast. Note the vortex structure in the water vapour imagery associated with the high vorticity of this region. Precipitation was first seen in the Radarnet radar derived rain rates (‘radar data’) at 16Z. The precipitation organised into cells along a SW-NE orientated line as shown at 17Z in Figure 5.1g. This line persisted until 19Z after which showers became less organised and, after 20Z, little precipitation remained.



(a) Analysis 17/07 12Z (b) Analysis 23/07 06Z (c) Analysis 27/07 18Z



(d) WV imagery 17/07 12Z (e) WV imagery 23/07 06Z (f) WV imagery 27/07 18Z



(g) Radar 17/07 17Z (h) Radar 23/07 06Z (i) Radar 27/07 23Z

Figure 5.1: Met Office surface analysis (top, courtesy of the Met Office), Meteosat water vapour imagery (NERC Satellite Receiving Station, 2015) (centre; (d) channel 5, (e) and (f) channel 6) and Radarnet radar derived rain rates (bottom). Three cases are shown: the 17/07 (left), 23/07 (centre) and 27/07 (right). The cross-hairs on the radar imagery show the horizontal locations of vertical cross sections used when analysing the vertical structure (Section 5.4; Chapters 6 and 7).

23/07

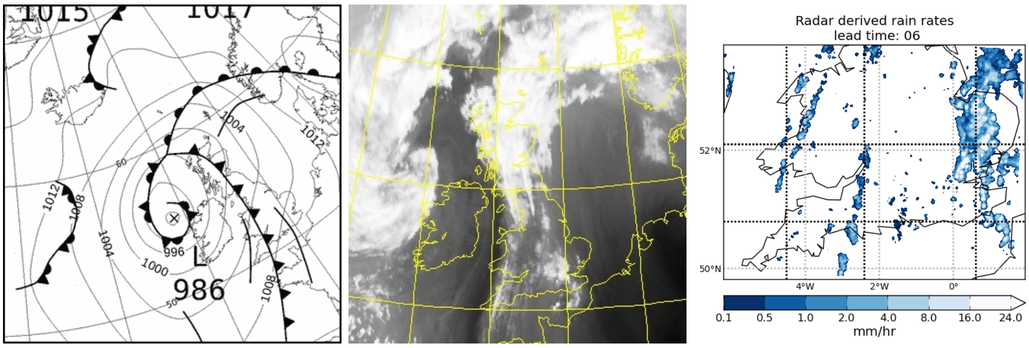
The synoptic situation on the 23/07 is dominated by a mature cyclone to the west of the UK, and high pressure over Scandinavia. At 00Z two troughs lie over the UK; it is in these locations that convection later develops. By 06Z (Figure 5.1b) frontogenesis occurs along the western trough and the analysis shows a cyclone over northern England which moves north throughout the day. Convection associated with the cyclone reached upper tropospheric levels as shown by the Meteosat water vapour imagery (Figure 5.1e). Two bands of heavy precipitation are also clearly visible in the radar imagery at this time as shown in Figure 5.1h. The easterly precipitation band broadens out throughout the day but the westerly band remains narrow.

27/07

On the 27/07 the UK sat in an area of low pressure associated with mature interconnected cyclones to the southwest, northwest and northeast. A frontal feature from the northeast cyclone extended south across the North Sea and down into France. A frontal wave formed along this feature, initially over northern France, and developed into an MCS. From 06Z-12Z the wave moved north over the English channel and a cyclonic circulation developed. By 18Z (Figure 5.1c,f) the MCS reached southern England and continued northward. Precipitation associated with the MCS is clearly seen in the Radarnet derived rain rates shown in Figure 5.1i at 23Z. The MCS and associated precipitation moved north-northeast throughout the night. Initially the convective region was seen along the leading edge of the MCS, followed by a broad area of stratiform rain, agreeing with the conceptual model of a convective line with trailing stratiform cloud (Houze, 2004).

02/08 (morning)

On the 02/08 a mature cyclone was positioned over the Atlantic to the north of Ireland. A cold front associated with this system extended over the UK (north-south) as shown in the Met Office surface analysis at 06Z (Figure 5.2a). At this time cloud was seen ahead of the cold front (Figure 5.3b). Also of interest is the area of deep convective cloud situated over London which resulted in heavy precipitation and many lightning strikes. This area of convection initiated over France before propagating north to the UK and is clearly seen in the radar data at 09Z (Figure 5.3c).



(a) Analysis 06Z (b) WV imagery 06Z (c) Radar 09Z

Figure 5.2: Observations for the 02/08 case. (a) Met Office surface analysis at 06Z (courtesy of the Met Office), (b) Meteosat channel 6 water vapour imagery at 06Z (NERC Satellite Receiving Station, 2015) and (c) Radarnet radar derived rain rates at 09Z. The cross-hairs on the radar imagery show the horizontal locations of vertical cross sections used when analysing vertical structure (Section 5.4; Chapters 6 and 7).

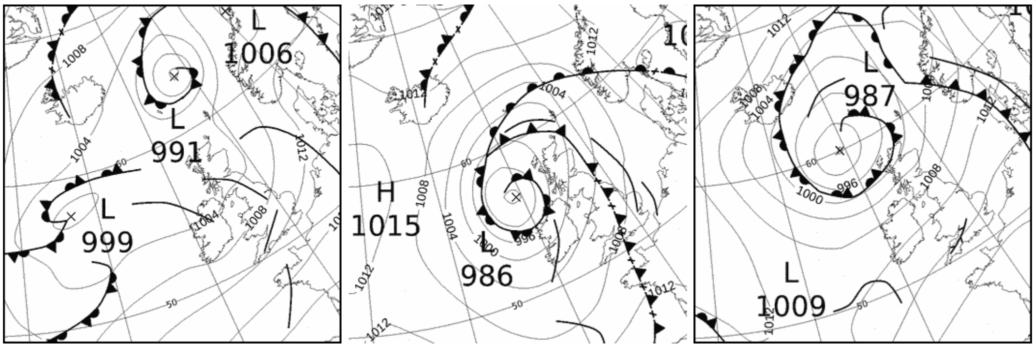
5.3.2 Peninsula convergence cases

29/07

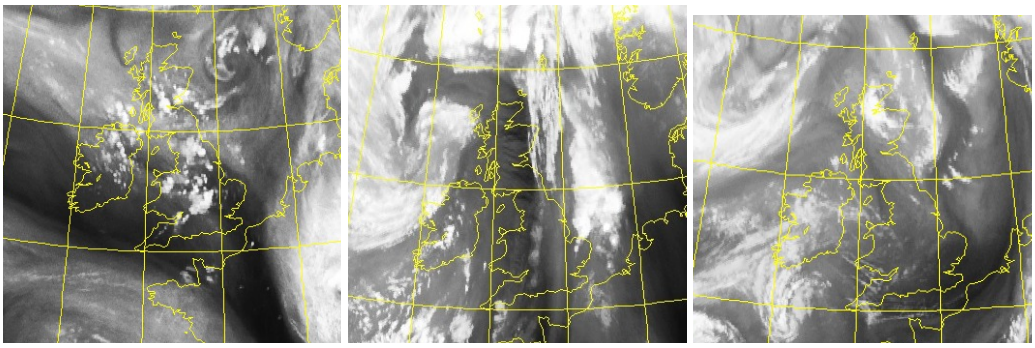
On the 29/07 a mature low pressure system was located to the north of the UK with high pressure over France and a southwesterly flow. Several troughs lay over the UK as shown in the Met Office analysis at 12Z (Figure 5.3a). From 03Z – 08Z some precipitation was seen to the west of England, over the Welsh hills and the SW peninsula. From 08Z onwards scattered convection also initiated over the eastern UK and organised lines formed over the SW and Welsh peninsulas. These lines of precipitation were seen throughout the day (e.g. at 15Z, Figure 5.3g). From 12Z convection reached upper tropospheric levels and was seen in the Meteosat water vapour imagery (Figure 5.3d). Over the SW peninsula the convection remained shallow.

02/08 (afternoon)

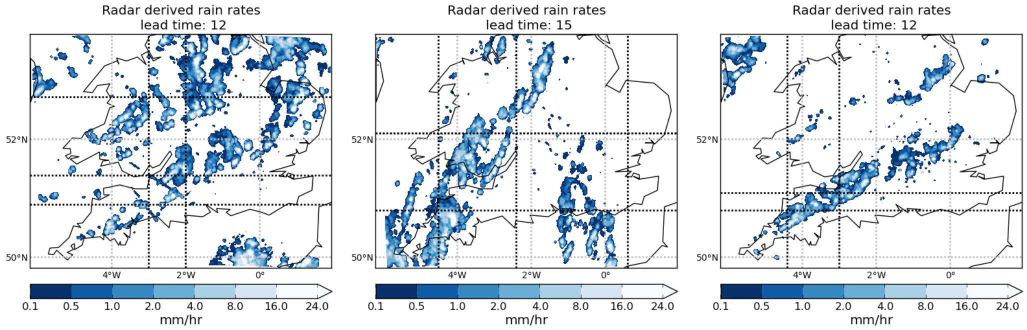
Following the deep convection seen on the morning of 02/08 (Figure 5.2), a front moved eastward across the UK during the day as shown in the 12Z analysis (Figure 5.3b). This front is still present in the 18Z and 21Z analysis. Meteosat water vapour imagery reflects the cyclone to the north of the UK, clearly showing a hook of cloud around the cyclone centre and the dry intrusion (Figure 5.3e), and also shows cloud along the frontal feature. This convective cloud was associated with a line of heavy precipitation, initiating around



(a) Analysis 29/07 12Z (b) Analysis 02/08 12Z (c) Analysis 03/08 12Z



(d) WV imagery 29/07 12Z (e) WV imagery 02/08 12Z (f) WV imagery 03/08 12Z



(g) Radar 29/07 15Z (h) Radar 02/08 18Z (i) Radar 03/08 15Z

Figure 5.3: Met Office surface analysis (top, courtesy of the Met Office), Meteosat water vapour imagery (NERC Satellite Receiving Station, 2015) (centre; (d) channel 5, (e) and (f) channel 6) and Radarnet radar derived rain rates (bottom). Three cases are shown: the 29/07 (left), 02/08 (centre) and 03/08 (right). The cross-hairs on the radar imagery show the horizontal locations of vertical cross sections used when analysing the vertical structure (Section 5.4; Chapters 6 and 7).

10Z, and continuing throughout the day. The band of precipitation was formed of several lines of cells (e.g as shown in the radar data at 18Z, Figure 5.3h), giving a wider area of precipitation than seen for the other peninsula convergence cases.

03/08

On the 03/08 the southwest of England was situated in a southwesterly flow associated with a mature cyclone positioned off the northwest coast of Scotland as shown in the Met Office analysis at 12Z (Figure 5.3c). As highlighted in the analysis, a convergence line formed along the SW peninsula. Cloud associated with the convergence line can just be seen in the Meteosat water vapour imagery at this time (Figure 5.3f). Data from the National Centre for Atmospheric Science (NCAS) mobile radar, situated at Davidstow Cornwall for the COPE field campaign, showed that convective clouds reached heights of 5 km. Precipitation associated with the convergence line was seen from 09Z to 18Z and extended as far as Hertfordshire (52°N, 0.5°E) in a tight organised line from southwest to northeast. Example radar rain rates at 15Z are shown in Figure 5.3i.

5.4 Model meteorology

5.4.1 Deep convective cases

Overall, at the lead times considered here, the MOGREPS-UK members captured the main characteristics of the convective events. On the 17/07 (left column in Figure 5.1) all ensemble members captured some deep convection in the correct region with slightly different line orientations and spatial displacements. Example ensemble members at 17Z on 17/07 are given in Figure 5.4a. This is the case study with the highest spatial uncertainty (to be discussed in Chapter 6); there is also a timing error with the forecast precipitation initiating early. On the 23/07 and 27/07 (Figure 5.4b,c) the members again capture precipitation close to the region in which it is observed. However, some of the small scale structure is missed.

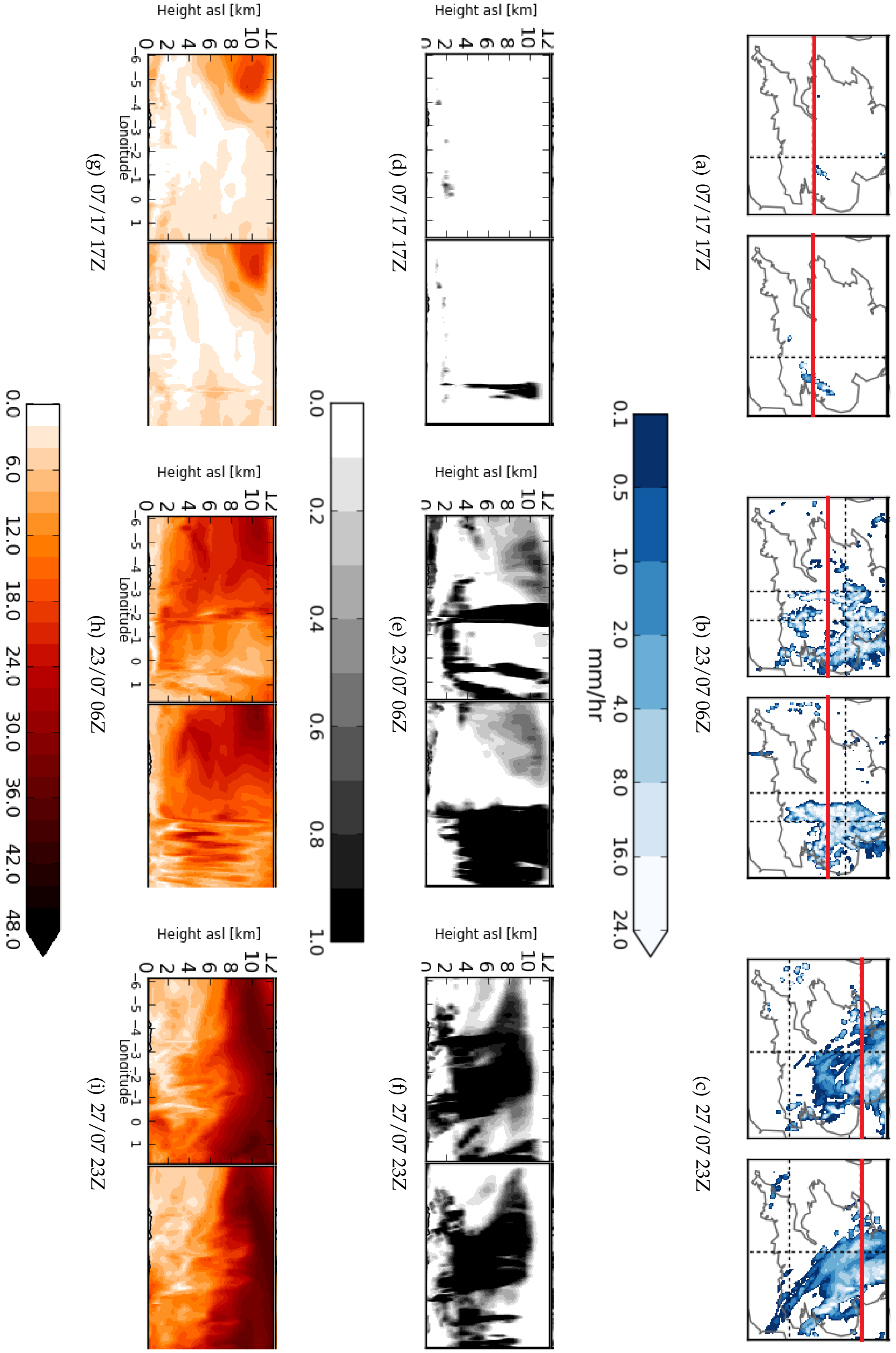


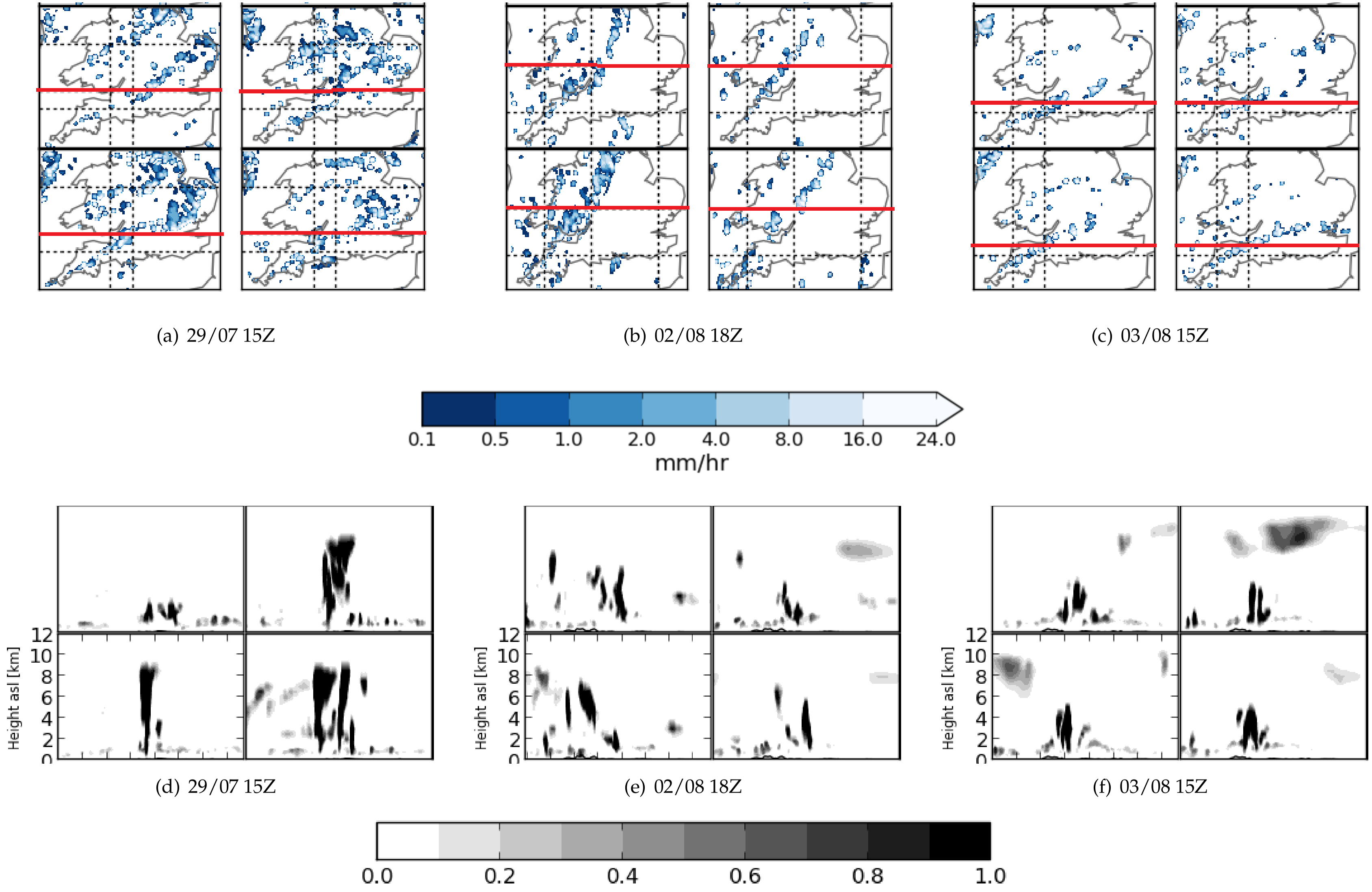
Figure 5.4: Example member surface rain rates (a-c), and cross sections of cloud fraction (d-f), and horizontal wind speed (g-i), for the first three deep convective cases (17/07, 23/07 and 27/07). The horizontal locations of the vertical cross sections for each case are shown as red lines on the corresponding surface rain rate maps. The same cross sections are shown for cloud fraction and horizontal wind speed.

For all cases the model does not produce enough light rain and the heavy rain is too heavy. This is a known problem with the model as found by e.g Hanley et al. (2015); McBeath et al. (2014). For all three cases the model cloud reached the tropopause as shown by vertical cloud fraction cross sections (Figure 5.4d-f). On the 17/07 (Figure 5.4d) the convective cells are narrow and little organisation is seen. For the other cases the cloud forms broad organised structures with large horizontal extent (Figure 5.4e,f). The cloud fields are more horizontally continuous than the precipitation fields and "fill in the gaps" between areas of heavy precipitation. This suggests that the lack of light rain may be due to microphysical processes in the model as opposed to deficiencies in the modelled dynamics failing to produce cloud. For the deep convective cases the three dimensional horizontal wind speed fields show clear signals associated with the convection. In these instances the convection is associated with, and locally modifies, the horizontal wind speed fields. Example vertical cross sections of horizontal wind speed are given in Figure 5.4g-i where convective updraughts and downdraughts can be inferred from the vertical streaks of higher and lower than average horizontal wind speeds to the east (right) of the cross sections.

5.4.2 Peninsula convergence cases

Similar conclusions can be drawn about the model performance for the peninsula convergence cases. In particular, although precipitation is captured by the model for all three cases, the cells are more separated than they should be, too heavy, and there is too little light rain. Examples of individual member rain rate fields are given for the 29/07, 02/08 and 03/08 in Figure 5.5a-c. These can be compared with the radar rain rates (Figure 5.3g-i). Overall, the members capture the position and orientation of the precipitation lines. However, the model tends to prefer a single continuous unbroken line of cells, whereas in reality the lines are more broken up. This is seen, for example, on the 02/08 at 18Z and 03/08 at 15Z by comparing Figure 5.5b,c (ensemble data) with Figure 5.3h,i (radar data). On the 29/07, in the majority of ensemble members, narrow convective plumes extend up to the tropopause as shown by vertical cloud fraction cross sections (Figure 5.5a). This contrasts with the 02/08 and 03/08 when convection was capped at 6 km and 4 km respectively (Figure 5.5e,f).

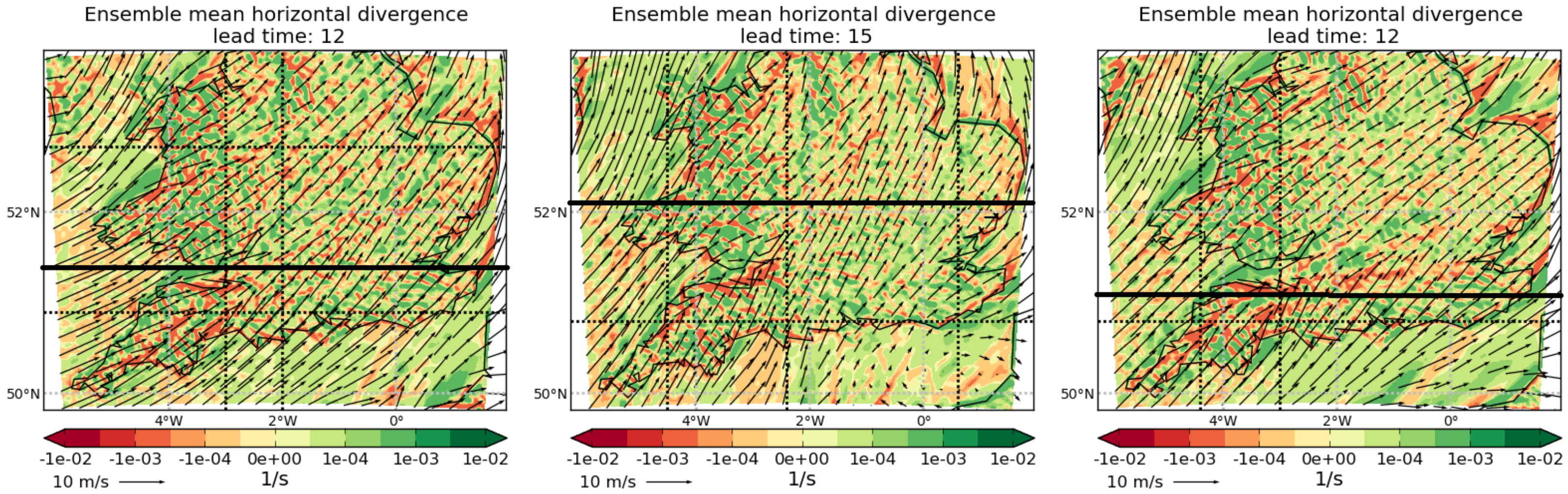
As expected, the location of precipitation on 29/07, 02/08 and 03/08 is related to the position of convergence lines along the SW peninsula. Convergence lines are seen before the convection (playing a role in initiation), and also during the convection. Horizontal maps and vertical cross sections of ensemble mean horizontal divergence for these cases are given in Figure 5.6, at the same times as the precipitation maps in Figure 5.5. For all three cases convergence is seen along the north coast of the peninsula, and on the 29/07 and 03/08, there is also a hint of convergence from the south coast of the peninsula. The vertical cross sections of horizontal divergence clearly show the location of the convergence lines and depth reached by the convection. A feature of the 03/08 case is the vertically stacked sloping layers of horizontal convergence and divergence that extend from the surface up to 12 km, which are seen throughout the day. The layers are clear in the cross sections (Figure 5.6f) starting at ground level and extending upwards with a westerly tilt. It is thought that these layers are caused by orographic gravity waves. A schematic showing how gravity waves may lead to a layered structure is shown in Figure 5.7. From this we predict that, for gravity wave structures, the layers of horizontal divergence/convergence are associated with ascent/descent. This was confirmed by inspection of the UKV fields (not shown).



(a) 29/07 15Z (b) 02/08 18Z (c) 03/08 15Z

(d) 29/07 15Z (e) 02/08 18Z (f) 03/08 15Z

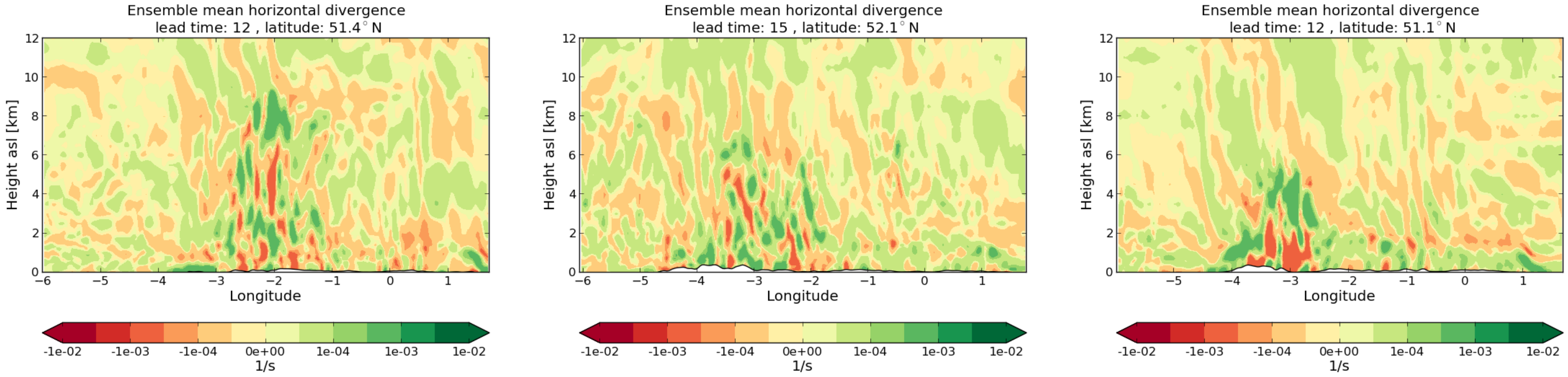
Figure 5.5: Example member surface rain rates (top) and cloud fraction cross sections (bottom) for the three peninsula convergence cases (29/07, 02/08 and 03/08). The horizontal locations of the vertical cross sections for each case are shown in red on the corresponding surface rain rate map.



(a) 29/07 15Z

(b) 02/08 18Z

(c) 03/08 15Z



(d) 29/07 15Z

(e) 02/08 18Z

(f) 03/08 15Z

Figure 5.6: Ensemble mean horizontal divergence at 10 m (a-c) and for a vertical cross section (d-f). Results are shown for the 29/07 (a,d), 02/08 (b,e) and 03/08 (c,f). The locations of the cross sections are highlighted in black on the corresponding horizontal map.

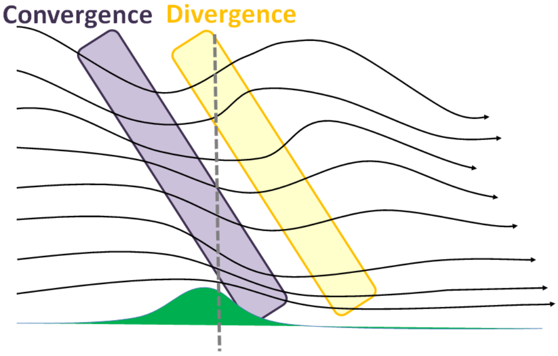
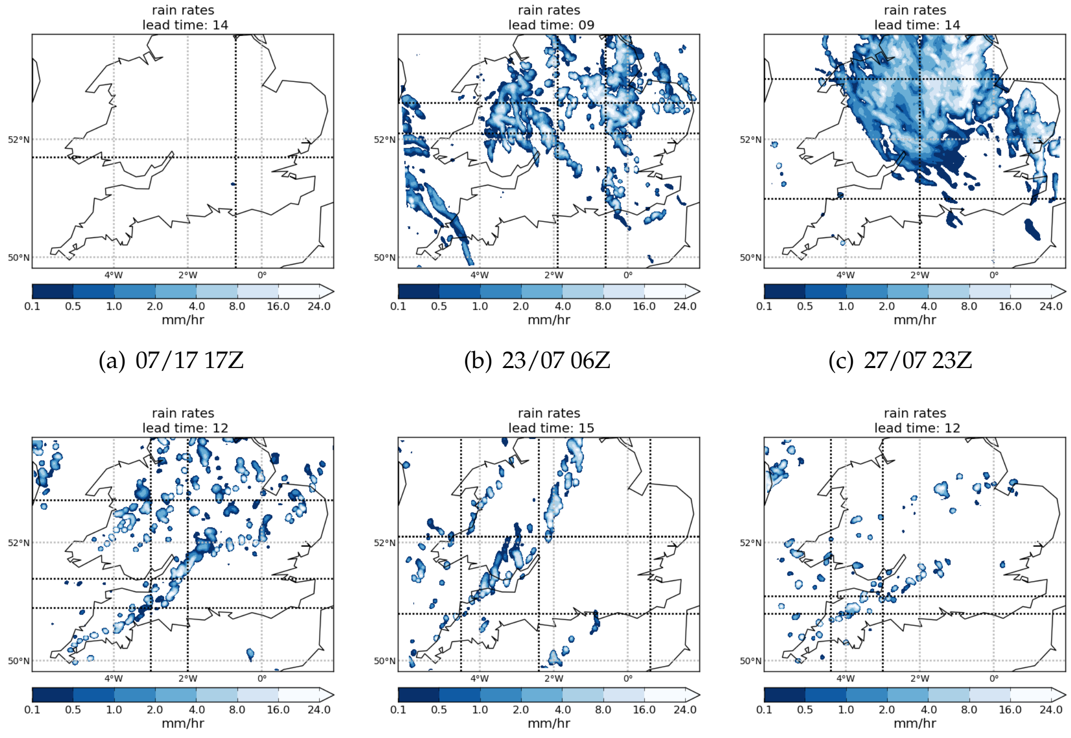


Figure 5.7: Schematic showing a simplified view of gravity waves over orography. Flow lines are shown in black and areas of horizontal convergence (purple grey) and divergence (yellow) are highlighted. Note that, to highlight the changes in the flow lines, vertical differences have been exaggerated in this schematic.

5.5 Comparison with the UKV

Overall, the UKV forecasts look similar to those from the ensemble. This is expected as both models take their boundaries from the global MetUM forecast and, apart from a decrease in resolution for the ensemble, are run with the same physics. There are however differences in the data assimilation of the two systems: MOGREPS-UK takes its initial conditions by direct downscaling from the global ensemble, whereas the UKV is initialised from an UK analysis. UKV instantaneous rain rate plots are given in Figure 5.8 for comparison with the ensemble rain rates (Figures 5.4a-c and 5.5a-c). There are some cases where the differences between the UKV and MOGREPS-UK influence the forecasts. For example, unlike all MOGREPS-UK forecasts, the UKV 03Z run on the 17/07 did not capture convective precipitation. Later UKV runs did slightly better, capturing several intense showers, but still missing the convective organisation. An example of instantaneous rain rates at 17Z for the UKV 17/07 03Z run is given in Figure 5.8a. Similar conclusions can be drawn from the hourly precipitation accumulations. It is believed that, for the case considered here, convection was strongly related to the large scale structure and errors introduced in the UKV analysis (in this instance) resulted in the UKV failing to capture the convection.

An example where the UKV performs better than the ensemble is found for the 23/07. Early in the forecast the UKV does not show any linear organisation of the precipitation, contrasting with the ensemble members and observations. However, the location of the precipitation is more accurate in the UKV than in any ensemble member, suggesting that the data assimilation is helping to correct the location of the precipitation but in doing so destroys the linear organisation. At longer lead times (e.g. at 06Z, Figure 5.8b) the UKV precipitation does form two organised lines whose structure is maintained for longer than that of any ensemble member. Overall, the UKV precipitation has the same problems as the ensemble with too many heavy cells and not enough light rain.



(a) 07/17 17Z (b) 23/07 06Z (c) 27/07 23Z

(d) 29/07 15Z (e) 02/08 18Z (f) 03/08 15Z

Figure 5.8: Example UKV surface rain rates for the six cases considered. The times shown are the same as those in Figure 5.4 (MOGREPS-UK member rain rates) and Figures 5.1, 5.2 and 5.3 (radar rain rates). The cross-hairs on the radar imagery show the horizontal locations of vertical cross sections used when analysing vertical structure (Chapters 6, 7).

5.6 Chapter conclusions

This chapter introduced the six case studies that form the basis of Chapters 6 and 7. The cases were selected to represent a range of different convective forcings and levels of predictability (to be discussed further in Chapter 6). Three cases (17/07, 23/07 and 27/07) had deep convection associated with large scale forcing. Two of these, 23/07 and 27/07, had large areas of organised convection. In contrast, the 17/07 was characterised by a number of individual cells positioned in a line. In the remaining three cases (29/07, 02/08 (afternoon) and 03/08) peninsula convergence played a role in the initiation of convection. Although all three peninsula convergence cases showed a line of precipitation along the SW peninsula, differences were seen in the depth of convection, and line orientation. For example, on the 29/07 scattered convection was seen away from the main line, whereas on the 03/08 convection was strongly tied to the convergence line. In addition to the line of precipitation along the peninsula, the 02/08 also featured less predictable deep convection moving north from France, which provided a scenario for predictability differences across the domain to be investigated (see Chapter 6).

In addition to the meteorology of each case, this chapter also discussed the behaviour of the MOGREPS-UK and the UKV models. For all cases the ensemble did a reasonable job at predicting convection, with the exact structure and location varying between ensemble members. The UKV forecasts were, in general, found to be of similar quality to those from the ensemble. However there were instances when the UKV precipitation forecasts had structures that were less (e.g 17/07), and more (e.g 23/07), like the observations. Overall, neither model configuration produced enough light rain and in both models the heavy rain was too heavy.

Chapter 6

A location-dependent method for spatial ensemble evaluation

6.1 Introduction

Chapter 4 introduced methods for summarising the spatial ensemble characteristics using the FSS to provide a compact summary of the ensemble performance over the whole domain. However, as discussed in Chapter 4, by using only one value to represent the domain, different meteorological phenomena, such as convective and frontal precipitation, are considered together, when individually each may have an inherently different predictability and ensemble spread. Hence, it is informative to examine the ensemble spatial spread and skill in a manner that preserves location-dependent information.

Using similar principles to the FSS, this chapter introduces a new, location–dependent measure of the scales over which precipitation fields, either forecasts or observations, are deemed to be acceptably similar (to be defined in Section 6.3.2). When calculated using ensemble members from a well spread ensemble system, these agreement scales, denoted as $S_{ij}^{A(\overline{mm})}$, give a measure of the smallest believable scales for the ensemble at a given point $x = i, y = j$ in the domain. When ensemble members are compared with radar observations, the agreement scales, denoted $S_{ij}^{A(\overline{mo})}$, give a measure of the smallest skillful scale for the ensemble at a given point $x = i, y = j$ in the domain. Using the location-dependent agreement scales, this chapter addresses the following questions:

1. Over what spatial scales should the ensemble be interpreted and evaluated? (Thesis question 2; Section 1.2)

2. How can these spatial scales be defined? (Thesis question 2; Section 1.2)

3. What information does this spatial approach provide for forecasting?(Thesis question 3; Section 1.2)

In Section 6.2 the neighbourhood approach is introduced, and spatial predictability is defined. With the aim of answering Question 1, the methods used to calculate the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are presented in Section 6.3, and compared to and contrasted with the FSS. In Section 6.4 an idealised ensemble is introduced and used to investigate properties of the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$. This is important for the correct interpretation of these metrics, and is necessary to obtain meaningful answers to Questions 2 and 3. Sections 6.5 and 6.6 calculate the agreement scales for the six convective cases introduced in Chapter 5. Using these examples, Questions 2 and 3 are addressed. The $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ for instantaneous rain rates are investigated first (Section 6.5) followed by an application of $S_{ij}^{A(\overline{mm})}$ to 3D variables (Section 6.6). The aim is to understand how the agreement scales behave in an operational ensemble system, what information they can provide about the ensemble spatial spread and error, and how this relates to physical processes. The overall conclusions are presented and discussed in Section 6.7.

6.2 Spatial predictability and the neighbourhood approach

Here, and for the remainder of this thesis, the term "spatial predictability" refers to differences in the location of precipitation in the ensemble member forecasts. Cases where the member forecasts are in close agreement about the location of precipitation are termed spatially predictable, and cases where the location of precipitation is uncertain (i.e. when ensemble members produce rain at different places in the domain) are termed less spatially predictable. Examples with different spatial predictability are shown schematically in Figure 6.1. Note that this definition of spatial predictability only refers to the positional differences between the ensemble member forecasts (i.e. amplitude errors are not included).

A neighbourhood based approach is used to quantify differences between precipitation forecasts. In the neighbourhood approach, forecasts are compared over differently sized areas (neighbourhoods). Summary measures are then used to compare the forecasts over these areas. For example, the amount of precipitation exceeding a specified threshold, the maximum or average precipitation value of all raining points, or the aver-

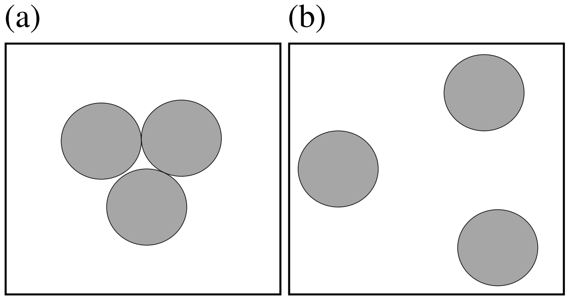


Figure 6.1: Schematic representing precipitation forecasts from three different ensemble members (grey circles, one per member). Each grey circle represents an area of forecast precipitation, say with a uniform rain rate of 0.1mm hr^{-1}. Events are shown with different levels of spatial predictability: (a) a spatially predictable event and (b) a less spatially predictable event.

age precipitation over all points in the neighbourhood, could be considered.

In this chapter, the average is taken of all points in the neighbourhood, including points with zero precipitation. Hence no distinction is made between amplitude, timing and spatial structural differences. This method was chosen to be as generally applicable as possible, giving an overview of the forecast differences, and keeping the number of parameters to a minimum. The aim is to provide a single location-dependent summary measure of the spatial forecast differences. Of course, this comes at the cost of providing less detailed information about individual components such as timing errors, although some timing errors due to advection (rather than initiation or decay) are naturally included in the spatial approach.

It is informative to relate the neighbourhood approach used in this chapter (which calculates the spatial agreement between fields) to the spatial predictability as defined above. First consider the comparison of two binary fields, for example created by setting precipitation values to zero/one dependent on whether they are below/above a predetermined threshold. In this case, any differences in the neighbourhood averaged values of the two fields will relate to differences in the location of precipitation. Hence, in this situation, the forecasts will agree over a smaller/larger neighbourhood for cases with higher/lower spatial predictability. Thus, when binary fields are considered, the spatial predictability relates directly to the neighbourhood size. Next consider precipitation fields where no threshold has been applied. The spatial predictability will still influence the neighbourhood size over which the forecasts agree, but any difference in the magnitude of the two fields will also contribute. This is also true for other fields which, like precipitation, have high small scale variability and a background value of zero. Fields which vary smoothly over large scales (larger than the neighbourhood sizes being considered), with large scale gradients, require a different interpretation. In these instances there is no longer a direct link between neighbourhood size and spatial predictability. Hence, for smoothly varying large scale fields the neighbourhood approach compares only the bias between the two fields.

6.3 Calculation of location-dependent agreement scales between two fields

6.3.1 Overview of method

First we focus on calculating location-dependent agreement scales for two different fields, say two ensemble member precipitation forecasts, denoted f_1 and f_2 . At each grid point in the domain, we search for the minimum neighbourhood size (hereafter the scale) over which suitable agreement between f_1 and f_2 is obtained. Here, and for the remainder of this thesis, the scale is defined as the number of grid points from the centre to edge of the neighbourhood (excluding the central grid point). For example a 3 by 3 neighbourhood would have a scale of 1. The scale at which suitable agreement is obtained between the forecasts f_1 and f_2 at this point x,y=i, j will be referred to as the agreement scale $S_{ij}^{A(f_1 f_2)}$. The calculation of $S_{ij}^{A(f_1 f_2)}$ proceeds as follows:

1. One grid point in the domain is selected. Call this point P at i, j.
2. The precipitation values from the two forecasts are compared at point P, and their similarity assessed using the methods presented in Section 6.3.2.
3. If the forecasts are found to be suitably similar (to be defined in Section 6.3.2), then the agreement scale at point P, $S_{ij}^{A(f_1 f_2)}$, is the grid scale. If the fields are not suitably similar, then a square neighbourhood of scale 1 (3 by 3 grid points), centred upon the point P, is considered.
4. The spatial average precipitation amount over this neighbourhood is calculated separately for f_1 and f_2, as discussed in Section 6.2. Forecasts f_1 and f_2 are again compared, this time using the average precipitation amount over the neighbourhood, and their similarity assessed.

5. If, this time, the forecasts are found to be suitably similar, then a neighbourhood of size 1 is the agreement scale. If the fields are not suitably similar, then the scale is increased by 1 (i.e. to give a 5 by 5 grid point neighbourhood).
6. Steps 4 and 5 are repeated, for incrementally larger scales, until a scale has been found for which the forecasts are suitably similar around point P. Note that this is defined as the minimum agreement scale for comparing these forecasts: generally, it would be expected that the forecasts would also be in agreement over larger neighbourhoods.
7. Steps 1 to 6 are repeated for each grid point in the domain.

This process is shown schematically in Figure 6.2.

6.3.2 Criterion for assessing forecast similarity

It remains to define how the forecast similarity is assessed and how "suitably similar" is defined. Consider the comparison of two fields f_1 and f_2 for a given neighbourhood size (scale) S, and at grid point i, j. For both fields, the average over all points in the neighbourhood is taken: we denote these averages as f^S_{1ij} and f^S_{2ij}. The fields (assuming at least one average is non zero) are compared by taking the ratio of the squared difference between these averages and the sum of their squares:

$$D^S_{ij} = \begin{cases} \frac{(f^S_{1ij} - f^S_{2ij})^2}{(f^S_{1ij})^2 + (f^S_{2ij})^2} & \text{if } f^S_{1ij} > 0 \text{ or } f^S_{2ij} > 0. \\ 1 & \text{if } f^S_{1ij} = 0 \text{ and } f^S_{2ij} = 0. \end{cases} \tag{6.1}$$

The numerator is a direct measure of the difference between the fields; the denominator a normalising factor selected such that comparison between a forecast which captures

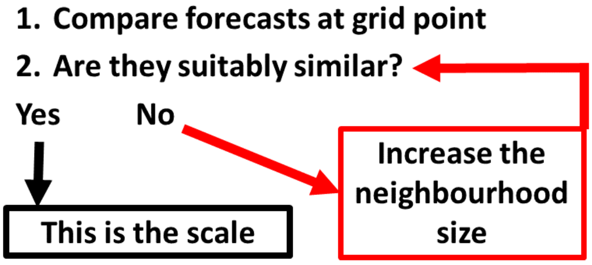


Figure 6.2: Schematic showing the method for assessing the similarity of two fields.

some precipitation ($f^S_{1ij} > 0$) and one with no precipitation ($f^S_{2ij} = 0$), gives a D^S_{ij} value of one. This is a convenient choice of normalisation: other normalisation factors are possible and would not change the overall method and conclusions presented here. Note that Equation 6.1 only holds for positive fields with the zero value acting like a threshold. Note also that Equation 6.1 is homogeneous: D^S_{ij} is unchanged by multiplying both f^S_{1ij} and f^S_{2ij} by a constant value.

The fields are then deemed sufficiently similar (i.e. to be in agreement) at scale S if

$$D^S_{ij} \leq D^S_{\mathrm{crit},ij}, \tag{6.2}$$

where

$$D^S_{\mathrm{crit},ij} = \alpha + (1 - \alpha)\frac{S}{S_{\mathrm{lim}}}. \tag{6.3}$$

The agreement scale between forecasts f_1 and f_2 at point i, j is denoted $S^{A(f_1 f_2)}_{ij}$, and is defined as the minimum scale S at which Equation 6.2 is met. The minimum possible $S^{A(f_1 f_2)}_{ij}$ is zero (showing agreement between the forecasts at the grid scale) and the maximum possible $S^{A(f_1 f_2)}_{ij}$ is S_{lim} (showing no agreement between the forecasts, or no rain in the neighbourhood for at least one forecast).

At the grid scale ($S = 0$) the second term on the right-hand side of Equation 6.3 is zero and the constant α controls the acceptable fractional difference between f^S_{1ij} and f^S_{2ij}. Different values of α can be selected: $0 < \alpha \leq 1$ where $\alpha = 0$ corresponds to no bias being tolerated at the grid scale and for $\alpha = 1$ any bias is tolerated. S_{lim} is a predetermined, fixed maximum scale and, by construction, Equation 6.2 is always satisfied at the scale S_{lim}.

This maximum scale is important for both computational and scientific reasons. Computationally, it is more expensive to make the necessary calculations at increasingly larger scales, which is an important consideration in an operational context. Scientifically, there is a scale above which it is no longer appropriate to consider high resolution forecasts: for example, when there also exists a lower resolution forecast (e.g from a global model), better placed to assess large scale errors.

Additionally, it is necessary to separate cases where two forecasts predict the same event, but at a different location, from those where each forecast predicts essentially different events. Consider the comparison of two forecasts which both produce precipitation, but at a different location in the domain. In some situations the forecasts will be predicting a region of precipitation with the same physical characteristics. In this case, we could say that the same event is predicted by both forecasts, but with uncertainty in the location. This is the location uncertainty that can be quantified using the agreement scales, $S_{ij}^{A(f_1 f_2)}$. However, it is also possible that the forecasts are predicting different events entirely, such as convective showers due to low level convergence in one, and convection associated with a frontal system in another. In this second situation, the differences between the forecasts are not representative of their spatial uncertainty, and hence the values of $S_{ij}^{A(f_1 f_2)}$ could be misleading. Thus, when the agreement scales are calculated, it is an underlying assumption that the same events are being forecast by the two fields, but at different locations. As the scale increases this assumption is likely to be less valid and the forecasts are more likely to be representing different physical phenomena. Note that this assumption is needed because of the spatial neighbourhood approach where forecasts at different locations in the domain are considered together: it is not needed in traditional measures which compare fields at the same grid point only.

For these reasons it is necessary to fix a maximum scale of interest, as done here using S_{lim}, above which forecast features are not expected to be related. The consequences of having a fixed maximum scale are assessed further in Section 6.4. Equation 6.3 is formulated so that, as forecast differences increase, the scales of acceptable agreement tend smoothly towards S_{lim}. Specifically, the fractional difference between the fields that is considered acceptable increases for increasing S until, at S_{lim} itself, any difference is accepted. The dependence of the acceptable fractional discrepancy between the fields as a function of spatial scale S is shown in Figure 6.3 for $\alpha = 0.5$ and S_{lim}=80 or 100. In the work presented here, values of $\alpha = 0.5$ and $S_{\mathrm{lim}} = 80$ have been used. For specific applications that require a more/less stringent match lower/higher values of α could be selected. For the forecasts analysed in Section 6.5, the maximum scale of 80 grid points corresponds to a square neighbourhood of 25921 grid points with total width 354.2 km (the model has a grid length of 2.2 km). It is believed that beyond this maximum scale any useful information from the convective-scale forecasts has already been extracted. Note that experiments were conducted with different values of both α and S_{lim} but, as these modifications did not affect the overall conclusions, they are not presented here.

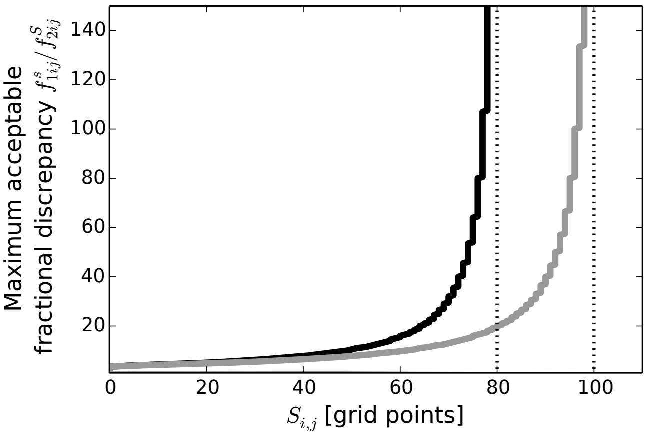


Figure 6.3: Maximum acceptable fractional discrepancy between f^S_{1ij} and f^S_{2ij} as a function of neighbourhood size S, for $\alpha = 0.5$ and $S_{\text{lim}} = 80$ (black) and 100 (grey).

6.3.3 Comparison with the Fractions Skill Score

It is informative at this stage to compare Equation 6.2, defining the agreement scale at a particular location, with the FSS believable scale as defined in Roberts and Lean (2008), and discussed in Chapter 4, Section 4.4.1. Using idealised and real examples, Roberts and Lean (2008) inferred that for small rainfall coverage, the FSS believable scale is obtained (Equation 4.2 is met) when the neighbourhood size is equal to twice the separation of forecast rainfall features. More recently, work by Skok (2015) has shown analytically that, for simple idealised configurations in an infinite domain, the neighbourhood size is twice the spatial separation of precipitation objects when the FSS has a value of 0.5. Thus, using the FSS, the scale can be found at which a forecast is, on average, skillful across the model domain.

There are some similarities between the method of calculating the agreement scales (Equation 6.2) and the FSS (Equation 4.2). For example, in both calculations, the difference of quantities squared is divided by the sum of their squares. However, there are also some important differences.

- The key difference in approaches is that the FSS gives a single domain-wide value for the spatial agreement, whereas the agreement scales provide a location-dependent map of the spatial agreement. Therefore, in the FSS, the squared difference between fields, and sum of the squares of the two fields, are further summed over all points in the domain. This is not the case for the agreement scales (Equation 6.2), where each location is considered separately. The denominator of the FSS equation (Equation 4.2) is the maximum possible difference that can be obtained from two fields of fractions, whereas in Equation 6.2 the denominator is a convenient normalisation factor.

- Scales of interest are obtained for $S_{ij}^{A(f_1 f_2)}$ and the FSS when a criterion exceeds a value of 0.5 plus an extra term. It should be stressed that these criterion do not have the same meaning. For the FSS, the value "0.5" relates directly to the spatial separation of precipitation features (Roberts and Lean, 2008; Skok, 2015), whereas in Equation 6.2 the value α (equal to 0.5 here) controls the bias considered acceptable. The additional terms in the criteria also have different functions in each of the two measures: that used for the FSS relates to the coverage of precipitation in the domain, whereas that in Equation 6.2 ensures that the search algorithm always returns a meaningful scale.
- Although both equations consider errors both in precipitation location and precipitation amount, these are treated differently. In particular, the FSS is applied to precipitation fields that have undergone thresholding to produce binary fields. In contrast, the agreement scales compare the precipitation amounts themselves (Equation 6.2). This is a more general approach which does not require a threshold to be defined, and directly considers the scale-dependent bias between the fields.

6.3.4 Calculations for an ensemble

In Chapter 4 the FSS was used to estimate the domain-averaged spatial ensemble spread and skill by comparing all independent pairs of ensemble members, and all ensemble member-radar pairs. These were also considered as options in Leoncini et al. (2016, submitted). Here a similar approach is applied to the agreement scales $S_{ij}^{A(f_1 f_2)}$ to calculate how the ensemble agreement and skillful scales vary with location across the domain.

To give a measure of the location-dependent spatial ensemble spread, the agreement scales $S_{ij}^{A(f_1 f_2)}$ are calculated separately for each independent pair of ensemble member forecasts. This gives

$$N_p = \frac{N(N-1)}{2} \tag{6.4}$$

fields of agreement scales for an ensemble of N members. It is necessary to provide a summary value of all these fields to quantify the overall spatial uncertainty of the ensemble at each point in the domain. Here, to get an agreement scale representative of the ensemble, the average is taken, at each grid point in the domain, over the N_p values of $S_{ij}^{A(f_1 f_2)}$. Hence, for an ensemble of twelve members, 66 agreement scales would be separately calculated ($N_p = 66$), and the average of these 66 fields would be taken at each grid point in the domain. As the distribution of the N_p agreement scales was found to be uni-modal (not shown), the mean is an appropriate value to characterise the distribution of individual scales. This mean field indicates the average agreement scale between the ensemble members at each grid point, and is denoted $S_{ij}^{A(\overline{mm})}$. It represents the scales over which the ensemble should be evaluated, and the area over which individual features seen in the member forecasts should be expected to occur. Mathematically, the $S_{ij}^{A(\overline{mm})}$ are given by

$$\mathrm{S}_{ij}^{A(\overline{mm})} \equiv \frac{1}{N_p} \sum_{f_1=1}^{N-1} \sum_{f_2=f_1+1}^{N} \mathrm{S}_{ij}^{A(f_1 f_2)}. \tag{6.5}$$

The $S_{ij}^{A(\overline{mm})}$ show the local variations of the ensemble spatial dispersion which, for a well-spread ensemble system, should represent the true spatial forecast uncertainty. This spread-skill relationship has proved useful for the traditional analysis of ensembles (e.g. Buizza, 1997; Leutbecher and Palmer, 2008, and references therein), but is only recently being explored in a spatial context. To test the spatial spread-skill relationship we define a corresponding skill measure, by calculating the average agreement scales of all member fields compared with radar observations. These are the scales over which the ensemble was skillful in capturing the observed precipitation, and are denoted $S_{ij}^{A(\overline{mo})}$. It is necessary to use radar observations for this comparison due to their high spatial coverage. To calculate the $S_{ij}^{A(\overline{mo})}$, the mean is taken, at each grid point, over the fields of agreement scales calculated from comparing all N member-radar pairs:

$$\mathrm{S}_{ij}^{A(\overline{mo})} \equiv \frac{1}{N} \sum_{f=1}^{N} \mathrm{S}_{ij}^{A(fo)}. \tag{6.6}$$

Therefore, for an ensemble of twelve members, there are 66 pairs contributing to the $S_{ij}^{A(\overline{mm})}$, but only twelve pairs contributing to the $S_{ij}^{A(\overline{mo})}$. In addition to the different number of comparison pairs, there are other questions regarding the validity of directly comparing $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$. For example, similar to the comparison of dFSSmean and eFSSmean (Chapter 4), the agreement scale for each member-member or member-radar pair is calculated separately and normalised within Equation 6.1. Further, as for scalar Gaussian random variables, the mean euclidean distance between points is equal to twice the variance, it might be expected that there is a factor of two difference between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$. These questions are discussed in Section 6.4 in the context of an idealised experiment. The methods of calculating $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are summarised schematically in Figure 6.4 for an ensemble of four members.

The $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ measure the location-dependent spatial ensemble believable and skillful scales respectively. These measures differ in several key ways from the traditional spread-skill measures of ensemble standard deviation and RMSE as used, for example, by Buizza et al. (2005); Kong et al. (2007); Bouttier et al. (2012); Baker et al. (2014). In particular, the RMSE compares the ensemble mean to observations, and hence a minimum possible RMSE of zero can be obtained when the observations equal the ensemble mean. In contrast, the $S_{ij}^{A(\overline{mo})}$ compares the observations directly to each ensemble member. For any situation where the ensemble members differ spatially or in magnitude, the $S_{ij}^{A(\overline{mo})}$ will be non-zero. Hence the minimum $S_{ij}^{A(\overline{mo})}$ will also be non-zero: this is limited by, and related to, the ensemble spread. Note that this is a general feature of spatial analysis and would also be true of other spatial comparison methods; for example, any method which considers the differences in location of forecast features between observations and individual ensemble members.

6.4 Idealised experiment

To investigate the properties of the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$, an idealised experiment was employed. This allows links between the precipitation distribution and agreement scales to be explored for configurations with known properties. As discussed in the preceding

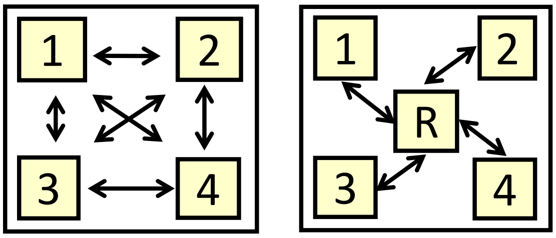


Figure 6.4: Schematic showing the comparisons used to calculate the mean over all member-member pairs (left) and mean over all member-radar pairs (right) for an ensemble of 4 members.

section, questions remain about using a direct comparison of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ to investigate the spatial spread-skill relationship. In the absence of a rigorous mathematical proof, the idealised experiment allows some of these questions to be addressed. Ensembles were created that were either spatially well spread, over spread or under spread, allowing the validity of the spatial spread-skill comparison between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ to be tested.

6.4.1 Overall setup

To mirror the analysis of the real MOGREPS-UK cases (introduced in Chapter 5 and to be discussed in Section 6.5) a domain of 193 by 242 grid points was created. Initially all points in the domain were set to zero, representing the case of no rain. To simulate precipitation, circular areas ('rain blobs') within the domain were each set to an arbitrary value (greater than 0). To represent an ensemble of N forecasts at a given time t_1, the centres of N rain blobs were randomly positioned within a square 'rain area' of side L and lower left corner (LLC) at point (X, Y). Similarly, to represent the radar observation at t_1, one rain blob was positioned within a square 'radar rain area' of length L_o and LLC at point (X_o, Y_o). To represent different draws from the ensemble distribution, or equivalently different forecasts of the event, multiple random draws were made for the ensemble and radar positions.

The standard ensemble configuration considered 13 different draws of a 12 member ensemble in order to mirror the number of times and members considered for the MOGREPS-UK cases (analysed in Section 6.5). In the standard setup the ensemble member and radar rain areas were set to $L = L_o = 50$ grid points positioned towards the centre of the domain with the LLC at $(60, 60)$. The standard rain blob radius was 8 grid points. An example of the ensemble member positions, from one random draw of the standard configuration, is given in Figure 6.5.

6.4.2 Agreement scale maps

An example map of $S_{ij}^{A(\overline{mm})}$ from the standard ensemble configuration at one time is shown in Figure 6.6a. Near the centre of the rain area the scales are smallest, around 10 grid points. This scale is representative of the average separation of the rain blobs.

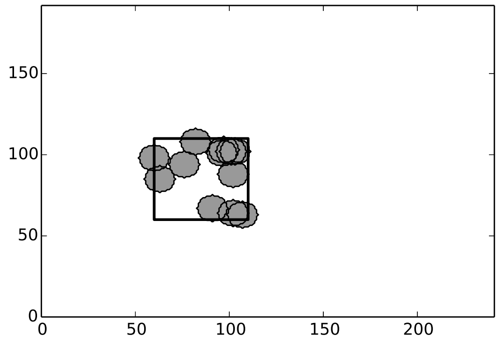


Figure 6.5: Example of ensemble member rain blobs (grey circles, one per member) positioned within a rain area (black square) for one random draw of the standard idealised setup.

Moving away from the precipitation area the $S_{ij}^{A(\overline{mm})}$ increases as the distance from the rain area dominates the $S_{ij}^{A(\overline{mm})}$. This is an important feature of the $S_{ij}^{A(\overline{mm})}$: outside the rain area the scales are increasingly representative of the distance from the precipitation. This makes sense when considering the $S_{ij}^{A(\overline{mm})}$ to be the scales over which the fields should be evaluated: far from a rain area any small variations in the precipitation field should be smoothed out over a large neighbourhood. Figure 6.6b shows a second example $S_{ij}^{A(\overline{mm})}$ map, at a different time to that shown in Figure 6.6a. Overall the results between the two times are similar. There are however slight differences in the location of the minimum $S_{ij}^{A(\overline{mm})}$ values, due to the rain blobs being at different locations in the rain area. Note that in the situation of no rain, the $S_{ij}^{A(\overline{mm})}$ will be equal to the maximum scale considered, S_{lim}.

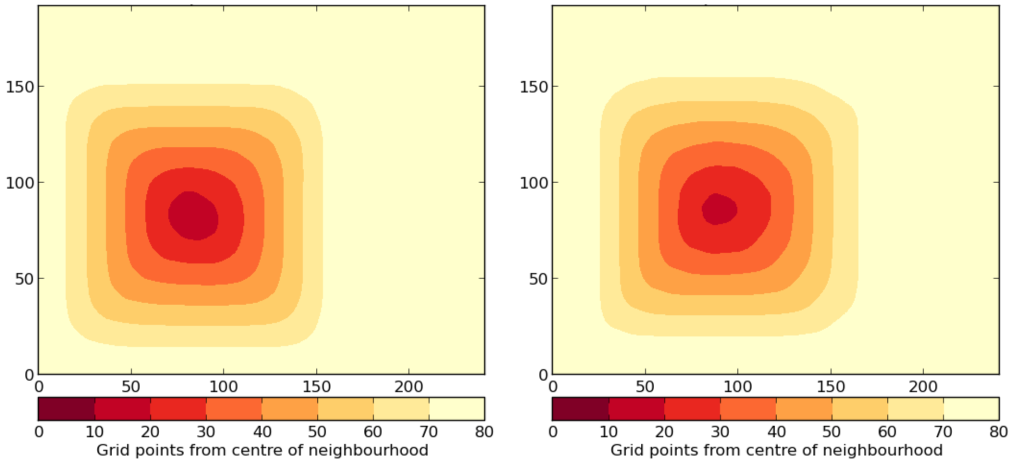


Figure 6.6: $S_{ij}^{A(\overline{mm})}$ from the idealised experiment standard configuration (a blob radius of 8 grid points) at two different times. All points in the idealised domain are included.

6.4.3 Different configurations

Agreement scale maps are useful for visualising and understanding spatial predictability differences across the domain. To compare different configurations, histograms of all points from the $S_{ij}^{A(\overline{mm})}$ maps are considered. An example is given in Figure 6.7 for experiments with different rain blob radii (the other parameters were unchanged from the standard configuration). It is expected that larger/smaller rain blobs will have more/fewer locations with small $S_{ij}^{A(\overline{mm})}$ as they represent situations that are more/less spatially predictable. The histogram for configurations with different rain blob radii is given in Figure 6.7, the other parameters were unchanged from the standard configuration. From Figure 6.7 it can be seen that the $S_{ij}^{A(\overline{mm})}$ are behaving as expected: the experiment with a radius of 30 grid points has a minimum spatial scale 18 times smaller than that seen for the experiment with a radius of 1 (a single point). The experiments with larger radii have more points at all scales below 65. Above 65 this behaviour changes and the experiments with smaller radii have more points. Note that, as all experiments have the same total number of points, those experiments with more points at small scales must have fewer points at the largest scales: the fact that this crossover happens around 65 is due to the relative sizes of the rain area and the domain.

Experiments with varying numbers of ensemble members (from 4–20 members) gave very similar $S_{ij}^{A(\overline{mm})}$. This suggests that, at least for this simple idealised setup where the ensemble spread is predefined, the $S_{ij}^{A(\overline{mm})}$ are not overly sensitive to the number of ensemble members. Of course, this result may not hold for real case studies and operational ensemble systems. The investigation of the effect of ensemble size for real case studies would require the consideration of a large number of cases, members, and weather regimes and is beyond the scope of this thesis.

6.4.4 Different spread-skill relationships

In this subsection the relationship between $S_{ij}^{A(\overline{mm})}$ (representing spatial ensemble spread) and $S_{ij}^{A(\overline{mo})}$ (representing spatial ensemble skill) is illustrated. If these measures are to provide useful information about an ensemble, they must differentiate between ensembles that are spatially well spread, over spread, and under spread. Here this is tested by calculating $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ for idealised ensembles with known spread-skill

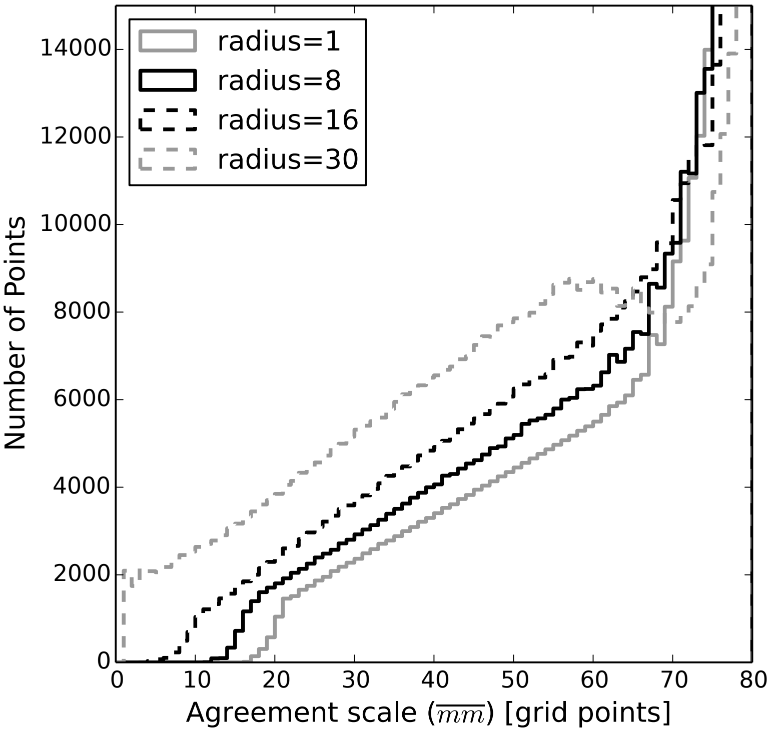


Figure 6.7: Histogram of $S_{ij}^{A(\overline{mm})}$ for all points in the domain. Idealised experiments are shown with different blob radii (r): grey solid line, $r = 1$; black solid line, $r = 8$ (standard radius); black dashed line, $r = 16$; grey dashed line, $r = 30$. Other parameters were unchanged from the standard configuration.

properties: well spread, over spread, under spread and wrongly located precipitation (biased). To generate a spatially well spread ensemble both members and radar were selected from the same area (i.e $L = L_o$ and $(X, Y) = (X_o, Y_o)$). To generate an over/under spread ensemble the radar rain area was defined to be smaller/larger than the member rain area. An additional case, where the ensemble was under spread due to a spatial displacement between the ensemble and observations was also considered with $L = L_o$ but $(X, Y) \neq (X_o, Y_o)$. The radar and member rain areas for these different ensemble configurations are shown in Figure 6.8, and the settings for these idealised setups are given in Table 6.1.

6.4.5 Method for comparing $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$

Although histograms of all points from the $S_{ij}^{A(\overline{mm})}$ maps allow the differences between configurations to be visualised (e.g Figure 6.7), in order to fully compare the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ it is necessary to chose a method that enables a scale selective comparison, whilst

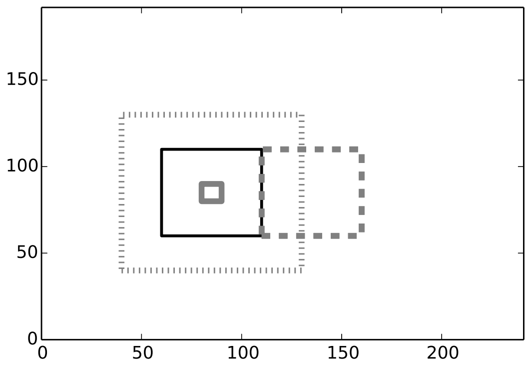


Figure 6.8: Positions of radar rain areas for cases with different spread-skill: well spread (black solid), over spread (grey solid), under spread (grey dotted) and under spread due to misplaced precipitation (grey dashed). For all experiments the ensemble members were selected from the black square.

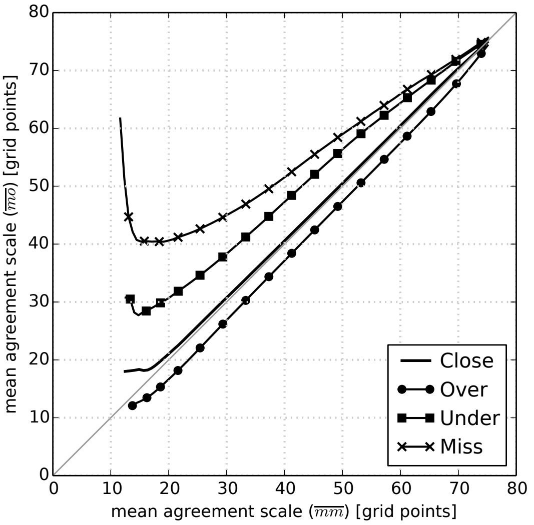


Figure 6.9: Binned scatter plot for idealised ensembles with different spread-skill characteristics: over spread (black with circles), well spread (black with no markers), under spread (black with squares), and missed precipitation (black with crosses). A bin size of 10 grid points was used.

preserving the point-to-point relationship between the S_{ij}^{A} fields. One way to do this would be a simple scatter plot of the $S_{ij}^{A(\overline{mo})}$ against the $S_{ij}^{A(\overline{mm})}$. However, this would give a noisy result. To enable simpler comparison, we bin the scatter plot based on the $S_{ij}^{A(\overline{mm})}$ value.

First, a bin size is selected, for example to select scales with values between 0 and 9 grid points. The locations for which the $S_{ij}^{A(\overline{mm})}$ value lies within the bin are then found and the mean $S_{ij}^{A(\overline{mm})}$ over these locations calculated. By definition this mean value must lie within the selected bin. Next, the $S_{ij}^{A(\overline{mo})}$ mean value *over the same locations*

Table 6.1: Idealised ensemble settings for ensembles with different spread-skill relationships.

Spread-skill	L	L_o	X, Y	X_o, Y_o
Well spread (‘close’)	50	50	60,60	60,60
Over spread (‘over’)	50	10	60,60	80,80
Under spread (‘under’)	50	90	60,60	40,40
Displaced precipitation (‘miss’)	50	50	60,60	110,60

is computed. If the ensemble is well spread this will equal the $S_{ij}^{A(\overline{mm})}$ mean value; if the ensemble is over/under spread then the $S_{ij}^{A(\overline{mo})}$ mean value will be smaller/larger than that of the $S_{ij}^{A(\overline{mm})}$. In the binned scatter plot, for each bin considered, the $S_{ij}^{A(\overline{mo})}$ mean value is plotted against the $S_{ij}^{A(\overline{mm})}$ mean value. Hence, on the binned scatter plot, a well spread ensemble should lie on the diagonal, and under/over spread ensembles should lie above/below the diagonal.

These interpretations have been checked using various idealised ensembles with pre-defined spread-skill characteristics, such as those specified in Figure 6.8 and Table 6.1. An example binned scatter plot is shown in Figure 6.9 for a bin size of 10 grid points. A running bin is used giving bins of agreement scale values from 0 to 9, 1 to 10, 2 to 11, $\cdots$, 71 to 89 grid points.

As expected, the over spread case lies below the diagonal with $S_{ij}^{A(\overline{mo})}$ smaller than $S_{ij}^{A(\overline{mm})}$ for a given bin. The two experiments with under spread ensembles ('under' and 'miss') both lie above the diagonal, with $S_{ij}^{A(\overline{mo})}$ larger than $S_{ij}^{A(\overline{mm})}$ for a given bin. This confirms that $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are differentiating between the different ensembles and provide useful information about the spatial spread-skill. The 'miss' experiment lies furthest from the diagonal as, in this case, the radar and ensemble rain blobs are never in the same rain area. Notice also that the lines do not lie uniformly away from the diagonal; the difference between the $S_{ij}^{A(\overline{mo})}$ and $S_{ij}^{A(\overline{mm})}$ is most noticeable for small agreement scales. This is seen particularly for the 'miss' experiment. Here the ensemble members place the precipitation at a different location to the observations and hence the minimum $S_{ij}^{A(\overline{mm})}$ will be located at a different point in the domain to the minimum $S_{ij}^{A(\overline{mo})}$. Thus, as both minima are surrounded by larger agreement scales, the maximum difference between the two fields will occur in at the locations of the minima, i.e. for the smallest agreement scales.

The 'close' experiment is shown in black and lies on the diagonal as expected: the average of all $S_{ij}^{A(\overline{mm})}$ points within a given bin is equal to the $S_{ij}^{A(\overline{mo})}$ averaged over the same points. Notice that there is some departure from the diagonal at scales 10-18: the average $S_{ij}^{A(\overline{mo})}$ over these points is larger than the average $S_{ij}^{A(\overline{mm})}$. This is due to our simple method of defining the idealised ensemble: randomly selecting a modest number of ensemble members within a given area results in a non-uniform member distribution over that area, which would for an ideal ensemble represent an uneven radar spatial probability distribution across the area. However, the radar distribution was assumed to be uniform. This interpretation was confirmed by experiments in which the rain blobs for the ensemble members were positioned not randomly but at fixed, uniformly-distributed, locations.

The results from this section show that the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ can successfully be used to determine the spatial spread-skill characteristics of an ensemble system, and that the binned scatter plot provides a particularly clear method of viewing these results. It has been demonstrated that $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ can be compared directly (i.e. there is no factor of two, which would be found in the case of euclidean distances between Gaussian random variables). This gives some confidence in these measures which, in Section 6.5, will be applied to real convective cases. However, other questions, relating to the effects of non-Gaussianity of the rainfall distribution, or bias between ensemble members and observations have not been addressed by the idealised experiment: only uncertainties in the location of precipitation was considered here. Future work should investigate the effects of different precipitation distributions and biases using more complex idealised simulations.

6.5 Convective cases from MOGREPS-UK: instantaneous rain rates

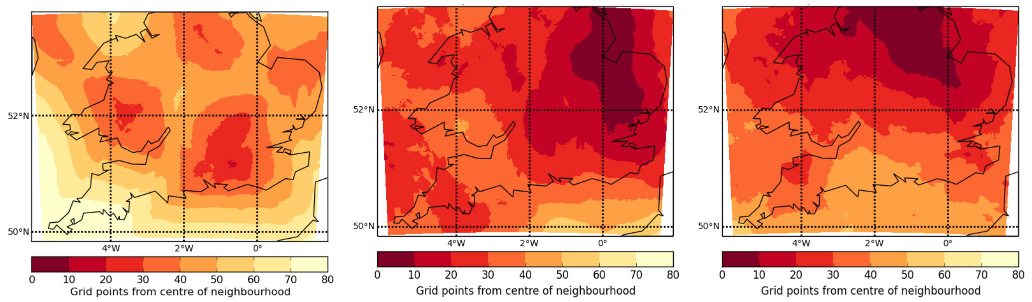
6.5.1 Results: spatial maps

To investigate the spatial predictability of the six MOGREPS-UK cases discussed in Chapter 5, the spatial differences between ensemble members were examined. To do this, the $S_{ij}^{A(\overline{mm})}$ were calculated hourly for each case using instantaneous rain rates. Example $S_{ij}^{A(\overline{mm})}$ maps are given in Figure 6.10 at the same times as shown in Figures 5.4a-c and 5.5a-c for the MOGREPS-UK rain rate data. Comparison of Figures 5.4a-c, 5.5a-c and 6.10 shows that the smaller $S_{ij}^{A(\overline{mm})}$ tend to be linked to areas of precipitation: in these areas, the spatial differences in the placement of precipitation between members are smallest. However, there are additional aspects of the pattern of $S_{ij}^{A(\overline{mm})}$ that are highly case dependent.

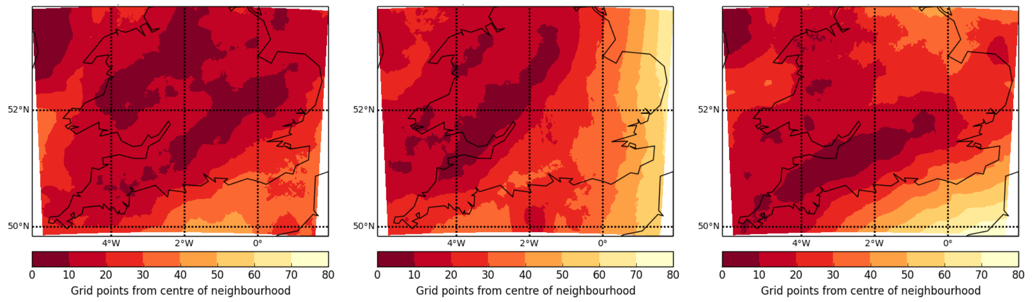
For the cases showing peninsula convergence (29/07, 02/08 and 03/08, Figure 6.10d-

f) smaller scales are seen along the peninsula where the precipitation is highly spatially predictable. This predictability is a result of the topographical influence in these cases. In contrast, the precipitation on the 17/07 (Figure 6.10a) is less predictable with a minimum $S_{ij}^{A(\overline{mm})}$ of around 20 grid points. The large spatial differences between ensemble members for this case are believed to be due to subtle differences in the large scale forcing: small variations in the mesoscale flow led to large variations in triggering locations for convection. For this case the model consistently predicts localised thunderstorms, but their location is uncertain. The $S_{ij}^{A(\overline{mm})}$ map allows this valuable information to be easily visualised. The same conclusions can, of course, be drawn from close inspection of the individual member rain rate fields but that is a more cumbersome and subjective process.

A further example of precipitation with lower spatial predictability is seen on the 02/08 to the east of the domain (Figure 6.10e). Here the observations show convective storms moving north from France (e.g Figure 5.3h) throughout the day. Similar behaviour is captured by a small number of ensemble members (the particular members, and the number of members is time dependent). In this region (lower left corner 50.0°N,



(a) 17/07 at 17Z (b) 23/07 at 06Z (c) 27/07 at 23Z



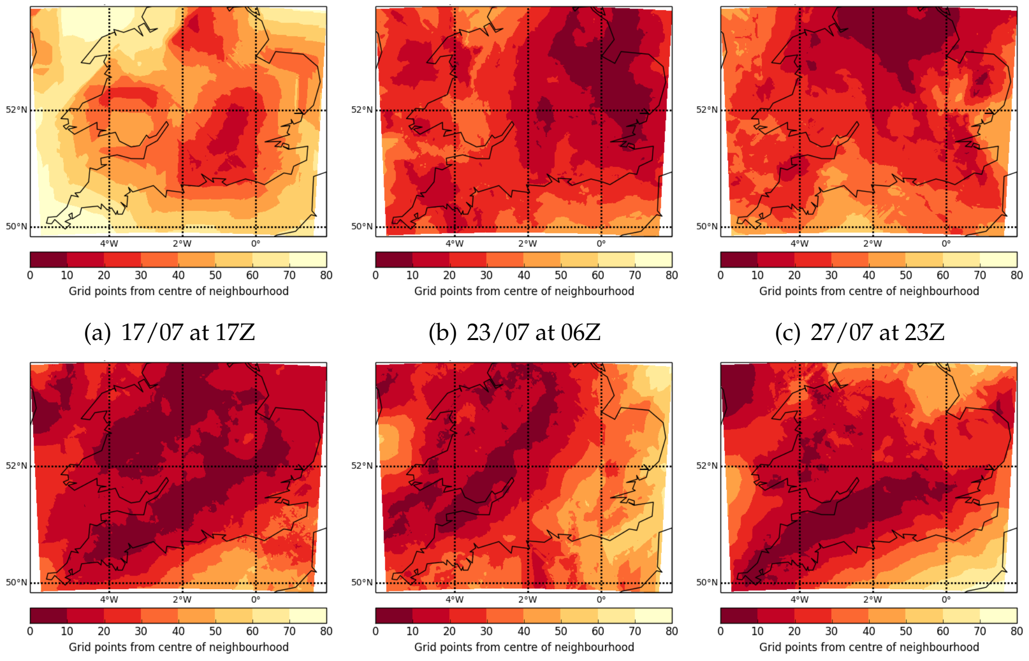
(d) 29/07 at 15Z (e) 02/08 at 18Z (f) 03/08 at 15Z

Figure 6.10: $S_{ij}^{A(\overline{mm})}$ for the (a) 17/07 at 17Z, (b) 23/07 at 06Z, (c) 27/07 at 23Z, (d) 29/07 at 15Z, (e) 02/08 at 18Z and (f) 03/08 at 15Z. These are the same times as those shown for the radar rain rate data in Figures 5.1 and 5.3, and MOGREPS-UK rain rates in Figures 5.4 and 5.5.

-1.5°E, upper right corner 52.0°N, 1.0°E) the $S_{ij}^{A(\overline{mm})}$ vary from 30 to 60 grid points, suggesting that precipitation could occur within a broad region. This information, in conjunction with a single reference ensemble member, or a deterministic forecast, would help the risk of heavy rain to be more accurately assessed.

The two cases with the most widespread precipitation, 23/07 and 27/07 (Figures 6.10b and 6.10c respectively), both have $S_{ij}^{A(\overline{mm})}$ of less than 20 grid points over the regions where precipitation occurred. For these cases the spatial uncertainty in the location of precipitation was much smaller than the size of the precipitation area, and hence there was a high degree of agreement and overlap between the individual member forecasts.

The results from Figure 6.10 provide a summary of the spatial uncertainty within the ensemble for a particular time on each day, in one single image. This is useful for model interpretation, and would be valuable in an operational forecasting context. However, it is also important to consider whether these scales are representative of the true spatial predictability for each case. To assess the 'spatial spread-skill relationship' the $S_{ij}^{A(\overline{mo})}$ were also calculated hourly for all cases, as described in Sections 6.3 and 6.4. Exam-



(a) 17/07 at 17Z

(b) 23/07 at 06Z

(c) 27/07 at 23Z

(d) 29/07 at 15Z

(e) 02/08 at 18Z

(f) 03/08 at 15Z

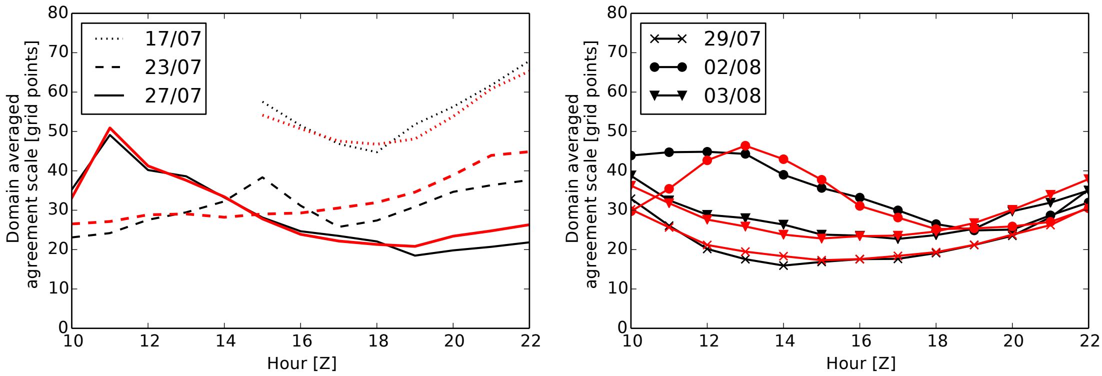
Figure 6.11: $S_{ij}^{A(\overline{mo})}$ for the (a) 17/07 at 17Z, (b) 23/07 at 06Z, (c) 27/07 at 23Z, (d) 29/07 at 15Z, (e) 02/08 at 18Z and (f) 03/08 at 15Z. These are the same times as those shown for the radar rain rate data in Figures 5.1 and 5.3, and MOGREPS-UK rain rates in Figures 5.4 and 5.5.

ple $S_{ij}^{A(\overline{mo})}$ maps are shown for all six cases in Figure 6.11, at the same times used for the $S_{ij}^{A(\overline{mm})}$ maps (Figure 6.10). Overall the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ look qualitatively similar. There are however some differences. In particular, the $S_{ij}^{A(\overline{mo})}$ have larger areas of both the smallest and largest scales, and are more noisy. These differences will be quantified in the following sections.

6.5.2 Results: domain average

To investigate the overall, and time evolution, of spatial agreement for each of the six cases, we now consider the domain average $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$. These are shown in Figure 6.12a for the 17/07, 23/07, 27/07 and Figure 6.12b for the 29/07, 02/08 and 03/08. The cases are shown from 10Z to 22Z, a period which covers the convective events of interest. The 17/07 case is only shown from 14Z onwards when convection occurred: before this time there was no simulated precipitation over the domain and, additionally, problems with the radar data.

Note that, for the results shown here, the forecast start time varies between cases, consistent with work presented in Chapters 5 and 7. Considering different forecast start times allowed the ensemble to be run from the most recent, and hence most accurate, global analysis, whilst also ensuring that the model had fully spun up before the convective events of interest. The forecast start times are listed in Table 5.1. For a direct comparison of the spread skill between cases, it is necessary to consider forecasts with



(a) (b)

Figure 6.12: Time series of domain averaged $S_{ij}^{A(\overline{mm})}$ (red) and $S_{ij}^{A(\overline{mo})}$ (black). (a) 17/07, dotted; 23/07, dashed; 27/07, solid; (b) 29/07, solid with crosses; 02/08, solid with circles; 03/08, solid with triangles.

the same start time as ensemble spread is expected to grow with forecast lead time (due to increasing errors). Hence, the calculations in this section were also repeated for all six cases with UK forecast start time 03Z (i.e. using different forecasts for the 23/07 and 27/07 cases to those presented here). The 03Z start time cases lead to similar overall conclusions and can be found in Dey et al. (2016a, under review). Although the 03Z results are not presented here, any differences between them and the runs with varying start time will be highlighted and discussed.

The 17/07 case has the lowest spatial agreement with a minimum domain average $S_{ij}^{A(\overline{mm})}$ of around 50 grid points. This agrees with the qualitative analysis of the agreement scale maps (Figure 6.11a). Interestingly, despite such large spatial uncertainties, the model has captured the spatial agreement well on this day: the domain averaged $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are similar, with the black and red lines lying close to each other. The two cases with widespread precipitation, 23/07 and 27/07, both have a domain-averaged $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ in the range of 20 to 40 grid points, showing that these cases are more spatially predictable than the 17/07. Similar results are found for the 23/03 03Z run (not shown). On the 27/07, the results in Figure 6.12a show that from 10Z to 19Z the domain averaged $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are similar: the ensemble forecast of the MCS is spatially well spread. This was not the case for the earlier 27/07 03Z forecast (not shown) where the domain averaged $S_{ij}^{A(\overline{mo})}$ were larger than the domain averaged $S_{ij}^{A(\overline{mm})}$: the ensemble members were closer to each other than they were to observations and the ensemble forecast of the MCS was spatially under spread. This was due to a timing error between the model and observations, and was one motivation for choosing the later forecast start time for the 27/07 case.

Precipitation on the 29/07 was the most spatially predictable of the six cases, with high spatial agreement and domain average $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ dropping below 20 grid points from 12Z to 18Z. Again, this agrees with the qualitative analysis. The other cases with peninsula convergence (02/08 and 03/08) behave similarly to the 29/07 with domain average scales below 30 grid points, and similar values of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$. Before 15Z on the 02/08, the domain-average agreement scale is dominated by the spatially unpredictable precipitation to the east of the domain, and larger $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are seen. At these times the ensemble is, at least in a domain-averaged sense, under spread with a difference of over 10 grid points between the domain averaged $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$.

6.5.3 Results: location-dependent comparison

Although the domain average is a useful indicator of the overall spatial agreement for a given case, it does not say anything about the local spatial behaviour and hence it is also important to compare the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ in a manner that preserves the spatial information locally. For this purpose the binned scatter plot is employed, as applied to the ideal ensemble in Section 6.4. Results are shown for all cases except the 17/07 at 10Z and 14Z (Figure 6.13a,b) and for all cases at 18Z (Figure 6.13c). A bin size of 10 grid points has been used for these plots. This bin size was chosen as it allows low agreement scales to be represented, whilst still considering a large enough scale range to give robust results. Similar conclusions are obtained from bin sizes in the range of 4 to 20 grid points, and are not presented here.

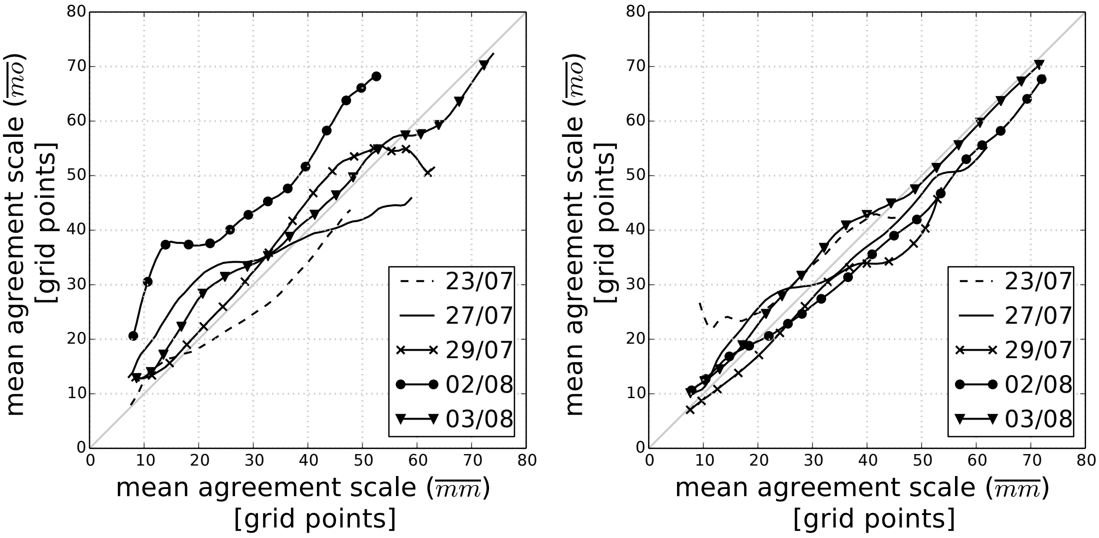
At both 10Z and 14Z (Figure 6.13a and b respectively) the spatial spread-skill relationship is case dependent. This can be related to the different physical processes occurring for each case and time, and also to differences between the forecast and observations.

23/07

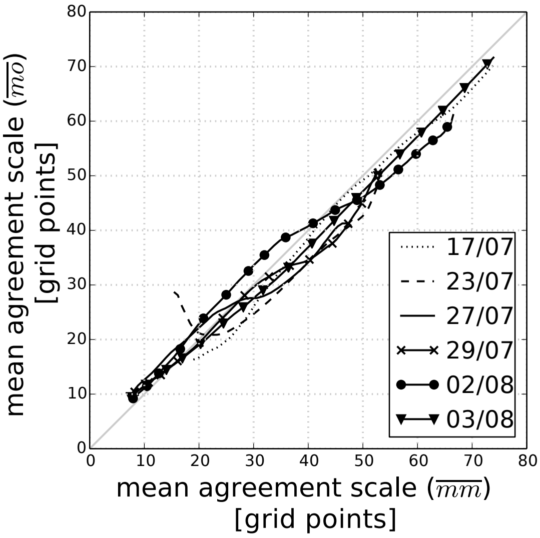
On the 23/07 at 10Z (dashed line) the ensemble members were further from each other than they are from observations ($S_{ij}^{A(\overline{mm})}$ greater than $S_{ij}^{A(\overline{mo})}$): the observed precipitation fell towards the centre of the region where the individual ensemble members forecast precipitation. By 14Z the ensemble was spatially under spread with $S_{ij}^{A(\overline{mm})}$ less than $S_{ij}^{A(\overline{mo})}$: at this time all the ensemble members were still predicting a band of convection to the west of the the domain, whereas in reality the main precipitation band had moved northwest out of the domain.

27/07

On the 27/07 (solid line) similar behaviour is seen at 10Z and 14Z with $S_{ij}^{A(\overline{mm})}$ smaller than $S_{ij}^{A(\overline{mo})}$ for scales below 30 grid points: when the ensemble is confident in the positioning of precipitation (e.g light rain at 10Z or the MCS at 14Z) it is spatially under spread (overly confident in the location of precipitation). At scales above 30 grid points a different behaviour is seen with $S_{ij}^{A(\overline{mm})}$ greater than $S_{ij}^{A(\overline{mo})}$. This is due to the ensemble



(a) 10Z (b) 14Z



(c) 18Z

Figure 6.13: Binned scatter plot for the 23/07, 27/07, 29/07, 02/08, 03/08 at (a) 10Z, (b) 14Z, and for all cases at (c) 18Z. Individual traces are plotted for each case at the specified time: 17/07, dotted; 23/07, dashed; 27/07, solid; 29/07, solid with crosses; 02/08, solid with circles; and, 03/08, solid with triangles.

members having a smaller area of precipitation than was observed.

29/07

On the 29/07 (solid line with crosses) the ensemble was spatially under spread at 10Z for scales below 55 grid points and the ensemble was too confident in the placement of precipitation. At 10Z the largest $S_{ij}^{A(\overline{mm})}$ (greater than 50 grid points) are found to the east of the domain where the ensemble forecast little precipitation. The observations showed more precipitation in this region resulting in the $S_{ij}^{A(\overline{mo})}$ being smaller than the $S_{ij}^{A(\overline{mm})}$. At 14Z the ensemble was spatially over spread with ensemble members lying on average further from observations than from each other; $S_{ij}^{A(\overline{mo})}$ greater than $S_{ij}^{A(\overline{mm})}$.

02/08

The 02/08 case (solid line with circles) is spatially under spread at 10Z but spatially over spread by 14Z. The spatial predictability for this case varied throughout the day as discussed for the domain-averaged values (Figure 6.12). These results agree with those from the domain average: the ensemble is under spread at 10Z suggesting that the uncertainty in the convection moving in from France was difficult for the ensemble to quantify. This is possibly due to the convection initiating outside the MOGREPS-UK domain and so relying on the global model's convective parametrisation. Later in the day, when precipitation was mainly in an organised spatially-predictable line over the SW peninsula (as discussed with reference to Figures 5.3h and 6.10), $S_{ij}^{A(\overline{mo})}$ values were smaller than $S_{ij}^{A(\overline{mm})}$: the radar precipitation fell within the ensemble distribution. This could indicate that the ensemble was too pessimistic about spatial accuracy at this time.

03/08

On the 03/08 (solid line with triangles) the ensemble is spatially under spread at both 10Z and 14Z with $S_{ij}^{A(\overline{mo})}$ greater than $S_{ij}^{A(\overline{mm})}$. This is particularly noticeable at the earlier time, and is related to the ensemble members producing showers in a different area of the domain to where they were seen in reality. Later in the day both model and observations produced precipitation associated with convergence lines from the SW and Welsh peninsulas and the spread-skill relationship improved.

At 18Z (Figure 6.13c) the case-to-case differences in spread-skill, seen at 10Z and 14Z, are much reduced. By this time the agreement scales are reflecting the convection that has developed and evolved over the course of the day. Earlier in the day initiation errors degrade the spatial spread-skill. However, as the precipitation remains for a number of hours, once initiation has occurred in both model and observations, there is a large degree of overlap. This highlights the link between spatial and temporal errors: a timing error will also result in a spatial error between fields. Note, however, that this result may also be linked to the choice of only a limited number of convective cases, where convection was reasonably captured by the model. In Chapter 8 the spatial spread-skill relationship will be investigated for 3 months of data.

6.6 Convective cases from MOGREPS-UK: 3D variables

So far, this chapter has focused on the calculation of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ for forecasts of surface instantaneous rain rates. Rain rates were selected for this study to avoid any temporal smoothing from using precipitation accumulations, and hence to focus on the spatial features. Of course, the methods presented here could also be used to evaluate precipitation accumulations. More generally, although precipitation forecasts are a key application of these methods (due to their high spatial uncertainty), the $S_{ij}^{A(\overline{mm})}$ can also be calculated for other fields. In this section example results of the $S_{ij}^{A(\overline{mm})}$ for cloud fraction, horizontal convergence, horizontal divergence, specific humidity, temperature, and horizontal wind speed are presented. For each variable the $S_{ij}^{A(\overline{mm})}$ were calculated at every model level, and interpolated linearly onto constant height levels, to quantify the vertical variation in horizontal spatial uncertainty. Note that spatial uncertainties in the vertical are not considered here as errors in vertical positions are expected to be smaller than those in the horizontal due to smaller scale lengths for processes in the vertical. Investigation of the vertical spatial differences would be an interesting extension to the work presented here (discussed further in Chapter 9, Section 9.3).

6.6.1 Cloud fraction

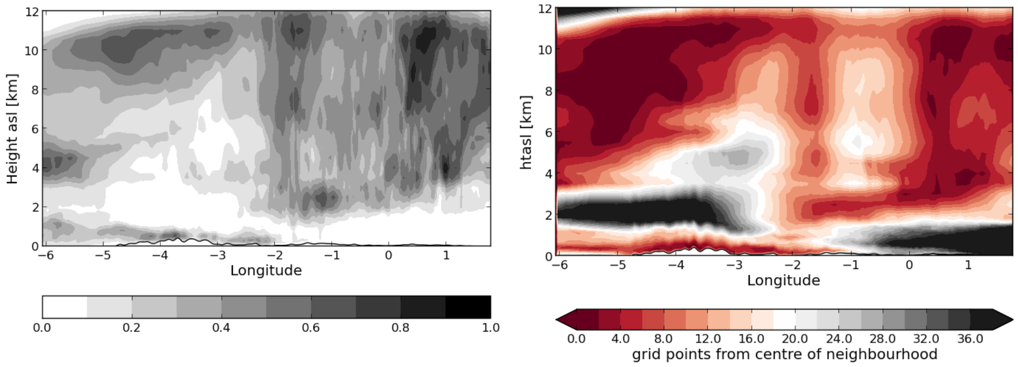
Cloud fraction fields are often highly inhomogeneous making this an interesting variable to consider from a spatial perspective. Using satellite observations it would also be possible to verify the cloud fraction $S_{ij}^{A(\overline{mm})}$ (as done using the FSS in Harvey and Dacre (2015)). Here we focus on the cloud fraction $S_{ij}^{A(\overline{mm})}$ to quantify the cloud fraction spatial uncertainty predicted by the ensemble. Two examples are provided that highlight the main features of cloud fraction $S_{ij}^{A(\overline{mm})}$ seen across the six COPE cases.

As discussed in Chapter 5, on the 23/07 two precipitation bands extended N-S across the UK associated with upper level forcing. Both precipitation bands were clearly defined in the radar observations (e.g. Figure 5.1h), but were less clear in the ensemble members (e.g. Figure 5.4b). A vertical cross section of cloud fraction $S_{ij}^{A(\overline{mm})}$ at 06Z, for a constant latitude of 52.1°N, is given in Figure 6.14b. This cross section intersects the two precipitation bands at -2.0°E and 0.5°E and is highlighted in red in Figure 5.4b. The individual members all had deep convective cloud along this cross section but the exact horizontal positions of the plumes were uncertain (e.g. Figure 5.4e). In the cloud fraction $S_{ij}^{A(\overline{mm})}$ three regions of very small agreement scales (2 grid points) are seen: to the west (left) of the cross section, and in the regions of the two precipitation bands. Elsewhere in the cross section the $S_{ij}^{A(\overline{mm})}$ is greater than 30 grid points reflecting the greater horizontal distance from cloudy points.

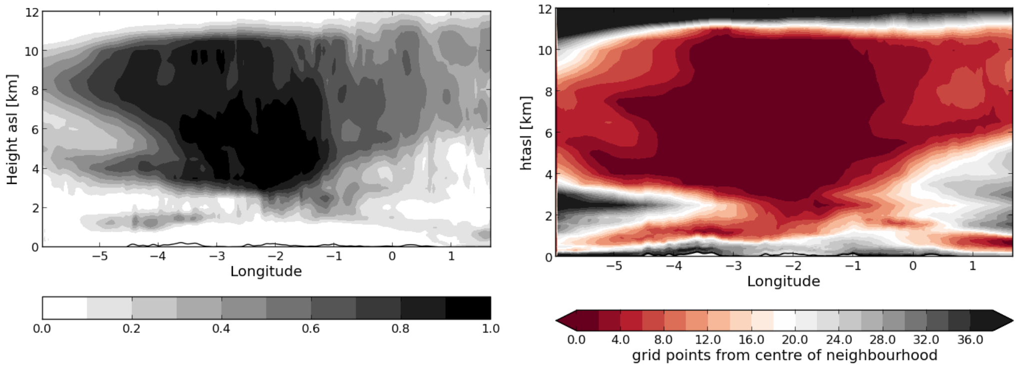
It is informative to compare the cloud fraction $S_{ij}^{A(\overline{mm})}$ cross section with that of the corresponding ensemble mean cloud fraction. The ensemble mean is typically what would be used to summarise the ensemble information in this situation, and is shown in Figure 6.14a. Higher ensemble mean cloud fractions are seen in the region of convective storms from latitudes -2°E to 2°E. Although there is a slight indication of higher cloud fraction bands around -2°E and 1.5°E these are not obvious, and may not be picked out without the knowledge that bands were indeed seen in the observations. However, from the $S_{ij}^{A(\overline{mm})}$ the banding is clear, with an area of reduced predictability between the bands around -1°E. This suggests that, although the two bands of precipitation are not so evident from the ensemble mean at this time, the ensemble is in fact confident at placing cloud in the locations of the bands.

The second example of cloud fraction $S_{ij}^{A(\overline{mm})}$ is taken from the 27/07, when an MCS moved north across the UK. At 15Z the MCS was located over southern England characterised by a large region of precipitation and cloud. The extent of this cloud, and the high level of agreement in its horizontal position between ensemble members, is highlighted in Figures 6.14c,d which show vertical cross sections at a constant latitude of 51.0°N for ensemble mean cloud fraction and cloud fraction $S_{ij}^{A(\overline{mm})}$ respectively. The location of this cross section is highlighted in Figure 5.4c (red line). The cloud fraction field has high spatial agreement in the large area of cloud associated with the MCS.

In general, smaller scales (decreasing to 2 grid points) are seen in areas of higher



(a) 23/07 Ensemble mean (b) 23/07 Agreement scales



(c) 27/07 Ensemble mean (d) 27/07 Agreement scales

Figure 6.14: Ensemble mean cloud fraction (left) and cloud fraction $S_{ij}^{A(\overline{mm})}$ (right) for the 23/07 at 06Z (a,b) and 27/07 at 15Z (c,d). The cross sections are at a constant latitude of 52.1°N (23/07) and 51.0 °N (27/07).

mean cloud fraction. This makes sense if we consider properties of the ensemble mean: the largest mean values are seen when all members place high values at the same spatial location. Low values of the ensemble mean have a less direct link with the spatial agreement: low mean values can be due to either large individual member values that occur at different spatial locations (low spatial agreement), or small member values at the same location (high spatial agreement). An example of high spatial agreement for low mean cloud fractions is seen around a height of 1.5km in Figure 6.14d.

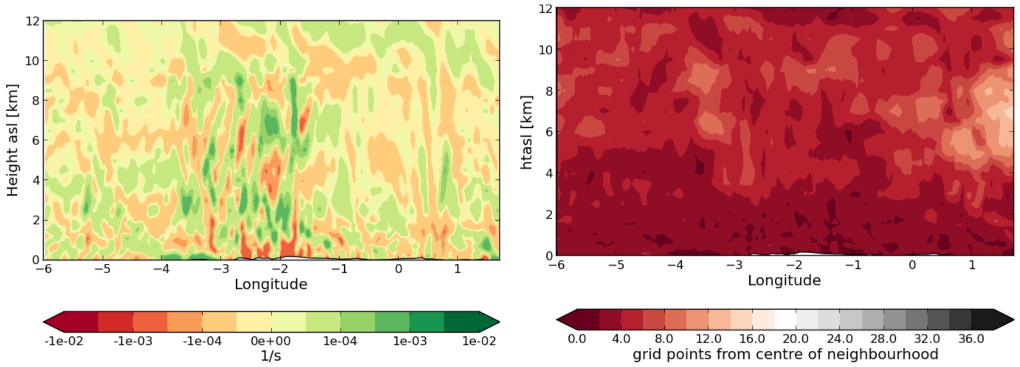
6.6.2 Horizontal divergence

As a derivative quantity of the forecast horizontal winds, the horizontal divergence has high spatial variability. This, and the link between horizontal divergence, vertical motion and convection, make horizontal divergence an important variable to consider when assessing spatial uncertainty. Note that, in this section, the term divergence always refers to the horizontal component only. Divergence, unlike precipitation and cloud fraction, presents a problem for the agreement scale calculation because it can be both positive and negative. Therefore, as the method of calculating the $S_{ij}^{A(\overline{mm})}$ assumes positive fields (Section 6.3.2), it is necessary to consider separately the positive and negative parts of the divergence field. This is done by setting either all the negative, or all the positive, values in the divergence field to zero before calculating the $S_{ij}^{A(\overline{mm})}$. Thus the $S_{ij}^{A(\overline{mm})}$ are calculated twice for the divergence field, once for the positive part (referred to as the P-divergence field) and once for the negative part (referred to as the N-divergence field).

In general, the sign of the divergence field is dependent on the horizontal location considered. For some meteorological situations this change in sign occurs over short horizontal distances, for example on the 29/07 at 15Z as shown in Figure 6.15a (ensemble mean divergence). For this case, the individual ensemble member divergence forecasts (not shown) have similar overall structures (i.e. the divergence field changes sign over similar horizontal distances) although the locations at which the divergence field changes sign differ between members. Hence, for the 29/07 at 15Z, all ensemble members will have P-divergence and N-divergence fields that alternate between values that are non-zero, and values that are zero, over small horizontal distances. As both the P-divergence and N-divergence fields change sign over small horizontal distances, a large amount of overlap is seen between the ensemble member fields for both P-divergence and N-divergence when considered over small neighbourhood sizes. Thus, small $S_{ij}^{A(\overline{mm})}$ are obtained for both P-divergence and N-divergence. An example is shown in Figure 6.15b for a vertical cross section of $S_{ij}^{A(\overline{mm})}$ calculated for P-divergence at constant latitude of 51.4°N (similar conclusions can be drawn from the N-divergence $S_{ij}^{A(\overline{mm})}$).

In contrast with the 29/07 at 15Z, some cases, such as the morning of 02/08, have well defined positive and negative layers in the divergence field, which extend over large horizontal distances. An example is shown in Figure 6.16 for vertical cross sections at a fixed longitude of 0.6°E on the 02/08 at 09Z. The layers where the divergence field is positive and negative are clearly visible from the ensemble mean field (Figure 6.16a), particularly to the south (left) of the cross section where convection had not yet developed.

The $S_{ij}^{A(\overline{mm})}$ for P-divergence and N-divergence are shown in Figure 6.16b and Figure 6.16d respectively. Consider locations in the cross sections where the ensemble mean



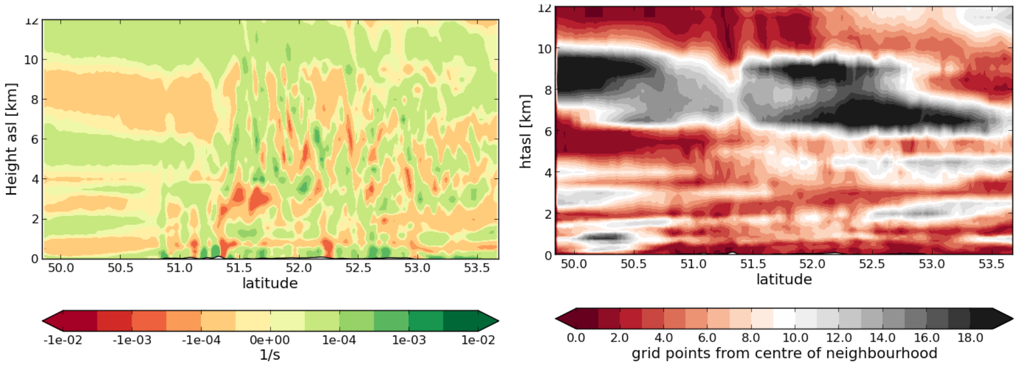
(a) 29/07 15Z Ensemble mean divergence

(b) 29/07 15Z P-divergence $S_{ij}^{A(\overline{mm})}$

Figure 6.15: Vertical cross sections for (a) mean horizontal divergence and (b) horizontal P-divergence $S_{ij}^{A(\overline{mm})}$ at 15Z on the 29/07 for a constant latitude of 51.4°N.

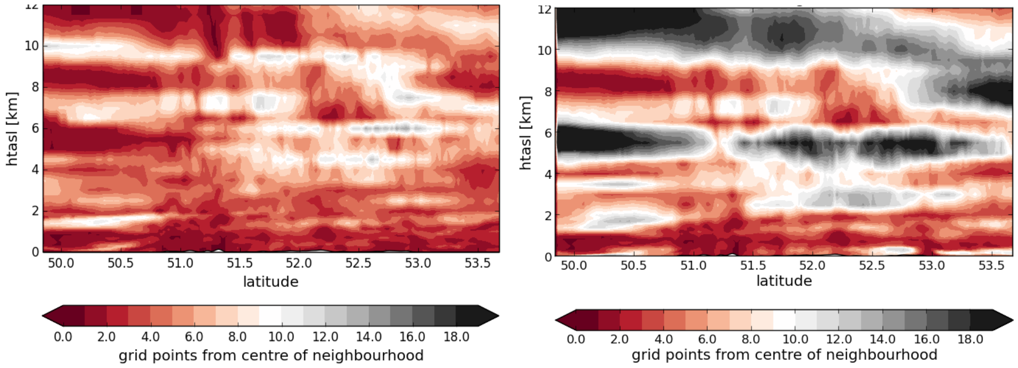
divergence field (Figure 6.16a) is positive. At these locations small $S_{ij}^{A(\overline{mm})}$ (2-6 grid points) are seen for P-divergence (Figure 6.16b). This makes sense as, within positive layers of the mean divergence field, there is high overlap in P-divergence, giving small P-divergence $S_{ij}^{A(\overline{mm})}$. Conversely, at these locations where the ensemble mean divergence field is positive, the N-divergence $S_{ij}^{A(\overline{mm})}$ (Figure 6.16d) are large (above 12 grid points). To obtain N-divergence fields that are sufficiently similar to meet the agreement scale criterion (Equation 6.2), larger horizontal scales must be considered so that locations with non-zero N-divergence are also included. Similar arguments apply for negative layers of ensemble mean divergence (where we have small large P-divergence $S_{ij}^{A(\overline{mm})}$ in Figure 6.16b and small N-divergence $S_{ij}^{A(\overline{mm})}$ in Figure 6.16d).

To obtain information about the total divergence field (i.e. both positive and negative parts of the divergence field) it is necessary to consider the $S_{ij}^{A(\overline{mm})}$ for both P-divergence and N-divergence. For cases such as the 29/07 (Figure 6.15) both components show similar scales, and hence both P-divergence and N-divergence $S_{ij}^{A(\overline{mm})}$ represent the spatial uncertainty in the total divergence. However, for cases such as the 02/08 (Figure 6.16) the $S_{ij}^{A(\overline{mm})}$ can be very different between the P-divergence and N-divergence as discussed above. In this situation the smaller of the two $S_{ij}^{A(\overline{mm})}$ results at each location is taken to be representative of the spatial uncertainty in the total divergence field. This is justified by considering the overall characteristics of the divergence field in this region: a region with small $S_{ij}^{A(\overline{mm})}$ for P-divergence/N-divergence will have divergence member forecasts that are predominantly positive/negative and hence the overall spatial uncertainty



(a) 02/08 09Z Ensemble mean divergence

(b) 02/08 09Z P-divergence $S_{ij}^{A(\overline{mm})}$



(c) 02/08 09Z minimum of P-divergence and N-divergence $S_{ij}^{A(\overline{mm})}$

(d) 02/08 09Z N-divergence $S_{ij}^{A(\overline{mm})}$

Figure 6.16: Vertical cross sections at a constant longitude of 0.6°E on the 02/08 at 09Z for (a) the ensemble mean horizontal divergence, (b) horizontal P-divergence $S_{ij}^{A(\overline{mm})}$ and (d) horizontal N-divergence $S_{ij}^{A(\overline{mm})}$. In (c) the minimum of the P-divergence and N-divergence $S_{ij}^{A(\overline{mm})}$ (i.e the minimum of sub-figures (b) and (d)) is taken at each point in the cross section.

will be determined by the P-divergence/N-divergence $S_{ij}^{A(\overline{mm})}$. Figure 6.16c shows the total divergence $S_{ij}^{A(\overline{mm})}$ for the P-divergence and N-divergence $S_{ij}^{A(\overline{mm})}$ shown in Figures 6.16b and d.

6.6.3 Temperature, specific humidity and horizontal wind speed

The variables temperature, specific humidity and horizontal wind speed are more spatially homogeneous, and vary more gradually, than cloud fraction, horizontal divergence, and rain rates. Hence, a large amount of overlap is seen between the different ensemble members. This leads to small $S_{ij}^{A(\overline{mm})}$ for these variables, and means that the agreement scales do not relate directly to the spatial predictability. For example, temperature $S_{ij}^{A(\overline{mm})}$ are always less than 1 grid point showing that there is high agreement between the fields even on the grid scale. However, as it is known that it is not possible to resolve features at the grid scale (e.g. Harris et al., 2001; Nogueira and Barros, 2014), these $S_{ij}^{A(\overline{mm})}$ are not considered to be realistic. In these situations, a minimum $S_{ij}^{A(\overline{mm})}$ of 2 grid points is taken. This gives a total neighbourhood width of 5 grid points, the minimum number of points required to resolve a wave, and is recommended as the minimum resolvable scale in Cullen and Brown (2009).

There are, of course, situations where the fields are more spatially variable, and larger $S_{ij}^{A(\overline{mm})}$ are seen. Examples include the specific humidity field surrounding a dry intrusion or around the tropopause, or for the horizontal wind speed field close to the surface. An interesting avenue of future investigation would be to assess whether, by modifying the $S_{ij}^{A(\overline{mm})}$ calculation, more information can be obtained about the location-dependent spatial uncertainty for large-scale smoothly varying fields. For example, a more stringent matching condition, or different normalisation method, could be used.

6.6.4 Summary values for 3D scales

In this section examples have been given of the $S_{ij}^{A(\overline{mm})}$, calculated horizontally, for the 3D variables cloud fraction, horizontal divergence, temperature, specific humidity and horizontal wind speed. It has been demonstrated that different variables have different spatial characteristics, with the $S_{ij}^{A(\overline{mm})}$ reflecting both horizontal and vertical physical structures. For all the variables considered here, the $S_{ij}^{A(\overline{mm})}$ showed vertically consistent structures (i.e representative of the meteorological situation and not noisy or broken up), despite the $S_{ij}^{A(\overline{mm})}$ being calculated separately at each vertical level. Due to this continuity in the vertical, similar $S_{ij}^{A(\overline{mm})}$ were often seen through large vertical depths in the atmosphere, e.g. cloud fraction $S_{ij}^{A(\overline{mm})}$ through the depth of the cloud, or specific humidity in the troposphere. This allows the 'overall' $S_{ij}^{A(\overline{mm})}$ for each variable over the domain (including both the horizontal and vertical) to be determined. These summary values were determined qualitatively from the examination of all forecast lead times and a selection of cross sections for each case (the cross section locations shown by the dotted black lines in Figures 5.1, 5.2 and 5.3). The summary $S_{ij}^{A(\overline{mm})}$ values are given in Table 6.2. A qualitative approach was used here to allow the characteristics of the different variables to be considered; for example, cloud fraction scales were only considered in regions of cloud. Of course there are situations where the $S_{ij}^{A(\overline{mm})}$ differ from the scales in Table 6.2: these scales are intended simply to provide an indication of the typical behaviour of the different variables. In Chapter 7 the $S_{ij}^{A(\overline{mm})}$ listed in Table 6.2 will be considered further in order to define suitable spatial scales for the calculation of multivariate correlations.

6.7 Chapter discussion and conclusions

This chapter has presented, with operational and idealised examples, a new spatial method for characterisation and evaluation of the location-dependent spatial agreement between members in convective-scale ensembles. In particular, this chapter addresses three research questions relating to the definition, meaning, and application of the scales over which members of a convective scale ensemble agree (thesis questions 2 and 3). The conclusions relating to each research question will be discussed in turn.

1. Over what spatial scales should the ensemble be interpreted and evaluated? (Thesis question 2; Section 1.2)

In Section 6.3 a new spatial method was presented for the characterisation and evaluation of the local spatial agreement between members in convective-scale ensembles. Based on a neighbourhood approach, the scales (neighbourhood sizes) over which ensemble members reached a specified level of agreement ($S_{ij}^{A(\overline{mm})}$) were calculated, at each grid point in the domain, to give a measure of location-dependent believable scales for an ensemble forecast. These are the location-dependent scales at which the ensemble members become sufficiently similar so that the forecast forms useful, trustworthy guidance. A method was also presented to verify the $S_{ij}^{A(\overline{mm})}$ by using radar observations to calculate the location-dependent ensemble skillful scales, $S_{ij}^{A(\overline{mo})}$. The interpretation assumes that differences between fields over this neighbourhood represent the spatial uncertainties (or errors) in the forecast. This assumption is good for small neighbourhoods, but becomes less valid as the neighbourhood size increases: events far apart in two different forecasts become more likely to represent different events rather than large displacement errors. This should be kept in mind when interpreting the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$.

To calculate the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$, ensemble members were compared, either pairwise against each other, or against observations. (This comparison method was also

Table 6.2: Summary values of the $S_{ij}^{A(\overline{mm})}$ for the different variables considered.

Variable	$S_{ij}^{A(\overline{mm})}$ **[grid points]**	**Features**
Cloud fraction	4	Low $S_{ij}^{A(\overline{mm})}$ in regions of cloud, rapid increase in $S_{ij}^{A(\overline{mm})}$ at edges of cloudy region.
Horizontal divergence	4	$S_{ij}^{A(\overline{mm})}$ of 4 to 6 grid points in regions that were, on average, divergent. Regions that were on average convergent had higher $S_{ij}^{A(\overline{mm})}$ ($S_{ij}^{A(\overline{mm})} \approx 20$)
Horizontal convergence	4	$S_{ij}^{A(\overline{mm})}$ of 4 to 6 grid points in regions that were, on average, convergent. Regions that were on average divergent had higher $S_{ij}^{A(\overline{mm})}$ ($S_{ij}^{A(\overline{mm})} \approx 20$)
Temperature	2	$S_{ij}^{A(\overline{mm})}$ less than 2 for all cases and times. $S_{ij}^{A(\overline{mm})}$ of 2 grid points used as minimum resolvable scale.
Specific humidity	2 to 6	Generally $S_{ij}^{A(\overline{mm})}$ =2. $S_{ij}^{A(\overline{mm})}$ of 6 –10 grid points were seen in regions of dry tropospheric air.
Horizontal wind speed	2 to 4	Generally $S_{ij}^{A(\overline{mm})}$ less than 2; $S_{ij}^{A(\overline{mm})}$ of 2 grid points used as minimum resolvable scale. In regions of convection, or near the surface, $S_{ij}^{A(\overline{mm})}$ of 4-6 are seen.

employed in Chapter 4 to calculate domain-wide ensemble spread and skill.) At each grid point in the domain, the agreement scale between the fields (i.e the believable scale for comparison of two forecasts, or the skillful scale for the comparison of forecasts and observations) was defined as the minimum neighbourhood size over which the fields were deemed to be acceptably similar. To decide whether the forecasts were acceptably similar, a criterion was defined based on two predetermined parameters. The first, α, controls the acceptable fractional difference between the fields, and the second, S_{lim}, is a fixed maximum scale at which the forecasts are always deemed to be sufficiently similar. For the examples presented in this chapter the values $\alpha = 0.5$ and $S_{\text{lim}} = 80$ were used: other values could also be chosen to give a more, or less, stringent criterion. Thus, the required level of agreement (or skill) is not fixed, and may be determined from the user's requirements.

In formulating the agreement scales, the aim was to present a simple, generally applicable, method of quantifying forecast differences. These measures are not designed to distinguish between temporal, amplitude, and structural components of forecast uncertainty and error. Other methods (such as those discussed in Gilleland et al. (2009)) do attempt to provide such information for the verification of high resolution deterministic forecasts, and could be developed for application to ensemble systems. This information would be complementary to that obtained using the methods presented in this chapter.

To investigate the vertical distribution of horizontal spatial differences, the $S_{ij}^{A(\overline{mm})}$ were calculated at different model levels for the 3D variables cloud fraction, horizontal divergence, horizontal wind speed, specific humidity and temperature. Despite the $S_{ij}^{A(\overline{mm})}$ being calculated separately at each vertical level, vertically consistent structures were found for all variables. This gives some confidence in the methods applied here, and will be important for Chapter 7, where vertical correlations are computed, taking horizontal uncertainties into account.

2. How can these spatial scales be defined? (Thesis question 2; Section 1.2)

A simple idealised system was created to investigate the properties of the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$. Each individual ensemble member, and the observations, were represented by a circular blob of rain, randomly positioned within a square region. Using this simple setup, it was shown that the $S_{ij}^{A(\overline{mm})}$ successfully represent spatial differences with larger spatial differences leading to larger $S_{ij}^{A(\overline{mm})}$. The method was found to be robust to changes in the number of ensemble members and to the position of the square rain region within the domain. Note that the effects of a non-Gaussian precipitation distribution, or bias between the ensemble and observations were not tested. The idealised ensemble was further used to asses the utility of comparing the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ to investigate the spatial spread-skill relationship of the ensemble. Through comparing the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ it was possible to differentiate between pre-determined scenarios in which the synthetic precipitation is set up to be either well spread, over spread, or under spread spatially. The spatial spread-skill relationship was visualised through histograms of all agreement scale data, and using binned scatter plots. It was found that binned scatter plots provide a particularly useful method for assessing the spatial spread-skill properties because the location-dependent character of convective-scale predictability is respected.

To demonstrate the utility of these techniques as an investigation tool for operational ensemble systems, the hourly instantaneous rain rate $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ were calculated for the six convective case studies introduced in Chapter 5. Maps of the $S_{ij}^{A(\overline{mm})}$ depicted the different levels of spatial predictability across the cases. For example, cases where precipitation was strongly linked to convergence along the SW peninsula showed high levels of spatial predictability with local $S_{ij}^{A(\overline{mm})}$ of less than 10 grid points. This high spatial predictability is expected from the topographic influence for these cases. In contrast, other cases, such as the 02/08, showed that precipitation could be highly unpredictable spatially.

3. What information does this spatial approach provide for forecasting? (Thesis question 3; Section 1.2)

Used in conjunction with a single ensemble member, or deterministic forecast, the $S_{ij}^{A(\overline{mm})}$ provides a useful visualisation for forecasting. The rainfall structures themselves can be viewed from an individual model run (perhaps the control) and the $S_{ij}^{A(\overline{mm})}$ map can be used to view the spatial uncertainty in that rainfall given by the ensemble. This provides a method of quickly assessing the spatial predictability obtained from the ensemble. It gives a more physically meaningful view of ensemble-member differences than using grid point measures, for example, the variance at each grid point.

To demonstrate how the location-dependent agreement scales can be used to diagnose ensemble performance, the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ were compared for the six summer convective cases. Note that the aim was to provide concrete examples of how these techniques can be applied and interpreted, not to provide a statistical verification of the operational ensemble system. It was found that, as well as having different levels of spatial agreement, the different cases showed different spatial spread-skill relationships. Poor spatial spread-skill consistency, measured by larger differences between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$, could be linked to differences between the model and observations, such as a timing error or precipitation incorrectly forecast by the model. For these six convective cases, the spatial spread-skill relationship improved in the afternoon, suggesting that it was the spatial characteristics during precipitation initiation that were most difficult for the model to handle in these instances. Once established, precipitation occurred for a number of hours and the spatial spread-skill improved. Through comparing the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$, these features of the ensemble performance were easily identified. This suggests that the agreement scales would provide a valuable diagnostic for characterising the spatial ensemble performance. In Chapter 8 such an investigation is conducted to investigate the performance of MOGREPS-UK for June, July and August 2013. The agreement scales could also be used to assess the impact of changes to the forecasting system, for example the use of stochastic increments to model systematic initiation uncertainties (e.g. Leoncini et al., 2010).

In this chapter the analysis of location-dependent spatial spread-skill focused on rain rates. Rain rates were selected for this study to avoid any temporal smoothing from using precipitation accumulations, and hence to focus on the spatial features. Of course, the methods presented here could also be used to evaluate precipitation accumulations. More generally, although precipitation forecasts are a key application of these methods (due to their high spatial uncertainty and the availability of radar observations for verification), the comparison of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ is equally applicable to other single-signed meteorological fields where gridded observations, or simulated observations, are available: for example, from satellite imagery.

Section 6.6 presented $S_{ij}^{A(\overline{mm})}$ for 3D variables at different vertical levels. These $S_{ij}^{A(\overline{mm})}$ were found to represent physical structures in the forecast, and differed between cases, times, and variables. Overall, it was found that smoothly varying, large scale fields, such as temperature showed agreement close to the grid scale. In contrast, other variables with variability on small scales, such as cloud fraction and positive divergence, had larger $S_{ij}^{A(\overline{mm})}$. This highlights the importance of taking multivariate aspects into account when calculating the spatial agreement between forecasts. Inter-variable relationships will be further discussed in Chapter 7.

Chapter 7

Application of the spatial approach to the calculation of correlations

7.1 Introduction

Chapter 6 introduced a new method for quantifying the scale (neighbourhood size), at each grid point in the domain, over which two fields (either forecast or observed), reach a required level of similarity. When calculated for ensemble members from a well spread ensemble system, these agreement scales, denoted as $S_{ij}^{A(\overline{mm})}$, give a measure of the smallest believable scale for the ensemble at a given point in the domain. These are the scales over which features in the forecast would be expected to occur, and over which the forecast should be evaluated. Hence, for the correct interpretation of ensemble output, forecasts should be considered on these scales (greater than the grid scale).

In Chapter 6 this reasoning is applied to the calculation of multivariate correlations. For the six summer convective cases introduced in Chapter 5, the $S_{ij}^{A(\overline{mm})}$ (calculated in Chapter 6) are used to define suitable neighbourhood sizes over which the correlations should be calculated. As discussed in Chapter 2, multivariate correlations are commonly used in data assimilation to control the propagation of information through space and between variables. Here these techniques are applied to the physical and meteorological interpretation of convective scale ensemble forecasts. In particular this chapter aims to answer the following questions:

1. How can correlations in the vertical be used to allow physical structures and multivariate relationships from the ensemble to be usefully summarised? To what extent do these correlations reveal properties of the convection and the convective environment? (Thesis question 4; Section 1.2)

2. What is the dependence of vertical correlation structure on the horizontal scales

used? How does this relate to the scales obtained from the spatial methods? (Thesis question 5; Section 1.2)

3. In what ways might these methods be useful in other areas such as data assimilation? (Thesis question 6; Section 1.2)

This chapter is structured as follows. Section 7.2 introduces the method used to calculate correlations on scales greater than the grid scale, whilst preserving vertical structures. In this chapter we focus on vertical correlations, that is correlations between variables at different vertical levels. Due to computational constraints vertical correlations were only calculated at preselected grid points: the selection of these grid points, and a summary of all the points considered, is given in Section 7.2. Section 7.3 discusses the interpretation of univariate and multivariate correlations and provides examples that relate correlation structures to individual member behaviour.

Vertical correlations from the six summer convective cases are presented in Sections 7.4 and 7.5, first for one variable (vertical correlations, VCs) and then for vertical correlations between different variables (vertical cross correlations, VCCs). Section 7.6 then discusses temporal correlations (TCs) between surface rain rates and three dimensional variables. The results presented in Sections 7.4 to 7.6 are calculated using the $S_{ij}^{A(\overline{mm})}$ to allow the horizontal uncertainty to be taken into account, and to address Question 1. In Section 7.7 results are presented for correlations calculated at different horizontal scales (i.e at scales other than the $S_{ij}^{A(\overline{mm})}$) and Question 2 is considered. The conclusions of this chapter are presented and discussed in Section 7.8 where Question 3 is discussed and Questions 1 and 2 are revisited.

7.2 Method

An introduction to correlations and covariances was given in Chapter 2. This section further discusses correlations between two fields, and the methods that can be used to include information from different horizontal points in the calculation of VCs. This is particularly important for the calculation of correlations at the convective scale as it allows horizontal uncertainty to be taken into account.

7.2.1 Different covariance formulations

In this section three methods are presented for calculating correlations using information from all points within a horizontal area, and their relative merits discussed. For clarity of argument, the discussion focuses on covariances: the comments are equally applicable to correlations. Writing out the sums in full, and ignoring the $(N-1)$ correction due to sampling error (as it cancels out for the calculation of correlations), the covariance over an ensemble of N_e members, between two quantities S_{xy} and T_{xy} at a horizontal point $(x,y) = (p,q)$, can be written as

$$cov(S_{pq}, T_{pq}) = \frac{1}{N_e} \sum_{l=1}^{N_e} \left(S_{pql} - S_{pq}^E \right) \left(T_{pql} - T_{pq}^E \right), \tag{7.1}$$

where here, and for the remainder of this section, the superscript E indicates an average over the ensemble members. Three different methods are considered for calculating covariances over an horizontal area of N_x by N_y grid points. Note that, in general, this area is centred around a specified point, say point (p,q). Such an area (here of 5 by 5 grid points) is shown schematically in Figure 7.1 for two quantities S and T. To focus on the key features, all references to the location of point (p,q) are removed from the following equations. Thus the equations in this section only consider the selected neighbourhood (area).

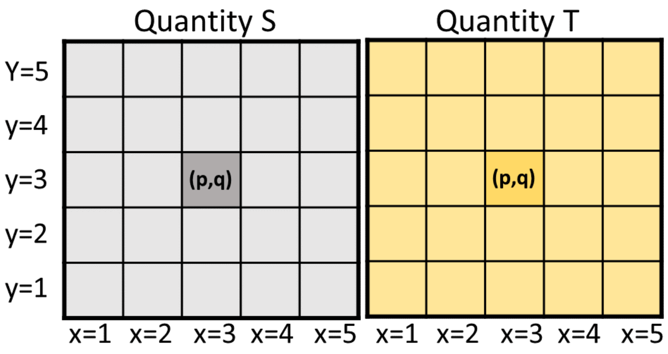


Figure 7.1: Schematic showing the grid for two example fields S and T. An area of 5 by 5 grid points is shown, centred around the point (p,q).

First we consider the covariances calculated separately at each point $x = 1, 2, ...N_x$, $y = 1, 2, ...N_y$ within the selected area N_x by N_y, and then averaged over that area. In the example shown schematically in Figure 7.1, the individual covariances would be calculated separately at each of the 25 grid points, then the average taken over these to get the averaged covariance. In the general case this averaged covariance $\overline{cov(S,T)}$ is given by

$$\overline{cov(S,T)} = \frac{1}{N_e N_x N_y} \sum_{i=1}^{N_x N_y} \sum_{l=1}^{N_e} \left(S_{il} - S_i^E\right)\left(T_{il} - T_i^E\right). \tag{7.2}$$

Alternatively, the area average can be taken for each ensemble member before the covariance is calculated. In this case, for the schematic example in Figure 7.1, the averages are first taken over all 25 grid points for each ensemble member forecast of quantities S and T. For an ensemble member l, these averages are denoted as $\overline{S_l}$, $\overline{T_l}$. Generally, the average would be calculated over an area of N_x by N_y grid points. The covariance is then calculated using the area average from each ensemble member. Mathematically, this covariance of the average $cov(S,T)_{av}$ is given by

$$cov(S,T)_{av} = \frac{1}{N_e} \sum_{l=1}^{N_e} \left(\overline{S_l} - \overline{S^E}\right)\left(\overline{T_l} - \overline{T^E}\right). \tag{7.3}$$

Finally, we can consider all the members within the area $N_x \times N_y$ to be part of the ensemble before calculating the covariances. Hence, for the example in Figure 7.1, the covariances will be calculated from an augmented ensemble with 25 times N_e members. In general, this covariances from an augmented ensemble of $N_x \times N_y \times N_e$ members, $cov(S,T)_{aug}$, is given by

$$cov(S,T)_{aug} = \frac{1}{N_e N_x N_y} \sum_{i=1}^{N_x N_y} \sum_{l=1}^{N_e} \left(S_{il} - \overline{S^E}\right)\left(T_{il} - \overline{T^E}\right). \tag{7.4}$$

Equations 7.2, 7.3 and 7.4 are different in two key ways.

1. Equations 7.3 and 7.4 consider the difference between individual members and the mean calculated over all members, and all grid points within the neighbourhood (i.e over $N_x \times N_y \times N_e$ ensemble members). Hence, all members in the neighbourhood are used to form the reference point and are given equal weight. Equation 7.2 instead uses the mean at each point separately, giving a different reference point at each horizontal location. Of course, if the N_e members at each individual grid

point were a good sample of the true PDF at that point, and the same PDF was true over the area $N_x \times N_y$, then the point mean and neighbourhood mean would be identical. That is to say, equations 7.2 and 7.4 would be the same within sampling error.

2. Equations 7.2 and 7.4 only make comparisons between fields S and T at the same points (x, y). This contrasts with Equation 7.3 which also includes comparisons of S and T at different points. This has important implications, for instance when S and T represent fields at different vertical levels; for example if S is a surface precipitation field, and T is cloud fraction at 10 km. In considering only the same horizontal points in Equations 7.2 and 7.4, vertical structure in the correlations between S and T is preserved, whereas when the points (x, y) are mixed (as in Equation 7.3) vertical structures are smoothed out.

Given that it is known that the ensemble severely under samples the PDF at the grid scale, and one of the aims of this work is to investigate vertical structure, $cov(S, T)_{aug}$ is the preferred method of calculating covariances. Note that it has been implicitly assumed that the vertical structure is correctly represented by the model, with errors due only to under sampling in the horizontal. Of course there are errors in the vertical, and these should be investigated in future work. However, due to scale lengths being larger in the horizontal than the vertical, it is believed that horizontal errors dominate.

7.2.2 Sample size and weighting

The sampling error in the correlation coefficient ρ, calculated from a sample of N independent points, can be estimated as

$$\sigma_\rho(N) = \frac{1-\rho^2}{\sqrt{N-1}} \tag{7.5}$$

(Barlow, 1989). Hence, for a 12 member ensemble, true correlations can only be separated from sampling error if they have values greater than $\sigma_\rho(12) = 0.27$. To allow the direct comparison of correlations calculated using different sample sizes, correlations with values less than 0.27 are not shown on any plots in this chapter.

Using the methods described in Section 7.2.1 the sample size used to calculate the correlations is artificially increased by including members from neighbouring grid points into the ensemble (a technique currently used to generate probabilistic output from deterministic forecasts e.g. Theis et al. (2005)). The simplest method treats all points within the neighbourhood equally. This assumes that the forecasts at all points in the neighbourhood are equally likely, irrespective of their distance from the original point at the centre of the neighbourhood. For the calculation of spatial scales in Chapter 6 this was assumed: forecast structures at scales below the $S_{ij}^{A(\overline{mm})}$ can be moved around spatially without affecting the overall message from the forecast. Hence, this method is also preferred for the calculation of correlations to preserve links with the $S_{ij}^{A(\overline{mm})}$. However, it is also possible to apply a weighting over the neighbourhood to account for the increased differences expected at larger horizontal separations. The effects of weighting were tested for the 17/07 and 03/08 case studies by applying a normalised Gaussian weighting over the neighbourhood before calculating correlations. This will be discussed in Section 7.7.

In the correlations calculation, the neighbourhood size is defined by the $S_{ij}^{A(\overline{mm})}$ at the central point where the correlation is calculated. For 3D variables a summary scale, chosen to represent the overall behaviour of the $S_{ij}^{A(\overline{mm})}$ for that variable across the vertical range of interest, was used. The calculation of these summary scales was discussed in Chapter 6, Section 6.6.4, and the values listed in Table 6.2. For the calculation of correlations between different variables with different values of $S_{ij}^{A(\overline{mm})}$, the larger of the two $S_{ij}^{A(\overline{mm})}$ values was used to calculate the correlation. Although this method may, in some instances, result in real information being discarded, it is considered to be preferable to the alternative: interpreting unreliable ‘noisy’ results as real.

7.2.3 Point selection

To allow the vertical structure of the convection to be investigated, correlations were calculated at selected points within the convective regions. These points were selected by visual examination of the radar derived precipitation rates, and were chosen to represent different regions in the convective environment. To allow temporal correlations with rain rates to be calculated (Section 7.6) it was also necessary that it rained at the selected points for several hours. For each case 2 to 3 points were considered to allow the structures at different locations to be compared. The positions of these points are listed in Table 7.1, and shown by red circles over the radar data in Figure 7.2.

Case	Point 1	Point 2	Point 3
17/07	51.7°N, -0.7°E		
23/07	52.1°N, -1.9°E	52.6°N, -0.6°E	
27/07	51.0°N, -2.0°E	53.0°N, -2.0°E	
29/07	50.9°N, -3.0°E	51.4°N, -2.0°E	52.7°N, -2.0°E
02/08	50.8°N, -4.5°E	52.1°N, 0.6°E	52.1°N, -2.4°E
03/08	50.8°N, -4.4°E	51.1°N, -3.0°E	

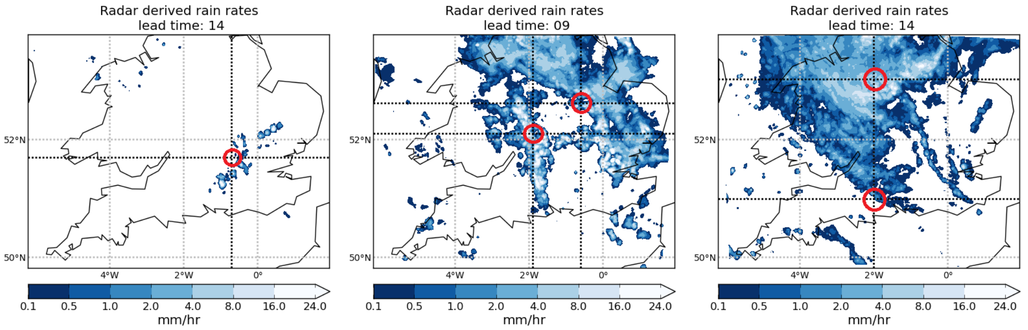
Table 7.1: Summary of the points used for the calculation of correlations.

7.3 Interpreting vertical correlations

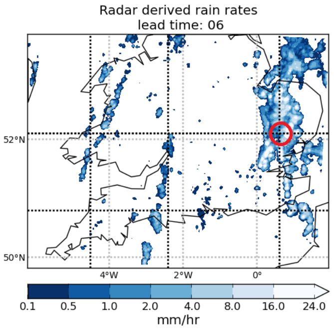
This section introduces the vertical univariate correlations (VCs) and vertical cross correlations (VCCs). Examples are taken from the case of 08/07/2011 (discussed in Chapter 4). Here we focus on the interpretation of correlations, with examples demonstrating how correlations may be understood by considering differences between ensemble members and the ensemble mean. The meteorological and physical processes leading to these correlations will be discussed in the following sections. For all the correlations presented in this chapter, physical interpretations were reached by investigating the individual ensemble member behaviour. In particular, maps and vertical cross sections for three-dimensional variables temperature, specific humidity, cloud fraction, horizontal wind speed and horizontal divergence, and maps and time series of surface rain rates, were visually examined.

7.3.1 Correlations for one variable

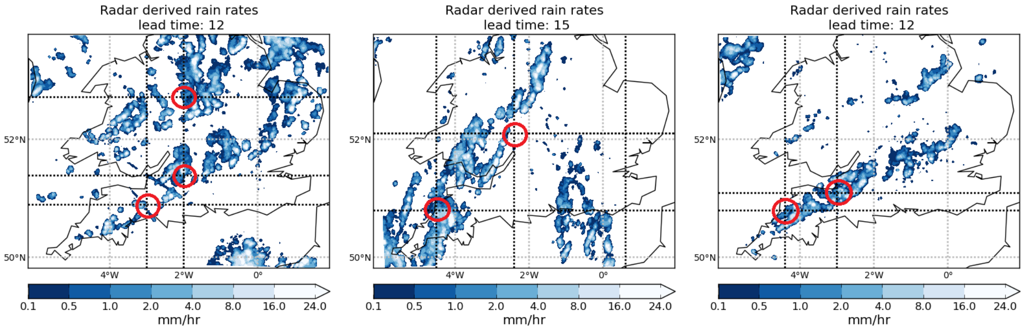
Firstly an example is made of VCs for temperature around the tropopause. Figure 7.3a shows VCs for temperature for the 08/07/2011 at 09Z. Along the diagonal correlations of one are obtained as expected (the ensemble is being correlated with itself at the same location). Away from the diagonal the correlation magnitude decreases as the ensemble members are less correlated at larger separations. Notice that Figure 7.3a is symmetric around the diagonal as expected. A sharp change from positive to negative correlation is often seen near the tropopause. Through examining individual member behaviour, this change in correlation sign was explained by the different member vertical temperature



(a) 17/07 17Z (b) 23/07 6Z (c) 27/07 23Z



(d) 02/08 9Z



(e) 29/07 15Z (f) 02/08 18Z (g) 03/08 15Z

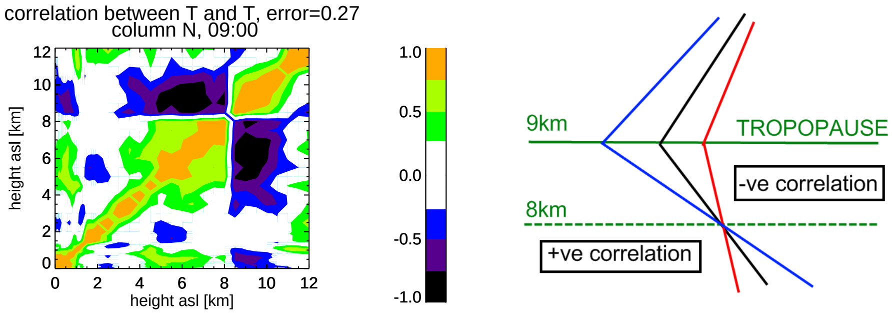
Figure 7.2: Radarnet radar derived rain rates for the six cases considered at times when convection occurred. The points where correlations were calculated are shown by the red circles. Note that these circles are for illustration only and do not represent the spatial scales considered. The 02/08 case is shown twice (subfigures d at 09Z and f at 18Z), to show points for two convective events of interest.

profiles: those members that are warmer at low levels have a stronger temperature gradient and colder tropopause, whereas the colder members at low levels have a weaker gradient and warmer tropopause. This is shown schematically in Figure 7.3b. From this we can conclude that the large change in correlation at 8km, 1km below the tropopause, is due to differences in the vertical gradients of temperature below the tropopause. Note that other ensemble configurations, for example where ensemble members have different tropopause heights, could give similar correlation patterns to that shown in Figure 7.3a.

Similar geometric arguments can be used to explain changes in correlation sign for other variables, for example, the strong negative region in the horizontal wind speed VCs of Figure 7.4a (again, 11/07/2011 at 09Z). Figure 7.4b schematically depicts the member behaviour at this time and shows how the change in correlation sign is due to different realisations of the upper level jet strength. Members with a weaker jet have stronger than average winds either side of the jet (i.e a broader jet); members with a stronger jet have weaker winds on the flanks due to a narrower jet. This suggests different regimes within the members. Note that, although Figures 7.3 and 7.4 are shown at the same point and time, it is not possible to diagnose the relationship between wind speeds and temperature from the VCs: the members that are faster/slower than average are not necessarily the same members that are warmer/colder than average at the tropopause. To compare different variables is is necessary to consider the VCCs.

7.3.2 Vertical correlations between different variables (VCCs)

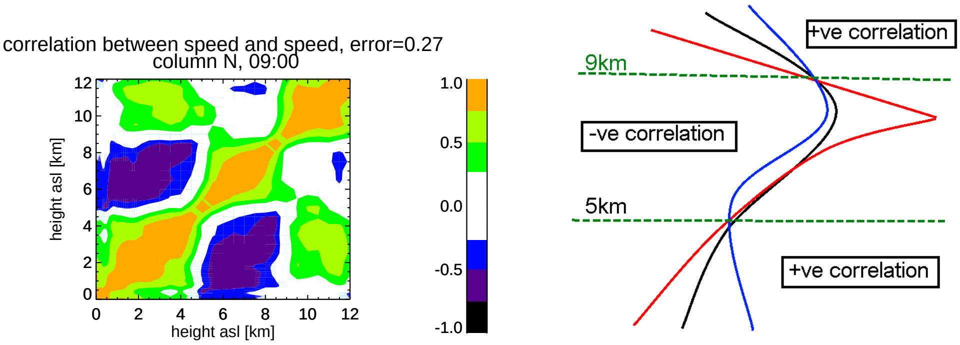
Now let us consider correlations between different variables. Often, the correlation structures of VCCs strongly resemble the vertical correlation structures of the individual variables. In a similar manner to the single variable correlations, cross correlations can be explained through interpreting the correlation changes as a change in position of members relative to the ensemble mean. An idealised example is given in Figure 7.5. In Figure 7.5a field A has positive VCs at all heights and field B has VCs that change sign at height H (positive below, negative above). Vertical cross correlations between fields A and B therefore have the same structure as the VCs for field B. If the blue and red lines of field A were reversed to give field C (Figure 7.5b) then the VCs of field C would be the same as those for field A. However, the VCCs between fields C and B would be the negative of the VCs for field B (i.e negative below height H and positive above height H). Examples of VCCs for the six cases discussed in chapters 5 and 6 are given in Section 7.5.



(a) VCs for temperature

(b) Schematic of member profiles

Figure 7.3: Vertical correlation structure of temperature profiles at the tropopause due to different temperature gradients: (a) VCs and (b) schematic representation. In the schematic the ensemble mean is represented by the black line. The red/blue lines represent members that have a warmer/colder tropopause temperature. This example was taken from 08/07/2011 at 09Z.



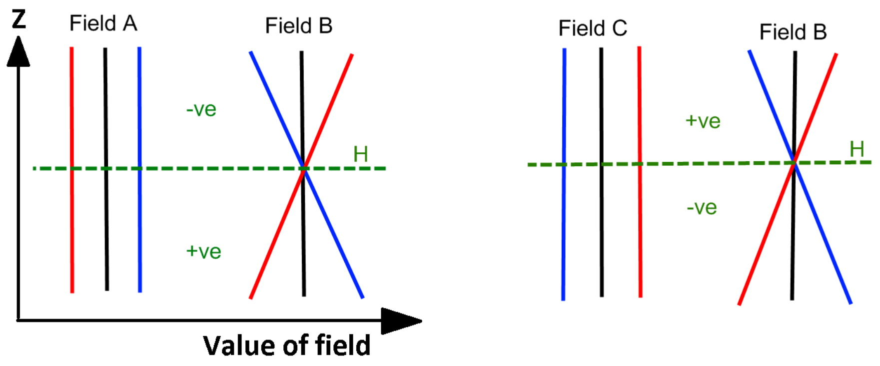
(a) VCs for horizontal wind speed

(b) Schematic

Figure 7.4: Vertical correlation structure of horizontal wind speed due different strength horizontal wind speed peaks: (a) VCs and (b) schematic representation. In the schematic the ensemble mean is represented by the black line. The red/blue lines represent members that have a faster/slower jet. This example was taken from 08/07/2011 at 09Z.

7.4 Vertical correlations (single variable)

In this section VCs, calculated using Equation 7.4, are investigated for the three dimensional variables horizontal divergence, cloud fraction, specific humidity, air temperature and horizontal wind speed. Correlations for these variables were calculated for all six cases at the points listed in Table 7.1. In the discussion below examples are given from specific cases to illustrate the range of observed correlation structures.



(a) Situation A (b) Situation B

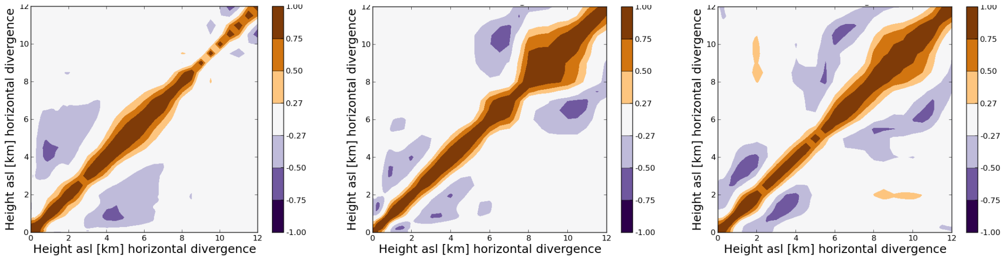
Figure 7.5: Schematic showing how VCCs between two different fields can relate to the VCs for the individual fields. The ensemble mean is represented by the black line and red/blue lines represent ensemble members either side of the ensemble mean. Height H is the vertical level at which the blue and red ensemble members of field B change from being less/greater than the ensemble mean to being greater/less than the ensemble mean.

7.4.1 Horizontal divergence

Horizontal divergence vertical correlation structures are expected be closely linked to the structure of convective cells, with convergence at the cell base and divergence around the cloud top. Horizontal divergence can also be used to indicate vertical motion with ascent/descent associated with convergence/divergence.

Vertical correlations of horizontal divergence are often characterised by an alternating correlation sign with height. This is due to the presence of convergent and divergent layers. Particularly clear examples are seen for the three peninsula convergence cases where low level convergence played a significant role in the convective forcing. Examples for the 29/07, 02/08 and 03/08 are given in Figure 7.6. For all three cases negative off diagonal correlations are seen in the convective region (up to 7km, 6km, and 5km for the 29/07, 02/08, and 03/08 respectively) as members that are more convergent/divergent near the cloud base are more divergent/convergent in the upper cloud region. This suggests that divergence VCs can be used to summarise the height of the convective cells, information that is particularly useful as it provides a meaningful summary of a physical process from all ensemble members at appropriate horizontal scales.

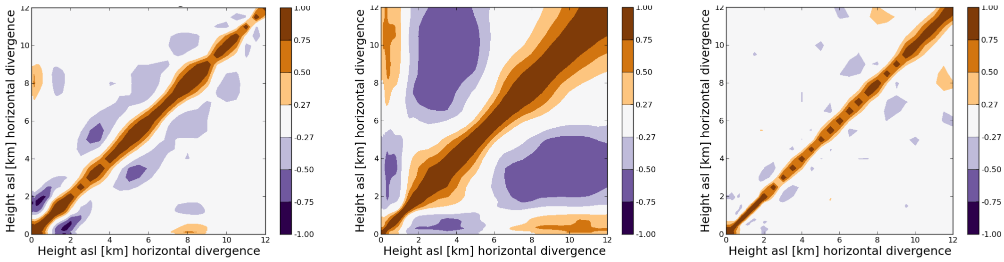
On the 29/07 correlations are mainly seen up to 6km contrasting with the 02/08 and 03/08 cases where divergence anti-correlations are also seen in the upper troposphere. It is thought that these upper level structures are linked to layers of convergence and divergence associated with large scale ascent and descent. On the 02/08 correlations are not seen between the surface and upper troposphere suggesting that the upper level divergence structure and peninsula convergence line are not directly linked. On 03/08 low magnitude correlations are seen between the lower and upper troposphere. However, as will be shown in Section 7.6, divergence at upper tropospheric levels does not correlate with surface rain rates in this case, suggesting that, similarly to the 02/08 case, the divergence VCs with levels above 4km are not linked to the peninsula convergence line.



(a) 29/07 11Z 51.4°N, -2.0°E (b) 02/08 14Z 50.8°N, -4.5°E (c) 03/08 15Z 51.1°N, -3.0°E

Figure 7.6: Divergence vertical correlations for the peninsula convergence cases: (a) 29/07, (b) 02/08 and (c) 03/08. Each case is shown at time when, and location where, convection was present. For all cases results have been calculated for a $S_{ij}^{A(\overline{mm})}$ of 4 grid points.

Divergence correlations for the deep convective cases also highlight other aspects of the model meteorology. As these cases have a variety of forcing mechanisms, greater variation is seen than for cases dominated by peninsula convergence. Example correlations are given in Figure 7.7 for the 17/07, 23/07 and 27/07. Both the 17/07 and 23/07 have off diagonal correlations extending throughout the troposphere. On the 17/07 correlations are low magnitude reflecting the high levels of uncertainty for this case. At 04Z on the 23/07 (Figure 7.7b) strong correlation structures are seen throughout the troposphere reflecting the deep convective cells that were present at this time. Later on the 23/07, when the model precipitation became increasingly wide-spread, the correlation magnitudes decreased. On the 27/07 convection occurred within an MCS. An MCS is formed of many interacting small cells, unlike the other cases which are dominated by individual cells. In particular, when the MCS is examined at the $S_{ij}^{A(\overline{mm})}$, the behaviour of a number of small cells is being considered. Hence, on the 27/07, correlations are averaged over both convergent and divergent regions, associated with several different cells. This results in weaker correlations compared to cases where only single updraughts and downdraughts are considered. This interpretation was verified by considering correlations calculated at the grid scale, where individual updraughts and downdraughts are considered. As expected, low magnitude off diagonal correlations were seen at the grid scale.

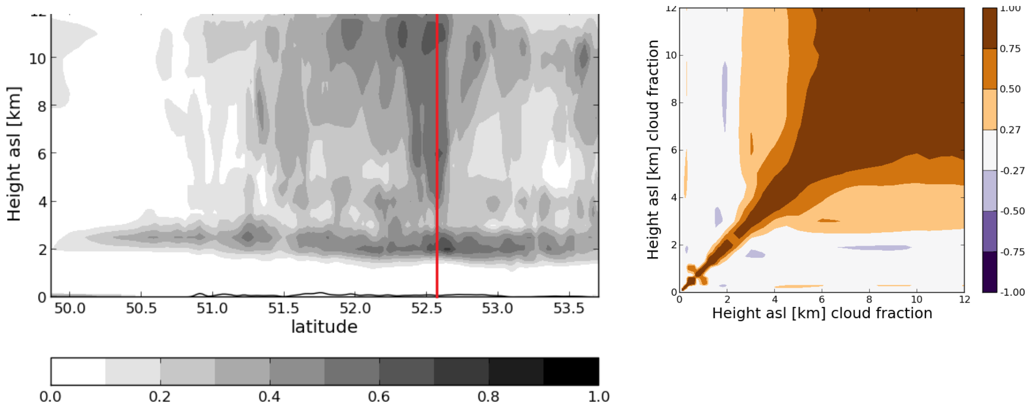


(a) 17/07 17Z 51.7°N, -0.7°E (b) 23/07 04Z 52.6°N, -0.6°E (c) 27/07 23Z 53.0°N, -2.0°E

Figure 7.7: Divergence vertical correlations for the deep convective cases (17/07, 23/07 and 27/07). Each case is shown at time when, and location where, convection was present. For all cases results have been calculated for a $S_{ij}^{A(\overline{mm})}$ of 4 grid points.

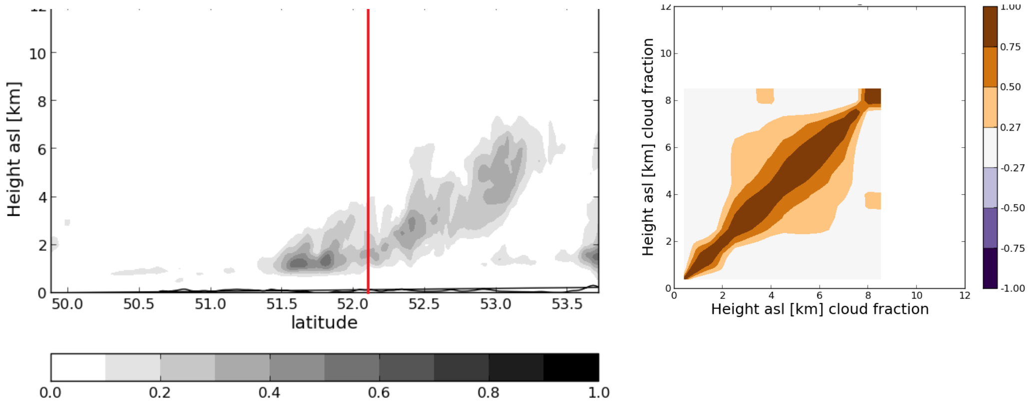
7.4.2 Cloud fraction

Cloud fraction VCs are positive within layers of convective cloud because members that have more cloud at cloud base also have more cloud throughout the cloud layer. Examples for two contrasting cases, the 23/07 and 02/08, are given in Figure 7.8 alongside ensemble mean cloud fraction vertical cross sections through the point where correlations were calculated. Note that the ensemble mean cloud fraction is shown to give an indication of the vertical levels over which the ensemble members produce cloud and does not represent the individual ensemble member behaviour. For example, on the 23/07 at 04Z, each ensemble member has narrow plumes of cloud, with a cloud fraction close to one, extending from 2 km to 12 km. However, as the exact location of these plumes is different for each member, the ensemble mean shows low cloud fractions from 2km–12 km extending over a broad horizontal area. On the 02/08 the cloud fraction fields are zero above 8km and correlations could not be calculated in this region. Below 6km, where the ensemble forms clouds, positive correlations are again seen. These results suggest that on the convective scale, at least for these six summer convective cases, information (for example in data assimilation) should be spread throughout, but not outside, the cloud layer. As the areas of positive correlations closely follow the regions of cloud these correlations, like those for horizontal divergence, give information about the convective processes produced by the model.



(a) Ensemble mean cloud fraction 23/07

(b) Correlation 23/07



(c) Ensemble mean cloud fraction 02/08

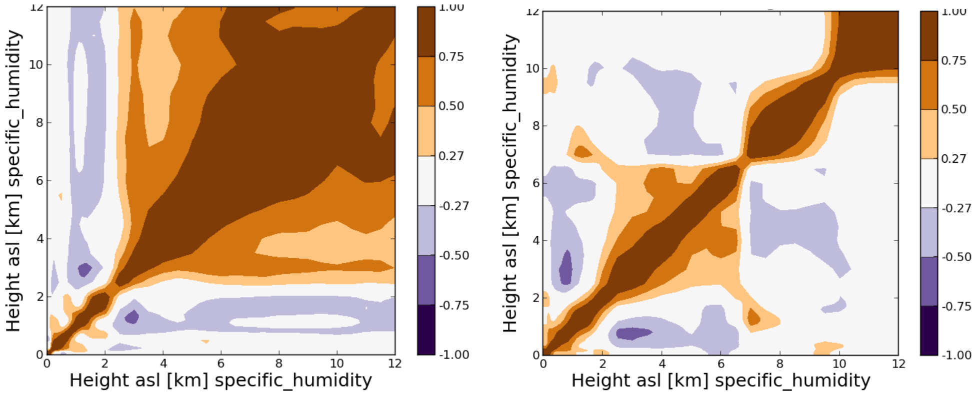
(d) Correlation 02/08

Figure 7.8: (b),(d) Cloud fraction vertical correlations for the 23/07 at 04Z 52.6°N, -0.6°E and 02/08 at 8Z 52.1°N, -2.4°E respectively. (a) and (c) show vertical cross sections of ensemble mean cloud fraction at the time when, and longitude corresponding to, the point where correlations were calculated. The vertical red line in (a,c) indicates the latitude position where correlations were calculated. For both cases results were calculated for a $S_{ij}^{A(\overline{mm})}$ of 4 grid points.

7.4.3 Specific humidity

Once convection has developed, specific humidity VCs are positive within cloud layers and closely resemble cloud fraction correlations. This is demonstrated by comparing Figure 7.9 (example specific humidity VCs for 23/07 and 02/03) to Figure 7.8 (cloud fraction VCs). This is expected physically as members with more/less cloud will be wetter/dryer. Of course, there are slight differences between the results of these two variables. In particular correlations with cloud fraction can only be obtained in regions of cloud. If only one of cloud fraction and specific humidity were to be used, these differences would have to be considered.

In addition to the positive correlations seen in cloud layers, negative off diagonal



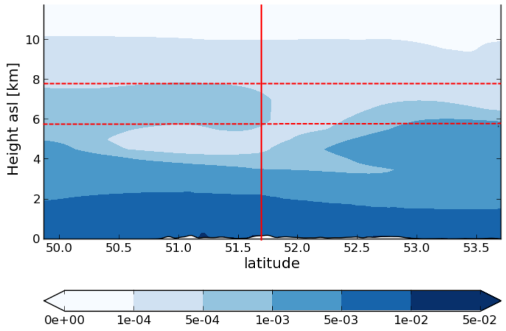
(a) 23/07 (b) 02/08

Figure 7.9: Specific humidity vertical correlations for (a) the 23/07 at 04Z 52.6°N, -0.6°E and (b) 02/08 at 8Z 52.1°N, -2.4°E. For both cases results were calculated for a $S_{ij}^{A(\overline{mm})}$ of 2 grid points.

correlations are often seen for specific humidity. These are associated with dry layers in the troposphere with members that are dryer/wetter at this level being wetter/dryer immediately above and below (i.e they have stronger humidity gradients). A particularly clear example is seen in the 17/07 case where convection occurs at the edge of a layer of dry air descended from the upper troposphere (dry intrusion). The specific humidity ensemble mean vertical cross section at 12Z clearly shows that there is a dry layer at 5km (Figure 7.10). From 07Z to 12Z the correlation structures resemble those shown in Figure 7.11a with changes in correlation sign at 5.5km and 8km. The heights of these changes can be related to the dry layer as shown in Figure 7.10. As convection develops the specific humidity increases in the convective area as moisture is transported upwards from lower to higher levels. This can seen in Figures 7.11b and 7.11c which show the specific humidity correlation structures with an area of positive correlations extending through the convective layer. At 13Z the convection reaches 8km and positive correlations also extend to this level. Above this level correlation structures agree with those from previous times as correlations are still representing the background vertical structure in the absence of convection. By 14Z convection extends up to 11km and positive correlations in specific humidity also extend up to this level.

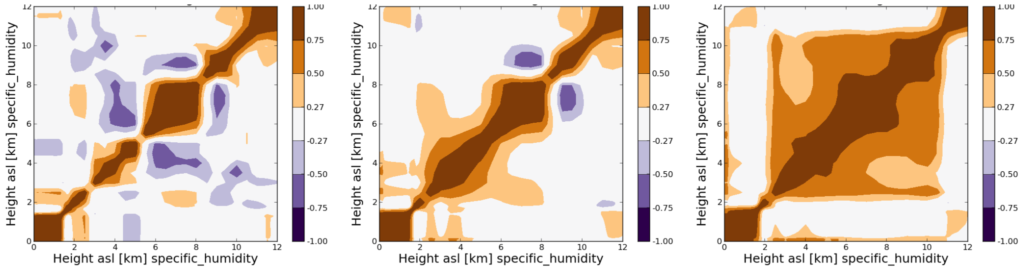
7.4.4 Temperature

Temperature VCs show rapid changes in sign between different vertical levels. This is commonly seen around the tropopause as discussed in Section 7.3 (with respect to Figure



(a)

Figure 7.10: Specific humidity vertical cross section at a constant longitude of -0.7°E on the 17/07 at 12Z. Heights where vertical correlations change sign are indicated by the red dashed lines and the latitude where correlations were calculated is indicated by the solid red line.



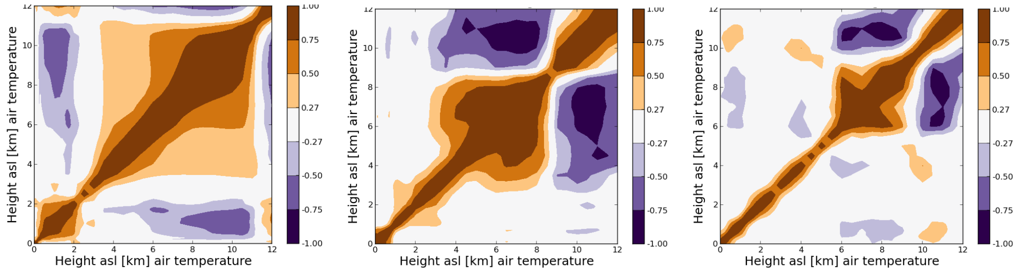
(a) 12Z

(b) 13Z

(c) 14Z

Figure 7.11: Specific humidity vertical correlations on the 17/07 for an $S_{ij}^{A(\overline{mm})}$ of 6 grid points at the point 51.7°N, -0.7°E. Three different times are shown: (a) 12Z (before storm), (b) 13Z (convection up to 8km) and (c) 14Z (convection up to 11km).

7.3): members that are warmer/colder in the upper troposphere have stronger/weaker temperature gradients and a colder/warmer tropopause. Further examples of the change in correlation sign around the tropopause are given in Figure 7.12 for the 23/07, 02/08 and 03/08. On the 23/07 correlations change sign at 3 km and 11 km corresponding to the cloud base and tropopause. For this deep convective case members that are colder/warmer near the surface are warmer/colder in the convective cloud and colder/warmer above the tropopause. Physically, this can be related to the convective structure: members with stronger convection have more latent heat release in the layer of cloud and possibly also have a stronger cold pool at low levels. Thus, the convection has stabilised the atmosphere. The two peninsula convergence cases, 02/08 and 03/08, have smaller convective cells that do not modulate the larger scale environment, and hence, for these cases, the convective structure is not seen in the temperature VCs. The change in correlation sign at the tropopause is still seen. The consistency and strength of this change in correlation sign around the tropopause suggests information should not be spread over this boundary, for example in convective scale data assimilation.

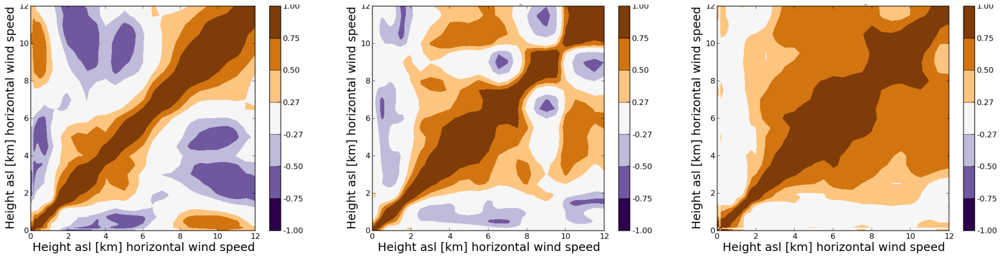


(a) 23/07 04Z 52.6°N, -0.6°E (b) 02/08 14Z 50.8°N, -4.5°E (c) 03/08 15Z 51.1°N, -3.0°E

Figure 7.12: Temperature vertical correlations for the (a) 23/07, (b) 02/08 and (c) 03/08. Each case is shown at time when, and location where, convection was present. For all cases results have been calculated for a $S_{ij}^{A(\overline{mm})}$ of 2 grid points.

7.4.5 Horizontal wind speed

In contrast to those of temperature, specific humidity, divergence and cloud fraction, VCs of horizontal wind speed do not show consistent patterns between the different cases. Three examples are given in Figure 7.13 for the 23/07, 02/08 and 03/08. On the 23/07 wind speed correlations closely resemble those of horizontal divergence (Figure 7.7b) changing sign at 2 km and 7 km as members that are faster/slower than average near the surface are slower/faster from 2 to 7 km and faster/slower above 7 km. From this it might be expected that divergence and wind speed are positively correlated, and that the wind speed correlations relate directly to the convective circulation (like those for divergence). However, as will be shown in Section 7.5 this is not the case. Horizontal wind speed correlations for 02/08 at 15Z are shown in 7.13b and show a clear change in correlation sign at 8 km and 10 km. Examination of the wind speed fields (not shown) suggests that these negative correlations are linked to different strengths of the jet: members with a slower/faster jet have faster/slower winds either side of the jet (i.e the jet may be more broad/narrow). This again shows the link between correlations and physical processes in the model, and agrees with the example given in Section 7.3. On the 02/08 negative correlations are also seen between the boundary layer and other levels throughout the troposphere because members with slower than average winds near the surface have faster than average winds in the troposphere. Finally, the 03/08 case again shows different wind speed correlation structures with positive correlations between all levels above 2 km. This shows that, for some meteorological situations, wind speed information can be spread through a large vertical extent. On 03/08 vertically stacked sloping layers of horizontal convergence and divergence were seen in the model, thought to be linked to gravity wave structures as discussed in Section 5.4.2 (Figures 5.6 and 5.7). It is believed that these gravity wave structures, extending through the tropopause, are the reason for the extended positive horizontal wind speed correlations for this case.



(a) 23/07 04Z 52.6°N, -0.6°E (b) 02/08 15Z 50.8°N, -4.5°E (c) 03/08 15Z 50.8°N, -4.4°E

Figure 7.13: Horizontal wind speed vertical correlations for (a) 23/07, (b) 02/08, and (c) 03/08. Each case is shown at time when, and location where, convection was present. For all cases results have been calculated for a $S_{ij}^{A(\overline{mm})}$ of 4 grid points.

7.5 Vertical cross correlations

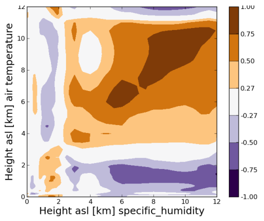
In this section VCCs (correlations between different variables at different vertical levels) are discussed. Often, the cross correlations between different variables are interrelated and show the consistency of the model meteorology. As was discussed in Section 7.3, cross correlations can also be related to the VCs of the individual variables when considered separately. The cleanest example demonstrating inter-variable relationships, from the 23/07 case study at 04Z, is presented first.

7.5.1 An example from the 23/07

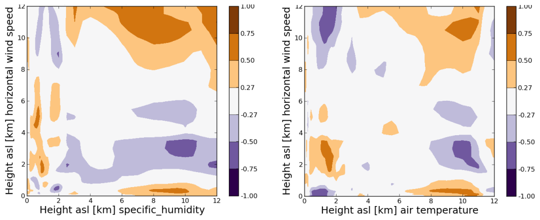
Figure 7.14 shows, in matrix form, correlations between the 3D variables temperature, horizontal wind speed, horizontal divergence, cloud fraction and specific humidity at 04Z on the 23/07. Focusing first on Figure 7.14a, it can be seen that temperature and specific humidity correlations are positive in the layer of convective cloud above 3km and negative below this level. The physical interpretation is that members that are wetter than average in the convective cloud are also warmer than average due to latent heat release. Notice also that the correlations are almost symmetric along the diagonal suggesting that these variables have a similar vertical structure. The similarity between the structure of these variables is also demonstrated by referring back to the VCs for specific humidity and temperature (Figures 7.9a and 7.12a respectively).

Given the similar vertical structure of specific humidity and temperature it is expected that the correlations between specific humidity and other variables will look similar to those between temperature and other variables. This is indeed the case as seen by comparing the first two columns of Figure 7.14. Figures 7.14b and 7.14c show correlations between horizontal wind speed and specific humidity/temperature. Following the specific humidity/temperature (X) axis, correlations change sign around 2km as expected from the temperature-humidity VCCs (Figure 7.14a) and VCs for these variables (Figures 7.12a and 7.9a). Following the wind speed (Y) axis correlations change sign at 2km and 7km. Referring back to the wind speed VCs (Figure 7.13a) we again see the same structure: members that are faster/slower than average near the surface are slower/faster from 2-7 km and faster/slower above 7 km. Notice how, in the cross correlation, the correlation sign changes when members from *either* variable swap sides of the ensemble mean (e.g. when members that are colder or dryer than average become warmer or wetter than average). The increased variation in correlation sign for correlations between variables with differing vertical structure also suggests that, for these variable combinations, information should be spread over a smaller vertical extent.

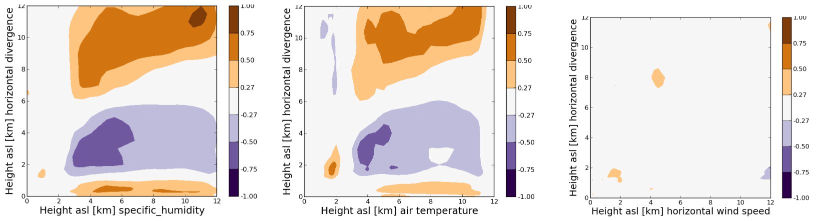
Additional information is provided by the horizontal divergence-specific humidity and horizontal divergence-temperature correlations (Figure 7.14d,e). Interestingly, for divergence, correlations are mainly seen with specific humidity or temperature above 3km in the region of convective cloud: divergence is not directly correlated with low level moisture or temperature for this case (shown by the white regions to the left of Figure 7.14d,e). Physically, this is related to the large scale forcing of this case: for a different meteorological situation where locally warm surface temperatures provide an important role in increasing instability, correlations between divergence and temperature/specific humidity would be expected at low levels (this is indeed seen for the peninsula convergence cases). Following the divergence (Y) axis correlations show a positive-negative-positive structure. This agrees with the divergence VCs (Figures 7.6, 7.7) and reflects the



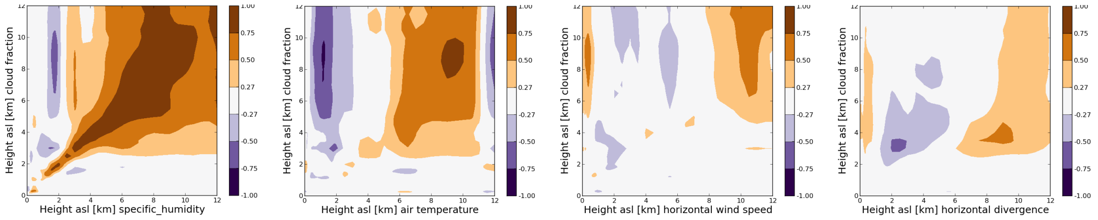
(a) T-SH



(b) Speed-SH (c) Speed-T



(d) Div-SH (e) Div-T (f) Div-speed



(g) Cloud-SH (h) Cloud-T (i) Cloud-speed (j) Cloud-div

Figure 7.14: Vertical cross correlations between five variables for the 23/07 case at 52.6°N, -0.6°E and 04Z. Rows from top to bottom: temperature (T), horizontal wind speed (Speed), horizontal divergence (Div), cloud fraction (Cloud). Columns from left to right: specific humidity (SH), temperature, horizontal wind speed, horizontal divergence. For correlations involving cloud fraction or horizontal divergence a neighbourhood of half width 4 grid points was used. For all other variables a neighbourhood of half width 2 grid points was used.

divergent-convergent-divergent vertical structure of the convective cells. Given the similarity between specific humidity/temperature and horizontal divergence/wind speed correlations it is perhaps surprising that horizontal divergence and wind speed are not themselves strongly correlated (Figure 7.14f). Although members that are more divergent below 2km are more convergent from 3-7km and more divergent above 7km, *and* members that are faster below 2km are slower from 3-7km and faster above 7km, these *are not* the same members.

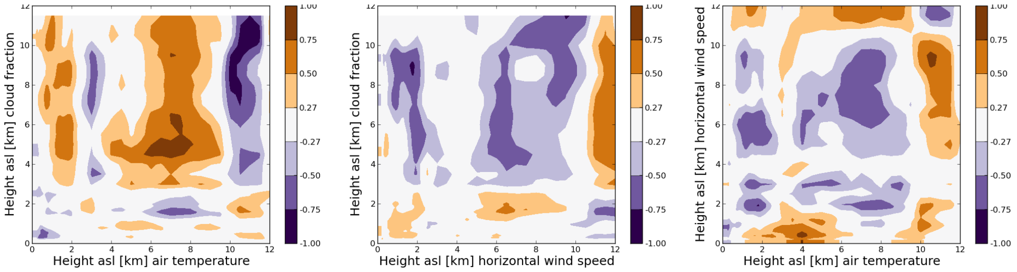
Finally we consider correlations with cloud fraction. Figure 7.14g,h show positive correlations between specific humidity-cloud fraction and temperature-cloud fraction in the convective layer: members that are warmer/colder than average in the cloud layer are also wetter/dryer than average and have more/less cloud than average. Below the cloud base negative correlations are seen with members that are than colder/warmer and dryer/wetter than average below cloud base, having more/less cloud in the cloud layer. Note that, although correlations between temperature and specific humidity are strong down to the surface, strong correlations with cloud fraction are only seen above 2km where there is cloud.

Although the inter-variable relationships discussed above are for one particular case and time only, the methodologies of linking the correlations can also be applied, to useful effect, for understanding the model meteorology of other cases. Two further examples of linked correlations are provided, one for the 27/07 and variables cloud fraction, horizontal wind speed and temperature, and the other for the 02/08 case for correlations between cloud fraction and specific humidity, horizontal divergence and horizontal wind speed.

7.5.2 Further vertical cross correlation examples

Cross correlations on the 27/07 between cloud fraction, horizontal wind speed and temperature are shown in Figure 7.15. Cloud fraction correlates positively with temperature in the upper troposphere (Figure 7.15a). As seen for previous cases members that are warmer below the tropopause are colder above the tropopause. These members also have more cloud throughout the cloud layer. Correlations between cloud fraction (again throughout the cloud layer) and temperature in the lower troposphere are positive below 2km and negative from 3-4km. This shows that the members with more cloud than average are warmer than average below 2km, and colder than average from 2-3km: these members have greater convective instability.

Correlations between cloud fraction and wind speed (Figure 7.15b) are opposite in sign to the equivalent correlations with temperature; negative below, and positive above the tropopause. Members that have slower horizontal wind speeds below/above the tropopause have more/less cloud from 0-12 km. The physical reasons for this could be related to several factors including: a combination of storm centred winds and the back-



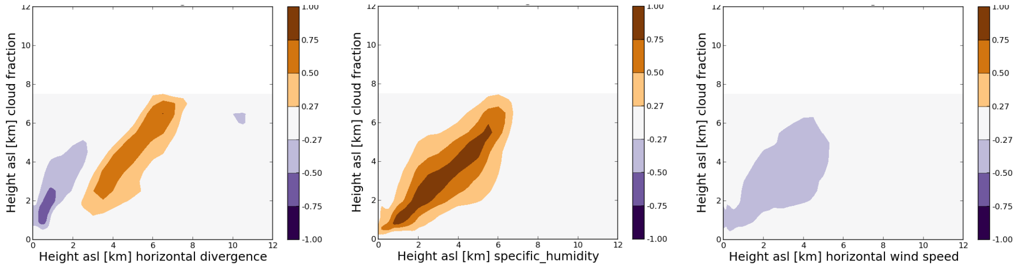
(a) Cloud-T (b) Cloud-speed (c) Speed-T

Figure 7.15: Vertical cross correlations on the 27/07 at 53.0°N, -2.0°E and 23Z . Correlations are shown between variables (a) cloud fraction and temperature, (b) cloud fraction and horizontal wind speed, and (c) horizontal wind speed and temperature. For all correlations an $S_{ij}^{A(\overline{mm})}$ of 2 grid points was used.

ground flow (e.g if the outflow of the MCS is in the direction of large scale flow, members with stronger convection will have a stronger outflow and faster horizontal wind speeds) and also the speed of storm propagation (storms with a faster background flow will clear the point of interest earlier in the day than slower storms). Using correlations at a fixed point in space, as done here, it is not possible to disentangle these different physical processes. However, these processes could be separated using a storm centred approach with correlations calculated using each member at a given position relative to the MCS (i.e a different point in the domain for each member). This storm centred approach was successfully used by Poterjoy and Zhang (2011) to investigate correlations between temperature, pressure, horizontal wind speeds and vapour mixing ratio in a developing hurricane.

The opposite signs of the cloud-temperature and cloud-horizontal wind speed correlations suggest that temperature and wind speed will also be negatively correlated when considered at the same vertical level. This is indeed seen in Figure 7.15c, where negative correlations occur along the diagonal. Note that correlations between temperature and horizontal wind speed calculated at vertical levels on different sides of a change in correlation sign of the VCs for each variable (e.g temperature at 8km and wind speed at 11km) are positive: the VCs for both variables change sign around the tropopause.

The final example of cross correlations is taken from the 02/08 at 14Z for a point situated within the precipitation band associated with the SW Peninsula convergence line. The cross correlation patterns seen here are also typical of both the 29/07 and the 03/08, and also other points along the convergence line on 02/08. Cross correlations between cloud fraction and divergence, specific humidity and wind speed are shown in Figure 7.16. Note that correlations were only calculated in regions of cloud (from the surface up to 7km). Cloud fraction correlates positively/negatively with divergence in the divergent/convergent layers at cloud base/top (Figure 7.16a). Correlations with specific humidity are positive as expected physically (Figure 7.16b). Wind speed correlates negatively with cloud fraction: slower/faster members have more/less cloud (Figure 7.16c).



(a) Divergence (b) Specific Humidity (c) Wind speed

Figure 7.16: Cloud fraction vertical cross correlations with (a) horizontal divergence, (b) specific humidity and (c) horizontal wind speed on the 02/08 14Z. An $S_{ij}^{A(\overline{mm})}$ of 10 grid points was used, centred upon the point 50.8°N, -4.5°E.

7.6 Temporal correlations with rain rates

In this section temporal correlations (TCs) are discussed between surface rain rates and horizontal divergence, horizontal wind speed, temperature, specific humidity, and cloud fraction at different vertical levels. First an introduction is given to the method used for calculating temporal correlations, along with an overview of the format used for plots in this section. Next, the different variable combinations are considered individually. Finally, the cases are compared and contrasted in order to investigate the use of TCs to infer properties of the convection.

7.6.1 Interpreting temporal correlations.

Due to the availability of rain rate data every 5 minutes, and three dimensional variables every hour, correlations were calculated relative to a particular time of the three dimensional variable (referred to as t_0). To allow the correlations to span the lifetime of convective cells, temporal correlations were computed from $t_0 - 3$ hours to $t_0 + 3$ hours. Hence, correlations were calculated between a given three dimensional variable and rain rates up to 3 hrs into the past, and 3 hrs into the future. The method used for presenting temporal correlations is summarised in Figure 7.17; all plots shown in this section will be of this form. At t_0, both rain rates and the 3D variable are considered at the same time and the correlations plotted at each height on the Y axis are VCs between the surface rain rates and the 3D variable at this height. Moving along the positive X direction from t_0 the surface rain rates are considered at times later than t_0 (i.e times in the future). For the propagation of cells along a line this corresponds to rain that has not yet reached the point of interest and may be associated with a cell that is upstream of the location of interest at time t_0. The opposite is true for movement along the negative X direction from t_0 where rain rates are at times earlier than t_0 (at t_0 the cells may have already passed the point of interest).

7.6.2 Horizontal divergence

In Sections 7.4 and 7.5 it was shown that horizontal divergence correlations, both VCs and VCCs, reflect the vertical divergence-convergence-divergence structure of the forecast fields. As this structure is physically related to the convective circulation, and a stronger convective circulation is expected to be associated with more convective precipitation, it is expected that horizontal divergence and surface rain rates are correlated. This is indeed the case as shown for all six cases at times when convection occurred in Figure 7.18.

Figure 7.18 shows correlations extending to different vertical levels, reflecting the different heights reached by convective circulation for each case. This allows the depth

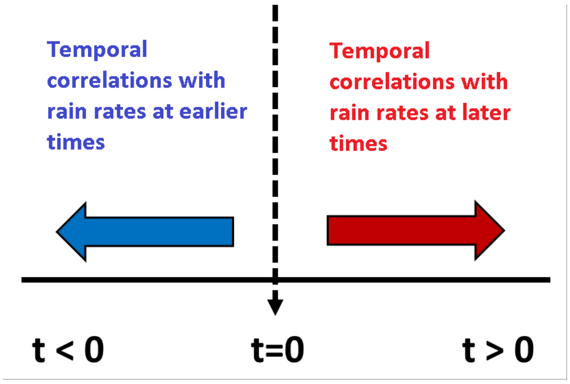
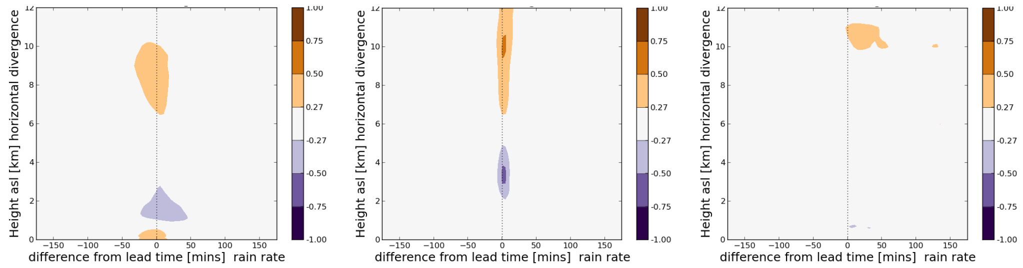
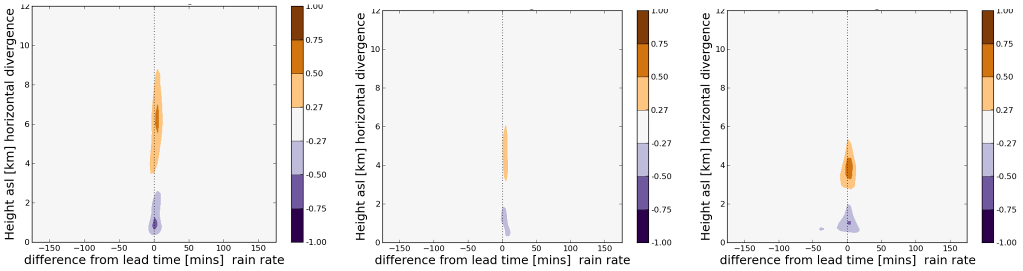


Figure 7.17: Schematic showing the methodology for presenting TCs. A three dimensional variable is considered at t=0 (t_0) and correlated with surface rain rates at $t < 0$ and $t > 0$.



(a) 17/07 17Z 51.7°N, -0.7°E (b) 23/07 04Z 52.6°N, -0.6°E (c) 27/07 23Z 53.0°N, -2.0°E



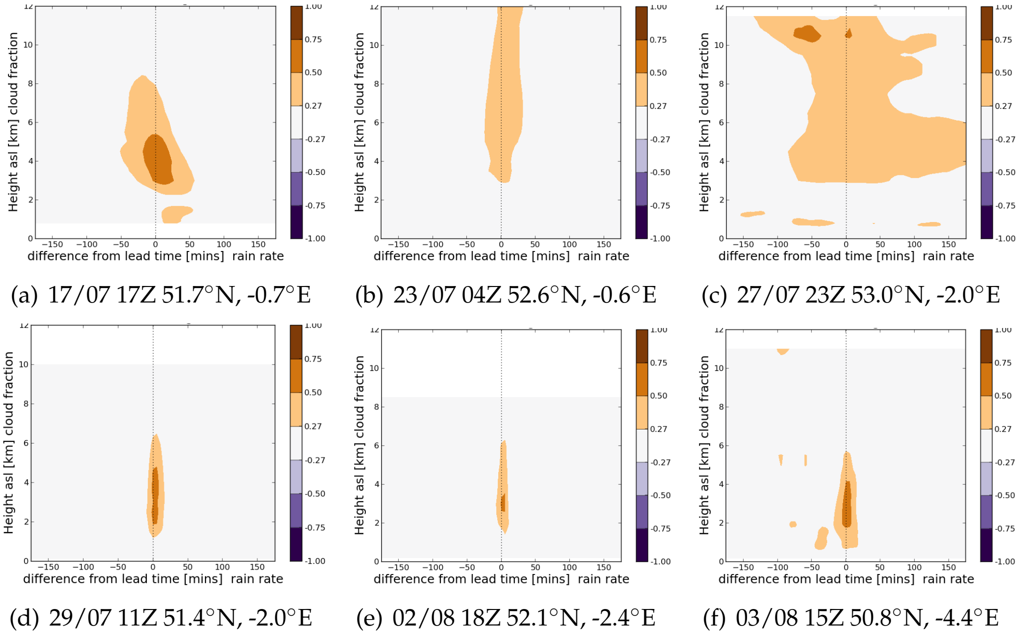
(d) 29/07 11Z 51.4°N, -2.0°E (e) 02/08 18Z 52.1°N, -2.4°E (f) 03/08 15Z 50.8°N, -4.4°E

Figure 7.18: Rain rate divergence correlations for the six different cases. Each case is shown at time when, and location where, convection was present. The correlations for each case were calculated over the $S_{ij}^{A(\overline{mm})}$ for rain rates at that point corresponding to (a) 20 grid points, (b) 22 grid points and (c)-(f) 10 grid points. The correlations are shown out to ±3 hours to highlight their short duration, and to maintain consistency with Figure 7.19.

of convective cells to be assessed from the ensemble correlations over appropriate horizontal scales. For example, on 17/07 and 23/07 (Figures 7.18a and 7.18b) convection reached upper tropospheric levels and these cells were predicted over an area of around (90 km)^2 whereas on the afternoon of 02/08 and on 03/08 convection only reached 5 to 6 km with cells of this nature likely over an area of (46 km)^2. Further information can be gained from the temporal extent of the correlations which reflect the length of time a cell remains over a particular point. This timescale is influenced by both the cell lifetime, and cell propagation speed, so should not be expected to give a direct measure of cell dimensions. The duration timescale does however give an indication as to the amount of rain that will fall over a given point, from a particular cell, which is useful information from a flood forecasting perspective. For example, on 03/08 the correlations are maintained for around 40 minutes suggesting that, despite being shallow, these cells may produce a considerable amount of rain over a given point. In contrast, on 02/08 correlations are maintained for 15 minutes and the cells quickly pass through. Obviously, this discussion refers to individual cells: a large amount of precipitation may fall at a given point if many short duration fast moving cells pass over. For the MCS case (27/07), there is little correlation seen between divergence and rain rates (Figure 7.18c). This is true throughout the day and, similarly to the lack of strong divergence VCs for this case (see Section 7.4.1), is explained by the MCS containing many interacting convective cells unlike the individual cells seen in the other cases. For all six cases, the horizontal divergence - rain rate correlations are not symmetric in time, with stronger correlations seen for positive times (i.e later times of day; rain that has not yet fallen at t_0). This makes sense as the convective circulation must build up before it produces precipitation. Additionally, microphysical processes will take some time to produce surface precipitation.

7.6.3 Cloud fraction and specific humidity

Sections 7.4 and 7.5 showed that VCs involving cloud fraction and specific humidity have similar structures. This was also found to be true for correlations between these variables and rain rates. Hence, for brevity, only rain rate - cloud fraction correlations are presented in this section. All the proceeding discussion is equally true for specific humidity temporal correlations with rain rates.



(a) 17/07 17Z 51.7°N, -0.7°E (b) 23/07 04Z 52.6°N, -0.6°E (c) 27/07 23Z 53.0°N, -2.0°E

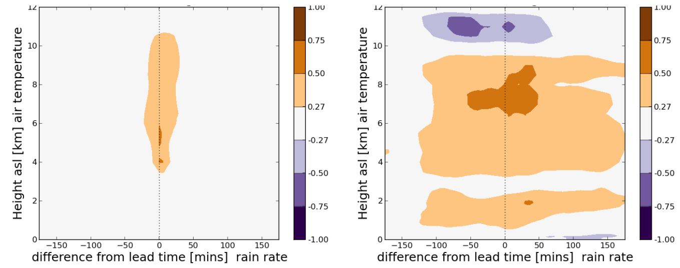
(d) 29/07 11Z 51.4°N, -2.0°E (e) 02/08 18Z 52.1°N, -2.4°E (f) 03/08 15Z 50.8°N, -4.4°E

Figure 7.19: Rain rate cloud fraction correlations for the six different cases. Each case is shown at time when, and location where, convection was present. The correlations for each case were calculated over the $S_{ij}^{A(\overline{mm})}$ for rain rates at that point corresponding to (a) 20 grid points, (b) 22 grid points and (c) to (f) 10 grid points.

Figure 7.19 shows rain rate-cloud fraction temporal correlations for all six cases at the same points and times as the rain rate - divergence correlations in Figure 7.18. Like horizontal divergence, cloud fraction is sensitive to the properties of the convection (stronger convection leading to more cloud) so it is not surprising that the temporal duration and vertical height of the cloud fraction-rain rate and divergence - rain rate correlations are similar. Interestingly the cloud fraction correlations tend to be slightly longer lived. An exception to this is the 27/07 case where rain rate - cloud fraction correlations extend up to 12km and are maintained for much longer than the rain rate - divergence correlations, at least 6 hours, contrasting with the very low magnitude divergence correlations. Physically, although individual updraughts and downdraughts are on small scales, this MCS case had a large area of cloud (both in terms of depth and horizontal extent) with many interacting cells. This leads to deep and long duration correlations with cloud fraction. For the other deep convective cases (17/07 and 23/07) cloud fraction- rain rate correlations are maintained for up to twice as long as the corresponding divergence correlations. For these larger scale events the cloud systems were larger than the individual convective circulations, i.e. cloud persists even after the circulation that produced it is no longer present.

7.6.4 Air temperature

As discussed in Section 7.4.4 convection only modified air temperature sufficiently to influence the correlation structures for deep convective cases. Hence, as expected, significant (here taken to mean with magnitude greater than ±0.27) correlations between rain rates and temperature are not seen for the 29/07, afternoon of 02/08, or 03/08. On the 17/07 correlations between temperature and rain rates were seen, but only at low magnitudes, and sometimes not at the believable scale (this example will be discussed further in Section 7.7.1). Example temperature-rain rate correlations for the remaining two cases, 23/07 and 27/07, are given in Figure 7.20. For both cases correlations are positive in the troposphere as expected physically (higher rain rates are associated with more cloud and more latent heat release). On the 27/07 correlations also extend above the tropopause where they change sign. As discussed in Section 7.4.4 members that are warmer in the troposphere tend to have a colder tropopause and also be colder above the tropopause. The correlation with surface rain rates at these tropospheric levels, and for over 2 hours either side of the time of interest, reiterates the large scale forcing of this case and also suggests that, at least for some convective cases, surface rain rates can be directly linked to larger scale features.



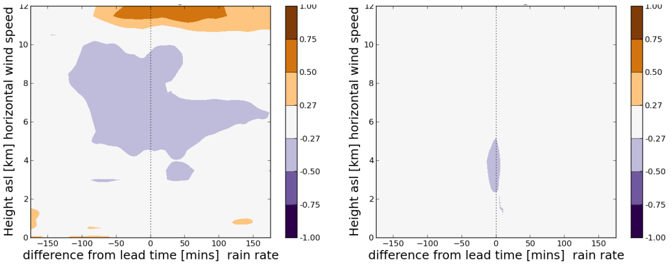
(a) 23/07 04Z 52.6°N, -0.6°E (b) 27/07 23Z 53.0°N, -2.0°E

Figure 7.20: Rain rate-air temperature correlations for (a) the 23/07 and (b) the 27/07. Each case is shown at time when, and location where, convection was present. The correlations for each case were calculated over the $S_{ij}^{A(\overline{mm})}$ appropriate for rain rates at that point corresponding to (a) 22 grid points and (b) 10 grid points.

7.6.5 Horizontal wind speed

The 27/07 is the only deep convective case to show TCs between rain rates and horizontal wind speed. These correlations are negative in the mid to upper troposphere and positive above 10 km as shown in Figure 7.21a. This correlation structure agrees with the cross correlation analysis for this case (Section 7.5.2): members with more cloud were warmer/colder below/above the tropopause with slower/faster winds.

A second example of surface rain rate - horizontal wind speed correlations is given in Figure 7.21b for the 02/08 case. Similar correlation structures were obtained for the other peninsula convergence cases (29/07 and 03/08). Low magnitude negative correlations are seen lasting for around 20 minutes with members that produce more rain having slower horizontal wind speeds: a slower propagating cell will stay in the same place for longer, and hence produce more precipitation at a specific location. Other physical factors may also contribute to this correlation structure. For example, as the peninsula convergence line is produced when the sea breezes from both sides of the peninsula collide, points directly on the convergence line will have lower horizontal wind speeds than further away points. The strongest convection and highest precipitation totals also occur along this convergence line, with less convection and less rain at other positions. Hence, the wind speed and precipitation may be expected to be negatively correlated. Another factor that could contribute would be a convective circulation that opposes the large scale flow.



(a) 27/07 23Z 53.0°N, -2.0°E (b) 02/08 18Z 52.1°N, -2.4°E

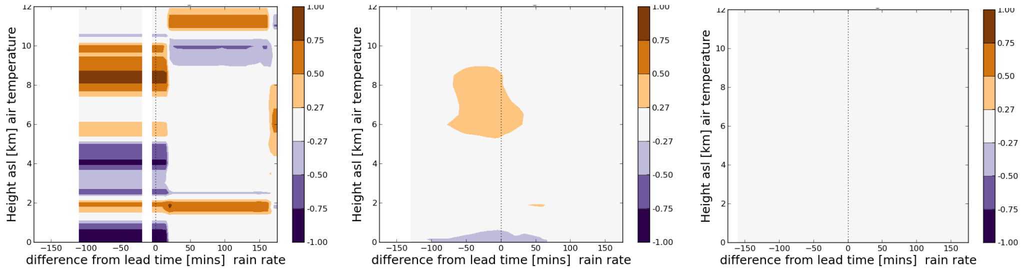
Figure 7.21: Rain rate horizontal wind speed correlations for (a) the 27/07 and (b) the 02/08. Each case is shown at time when, and location where, convection was present. The correlations for each case were calculated over the $S_{ij}^{A(\overline{mm})}$ for rain rates at that point which, for both cases, corresponded to 10 grid points.

7.7 Correlations at different spatial scales

Sections 7.4 to 7.6 considered correlations at the agreement scales, as defined in Chapter 6. Although the agreement scales are thought to provide the best compromise between smaller scales that preserve all information but are noisy, and larger scales where information is lost due to smoothing, correlations can be, and were, calculated over many plausible scales (i.e using differently sized areas or neighbourhoods). In this section the dependence of vertical correlation structure on the spatial scale is discussed. Here, consistent with the definition in Chapter 6, the spatial scale (S) is defined as the distance from the centre to the edge of the square area (neighbourhood) used to calculate the correlations. Hence, spatial scales less than, equal to, and greater than the the agreement scales $S_{ij}^{A(\overline{mm})}$ are considered. By calculating correlations at different scales it is possible to learn more about the underlying meteorological situation: examples of this are provided throughout this section. Variables with higher spatial variability, such as rain rates, are more sensitive to the spatial scale considered than other more smoothly varying fields. Hence, the first subsection will focus on temporal correlations with rain rates before giving examples from other variables.

7.7.1 Examples from temporal correlations with rain rates

Out of all the cases, the 17/07 had the largest spatial uncertainty with $S_{ij}^{A(\overline{mm})}$ for rain rates reaching 40 grid points in areas of convection. Figure 7.22 shows temporal correlations between rain rates and temperature for this case at 14Z for three different neighbourhood sizes. Correlations calculated at the grid scale (Figure 7.22a) are highly noisy, include short gaps in correlations where there was no rain, and cannot be reliably or physically interpreted. In contrast, the results from calculating correlations over a neighbourhood are physically sensible. Figure 7.22b shows correlations calculated at a scale of 8 grid points (37.4 km total neighbourhood width). Using larger spatial scales at this time (up to an scale of 40 grid points), gives similar correlation structures, although the correlation magnitudes differ. Two areas of strong correlation are seen in Figure 7.22b, negative correlations at low levels that can be related to the cold pool of the storm and positive correlations from 6 to 9 km that can be related to latent heat release in the convective cloud. It is interesting to note that these correlations are with rain rates at t_0 and preceding t_0: rain rates correlate with temperature at later times. Although Figure 7.22b appears to have sensible correlations, the neighbourhood used is still much smaller than that suggested by the $S_{ij}^{A(\overline{mm})}$ (40 grid points at 14Z). The correlations for this larger neighbourhood are very weak and do not show up in Figure 7.22c. This suggests that, although the correlations in Figure 7.22b are physically plausible, given the large spatial uncertainty at this time the details should be treated with caution. Similar results were obtained from analysis of the 17/07 rain rate - temperature correlations at 17Z.

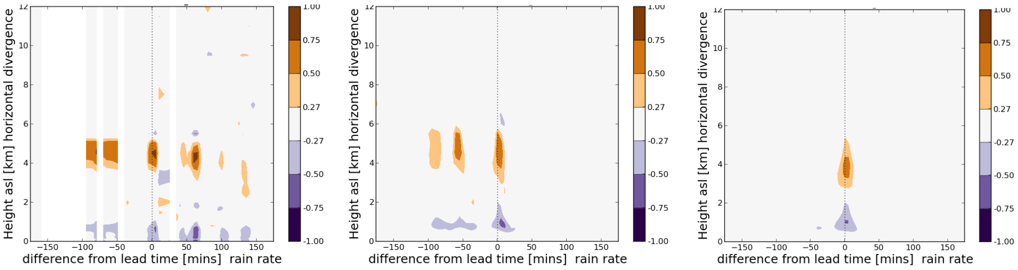


(a) Grid scale (b) S= 8 grid points (c) S=40 grid points

Figure 7.22: Rain rate-temperature TCs for 17/07 at 14Z for the point 51.7°N, -0.7°E. Results for are shown for different spatial scales: (a) grid scale (S=0), (b) S=8 grid points, and (c) S=40 grid points (corresponding to the $S_{ij}^{A(\overline{mm})}$ at t_0).

For the 03/08 case it is informative, meteorologically, to consider different spatial scales. (Similar conclusions can be drawn from the afternoon of 02/08). In Figure 7.23

different neighbourhoods are considered for rain rate - divergence temporal correlations calculated at 15Z on 03/08. For small spatial scales (e.g a scale of 2 grid points, Figure 7.23a) correlations are seen with rain rates at six different times, separated by around 50 minutes. At this spatial scale we see correlation structures for individual cells passing over the point of interest. Similar correlation magnitudes are seen at different temporal separations: members with a stronger convective circulation (stronger convective cell, associated with stronger rain rates) at 15Z also have a stronger circulation for other convective cells that move along the convergence line. As the spatial scale is increased there are a decreasing number of separate times where correlations are seen until, by a scale of 10 grid points (Figure 7.23c) correlations are only seen around t_0. This neighbourhood, corresponding to the agreement scale, is large enough to show effects from multiple cells and we no longer see their separation. This is exactly what is intended in defining the agreement scales: although the model is able to produce realistic cells that move in a sequence along the line we do not expect to be able to make specific predictions about the positioning and circulation of individual cells.

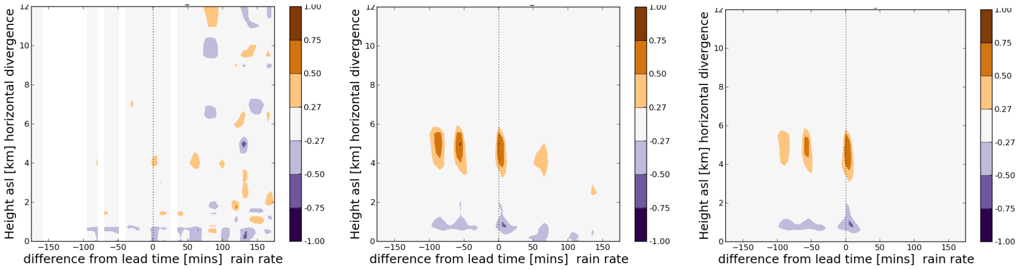


(a) S=2 grid points (b) S=6 grid points (c) S=10 grid points

Figure 7.23: Rain rate divergence TCs calculated at different spatial scales for the point 50.8°N, -4.4°E, on the 03/08 at t_0= 15Z.

All of the correlation plots shown so far were calculated using square neighbourhoods with no weighting given to account for expected weakening of correlations when moving further from the point of interest. As discussed in Section 7.2.2 this is the preferred method as it is consistent with the calculation of the agreement scales (Chapter 6). However, it is also possible to apply a weighting over the neighbourhood to account for the increased differences expected at larger horizontal separations (as discussed in Section 7.2.2). Figure 7.24 shows correlations for the same times and fields as Figure 7.23 but this time with a normalised Gaussian weighting with a half width of 2σ applied to the fields before calculating the correlations.

Comparing Figure 7.23 to Figure 7.24 we see that applying the weighting has an overall effect of smoothing less for a given neighbourhood size. Specifically, in Figure 7.24c, correlations calculated at a scale of 10 grid points with weighting resemble those in Figure 7.23b, correlations calculated at an scale of 6 grid points with no weighting. This suggests that, if a weighting were to be used, a larger neighbourhood would have to be considered to get the equivalent amount of smoothing as results with no weighting. Note that the correlation structures with and without the weighting are similar: applying a weighting function, at least for the cases considered here, does not alter the overall conclusions. Results from correlations between other variables (not shown) also support these arguments.



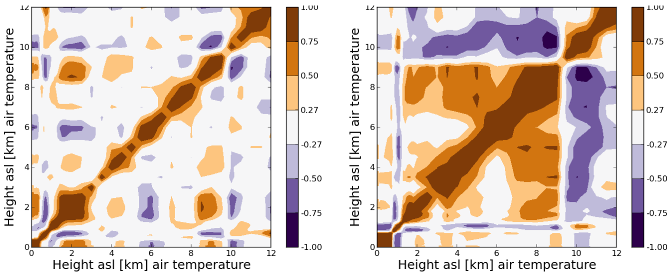
(a) $2\sigma = 2$ (S=2 grid points) (b) $2\sigma = 6$ (S=6 grid points) (c) $2\sigma = 10$ (S=10 grid points)

Figure 7.24: Rain rate divergence TCs calculated over different neighbourhood sizes for the points at 50.8°N, -4.4°E on the 03/08 at t_0= 15Z. A normalised Gaussian weighting of width 2σ (centred on the middle of the neighbourhood) was applied to the fields before calculation of the correlations.

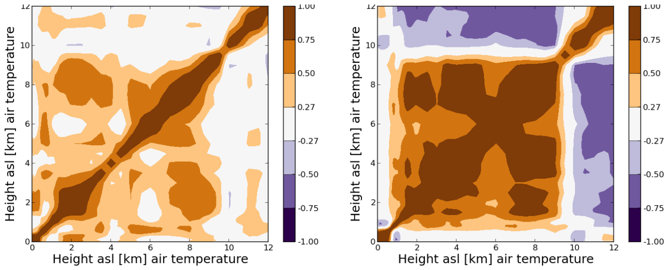
7.7.2 Examples from other variables

Finally, two examples are provided to illustrate the possible impact on other variables of calculating VCs over different horizontal scales. Generally, temperature VCs are insensitive to the choice of spatial scale, due to the smooth variation of temperatures over large horizontal separations. Here an unusual example from the 02/08 at 09Z, which does not follow this general pattern, is discussed.

On the 02/08 at 09Z local processes dominate the temperature correlation structure and, at the agreement scale (2 grid points), the correlation sign changes frequently with height as shown in Figure 7.25a. By 12Z local variations are no longer seen in the temperature correlations and the change in correlation sign at the tropopause, discussed in Section 7.4.4, is clearly seen (Figure 7.25b). It is expected that local variations will be smoothed out in correlations calculated over a larger spatial scale. This is indeed the case for temperature correlations at 09Z on 02/08 as shown in Figure 7.25c for a scale of 20 grid points. Correlations calculated over this large spatial scale (much larger than the agreement scale) vary less than those calculated at the agreement scale. Interestingly, the expected change in correlation sign at the tropopause is still not seen. Perhaps this is due to a very variable tropopause height resulting in the smoothing out of the correlations around this level. At 12Z where temperature correlations are not dominated by small scale processes, considering a larger scale has little effect (as seen by comparing Figures 7.25b and 7.25d). This is the general situation for temperature correlations.



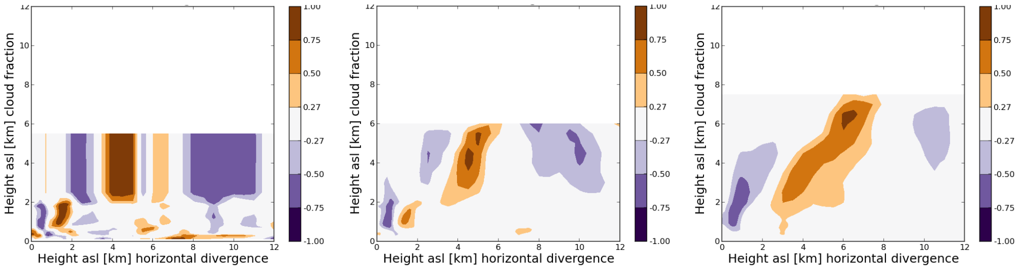
(a) 9Z (b) 12Z



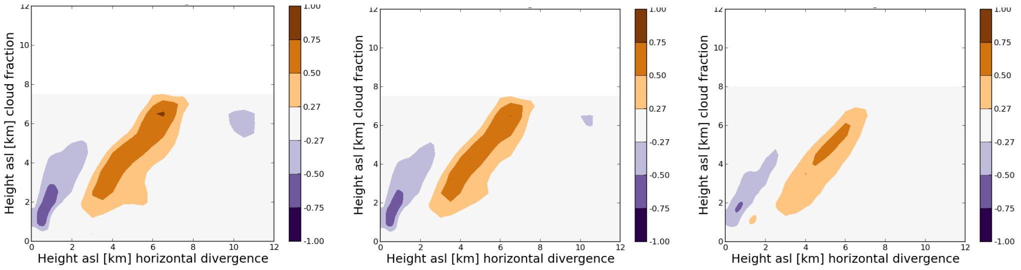
(c) 9Z (d) 12Z

Figure 7.25: Vertical correlations for temperature on 02/08 for the point 52.1°N, 0.6°E at (a,c) 09Z and (b,d) 12Z Results in the top/bottom rows were calculated for a scale of 2 grid points (the agreement scale) and 20 grid points respectively.

For the majority of cases noisy correlations are found at, and close to, the grid scale. As spatial scale increases the correlations rapidly smooth out due to the sample size increasing as $(2S+1)^2\times$ number of ensemble members. By the agreement scale correlations are physically meaningful and can be usefully interpreted, as demonstrated by the examples in this chapter. Beyond the agreement scale, the correlation structures are often maintained, although the correlation magnitude decreases with increasing spatial scale. Figure 7.26 shows an example of this for cross correlations between divergence and cloud fraction on the 02/08 at 14Z. In Figure 7.26 correlations are shown at a range of scales from the grid scale (Figure 7.26a) at which correlations are noisy, through the agreement scale (10 grid points, Figure 7.26e) to a scale of 20 grid points (Figure 7.26f). These correlations were discussed physically in Section 7.5.2: members with stronger convective circulations have more cloud. Beyond a scale of 20 grid points, correlation magnitudes would be expected to decrease as points outside the convective area are included.



(a) grid scale (b) S=2 grid points (c) S=6 grid points



(d) S=8 grid points (e) S=10 grid points ($S_{ij}^{A(\overline{mm})}$) (f) S=20 grid points

Figure 7.26: Cloud fraction VCCs with horizontal divergence for the 02/08 at 14Z and point 50.8°N, -4.5°E. Different spatial scales were used to calculate the correlations ranging from (a) the grid scale to (f) S=20 grid points.

7.8 Chapter discussion and conclusions

In this chapter a neighbourhood approach has been presented for calculating vertical correlations. Correlations calculated using this method were discussed for six convective cases from summer 2013 for the 3D variables horizontal divergence, horizontal wind speed, cloud fraction in each layer, specific humidity and air temperature and also for surface rain rates. Univariate, multivariate, and temporal correlations were considered. This chapter aimed to address three research questions (thesis questions 4-6) which will now be discussed in turn.

1. How can correlations in the vertical be used to allow physical structures and multivariate relationships from the ensemble to be usefully summarised? To what extent

do these correlations reveal properties of the convection and the convective environment? (Thesis question 4; Section 1.2)

The results presented in this chapter demonstrate that, by using a horizontal neighbourhood approach, meaningful and useful vertical correlations can be obtained. For example, while some fields such as cloud fraction and temperature had certain structures that were found in multiple cases, correlations for other variables, such as horizontal wind speed were highly case dependant. The vertical correlations were found to reflect the convective dynamics on each day. This was particularly true for correlations between variables such as rain rates or cloud fraction that were directly linked to the convection. For example, rain rate-divergence temporal correlations gave information on the height and structure of convective cells and also the length of time that cells were at a specific location. This information is contained in a small number of plots and hence provides an effective and physically meaningful summary of the ensemble information. For comparison, examining the points individually over a 10 by 10 neighbourhood would require 1200 plots. Obviously, the ensemble or spatial mean could be used to condense this information but this is known to be outside the model manifold (Ancell, 2013) and was shown in Chapters 5 and 6 to be unrepresentative of physical processes.

It is not feasible to calculate and study vertical correlations directly between all 3D model variables. Hence it is necessary to choose variables that provide the most useful information. Although the variables in this study were selected to represent a wide range of the system properties, specific humidity and cloud fraction correlations were found to provide overlapping information. This suggests that it may not be necessary to consider both cloud fraction and specific humidity. As it is only possible to calculate cloud fraction correlations in regions of cloud, it is suggested that for future investigations only specific humidity should be used.

From Sections 7.4 and 7.5 it was found that cross correlations could generally be interpreted, or even predicted, from the vertical correlations of the two variables treated separately. This suggests that it would be possible to gain a large amount of information from considering only one of the vertical uni-variate and vertical cross correlations. However, the interpretation of both types of correlation is sometimes involved and considering both types together does help in understanding the structures in general. Overall, correlations with rain rates are found to be the easiest to interpret and these would be recommended for use in the future.

2. What is the dependence of vertical correlation structure on the horizontal scales used? How does this relate to the scales obtained from the spatial methods? (Thesis question 5; Section 1.2)

When calculated at the grid scale correlations are found to be noisy for all variables due to under sampling from the 12 ensemble members. It is not recommended that any correlations calculated on the grid scale from the 12 member ensemble should be physically interpreted. Considering correlations over even a small neighbourhood of 5 by 5 grid points (representing the minimum scale resolvable by the model according to Cullen and Brown (2009)) significantly reduced noise in the correlations. There are disadvantages to using very large neighbourhoods including the smoothing of smaller scale structure and increased computational cost. Hence it is desirable to use a neighbourhood size that is large enough to remove noise but not so large as to unnecessarily smooth the correlations. For the summer convective cases considered here the $S_{ij}^{A(\overline{mm})}$ defined in Chapter 6 were found to be a good choice of neighbourhood size. Of course, further information can often be gained from also considering the behaviour over other spatial scales: an example was given in Section 7.7 where correlations calculated over small neighbourhoods showed the structures of individual cells whereas correlations calculated over the believable scale showed the properties of cells in general.

3. In what ways might these methods be useful in other areas such as data assimilation? (Thesis question 6; Section 1.2)

Overall the results from this chapter showed that there was a large variety of detailed correlation structures between the five variables considered. For some variables sharp transitions in correlation sign were seen e.g for temperature at the tropopause, or when moving between divergent and convergent layers. This suggests that data assimilation should not allow smoothing across these boundaries. Other variables, such as cloud fraction showed positive correlations extending throughout the convective cloud suggesting that, in these cases, information could be propagated over large vertical distances. Correlations, particularly those with surface rain rates, also showed that it is possible to get physically meaningful correlations that reflect the convective structure (even from a small 12 member ensemble), with a suitable choice of neighbourhood. This is encouraging for convective scale data assimilation using ensembles of limited membership.

Chapter 8

Location-dependent spatial predictability for a summer season

8.1 Introduction

Chapter 6 introduced a new location-dependent method (agreement scales, S^A) for quantifying the spatial characteristics of convective-scale ensemble forecasts. Results from six convective case studies demonstrated the utility of these techniques, for example to visualise spatial predictability variations across the domain and to evaluate the spatial spread-skill relationship. A simple idealised experiment was also used to verify that the agreement scales were being correctly interpreted.

The purpose of this chapter is to provide a more robust evaluation of the location-dependent spatial ensemble spread and skill. In order to do this, results are presented of agreement scales calculated for three months of hourly MOGREPS-UK rain rate data from June, July and August (JJA) 2013. Note that these results should not be interpreted as a summer climatology of the MOGREPS-UK spatial predictability: other years, with different meteorological situations, may have very different results. The aim is to go beyond the case study approach to better understand the spatial ensemble characteristics for one particular UK summer season. In particular, the chapter will quantify, and investigate factors influencing, the spatial predictability of precipitation over the UK. The period JJA 2013 was chosen to focus on summer convection, as it is particularly important to understand and quantify the spatial predictability of these events where the exact location of the precipitation is difficult (or even impossible) to forecast deterministically. The location-dependent spatial ensemble spread is quantified by calculating the average agreement scale, at each grid point, of all independent member-member comparisons ($S_{ij}^{A(\overline{mm})}$, introduced in Chapter 6). To quantify the spatial ensemble skill, the MOGREPS-UK output is compared with Radarnet rain rate observations (to give $S_{ij}^{A(\overline{mo})}$, introduced in Chapter 6). The work in this chapter aims to expand on the results of Chapter 6 to provide additional insight that allows us to answer the following questions (thesis questions 2 and 3; Section 1.2)

1. Over what spatial scales should the ensemble be interpreted and evaluated? How can these spatial scales be defined?
2. What information does this spatial approach provide for forecasting?

First, the method of calculating the agreement scales for three months of data is discussed in Section 8.2 including details of the MOGREPS-UK and Radarnet data. This is followed by a meteorological overview of JJA 2013 in Section 8.3, focusing on the variation of precipitation amount and frequency throughout the period. This section also includes a qualitative overview of the differences between ensemble and Radarnet data. Section 8.4 focuses on the spatial predictability given by the ensemble, by considering the average $S_{ij}^{A(\overline{mm})}$ over the three month period. The relationships between the $S_{ij}^{A(\overline{mm})}$ and the intensity and amount of precipitation are also discussed. Next, the ensemble spread is compared to the ensemble skill, first using traditional spread-skill measures (Section 8.5), and then using spatial measures (Section 8.6). The temporal evolution of the spatial spread is discussed along with the dependence of spatial spread-skill on precipitation extent and intensity. Finally, the conclusions are presented and discussed in Section 8.7.

8.2 Methods

The work presented in this chapter uses ensemble rain rate forecasts from the MOGREPS-UK ensemble that was operational in summer 2013. Details of this configuration can be found in Chapter 3, along with details of the Radarnet rain rate composite. Rain rate data were used here to focus on the spatial distribution of precipitation, and to avoid the temporal averaging associated with precipitation accumulations. Ensemble rain rate forecasts were considered over the constant resolution part of the MOGREPS-UK domain (light grey Figure 8.1, also Chapter 3 Figure 3.2). For comparisons with radar, in order to speed up processing and to allow comparison of different regions, two domains were considered. The northern domain covers northern England and southern Scotland, and the southern domain covers south and central England. The locations of the North and South domains are shown in Figure 8.1 in dark grey and mid-grey respectively. Both domains fall within the region of Radarnet coverage (dotted in Figure 8.1).

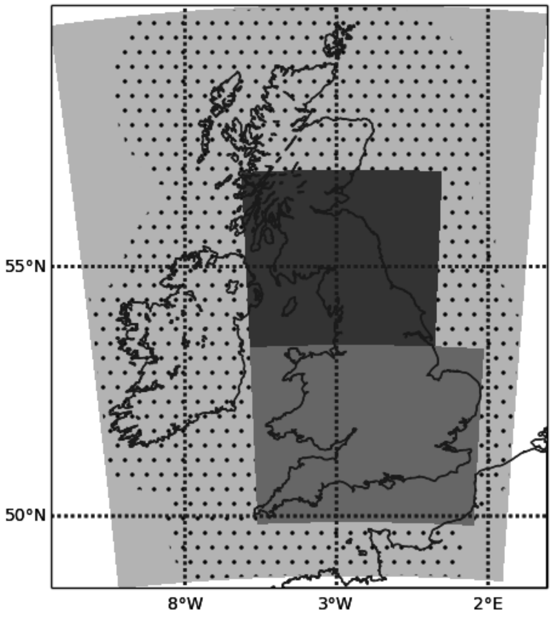


Figure 8.1: Domains considered: MOGREPS-UK domain (light grey), radar coverage (dotted), northern domain (dark grey) and southern domain (middle grey).

For all days in JJA 2013 the MOGREPS-UK forecasts starting at 03Z were used. These forecasts run out to a lead time of 36 hrs (i.e. ending at 15Z the following day). When the time-evolution of spatial predictability is investigated in Section 8.6, the whole 36 hour forecast period is considered. When results are averaged over the three month period forecasts are only considered for the 24 hour period from 09Z (T+6) on the forecast start day to 08Z the following day (T+29). The length of this averaging period was chosen so that each time of day is only included once in the average, minimising effects from the diurnal cycle. The start time (09Z) was chosen to be late enough in the forecast for convective showers to have spun up from the low resolution initial conditions, but still early enough to allow some of the large-scale information from the analysis to be retained.

8.2.1 Masking

Although quality controlled (see Section 3.3.1), radar data can still retain some unrealistic artifacts. Hence, to ensure the use of high quality radar data it was necessary to exclude some times and points from the analysis. In addition to times where there were no radar data due to problems with the Radarnet system (11/06 23Z, 12/06 00Z to 06Z, 02/07 08Z, 18/07 07Z to 13Z), or no model data (19/07 03Z to 20/07 15Z) times were also excluded where any point in the domain had rain rates exceeding $300 mm\, hr^{-1}$. In these instances the abnormally high rain rates suggested a problem with the Radarnet composite, and it was decided to mask all points in the domain. On a small number of occasions there were points within the normal region of radar coverage that had no radar data. These points were set to have rain rates of zero as they tended to be towards the edge of the domain, or in areas with low rain rates. Although setting these values to zero may increase errors in the rain rate values, it is preferable to discarding a large amount of data to allow for a few points in the domain. Recall that, because the agreement scales are calculated by considering increasingly large regions surrounding each given point, it is necessary to have data everywhere within the domain. Whenever masking occurred, the same mask was applied to both the ensemble and radar data to maintain consistency.

8.2.2 Thresholding

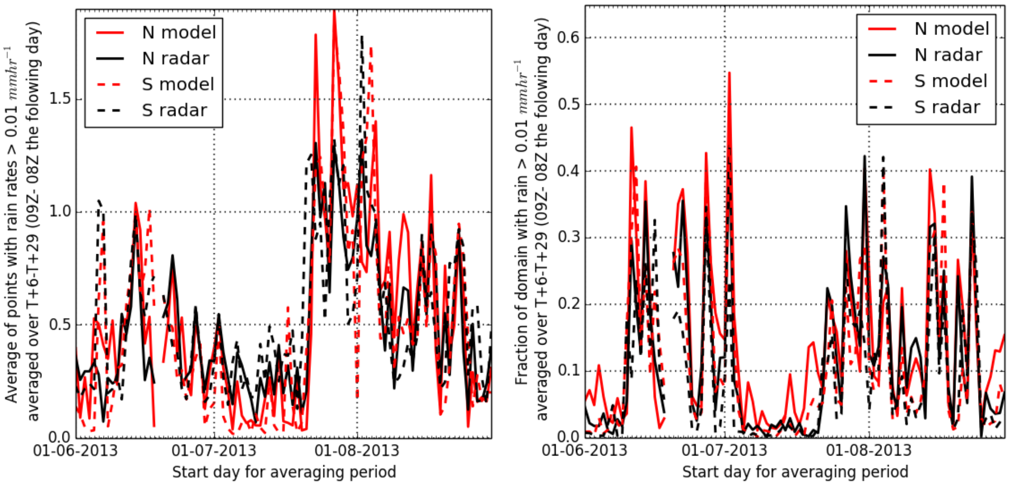
To concentrate on points with significant rain rates, points in the ensemble forecasts and Radarnet data with rain rates less than $0.01 mm\, hr^{-1}$ were set to zero for all the results presented in this chapter. This also removed a discrepancy between model and radar: the model archive does not have any data points with rain rates less than $0.01 mm\, hr^{-1}$, possibly due to the packing of files or cut off values.

A range of different thresholds are considered in Sections 8.4 and 8.6 to investigate how the spatial predictability of precipitation depends on the rain rates considered. Results are presented for thresholds of 0.01, 0.1, 1.0 and $4.0 mm\, hr^{-1}$ which summarise the range of threshold-dependent behaviour observed. Thresholds were applied by setting all points with rain rates less than the threshold rain rate value to zero (individually for each ensemble member and for the Radarnet observations). When considering a particular threshold (for either the model or the radar), only the times with rain rates greater than that threshold were included in the analysis. The radar observations are said to have rain rates greater than the threshold if any point in the radar rain rate map exceeds the threshold. For the model to exceed a threshold any one member has to have rain rates which exceed that threshold. Other methods of making the above/below threshold decision were considered (e.g all members must exceed the threshold, or several points must exceed the threshold). The overall results were not sensitive to the choice of method.

8.3 Overview of cases

Overall, summer 2013 was dryer and sunnier than average, with a short heat wave from 03–07/07/2013 (Met Office, 2013). To summarise how the precipitation varied over the three month period, several methods were used to characterise the precipitation across the domain. Two of these, the average rain rate of all points in the domain with rain rates exceeding $0.01 mm\, hr^{-1}$ (measuring precipitation intensity), and the fraction of points in the domain with rain rates exceeding $0.01 mm\, hr^{-1}$ (measuring the horizontal extent of the precipitation) are shown in Figure 8.2 for an ensemble member (the control, red) and radar observations (black). Results from both the North and South domains are shown. To focus on the variation in precipitation over the three month period the diurnal cycle is removed by averaging over forecast lead times T+6 to T+29 (09Z on the forecast start date to 08Z the following day). From Figure 8.2 it can be seen that the ensemble and radar have similar precipitation amounts overall, and both clearly identify the dry heat wave at the start of July, and another dry period around the beginning of June. Also noticeable is the wet period around the end of July/start of August containing the convective cases discussed in Chapters 5–7. Overall, the intensity and coverage measures give a similar impression of the season although there are some subtle differences. For example, in the dry period around the start of July the model underestimated the rain rates but overestimated the fractional coverage: on average the precipitation in this period was too light and widespread in the ensemble forecast. The opposite was true for the wet period around the start of August (intensities were overestimated by the ensemble but fractional coverage was underestimated).

Alongside the variation in precipitation across the three month period it is also informative to consider the average variation in precipitation throughout the day (diurnal cycle) and as a function of forecast lead time. Figure 8.3 shows the average and fraction of points with rain rates greater than $0.01 mm\, hr^{-1}$, averaged over all days in JJA 2013 at each forecast lead time. Results for the two precipitation measures are quite differ-

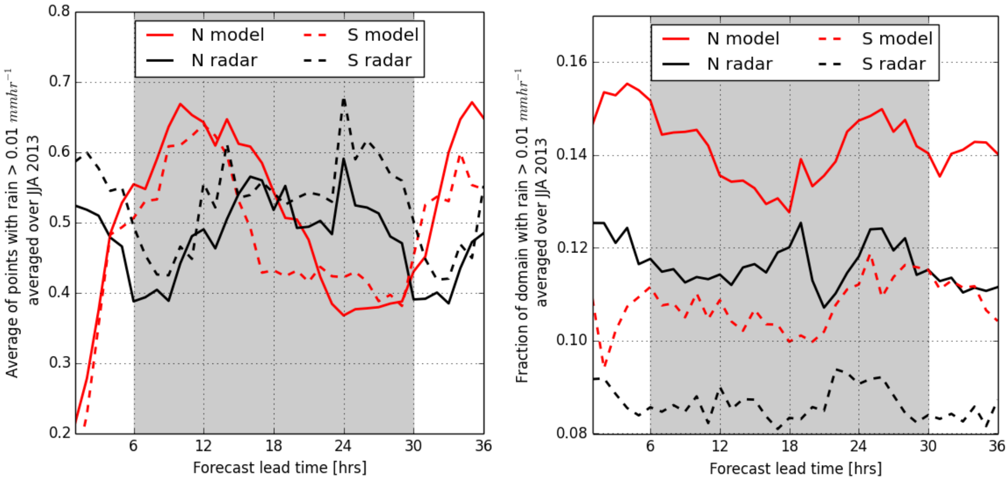


(a) Average of raining points

(b) Fraction of raining points

Figure 8.2: (a) The average rain rates of all points in the domain with rain rates exceeding $0.01 mm\, hr^{-1}$ and (b) the fraction of points in the domain with rain rates exceeding $0.01 mm\, hr^{-1}$ as a function of forecast start date. In both (a) and (b) the average value has been taken over forecast lead times T+6 to T+29 (09Z on the forecast start date to 08Z the following day). Results are shown for one ensemble member (the control, red) and for Radarnet data (black): the same averaging method was used for all results.

ent with the rain rate averages (Figure 8.3a) showing less difference between the North and South domains, and between ensemble member and radar data, than the fractional coverage (Figure 8.3b). The rain rate averages also show the spin-up of the model in the first 6 hrs of the forecast as convective showers are explicitly formed from the high resolution model dynamics. Both precipitation measures show a diurnal cycle, with slightly different timings. For the ensemble member forecast, heavier rain (on average) is seen in the late morning and early afternoon, with rain rates peaking around T+11 and T+34 (14Z), whereas the more widespread rain occurred in the early morning, with fractional coverage peaking around T+3/T+4 and T+27 (06Z/07Z). Notice that this cycle is more pronounced for the North domain, particularly for the fractional coverage. The Radarnet data shows a different cycle to the ensemble data, particularly for the averaged rain rates. In particular, the radar sees the heaviest rain rates later in the day (late afternoon and evening) and has a less pronounced cycle. This suggests that the model may be overemphasising and mistiming the diurnal cycle, possibly due to the convection only being partially resolved at the model grid spacing used here (2.2 km), and due to spin-up effects. The diurnal cycle will be further discussed in Section 8.6 in the context of the spatial analysis.



(a) Average of raining points

(b) Fraction of raining points

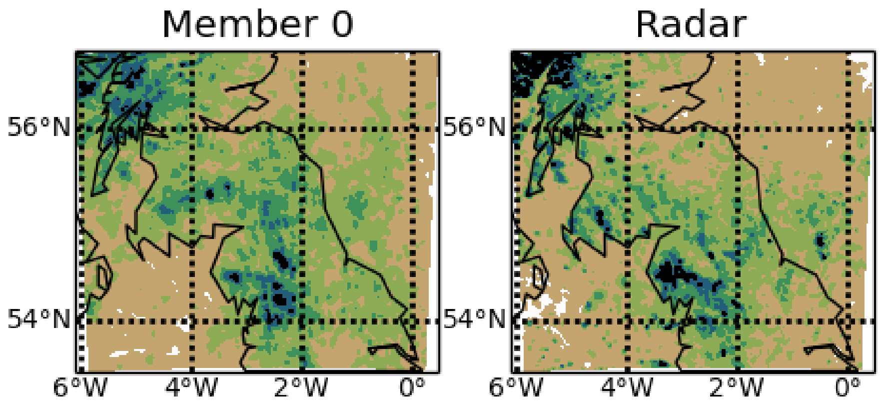
Figure 8.3: (a) The average rain rate of all points in the domain with rain rates exceeding $0.01 mm\,hr^{-1}$ and (b) the fraction of points in the domain with rain rates exceeding $0.01 mm\,hr^{-1}$ as a function of forecast lead time. In both (a) and (b) the average value has been taken over all cases in JJA 2013. Results are shown for one ensemble member (the control, red) and for Radarnet data (black): the same averaging method was used for all results. The area shown in grey corresponds to the 24 hour averaging period (T+6; 09Z to T+29; 08Z the following day) used in Figures 8.2, and 8.4 to 8.9.

To investigate the average spatial distribution of precipitation, Figure 8.4 shows maps, for the North and South domain, of the instantaneous rain rates averaged over all cases in JJA 2013 and forecast lead times T+6 to T+29 (09Z on day of forecast start to 08Z the following day). Note that here, to allow the relative amounts of precipitation at different locations in the domain to be compared, the average is take over all times in JJA 2013, including those times with no rain. This contrasts with the results shown in Figures 8.2 and 8.3 where only raining points are included in the average. Results are shown for an example ensemble member and the Radarnet data (other ensemble members give the same conclusions). Overall the North domain was wetter than the South domain, and more rain fell along the west coast and over the hills. A similar overall pattern is seen for ensemble member forecasts and the radar observations; there are however subtle differences, particularly in the positioning of the wettest areas. For example, the wet region over the Lake District is more pronounced, and extends further to the west in the Radarnet data. Overall, the Radarnet data has higher averaged rain rates than the ensemble members, particularly over mountainous regions such as southwest Scotland (Mull) and northwest Wales (Snowdonia). This could be due to an underestimation of the precipitation by the model, or alternatively due to radar errors over orography, for example due to the estimates made for orographic enhancement. Notice also that there are isolated regions of heavy precipitation, particularly in the radar data over the English Channel. During JJA 2013 there were several localised intense convective storms over the English Channel, which could be showing through in the average, given the relatively short averaging period. Alternatively, this could again indicate problems with the Radarnet composite in the Channel for example due to shipping (which can give false signals). These features do not show up in a 2008–2014 climatology of Radarnet data (Warren, 2014), or in spatial results presented in this chapter. Interestingly, Fairman et al. (2015) do see streaks of higher than average frequency of precipitation in a 2006-2014 Radarnet climatology of the frequency of precipitation exceeding $0.01 mm\, hr^{-1}$.

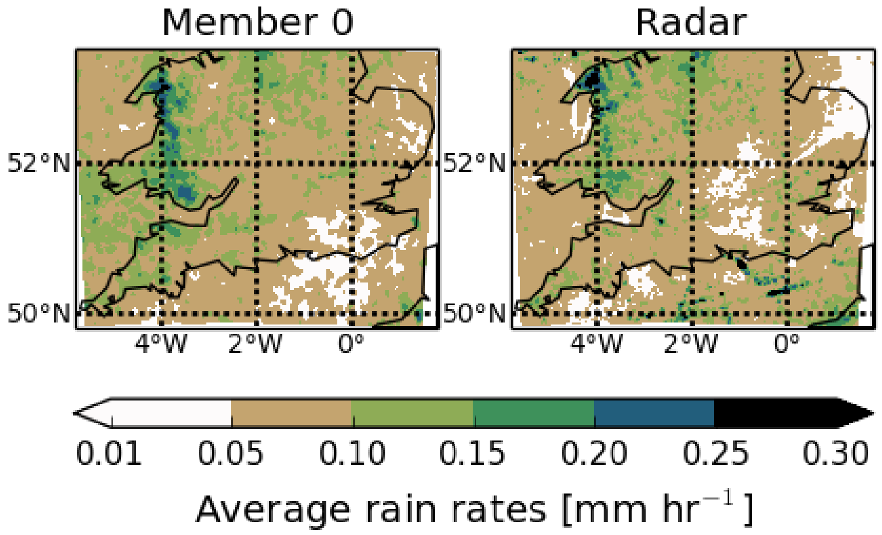
8.4 Spatial agreement from the ensemble

In this section the spatial predictability of precipitation, as forecast by the MOGREPS-UK ensemble, will be discussed. The spatial predictability is measured using the $S_{ij}^{A(\overline{mm})}$, calculated as described in Chapter 6 (Section 6.3). To give an overview of the spatial predictability over a summer season, the average is taken of all cases in JJA 2013. Figure 8.5a shows the $S_{ij}^{A(\overline{mm})}$ averaged over forecast lead times from T+6 to T+29 (09Z on day one to 08Z the following day), and calculated over the whole of the constant resolution part of the MOGREPS-UK domain (i.e. results for the north and south domains are not separately calculated). This provides an overview of the ensemble's confidence in the location of precipitation (spatial predictability) across the UK for that summer, and indicates the typical scales over which the model precipitation data should be interpreted as having spatial uncertainties (assuming a good spread-skill relationship). The time averaged $S_{ij}^{A(\overline{mm})}$ are smaller in the northwest, over mountainous regions, and along the west coast of both the UK and Ireland: in these regions the model is more confident about the location of precipitation. Overall, the ensemble is not confident about the location of precipitation close to the grid scale, with a minimum time-mean $S_{ij}^{A(\overline{mm})}$ of around 16 grid points (a total neighbourhood length of 72.2km).

For the six COPE cases discussed in Chapter 6 smaller agreement scales were often seen in regions of precipitation. Hence, it is possible that the pattern of scales shown in Figure 8.5a is simply caused by points being wetter to the north and west of the domain.



(a) North domain



(b) South domain

Figure 8.4: Rain rates averaged over all times in JJA 2013, including times with zero rain rates. A threshold of $0.01 mm\ hr^{-1}$ was applied, with all rain rates below the threshold set to zero. Forecasts were taken from T+6 (09Z) to T+29 (08Z the following day) inclusive.

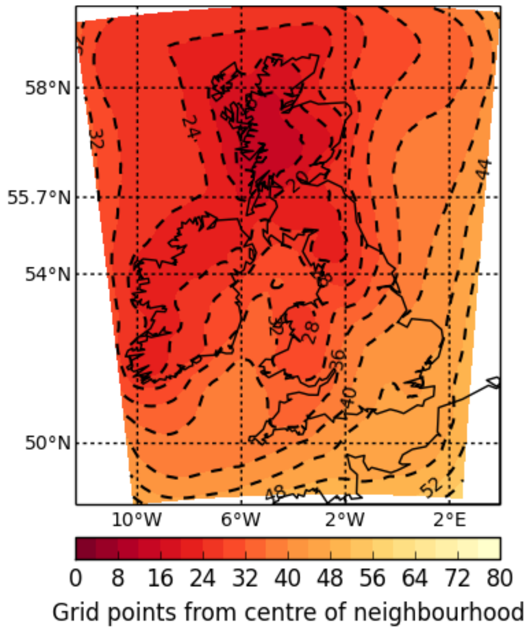
However, a scatter plot of all points in the average $S_{ij}^{A(\overline{mm})}$ map in Figure 8.5a against all points in the rain rate maps from one ensemble member averaged over the same period, shows that other factors are also important (Figure 8.5b). The points in Figure 8.5b cluster into two groups depending on their position north/south of a latitude of 55.7°N. Both of these clusters show that higher average rain rates (wetter points averaged over JJA 2013) are associated with smaller agreement scales. However, the splitting of the points into two clusters shows that the $S_{ij}^{A(\overline{mm})}$ does not simply indicate that rain is more predictable where there is more of it; there are other factors involved. In particular, high mean rain rates (greater than 0.15 $mm\ hr^{-1}$ in Figure 8.5b) north/south of 55.7°N cluster around an

$S_{ij}^{A(\overline{mm})}$ of 15 grid points/22 grid points respectively, suggesting that, in JJA 2013, the rain in north and central Scotland was more predictable than that in the rest of the UK. Possible reasons for this include the higher orography in this region (constraining where the rain falls, for example through up-slope enhancement or rain shadows) and the direct exposure to the Atlantic. Although these are the clearest clusters in the scatter plot, similar conclusions can be drawn from comparing other regions e.g. the Cumbrian hills and the Pennines (the precipitation over Cumbrian hills has higher spatial agreement than that over the Pennines). Although less clear cut, clustering is also seen in a scatter plot of orography height against $S_{ij}^{A(\overline{mm})}$ (Figure 8.5c). This confirms that, although higher rain rates (and smaller $S_{ij}^{A(\overline{mm})}$) are seen over the hills, this is not only due to the height of the orography: other factors, such as the orographic gradients perhaps, are also important. Similar clustering can be seen when comparing the $S_{ij}^{A(\overline{mm})}$ with JJA average radar rain rates, and when comparing with the ensemble mean. Similar results are also found for the $S_{ij}^{A(\overline{mo})}$.

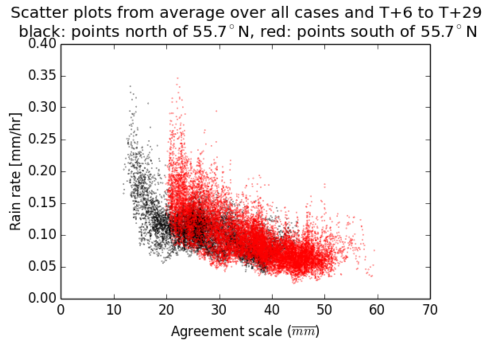
8.4.1 Dependence of spatial ensemble spread on precipitation amount

In Section 8.3 it was shown that the precipitation amount (both in terms of the precipitation intensity and precipitation extent) varied considerably over the three month period considered. In this section the effects of this variation on the spatial predictability are investigated. Here, similarly to the methods used in Section 8.3, the precipitation intensity is measured by the average of all points in the domain with rain rates greater than $0.01 mm\, hr^{-1}$, and the precipitation extent (fractional coverage) is measured by the fraction of points in the domain with rain rates greater than $0.01 mm\, hr^{-1}$.

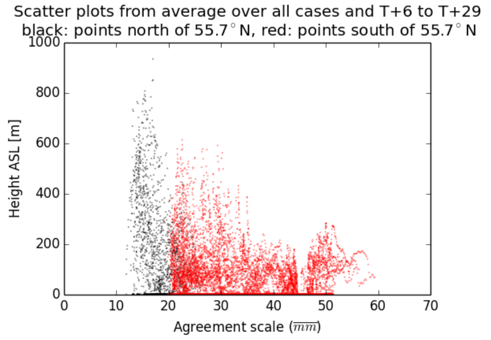
The $S_{ij}^{A(\overline{mm})}$ measure the spatial overlap between ensemble member forecasts, with a larger overlap (smaller spatial differences) giving smaller $S_{ij}^{A(\overline{mm})}$. Hence it is expected that, at times where the precipitation covers a large/small part of the domain, there will more/less overlap between the fields and smaller/larger $S_{ij}^{A(\overline{mm})}$. Additionally, large scale (frontal) precipitation tends to be more predictable at local scales than small scale precipitation such as scattered convection. This was found to be true for the six convective cases discussed in Chapter 6. Here the relationship between $S_{ij}^{A(\overline{mm})}$ and fractional coverage of precipitation across the domain is investigated by splitting the JJA 2013 $S_{ij}^{A(\overline{mm})}$ data by fractional coverage. Figure 8.6 shows $S_{ij}^{A(\overline{mm})}$ for the North domain



(a) $S_{ij}^{A(\overline{mm})}$ time mean



(b) Rain rate against $S_{ij}^{A(\overline{mm})}$

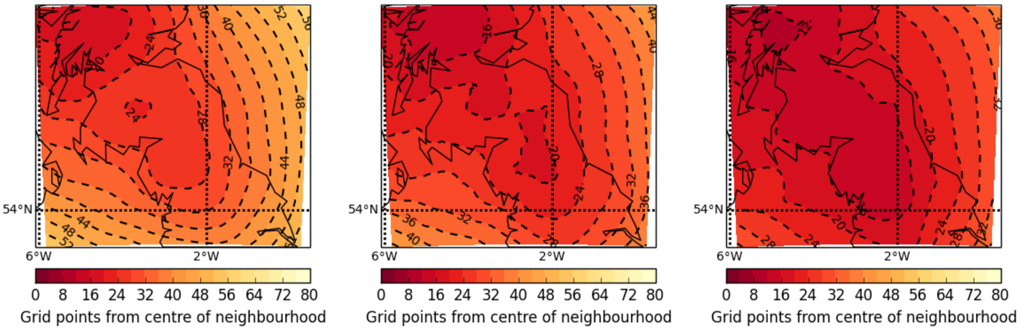


(c) Orography height against $S_{ij}^{A(\overline{mm})}$

Figure 8.5: (a) Map of $S_{ij}^{A(\overline{mm})}$ averaged for the 24 hours from T+6 to T+29 for JJA 2013. (b) Scatter plot of every point from the $S_{ij}^{A(\overline{mm})}$ map shown in (a) against the rain rate average from an ensemble member (the control) over the same period. (c) Scatter plot of every point from the $S_{ij}^{A(\overline{mm})}$ map shown in (a) against every point in the equivalent orography map. In (b), (c) points are coloured black/red according to their position north/south of a latitude of 55.7°N. This latitude is indicated in (a) and was subjectively chosen to provide the cleanest split between the two clusters of points in (b), (c) and lies along the Scottish lowlands.

from cases and times in JJA 2013 in which rain (greater than $0.01 mm\ hr^{-1}$) in an example ensemble member (here the control; the choice of ensemble member does not impact upon the results) had a fractional coverage of 1% to 10%, 10% to 20%, or 20% to 100% of the domain. Consistent with the discussion above, it is indeed seen that the $S_{ij}^{A(\overline{mm})}$ decrease as fractional coverage is increased. Nevertheless, the overall $S_{ij}^{A(\overline{mm})}$ pattern is retained for different fractional coverage. Similar results were obtained for the South domain (not shown).

Figure 8.7 shows maps of the $S_{ij}^{A(\overline{mm})}$ split by the average rain rate of points with rain rates greater than $0.01 mm\ hr^{-1}$ for both the North and South domains. Unlike the

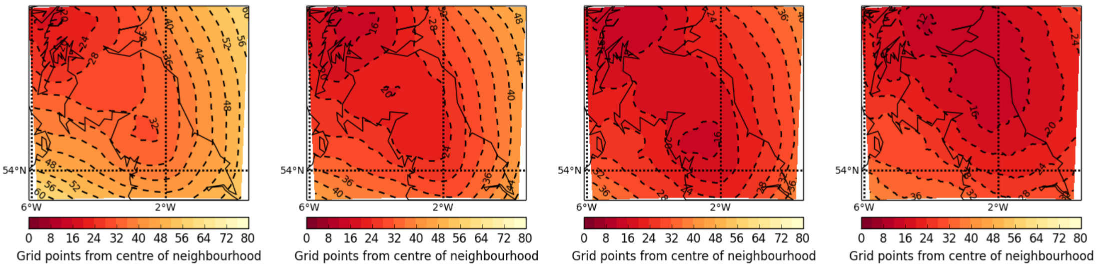


(a) Coverage from 1% to 10% (b) Coverage from 10% to 20% (c) Coverage from 20% to 100%

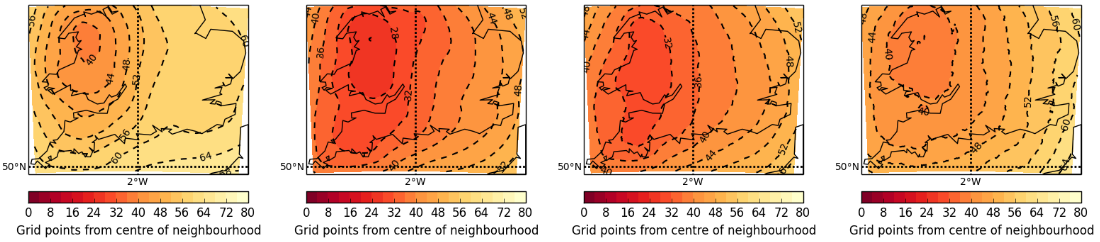
Figure 8.6: Maps of $S_{ij}^{A(\overline{mm})}$ for different fractional coverage of points in the domain with rain rates greater than 0.01 $mm\,hr^{-1}$ for the North domain averaged over times where rain was seen in the specified range. Only forecast lead times from T+6 to T+29 (09Z to 08Z the following day) were considered. Results for precipitation coverage between (a) 1% and 10% (b) 10% and 20%, and (c) 20% and 100% are shown with the coverage derived from the rain rates of a single ensemble member.

results for varying fractional coverage, there is not a monotonic trend in the $S_{ij}^{A(\overline{mm})}$ when different average rain rates are considered. For the lowest rain rate range considered (0.01 to 0.1$mm\,hr^{-1}$) relatively large $S_{ij}^{A(\overline{mm})}$ are seen for both domains (c.f. Figure 8.5), showing that the exact location of the precipitation for these low rain rate events is more uncertain than for higher average rain rates. For both domains higher spatial agreement is seen for rain rates in the range 0.1 to 1.0$mm\,hr^{-1}$: precipitation in these ranges is more spatially predictable. This is due to larger-scale frontal precipitation tending to fall into this category as shown by a scatter plot of the fractional coverage against average rain rates for each forecast time from T+6 to T+29 in JJA 2013 (Figure 8.8). In Figure 8.8 each point is coloured according to the average value of the rain rate spatial gradients from one ensemble member (the same member as that being used to calculate the average rain rates and fractional coverage) at that time. The gradients were calculated, at each point in the domain, by taking the maximum difference between the rain rate at that point, and at its neighbouring points, and provide an indication of the convective activity present at a given time: higher precipitation gradients are associated with convective (as apposed to stratiform) precipitation. This method was developed by Nigel Roberts, using a similar approach to Steiner et al. (1995), and was used in Warren (2014). To maintain consistency with the fractional coverage and average rain rates, in Figure 8.8 the rain rate gradients are averaged over all points in the domain with rain rates greater than 0.01$mm\,hr^{-1}$.

When considering rain rates above 1.0$mm\,hr^{-1}$ different behaviour is seen for the

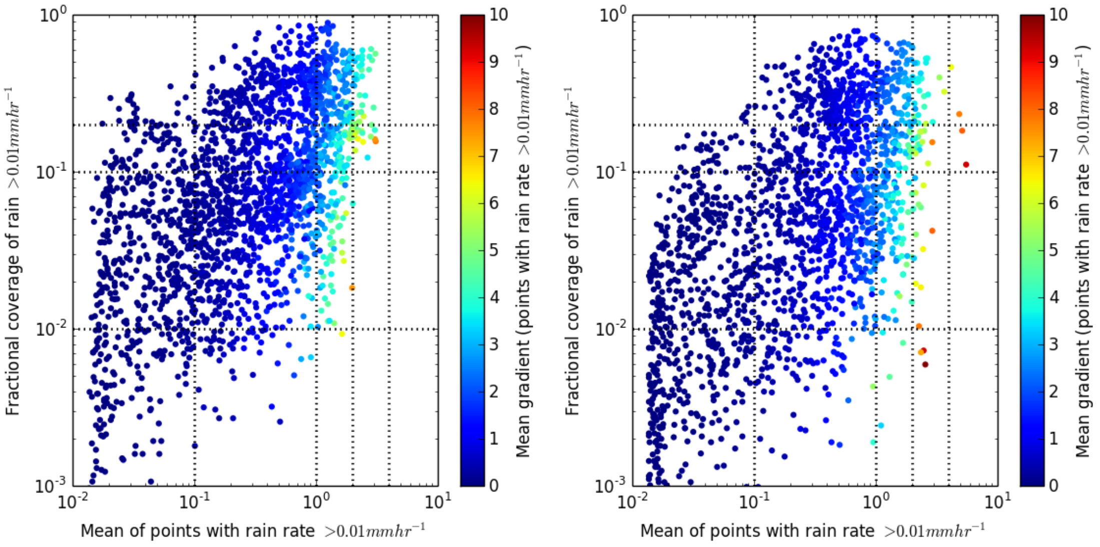


(a) 0.01 to 0.1$mm\,hr^{-1}$ (b) 0.1 to 1.0$mm\,hr^{-1}$ (c) 1.0 to 2.0$mm\,hr^{-1}$ (d) 2.0 to 4.0$mm\,hr^{-1}$



(e) 0.01 to 0.1$mm\,hr^{-1}$ (f) 0.1 to 1.0$mm\,hr^{-1}$ (g) 1.0 to 2.0$mm\,hr^{-1}$ (h) 2.0 to 4.0$mm\,hr^{-1}$

Figure 8.7: Maps of the $S_{ij}^{A(\overline{mm})}$ split by the average rain rate of points with rain rates greater than 0.01$mm\,hr^{-1}$. A single ensemble member (the control) was used for this categorisation. Results are shown for the North domain (top) and South domain (bottom). Data were averaged over cases where rain was seen in the specified rain rate range: forecast lead times from T+6 to T+29 were included from JJA 2013.



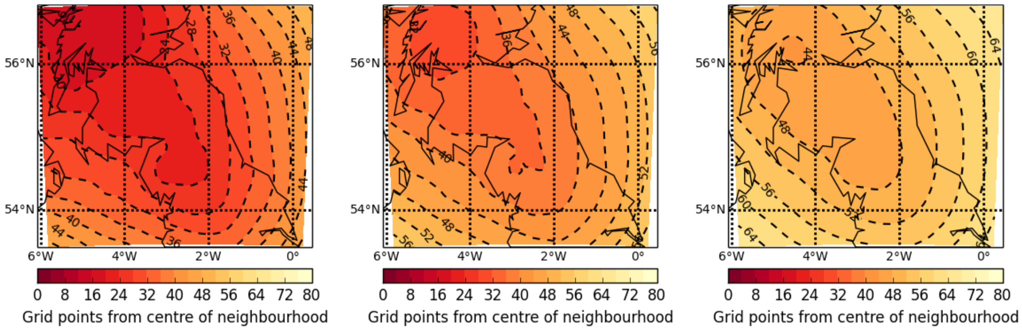
(a) North domain (b) South domain

Figure 8.8: Scatter plots of the fraction of points in the domain with rain rates greater than 0.01$mm\,hr^{-1}$ against the average of all points in the domain with rain rates greater than 0.01$mm\,hr^{-1}$ for forecast lead times T+6 to T+29 in JJA 2013. Points are coloured according to the average gradient of the rain rate field (over points with rain rates greater than 0.01$mm\,hr^{-1}$) at the time considered. The black dotted lines highlight the values used to group the cases in Figures 8.6 and 8.7: 0.1, 1.0, 2.0 and 4$mm\,hr^{-1}$; 1, 10 and 20%.

North and South domains. For the South domain the $S_{ij}^{A(\overline{mm})}$ increase when rain rates increase above $1.0 mm\,hr^{-1}$; for the North domain the $S_{ij}^{A(\overline{mm})}$ continue to decrease for rain rates above $1.0 mm\,hr^{-1}$. Hence, the rain rate ranges with the highest spatial agreement (lowest $S_{ij}^{A(\overline{mm})}$) are different for the two domains: 0.1 to $1.0 mm\,hr^{-1}$ for the South domain and 2.0 to $4.0 mm\,hr^{-1}$ for the North domain. Comparing Figures 8.8a and 8.8b (scatter plots for the North and South domains respectively) we see that the South domain has more points in the scatter plot with high rain rates and high gradients, but with low fractional coverage. For example, there are a number of points with rain rates from 2 to 4 $mm\,hr^{-1}$, with gradients above $6 mm\,hr^{-1} grid\ length^{-1}$, and fractional coverage from 1 to 10%. This suggests that, for the South domain, high rain rates tend to be associated with low coverage convective events, which are expected to be less spatially predictable. In contrast, for the North domain, the high rain rates (2 to 4 $mm\,hr^{-1}$) tend to have lower gradients (around $5 mm\,hr^{-1} grid\ length^{-1}$) and higher fractional coverage (above 20%). This suggests that heavy precipitation in the North domain tends to be embedded inside larger scale, more predictable, precipitation structures.

8.4.2 Varying precipitation thresholds

As discussed in Section 8.2.2 different thresholds were applied to the local instantaneous rain rates (i.e. at each grid point in the domain) to investigate how the spatial agreement of precipitation depends on the rain rates considered. Note that, in applying a local precipitation threshold in this subsection, we are considering both different precipitation coverage and different precipitation intensity. Hence, we would expect the $S_{ij}^{A(\overline{mm})}$ dependence on threshold to reflect the results of Section 8.4.1. Figure 8.9 shows the $S_{ij}^{A(\overline{mm})}$ calculated for the North domain with local thresholds of 0.01, 0.1 and $1.0 mm\,hr^{-1}$, averaged over JJA 2013 and forecast lead times T+6 to T+29. To focus on the $S_{ij}^{A(\overline{mm})}$ behaviour at times with precipitation, times with no precipitation above the threshold value (at any point in the domain) were not included in this average. The pattern of $S_{ij}^{A(\overline{mm})}$ stays the same when different thresholds are considered, with smaller scales to the northwest. For higher thresholds larger $S_{ij}^{A(\overline{mm})}$ are seen: these thresholds select heavier, localised values of precipitation, which have larger spatial separations and are therefore less spatially predictable. Similar results are found for the South domain.



(a) Threshold $0.01 mm\,hr^{-1}$ (b) Threshold $0.1 mm\,hr^{-1}$ (c) Threshold $1.0 mm\,hr^{-1}$

Figure 8.9: Maps of $S_{ij}^{A(\overline{mm})}$ for different local precipitation thresholds for the North domain averaged over forecast lead times T+6 to T+29 hrs where precipitation occurred over the specified threshold (at at least one grid point in the domain). Results are shown for rain rates greater than (a) $0.01 mm\,hr^{-1}$, (b) $0.1 mm\,hr^{-1}$, and (c) $1.0 mm\,hr^{-1}$.

8.5 Traditional spread-skill measures

To verify the ensemble performance, it is necessary to compare the spread of ensemble member forecasts with the skill of the ensemble in capturing the observed precipitation (the spread-skill relationship). Before discussing the spatial spread-skill relationship (Section 8.6) it is informative to consider first the traditional measures of ensemble spread and skill discussed in Chapter 2. Here the ensemble spread is taken to be the ensemble variance (Var) averaged over all points in the domain, and the ensemble skill is taken to be the mean squared error (MSE) of the ensemble mean compared to Radarnet observations. To account for the effect of using a limited sample size, the variance and MSE are normalised as suggested by Eckel and Mass (2005). This gives

$$\mathrm{MSE} = \left(\frac{N}{N-1}\right)\frac{1}{M}\sum_{m=1}^{M}\left(\overline{e_m} - o_m\right)^2 \tag{8.1}$$

and

$$\mathrm{Var} = \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{N-1}\sum_{i=1}^{N}\left(e_{m,i} - \overline{e_m}\right)^2\right] \tag{8.2}$$

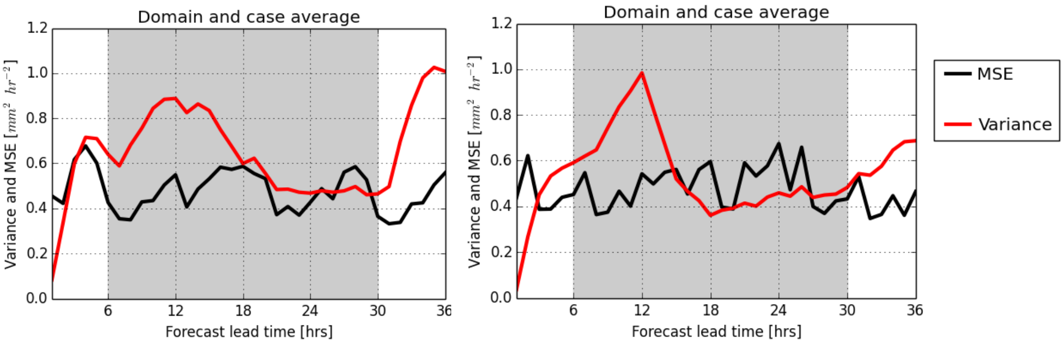
where M is the number of samples (here the number of points in the domain multiplied by the number of cases), N the number of ensemble members, $\overline{e}_m$ the ensemble mean at point m, and o_m the Radarnet value at point m. Here we assume that the Radarnet data is representative of the true meteorological situation: observation error is ignored.

The rain rate variance and MSE are shown as a function of forecast lead time in

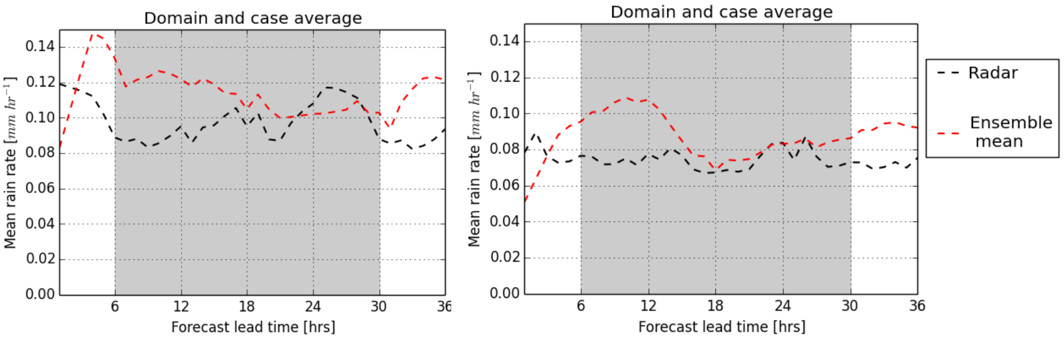
Figures 8.10a and 8.10b for the South and North domains respectively. Both the MSE and variance are noisy and, although there are clear differences between them (for example the ensemble variance shows a diurnal cycle not seen in the MSE) these are hard to interpret. Also shown in Figure 8.10 (sub-figures c and d) are the ensemble mean and Radarnet rain rate data, averaged over all points in the domain and all cases in JJA 2013. Note that, in Figure 8.10 the ensemble and Radarnet data are averaged over all points in the domain, including points with no rain, and hence the mean values are lower than those shown in Figure 8.3 in which the average was only taken over raining points. The variance and MSE follow variations in the ensemble mean and the radar data respectively; when the ensemble mean is greater than/less than the radar mean, the variance is greater than/less than the MSE. This is particularly clear in the first 6 hours of the forecast when the ensemble mean increases by around $0.025 mm\, hr^{-1}$ due to model spin-up. This suggests that the MSE and variance are strongly influenced by the mean state and are not giving independent information about the ensemble spread and skill. This is expected, and has been shown to be true statistically for precipitation distributions by e.g. Hamill and Colucci (1998).

8.6 Spatial spread-skill

Given the problems with traditional measures of ensemble spread and skill illustrated above in Section 8.5 and discussed in Chapter 2, this section provides an alternative spatial view of ensemble spread and skill over JJA 2013, using the techniques presented in Chapter 6. In particular, the $S_{ij}^{A(\overline{mm})}$ (a measure of the spatial ensemble spread) is compared with the $S_{ij}^{A(\overline{mo})}$ (a measure of the spatial ensemble skill). Spatial spread-skill results are presented, both for the domain average, and also using the binned scatter plot (introduced in Chapter 6, Section 6.4.5). Several key features of the spatial spread-skill relationship are identified and discussed in more detail in Sections 8.6.1 and 8.6.2. Finally, in Section 8.6.3 the effect of the amount of precipitation (both in terms of intensity and spatial extent) on the spatial spread-skill relationship is investigated.



(a) North domain (b) South domain

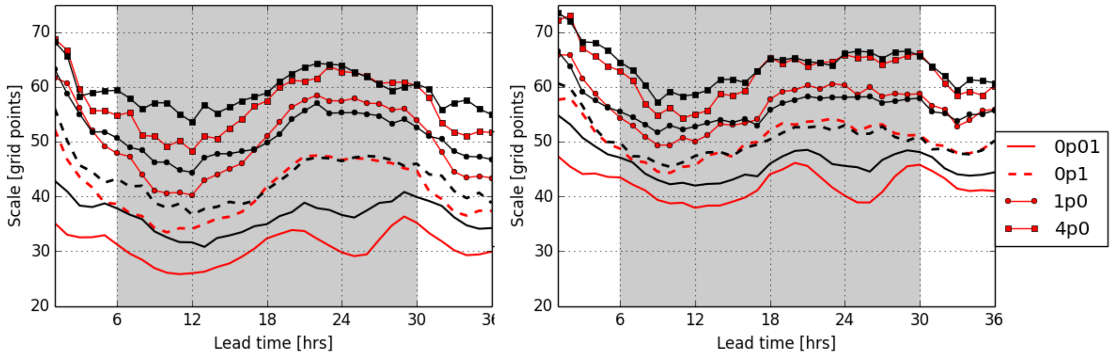


(c) North domain (d) South domain

Figure 8.10: Time series of (a, b) the domain-averaged MSE of the ensemble mean and the ensemble variance, and (c, d) the radar and ensemble mean. All quantities were calculated for the whole domain, and averaged over all cases in JJA 2013. Correction factors were applied to the MSE and variance calculations to account for the limited sample size as described in Eckel and Mass (2005). Results are shown for the North domain (left) and South domain (right). The area shown in grey corresponds to the 24 hour averaging period (T+6; 09Z to T+29; 08Z the following day) used in Figures 8.2, and 8.4 to 8.9.

Domain average

Figure 8.11 shows the domain averaged $S_{ij}^{A(\overline{mm})}$ (red) and $S_{ij}^{A(\overline{mo})}$ (black) for the North domain (Figure 8.11a) and South domain (Figure 8.11b). Four different local precipitation thresholds are used (0.01, 0.1, 1.0 and 4.0$mm\ hr^{-1}$), as was done in Section 8.4.2. Results have been averaged over all cases where rain, at some point in the domain (for both the radar data, and at least one ensemble member, see Section 8.2.2), exceeds the specified threshold, giving an average of between 50 (for the 4.0$mm\ hr^{-1}$ threshold) to 92 (for the 0.01$mm\ hr^{-1}$) cases. This gives a standard error in the mean agreement scale (standard deviation/$\sqrt{\text{number of cases}}$) of 1 to 2 grid points for the 0.01$mm\ hr^{-1}$ threshold and up to 4 grid points for the 4.0$mm\ hr^{-1}$ threshold. Hence, the differences seen in Figure 8.11 between different threshold results, and between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$, are significant. Consistent with the $S_{ij}^{A(\overline{mm})}$ behaviour for different precipitation thresholds (discussed in Section 8.4.2), higher thresholds give larger domain-averaged $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$; there is more spatial uncertainty and larger spatial errors for higher thresholds. The domain-average spatial ensemble spread ($S_{ij}^{A(\overline{mm})}$) and spatial ensemble skill ($S_{ij}^{A(\overline{mo})}$) are smoother and easier to interpret than traditional measures. As expected from the $S_{ij}^{A(\overline{mm})}$ maps (e.g. Figure 8.5a), the North domain has lower $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ than the South domain for a given threshold, again showing that the precipitation over the North domain is in higher spatial agreement than that over the South domain. This difference is larger than the standard error in the mean, and is not simply attributable to the occurrence of more widespread rain in the North domain: for example, smaller agreement scales were still obtained for the North domain than for the South domain when the fractional coverage of precipitation across the domain was restricted to be between 20% and 40%.



(a) North domain (b) South domain

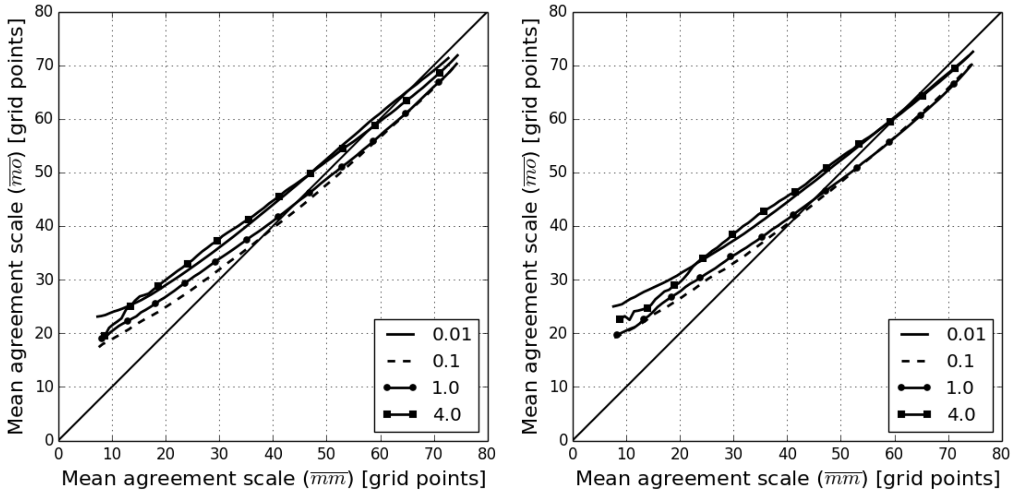
Figure 8.11: Domain-averaged $S_{ij}^{A(\overline{mm})}$ (red) and $S_{ij}^{A(\overline{mo})}$ (black) for the thresholds $0.01 mm\,hr^{-1}$ (solid), $0.1 mm\,hr^{-1}$ (dashed), $1.0 mm\,hr^{-1}$ (solid with circles) and $4.0 mm\,hr^{-1}$ (solid with squares) for (a) the North domain and (b) the South domain. At each forecast lead time, the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ were averaged over all cases in JJA 2013 where rain rates, at one or more points in the domain, exceeded the required threshold for both the radar data and at least one ensemble member. Errors associated with the mean over all cases are around 1 to 2 grid points for the $0.01 mm\,hr^{-1}$ threshold, increasing up to 4 grid points for the $4.0 mm\,hr^{-1}$ threshold. The area shown in grey corresponds to the 24 hour averaging period (T+6; 09Z to T+29; 08Z the following day) used in Figures 8.2, and 8.4 to 8.9.

Despite the difference between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ magnitudes for the North and South domains, the relationship between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ is similar for both. In particular, both domains show a threshold dependence of spatial spread-skill with results from the lowest precipitation threshold ($0.01 mm\,hr^{-1}$) showing a larger difference between the spatial spread and spatial skill ($S_{ij}^{A(\overline{mm})}$ less than $S_{ij}^{A(\overline{mo})}$) than the other thresholds. This result suggests that the ensemble is more spatially under spread when including low value rain rates in the agreement scales calculations. The reasons for this behaviour will be discussed in Section 8.6.1.

Another feature of Figure 8.11 is the temporal evolution of the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ magnitudes, and of the difference between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$. Initially, the agreement scale values (both $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$) decrease with time, reaching a minimum around T+12 hrs (15Z). After T+12 the agreement scales increase until around T+24 hrs (03Z) before decreasing again for the remainder of the forecast. This suggests a diurnal cycle, with smaller agreement scales (higher spatial agreement) in the afternoon and larger agreement scales (lower spatial agreement) at night. This cycle is seen for thresholds greater than $0.1 mm\ hr^{-1}$ and is more pronounced for $S_{ij}^{A(\overline{mm})}$ than $S_{ij}^{A(\overline{mo})}$, although these differences are smaller than the error in the mean values. Dey et al. (2016b, under review) show that (for thresholds greater than $0.1 mm\ hr^{-1}$) this diurnal cycle in the domain averaged agreement scales correlates negatively with the diurnal cycle in precipitation intensity or fractional coverage across the domain. Also, the diurnal cycle of differences between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ ($S_{ij}^{A(\overline{mm})}$ - $S_{ij}^{A(\overline{mo})}$), is shown to correlate negatively with the diurnal cycle of differences in precipitation intensity or fractional coverage between ensemble members and radar. These results show that the spatial spread-skill relationship is influenced by the amount of rain present in the domain, with the ensemble appearing more spatially under spread ($S_{ij}^{A(\overline{mm})}$ less than $S_{ij}^{A(\overline{mo})}$) when the amount of precipitation is overestimated by the ensemble. Interestingly, this suggests the opposite dependency to that found for the traditional spread-skill relationship, where an overestimation of precipitation by the ensemble is linked to the ensemble being over spread (Hamill and Colucci, 1998). Future work should focus on quantifying these links between the bias and the spread-skill relationship for both traditional and spatial measures. The effect of using convection permitting ensembles (as used for the spatial measures presented in this thesis), compared to larger-scale or synthetic ensembles (as used by Hamill and Colucci (1998)) should also be investigated. The location-dependence of the temporal evolution of the spatial spread and spatial skill will be discussed in more detail in Section 8.6.2.

Binned Scatter Plots

Although the domain average provides an overview of the relationship between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ it is also useful to consider a scale selective comparison method, where the scale dependence of the spatial spread-skill can be investigated. Here this is done using the binned scatter plot (Chapter 6, Section 6.4.5), which has the additional advantage of preserving the point-to-point relationship between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ (i.e the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are compared at the same locations). Binned scatter plots, averaged over all cases in JJA 2013 and forecast lead times T+6 to T+29, and with a bin size of 10 grid points, are given in Figures 8.12a and 8.12b for the North and South domains respectively. Note that the binned scatter traces were separately calculated for each individual time before averaging over cases and lead times. Results are shown for the precipitation thresholds 0.01, 0.1, 1.0 and 4.0$mm\,hr^{-1}$ (the same sample as used for the domain average results in Figure 8.11).



(a) North domain (b) South domain

Figure 8.12: Binned scatter plots averaged over JJA 2013 and lead times T+6 to T+29 for thresholds 0.01 (solid), 0.1 (dashed), 1.0 (solid with circles) and 4.0 (solid with squares) $mm\,hr^{-1}$ for (a) the North domain and (b) the South domain.

Similar behaviour is seen for all four thresholds and for both domains, with all traces lying close to the diagonal: in general, for this three month period, the ensemble was reasonably well spread spatially. There are however some deviations from the diagonal, with results from all precipitation thresholds showing that the ensemble is spatially under spread ($S_{ij}^{A(\overline{mm})}$ less than $S_{ij}^{A(\overline{mo})}$). Thus there is higher spatial agreement between ensemble member forecasts than there is between ensemble members and the Radarnet data. This is particularly noticeable for agreement scales less than 45 grid points: when the ensemble is most confident about the location of precipitation (small $S_{ij}^{A(\overline{mm})}$) it is over confident (spatially under spread, $S_{ij}^{A(\overline{mm})}$ less than $S_{ij}^{A(\overline{mo})}$). Results from the $0.01 mm\,hr^{-1}$ threshold are the most under spread with an average $S_{ij}^{A(\overline{mo})}$ of 25 grid points for an average $S_{ij}^{A(\overline{mm})}$ of 10 grid points. The 0.1 and $1.0 mm\,hr^{-1}$ thresholds move closer to the 1–1 line and have a better spread skill relationship. For the highest threshold considered, $4.0 mm\,hr^{-1}$, the ensemble is again more under spread at small scales as the uncertainty in higher intensity values is harder to quantify. This threshold dependence of spatial spread-skill agrees with the domain average results (Figure 8.11).

8.6.1 Sensitivity of spatial spread-skill to low precipitation thresholds

It has been shown that the relationship between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ is different when using a $0.01 mm\,hr^{-1}$ precipitation threshold, to that obtained using higher thresholds ($0.1 mm\,hr^{-1}$ or greater). This behaviour was further investigated by comparing the differences between ensemble member and Radarnet rain rate fields with different thresholds applied. In particular, the focus was on points with rain rates between 0.01 and $0.1 mm\,hr^{-1}$, as these were the points set to zero when calculating the $0.1 mm\,hr^{-1}$ threshold results, but not for calculating the $0.01 mm\,hr^{-1}$ threshold results. Figure 8.13 shows the total number of points with rain rates in the range $0.01 mm\,hr^{-1}$ to $0.1 mm\,hr^{-1}$ for an example ensemble member (here the control, the results are not sensitive to the choice of ensemble member) and for the Radarnet data. It can be seen that the ensemble member has many more points with rain rates from $0.01 mm\,hr^{-1}$ to $0.1 mm\,hr^{-1}$ than the Radarnet data. Interestingly, these differences are seen across all cases in JJA 2013, with particularly large differences in the wet period around the end of July (identified with reference to Figure 8.2). This suggests that it is predominantly the wet cases that are causing the differences in spatial spread skill between the $0.01 mm\,hr^{-1}$ to $0.1 mm\,hr^{-1}$ threshold results.

Visual inspection of precipitation maps showed that it was differences in the precipitation structures of the ensemble forecasts and Radarnet observations which caused the difference in spread-skill between the $0.01 mm\,hr^{-1}$ to $0.1 mm\,hr^{-1}$ threshold results. In particular, it was found that the ensemble forecasts had more fragmented regions of

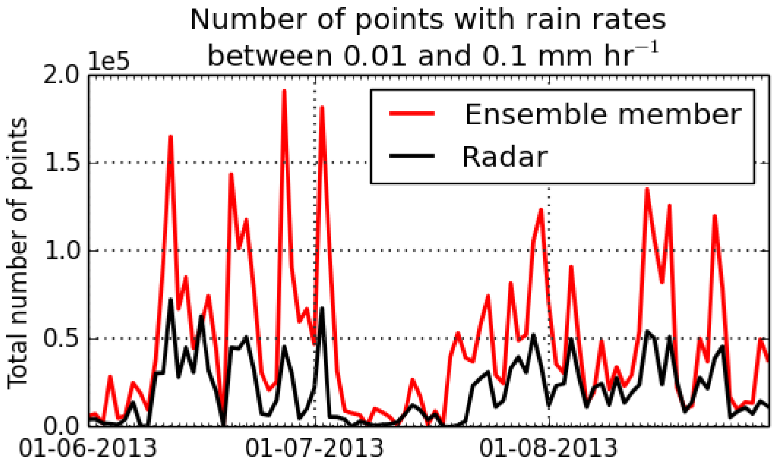
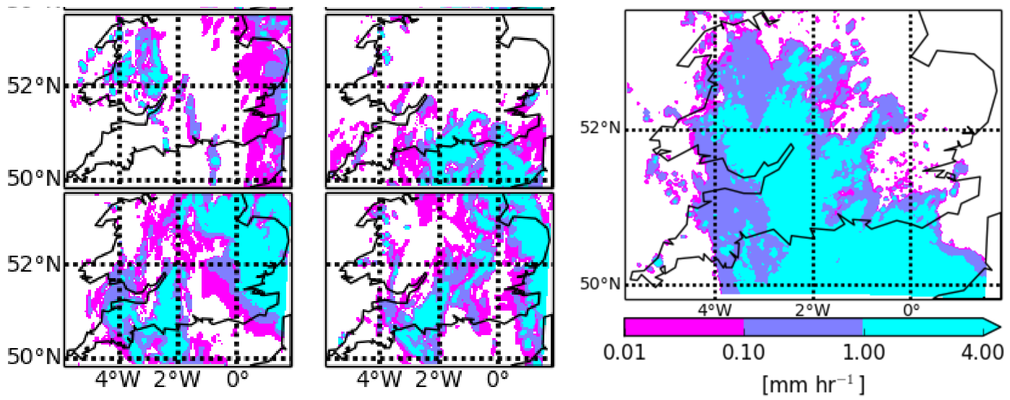


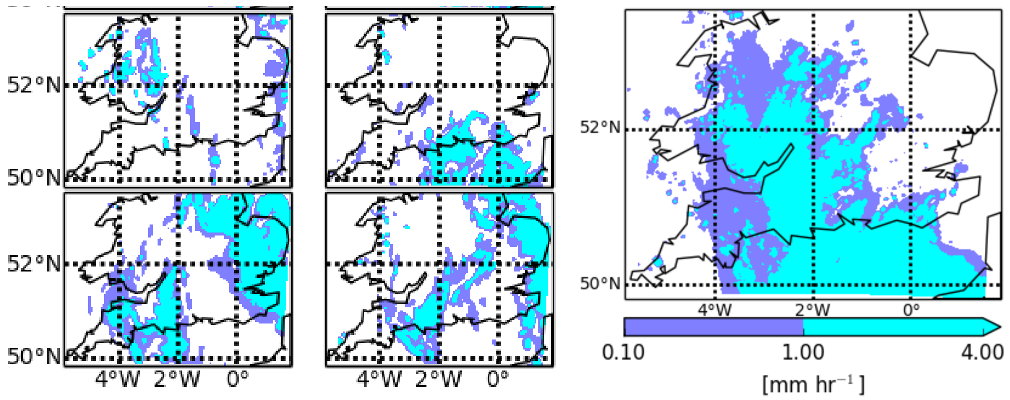
Figure 8.13: Total number of grid points with rain rates in the range $0.01 mm\, hr^{-1}$ to $0.1 mm\, hr^{-1}$ for each case in JJA 2013 and forecast lead times T+6 to T+29. Results are shown for an example ensemble member (red) and radar data (black) for the South domain.

precipitation than the observations, with several peaks occurring within a given precipitation region. In contrast, the Radarnet rain rates tended to have one peak within each precipitation region. An example is given in Figure 8.14 for the 27/07 (one of the cases considered in Chapters 5, 6 and 7) at forecast lead time T+14 (17Z). Figure 8.14a shows four example ensemble member forecasts at this time, each with several maxima in the precipitation field, located at different points in the domain. The Radarnet rain rates at this time (Figure 8.14b) show one broad precipitation region with a single peak in rain rates. Note that, despite their different structures, the ensemble member rain rates are generally in a similar region to the observed rainfall.

Applying a threshold of $0.1 mm\, hr^{-1}$ to the results in Figure 8.14 (removing the pink points in sub-figures a and b to give sub-figures c and d) results in the ensemble member rain rate structures becoming more fragmented as points linking the different maxima are removed. This increases the spatial differences between ensemble member forecasts, and the $S_{ij}^{A(\overline{mm})}$ increases by 3.0 grid points. However, applying a threshold of $0.1 mm\, hr^{-1}$ to the radar rain rates does not alter the overall precipitation structure, as the points with rain rates from $0.01 mm\, hr^{-1}$ to $0.1 mm\, hr^{-1}$ lie on the edge of the precipitation region. Hence, when a threshold of $0.1 mm\, hr^{-1}$ is applied to both the ensemble member forecasts and the Radarnet observations in this example, the overlap between ensemble member forecasts and the observations is only changed a small amount (due to the member forecasts being more fractured) and the $S_{ij}^{A(\overline{mo})}$ increases by only 1.4 grid



(a) Example members, threshold $0.01 mm\,hr^{-1}$ (b) Radar, threshold $0.01 mm\,hr^{-1}$



(c) Example members, threshold $0.1 mm\,hr^{-1}$ (d) Radar, threshold $0.1 mm\,hr^{-1}$

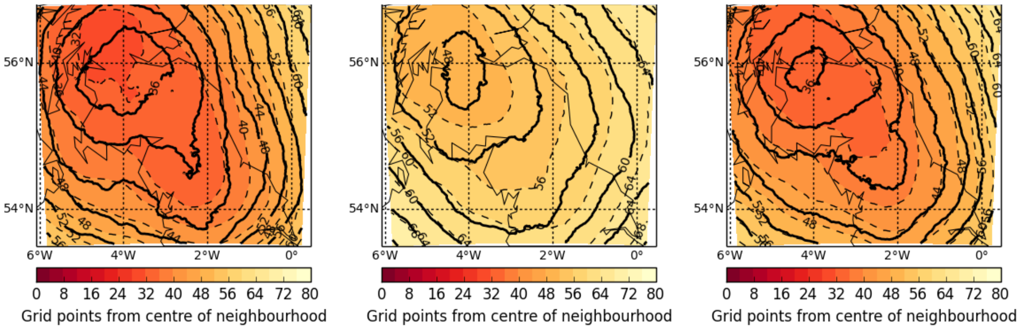
Figure 8.14: Rain rates for example ensemble members (a, c) and Radarnet observations (b, d) at T+14 (17Z) on 27/07/2013 for rain rate thresholds of (a, b) $0.01 mm\,hr^{-1}$ and (c, d) $0.1 mm\,hr^{-1}$. To allow comparison of the different threshold results the same contours are shown for all plots.

points. Thus, when applying a rain rate threshold of $0.1 mm\,hr^{-1}$, the $S_{ij}^{A(\overline{mo})}$ increase slightly (when compared to results with a threshold of $0.01 mm\,hr^{-1}$) but not as much as the $S_{ij}^{A(\overline{mm})}$. Hence, as the ensemble is spatially under spread for a precipitation threshold of $0.01 mm\,hr^{-1}$ ($S_{ij}^{A(\overline{mm})}$ less than $S_{ij}^{A(\overline{mo})}$), the spatial spread skill relationship improves (the difference between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ reduces) when the $0.1 mm\,hr^{-1}$ threshold is applied, consistent with what was found in Figures 8.11 and 8.12.

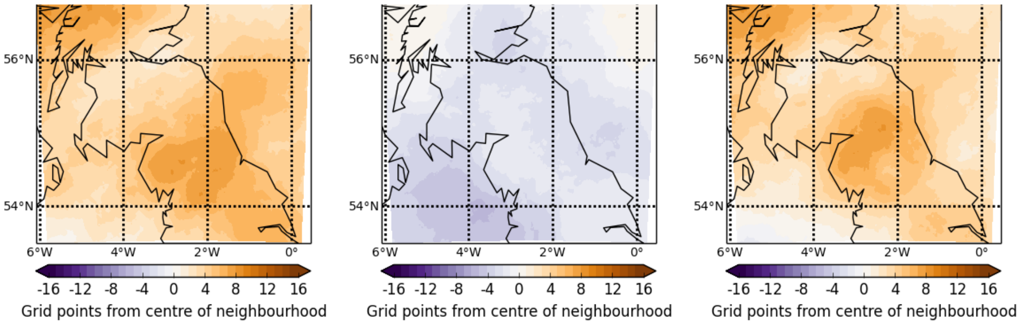
8.6.2 Temporal evolution

A strong diurnal signal was seen for the domain-average $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ (Figure 8.11) with the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ varying by up to 18 grid points and 12 grid points respectively. The location dependence of this temporal variation in agreement scales is shown in Figure 8.15 for maps of the $S_{ij}^{A(\overline{mo})}$ superimposed upon maps of the $S_{ij}^{A(\overline{mm})}$ for the

North domain. To highlight temporal differences in the spatial spread-skill relationship, the difference between the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ at each grid point in the domain is also plotted (Figures 8.15d-f).



(a) T+12 (15Z) (b) T+24 (03Z) (c) T+36 (15Z)



(d) T+12 (15Z) (e) T+24 (03Z) (f) T+36 (15Z)

Figure 8.15: $S_{ij}^{A(\overline{mm})}$ (filled contours and black dashed) and $S_{ij}^{A(\overline{mo})}$ (black solid) at (a) T+12; 15Z, (b) T+24; 03Z and (c) T+36; 15Z the next day. Results are shown for the North domain and are averaged over all cases in JJA 2013. A threshold of $1.0 mm\, hr^{-1}$ was used. Sub-figures (d), (e), and (f) show the difference $S_{ij}^{A(\overline{mo})} - S_{ij}^{A(\overline{mm})}$ at the times shown in (a), (b) and (c) respectively.

The behaviour seen in the maps of Figure 8.15 is consistent with that seen for the domain average (Fig. 8.11). During the day (e.g 15Z; T+12 or T+36, Figures 8.15a and d), the $S_{ij}^{A(\overline{mm})}$ is lower than the $S_{ij}^{A(\overline{mo})}$ (the members are closer to each other than they are to observations) and the model is spatially under spread. The $S_{ij}^{A(\overline{mm})}$ minimum is located to the northwest of the $S_{ij}^{A(\overline{mo})}$ minimum in Figure 8.15a and more elongated from north to south, possibly due to the model precipitation being too closely tied to the orography.

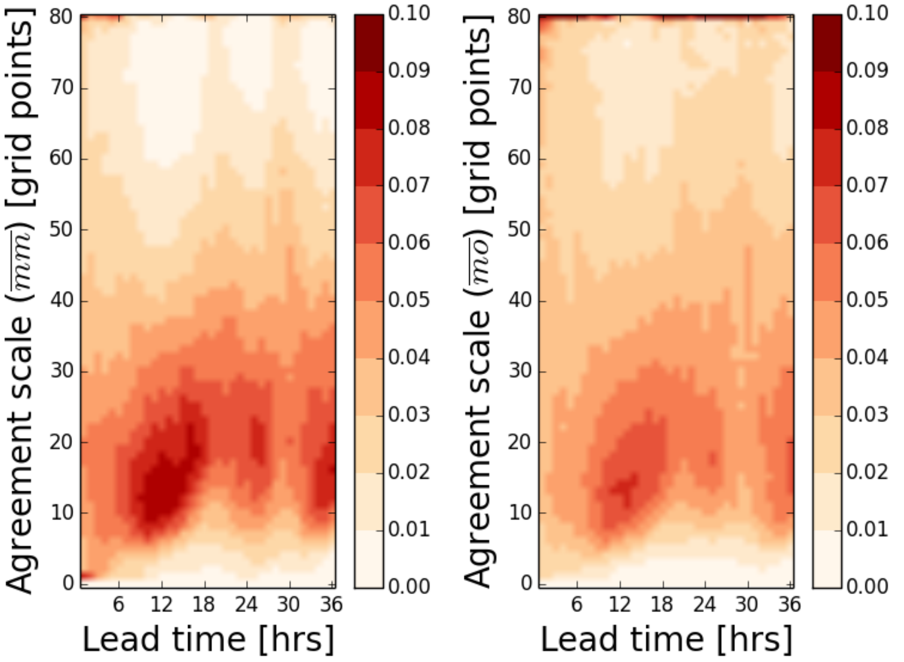
At night (e.g. 03Z; T+24 Figures 8.15b and e) the agreement scales have increased at all points in the domain and the minimum is less pronounced. The $S_{ij}^{A(\overline{mm})}$ has increased more than the $S_{ij}^{A(\overline{mo})}$ and, at this time the $S_{ij}^{A(\overline{mm})}$ is greater than the $S_{ij}^{A(\overline{mo})}$: the model is spatially over spread, particularly off the Cumbrian/Lancashire coast (around 54° N, 4° W). At lead times 24hrs to 36hrs both $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are similar to their corresponding values for lead times up to 12hrs with the ensemble being again under spread. Agreeing with the domain average results (Figure 8.11), there is little indication of the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ increasing with forecast lead time as may be expected due to error growth. This suggests that, at least for the short 36 hour rain rate forecasts considered here, and for a single summer season, local processes dominate spatial uncertainty and error rather than the temporal growth of errors.

Similar conclusions can be drawn for results from the South domain. However there are subtle differences. For example the diurnal cycle is less pronounced (possibly due to lower orography in the south domain), and differences between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are slightly smaller. Again this agrees with the domain average results.

To further investigate the changes in $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ throughout the forecast, Figure 8.16 shows two-dimensional (2D) histograms of the fraction of points (over all cases) at each agreement scale, as a function of forecast lead time for the North domain for a precipitation threshold of $0.01 mm\,hr^{-1}$. Hence, in Figure 8.16 we see how the PDF of the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ for JJA 2013 varies as a function of forecast lead time. Although the domain average did not show a trend of agreement scales increasing with time (variation in the domain average is dominated by the diurnal cycle as discussed above), the 2D histograms do show that fewer points have small $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ as the forecast lead time increases. This is particularly noticeable in the first 18 hrs of the forecast, and is what is expected due to the growth in forecast uncertainty with increasing forecast lead time. At the start of the forecast the distributions of both $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are flatter than at later times. In the first 6 hrs of the forecast the range of agreement scale values decreases and a tighter distribution is seen. This suggests that spin up in the model effects both the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$, with a larger variety of scales seen near the start of the forecast. This agrees with the results of Chapter 4 (Figure 4.8). Similar results were also found for the South domain.

Overall, the distributions for $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ are similar. However, there are some differences: the $S_{ij}^{A(\overline{mo})}$ distribution has a less pronounced peak agreement scale at around 15 grid points, and more points at the maximum agreement scales value of 80 grid points. This is consistent with the domain average $S_{ij}^{A(\overline{mo})}$ being greater than the domain average $S_{ij}^{A(\overline{mm})}$ for this threshold ($0.01 mm\,hr^{-1}$). This may be due to cases where the observations show very little precipitation across the domain (giving large $S_{ij}^{A(\overline{mo})}$

at most points) while the ensemble members have more precipitation (giving smaller $S_{ij}^{A(\overline{mm})}$ values). For the analysis presented in this chapter, only times with some precipitation in the ensemble forecasts and the observations have been included, removing false positives and missed events from the analysis. However, cases which were close to these limits have been included.



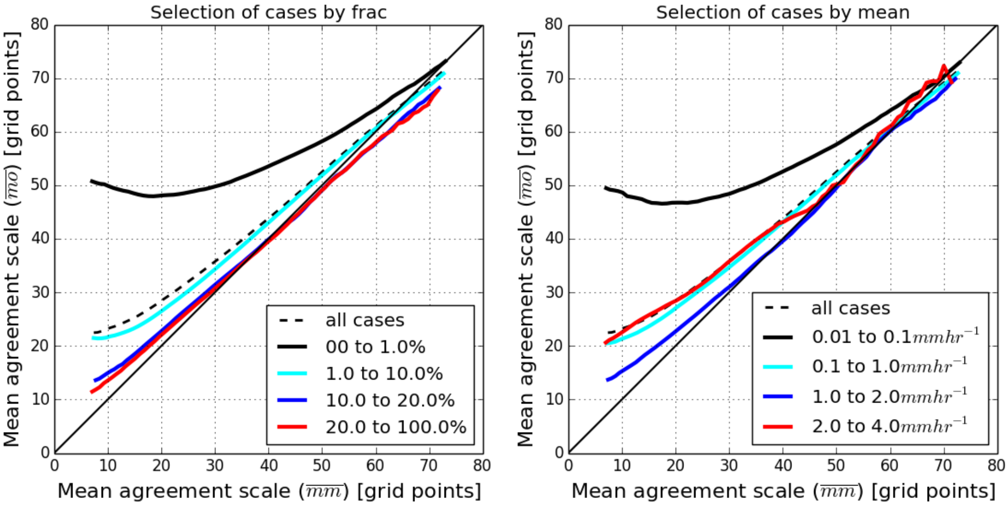
(a) $S_{ij}^{A(\overline{mm})}$ (b) $S_{ij}^{A(\overline{mo})}$

Figure 8.16: Histograms of the fraction of points at each scale (over all cases) to the total number of points, as a function of scale and lead time for the North domain. Bins for the scales are $i_{bin} - 0.5 <=$ data$<$ $i_{bin} + 0.5$ for i_{bin} in [0,1,...,80]. Results are shown for a threshold of $0.01 mm\,hr^{-1}$; the other thresholds lead to similar conclusions.

8.6.3 Dependence of spatial spread-skill on precipitation amount

In sections 8.3 and 8.4.1 precipitation characteristics were discussed using two summary measures: the average rain rate of all points in the domain with rain rates exceeding $0.01 mm\,hr^{-1}$ (measuring precipitation intensity), and the fraction of points in the domain with rain rates exceeding $0.01 mm\,hr^{-1}$ (measuring the horizontal extent of the precipitation). In this section these measures are used to investigate how the spatial spread-skill depends fractional coverage and intensity of precipitation.

First, values of the summary measure (say fractional coverage) were calculated for each forecast lead time from T+6 to T+29, and for all cases in JJA 2013. For the examples presented in this section the summary measures were calculated from the radar data; similar results were obtained when using an ensemble member forecast. Next the cases were separated into groups based on a specified range of the summary measure.



(a) Fractional coverage

(b) Average of raining points

Figure 8.17: Binned scatter plots for a threshold of 0.01$mm\,hr^{-1}$ averaged over cases with fixed ranges of (a) fractional coverage of points in the Radarnet data with rain rates greater than 0.01$mm\,hr^{-1}$, and (b) average rain rate of points in the Radarnet data with rain rates greater than 0.01$mm\,hr^{-1}$. The North domain and forecast lead times T+6 to T+29 were considered.

Figure 8.17 shows binned scatter plots (bin size 10 grid points) for data with a 0.01$mm\,hr^{-1}$ threshold, split by fractional coverage of rain rates exceeding 0.01$mm\,hr^{-1}$ (Figure 8.17a) and by the average rain rate of points with rain rates exceeding 0.01$mm\,hr^{-1}$ (Figure 8.17b). Examples from the North domain are given here; similar results are obtained for the South domain. Ranges of 0 to 1%, 1 to 10%, 10 to 20% and 20 to 100% were considered for the fractional coverage, and 0.01 to 0.1$mm\,hr^{-1}$, 0.1 to 1.0$mm\,hr^{-1}$, 1.0 to 2.0$mm\,hr^{-1}$ and 2.0 to 4.0$mm\,hr^{-1}$ for the average rain rates. Notice that the results with rain rates between 2.0 and 4.0$mm\,hr^{-1}$ (shown in red; Figure 8.17b) are more noisy. This is due to fewer grid points having large agreement scale values for rain rates in this range.

These plots show that the times with the lowest fractional coverage of rain, and lowest rain rates, were the most spatially under spread (lying the furthest from the diagonal in Figure 8.17). This suggests that forecasts from the ensemble in cases of low fractional coverage and intensity should be viewed as being most likely to be over confident about the location of precipitation. In contrast, Figure 8.17a indicates that the spatial uncertainty for cases with high fractional coverage was, on average, well captured. In Figure 8.17b, results for the highest rain rate range (2.0 to 4.0$mm\,hr^{-1}$), lie further from the diagonal than than those with rain rates from 0.1 to 2.0$mm\,hr^{-1}$: the ensemble is more under spread for these high rain rate cases. As discussed in Section 8.4.1 this high rain rate range selected the more convective cases, which were less spatially predictable (giving larger $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ than the lower thresholds). As well as being less spatially predictable, the spatial predictability for these more convective cases was also harder for the model to capture.

Given the dependence on precipitation amount and coverage shown in Figure 8.17, it is useful to consider again the binned scatter plots for different rain rate thresholds shown in Figure 8.12. In Figure 8.12 it was perhaps surprising that the different precipitation thresholds gave such similar results. Considering the results in Figure 8.17, cases with the lowest fractional coverage, and the lowest average rain rates have the poorest spatial spread-skill relationship. When selecting a higher rain rate threshold in Figure 8.12 a lower precipitation coverage is considered, suggesting a poorer spatial spread-skill relation, but higher rain rates are considered suggesting a better spatial spread-skill relationship. Hence, these two effects somewhat cancel out, and the spatial spread-skill relationship for different thresholds appears similar.

8.7 Chapter discussion and conclusions

In this chapter the measures of spatial ensemble spread and spatial skill introduced in Chapter 6 have been applied to three summer months of rain rate MOGREPS-UK data. The aim was to investigate the spatial ensemble characteristics in a more statistical manner and to provide additional insight into answering the questions (thesis questions 2 and 3; Section 1.2):

1. Over what spatial scales should the ensemble be interpreted and evaluated? How can these spatial scales be defined?
2. What information does this spatial approach provide for forecasting?

First the overall spatial predictability, forecast by the ensemble over JJA 2013, was considered. This gave an indication of the average spatial differences between ensemble member precipitation forecasts and how these varied across the UK. It was found that the $S_{ij}^{A(\overline{mm})}$ varied considerably across the UK, with smaller scales in the northwest (down to 16 grid points, a total neighbourhood length of 66km) and larger scales in the southeast (up to 52 grid points, a total neighbourhood length of 231km). Hence, overall for this summer period, the ensemble was more confident in the location of precipitation to the north and west of the UK, over regions of high orography and with direct exposure to the Atlantic. Note that, even in the regions of higher predictability, the ensemble members still only agree on average about the location of precipitation on scales much larger than the grid scale: in general the ensemble precipitation forecasts should not be considered quantitatively at, or close to, the grid scale. This is important information for forecasting and model interpretation. The scales over which the ensemble precipitation forecasts should be evaluated can easily be calculated using the methods applied in this chapter.

Alongside the spatial differences between ensemble member forecasts, it is also important to consider the spatial differences between ensemble forecasts and observations to verify the ensemble performance. In this chapter this was done by comparing the agreement scales calculated between ensemble members ($S_{ij}^{A(\overline{mm})}$) to those calculated between ensemble members and radar observations ($S_{ij}^{A(\overline{mo})}$). Overall, these measures of spatial ensemble spread and spatial ensemble skill were found to have similar magnitudes and variations across the domain. This shows that, for JJA 2013, the ensemble did a reasonable job at capturing the spatial predictability of precipitation, and gives some confidence in its use for forecasting, provided that the ensemble forecasts are interpreted over suitable spatial scales. There were, however, some differences between the spatial ensemble spread and spatial ensemble skill, with the ensemble tending to be slightly spatially under spread. This was particularly noticeable when the ensemble was confident about the location of precipitation (small agreement scales), and suggests that ensemble forecasts in which precipitation has a high degree of agreement should be treated cautiously.

In addition to the overall features described above, the spatial spread-skill relationship was found to depend on the threshold applied to the precipitation field. In particular, the ensemble had a poorer spatial spread-skill relationship when a $0.01 mm\ hr^{-1}$ threshold was applied, to when thresholds of $0.1 mm\ hr^{-1}$ or above were used. This was linked to the ensemble precipitation forecasts having more fragmented precipitation structures than the radar data. Thus, applying a threshold of $0.1mm\,hr^{-1}$ to the ensemble member forecasts resulted in them becoming more spatially different from each other and increasing the $S_{ij}^{A(\overline{mm})}$ relative to the $S_{ij}^{A(\overline{mo})}$. This highlights the importance of a correct physical structure to the precipitation for obtaining a good spatial spread-skill relationship.

The temporal evolution of the agreement scales (both $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$) was dominated by a diurnal cycle, with smaller agreement scales (higher spatial predictability) seen in the afternoon and larger agreement scales (lower spatial predictability) seen over night. The diurnal cycle was found to be slightly stronger for the comparison of ensemble members ($S_{ij}^{A(\overline{mm})}$) than when comparing ensemble members with radar observations. This resulted in temporal variations in the spatial spread-skill relationship, with the ensemble being spatially under spread during the day, possibly due to being too tied to the orography, and spatially over spread at night. There was little indication of forecast spatial differences growing with lead time. This was unexpected as, statistically, forecast errors are found to grow with lead time. Perhaps this may be due to the short forecast length studied here (36 hours) or the focus on rain rates for a summer season. It is possible that other, less locally dependant, variables would show a growth in differences with forecast lead time.

To investigate the relationship between the agreement scales and the precipitation structure, the data from JJA 2013 were grouped according to the fractional coverage of precipitation over the domain, and average rain rate of raining points. It was found that cases with lower fractional coverage were less spatially predictable than those with more wide-spread precipitation (larger $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$). As there is less likely to be overlap between forecasts when only a small area of precipitation is predicted, and additionally small scale convective precipitation is often less predictable than large scale frontal precipitation (e.g. Walser and Schär, 2004), this result was expected. When grouping the cases based on the average rain rate of raining points for the south domain, larger than average agreement scales (lower spatial predictability) were found both for the lowest and for the highest rain rate ranges considered (0.01 to $0.1mm\,hr^{-1}$ and 2.0 to $4.0mm\,hr^{-1}$ respectively). This was linked to the low precipitation rates being associated with spatially unpredictable showers or patches of drizzle with low fractional coverage, and high rain rates being associated with convective cases. The dependence of the spatial spread-skill relationship on precipitation structure was investigated and it was found that cases with lower fractional coverage of precipitation, and lower average rain rates, were the most spatially under spread with the poorest spatial spread-skill relationship. This shows that caution is needed when interpreting ensemble forecasts of low intensity or low fractional coverage precipitation.

To summarise, the methods introduced in Chapter 6 have been applied in a more statistical manner to a summer season of rain rate data. This has been found to give useful information for forecasting and model evaluation as described above. However, the conclusions from this work are limited by the consideration of only one season (summer 2013); it is likely that some differences in the results would be obtained from other seasons. Hence it is suggested that future investigation should focus on applying these techniques to a larger data sample, perhaps as part of ongoing monitoring of the ensemble system. For the comparison with Radarnet data there are additional questions that have not been addressed in this chapter, in particular regarding the nature of Radarnet errors. Here, for the limited three month sample visual quality checks were performed on the Radarnet data; this process would have to be automated for a long-term or routine operational analysis. Additionally, radar errors should be taken into account in the calculation of the $S_{ij}^{A(\overline{mo})}$. The best methodology for incorporating this into the analysis would have to be investigated.

Chapter 9

Conclusions and future work

9.1 Overview

One of the key issues in forecasting today is how to make best use of convection-permitting ensembles. For example, what do they tell us and how can they be usefully evaluated? The aim of this PhD was to develop new spatial approaches for the evaluation and characterisation of convective scale ensemble systems. Forecasts were considered not only at single grid points, but also over the regions or neighbourhoods surrounding these points. Three fundamental requirements were identified for these methods: firstly, they must give physically meaningful results (for example when considering the relationship between different times and locations), secondly they should not be overly sensitive to changes in ensemble membership, and finally they must be capable of considering multivariate relationships. Six research questions were formulated to guide the work of this thesis:

1. How can the spatial variability of ensemble member forecasts be meaningfully summarised? What are suitable summary measures, and what do these tell us about ensemble performance and spatial predictability?
2. Over what spatial scales should the ensemble be interpreted and evaluated? How can these spatial scales be defined?
3. What information does this spatial approach provide for forecasting?
4. How can correlations in the vertical be used to allow physical structures and multivariate relationships from the ensemble to be usefully summarised? To what extent do these correlations reveal properties of the convection and the convective environment?
5. What is the dependence of vertical correlation structure on the horizontal scales used? How does this relate to the scales obtained from the spatial methods?

6. In what ways might these methods be useful in other areas such as data assimilation?

To answer the six thesis questions, research was split into four areas: domain-wide spatial ensemble evaluation using the FSS (Chapter 4), development of a new location-dependent method for spatial ensemble evaluation (Chapter 6), application of the spatial approach to calculate multivariate correlations (Chapter 7) and the assessment of spatial predictability across the UK for a three month summer period (Chapter 8). Conclusions relating to the thesis questions, based on the findings from Chapters 4 to 8, are presented and discussed in Section 9.2. Following on from this, Section 9.3 discusses avenues of future investigation. Finally, Section 9.4 summarises the key findings from this work.

9.2 Discussion and conclusions

The aim of this PhD thesis was to answer the six research questions listed in Section 9.1; these are now addressed in turn.

1. How can the spatial variability of ensemble member forecasts be meaningfully summarised? What are suitable summary measures, and what do these tell us about ensemble performance and spatial predictability?

To summarise the spatial characteristics of ensemble forecasts, a novel method was developed based on the Fractions Skill Score (FSS Roberts and Lean, 2008). Originally, the FSS was formulated to quantify the domain-wide agreement between a deterministic precipitation forecast and radar observations over a predefined neighbourhood size (or spatial scale). However, the FSS can also be applied more generally to calculate the agreement between any two gridded fields (such as two ensemble member forecasts). This was the basis for the method of calculating the spatial variability of ensemble forecasts, presented in Chapter 4 of this thesis. When calculated over pairs of ensemble member forecasts, the FSS quantifies the differences between those forecasts for a given scale. Hence, if the FSS is calculated for all independent pairs of ensemble member forecasts, we get an impression of the overall ensemble spread at that scale.

In this thesis, the mean of the FSS values calculated over all independent ensemble member pairs was used as a summary measure to represent the ensemble dispersion at a given spatial scale (dFSSmean). The standard deviation of the FSS values from independent member-member pairs, dFSSstdev, was also calculated to give further information about the range of FSS values used in the calculation of dFSSmean. When calculated for a range of different neighbourhood sizes the dFSSmean and dFSSstdev provide a means of quantifying the spatial ensemble variability. It is also possible to quantify the skill of the ensemble in spatially matching the observations. This was done by taking the mean of the FSS values calculated over all ensemble member-radar pairs, to give a measure of the spatial error, eFSSmean. Having a measure of both the ensemble spread and the ensemble error provides useful information about the ensemble's ability to capture uncertainty over a range of neighbourhood sizes. It should be noted that dFSSmean and eFSSmean were developed from meteorological and physical, as apposed to mathematical, arguments. In particular, it has not been proved that comparing dFSSmean and eFSSmean gives quantitatively useful information about the spread-skill relationship of the ensemble. Instead, in this thesis, eFSSmean was used to put dFSSmean into context. Hence, when used with reference to dFSSmean and eFSSmean the terms "over spread" and "under spread" should be taken to have this qualitative meaning.

To investigate what these methods tell us about the ensemble performance, the dFSSmean, dFSSstdev, and eFSSmean were calculated for hourly precipitation accumulations from two convective case studies. For one case study horizontal 10m wind speeds were also considered. It was concluded that, by using dFSSmean and dFSSstdev, a detailed picture could be built up of how ensemble differences vary in time, and with spatial scale (neighbourhood size). For the case studies investigated here, it was found that both dFSSmean and eFSSmean were highly dependent on the scales considered for forecast evaluation, with larger spatial differences seen on smaller scales (i.e. over smaller neighbourhood sizes). Hence, at smaller scales the ensemble forecasts were less predictable, and less skillful, than at larger scales. In particular, the scale at which the ensemble had skill (as defined in Roberts and Lean (2008) and discussed in Section 4.4.1), generally varied from 50 km to 300 km, much larger than the grid scale (2.2 km). Hence, these results highlight that convective-scale forecasts should not be evaluated at the grid scale. Overall, for the cases considered here, comparison of dFSSmean and eFSSmean showed that the ensemble spread was a reasonable predictor of forecast uncertainty. There were times where the ensemble was under-spread due to timing errors between the simulations and the observations, or early on in the forecast when showers had not yet spun up.

The FSS evaluates the spatial differences between points in two fields whose values exceed a pre-determined threshold. In this thesis percentile thresholds were used (e.g. the 99^{th} percentile threshold which considers the top 1% of precipitation in the domain) to ensure that the same number of points were used in all comparisons. It was found that different percentile thresholds selected precipitation from different physical regimes (such as convective only or convective and stratiform). In particular, higher percentile thresholds selected the heavier, less-widespread precipitation, which was less spatially predictable. The temporal evolution of dFSSmean and eFSSmean was also found to be threshold dependent, with higher thresholds showing greater temporal variability: local effects dominated the ensemble spread and skill in these instances.

The dFSSmean can also be used to compare different ensemble configurations. This was demonstrated for one case study, where the spatial ensemble spread created using different physics configurations was compared against that generated using different lateral boundary condition and initial condition perturbations. It was found that dFSSmean allowed the spatial impact of changes to the ensemble configuration to be thoroughly investigated. The physics perturbations were chosen as plausible alternative model formulations, and were not found to influence the ensemble evolution at scales where the forecast had skill. It should be emphasised that other work (e.g. Stensrud et al., 2000; Keil et al., 2014) has shown a positive impact from physics perturbations. The methods presented in this thesis would provide a useful method of assessing the impact from further studies of this kind.

2. Over what spatial scales should the ensemble be interpreted and evaluated? How can these spatial scales be defined?

Thesis Question 1 asked how the spatial variability of ensemble forecasts could be summarised; i.e. if we pick a scale over which to evaluate the ensemble, how will the ensemble members differ on average over the domain (what will be the spread)? This second thesis question turns Question 1 around to ask instead: if we require a specific level of agreement between ensemble members, at what scale is this achieved? As spatial uncertainties vary with location in the domain (due to different meteorological phenomena being associated with different levels of predictability), it is important that the methods developed to answer Question 2 are location-dependent. Thus, instead of a domain-wide summary measure, we seek a measure quantifying the scale of spatial uncertainties at each grid point.

In Chapter 6, a new method was presented for evaluating the location-dependent scales over which ensemble members reach a specified level of agreement. Like the FSS, this method considers pairs of fields, either two ensemble member forecasts, or a single forecast and observations. At each grid point in the domain, the agreement scale between the fields is defined as the minimum neighbourhood size over which the fields meet a predefined similarity criterion. This criterion compares the total amount of precipitation within the selected neighbourhood for each of the two fields, and includes two parameters, α a measure of the acceptable bias, and S_{lim}, a fixed maximum scale. These parameters can be adjusted to make the criterion more/less stringent according to the users' requirements.

The location-dependent agreement scale for the ensemble is defined as the average (calculated separately at each grid point i, j) of the agreement scales from all independent ensemble member comparisons. These average agreement scales for the ensemble, denoted $S_{ij}^{A(\overline{mm})}$, give a measure of the believable scales according to the ensemble forecast at each grid point in the domain, and represent the scales at which the ensemble members become sufficiently similar so that the forecast forms useful, trustworthy guidance (assuming that the ensemble has a good spread-skill relationship). The location-dependent agreement scales between the ensemble and observations are defined as the average (calculated separately at each grid point) of the agreement scales calculated from all ensemble member-radar pairs. These scales, denoted $S_{ij}^{A(\overline{mo})}$, indicate the skillful scale of the ensemble at each grid point in the domain.

To be of use for ensemble interpretation it is necessary that the $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ can reflect changes in spatial predictability. Additionally, for these methods to be useful for verifying the ensemble spatial spread, it is necessary that a comparison of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ correctly identifies properties of the spread-skill relationship. For example, the measures must correctly differentiate between ensemble systems that are well spread, over spread, and under spread. To test these requirements, and hence to better understand what the location-dependent agreement scales tell us about ensemble spatial uncertainties, a simple idealised experiment was conducted in which an ensemble of synthetic precipitation forecasts, and synthetic observations, were generated with known spread-skill properties. Using this set-up it was shown that the location-dependent agreement scales behaved as designed for pre-defined spatial ensemble distributions: smaller/larger agreement scales were found for experiments with higher/lower spatial predictability. It was also found, for this simple experiment, that comparing $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ gave the correct interpretation of the ensemble spatial spread-skill relationship with under-spread ensembles showing $S_{ij}^{A(\overline{mm})} < S_{ij}^{A(\overline{mo})}$, overspread ensembles showing $S_{ij}^{A(\overline{mm})} > S_{ij}^{A(\overline{mo})}$ and well spread ensembles showing $S_{ij}^{A(\overline{mm})} \sim S_{ij}^{A(\overline{mo})}$. Binned scatter plots provided a particularly useful method for visualising and assessing the differences between $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ because the location-dependent character of convective-scale predictability was retained. It should be noted that there are still questions regarding the interpretation of $S_{ij}^{A(\overline{mm})}$ and $S_{ij}^{A(\overline{mo})}$ which are left unanswered by the idealised experiment. For example, the effect of bias between ensemble members and observations, or the effect of different precipitation distributions, has not been investigated. The conclusions presented below should be considered with these limitations in mind.

The member-member and member-radar agreement scales were calculated for hourly instantaneous rain rates for the six convective cases introduced in Chapter 5. Maps of the $S_{ij}^{A(\overline{mm})}$ depicted the different levels of spatial predictability seen across the cases. As well as having different levels of spatial agreement, the different cases showed different spatial spread-skill relationships, shown through comparing the $S_{ij}^{A(\overline{mm})}$ to the $S_{ij}^{A(\overline{mo})}$. The $S_{ij}^{A(\overline{mm})}$ were also used to investigate the horizontal structure of spatial differences at different vertical levels. The variables considered (cloud fraction, horizontal divergence, temperature, specific humidity and horizontal wind speed) were found to have a variety of spatial characteristics, with the $S_{ij}^{A(\overline{mm})}$ reflecting both horizontal and vertical physical structures.

3. What information does this spatial approach provide for forecasting?

As discussed in Chapter 4 the FSS can be successfully used to summarise the spatial domain-wide ensemble information. For example, the FSS can indicate the scale-dependence of forecast differences, and the variation of spatial errors with time. This is useful for forecasting as it provides a measure of the overall ensemble performance, and allows the fair comparison of different model configurations. However, the domain-wide approach also has the disadvantage that location-dependent information about the ensemble spatial characteristics is lost. This location-dependent information is important from a local forecasting perspective, and can be quantified using the the location-dependent agreement scales, presented in Chapter 6.

In Chapter 8, the location-dependent agreement scales were used to characterise the spatial predictability of precipitation forecasts for a three month summer season (June, July, and August 2013). Instantaneous rain rates from the MOGREPS-UK ensemble system were examined to quantify the spatial predictability of precipitation over this period, as forecast by the ensemble. These were then compared to radar derived rain rates from the Radarnet system to assess the spatial-skill relationship of the ensemble.

It was shown that, overall, MOGREPS-UK was reasonably well spread over summer 2013. There was a slight tendency for the model to be over confident in the location of precipitation, resulting in the ensemble being a little spatially under spread. This was particularly seen at small agreement scales where the ensemble had the highest confidence in the location of precipitation. These results agree with those from Chapter 6 and suggest that, on average for this particular summer, MOGREPS-UK provided useful forecast guidance. Local effects were found to dominate the spatial ensemble spread and skill and, surprisingly, there was little indication of forecast spread and error increasing with lead time, at least for the lead times considered here (up to 36 hrs). Instead, the spatial spread-skill relationship was found to vary with a diurnal cycle: the ensemble was slightly spatially under spread during the day (possibly due to the ensemble forecasts being too tied to orography), but slightly spatially over spread at night.

Factors influencing the spatial spread, spatial skill, and spread-skill relationship were investigated. It was found that all three are highly dependent on the precipitation amount, measured through the fractional coverage of rain, or the average rain rate of raining points. A poor spread skill relationship was associated with a low fractional coverage of rain, and extreme (either high or low) average rain rates; these cases also had larger agreement scales. This is important information for forecasting as it highlights when more or less confidence should be placed in the ensemble spatial uncertainty estimates. The spatial spread, spatial skill, and spread-skill relationship were also found to be location dependent. When averaged over the three month period, the ensemble had higher confidence in the location of precipitation (smaller agreement scales were obtained) over the northwest of the UK. It was found that this higher predictability in the northwest was partly due to the larger precipitation amounts and higher orography in the region. Other factors linked to the orography, such as higher orographic gradients, as well as non-orographic effects such as a more direct exposure to the Atlantic, could also be important.

In Chapters 6 and 8 it was demonstrated that the location-dependent agreement scales can provide a useful visualisation of the positional variability of rain for forecasting. For example, the rainfall structures themselves can be viewed from an individual model run (perhaps the control or the deterministic forecast) and the agreement scale map can be used to view the spatial uncertainty in that rainfall pattern, as given by the ensemble. This allows the spatial predictability from the ensemble to be quickly assessed in a physically meaningful manner, and could be used in operational forecasting.

4. How can correlations in the vertical be used to allow physical structures and multivariate relationships from the ensemble to be usefully summarised? To what extent do these correlations reveal properties of the convection and the convective environment?

Thesis questions 1-3 focused on quantifying the spatial uncertainty separately for different meteorological variables. However, one of the requirements of the spatial approach to be developed in this thesis is that it is capable of considering multivariate relationships. This requirement was addressed by applying the spatial approach to the calculation of multivariate correlations.

Vertical correlations were calculated for the 3D variables cloud fraction, horizontal divergence, temperature, specific humidity and horizontal wind speed. Correlations were calculated for one variable at different vertical levels (vertical correlations, VCs), and between different variables at different vertical levels (vertical cross correlations, VCCs). Additionally, temporal correlations (TCs) were also calculated between surface rain rates and three dimensional variables at different vertical levels.

To enable correlations to be calculated at the location-dependent agreement scales, a neighbourhood based approach was developed. To calculate correlations from an ensemble of 12 members, over a neighbourhood of N_x by N_y grid points centred on a grid point P, each ensemble member at each point in the neighbourhood was taken to be a possible realisation of the forecast at point P. All of the $N_x \times N_y \times 12$ realisations were then taken to form an augmented ensemble, from which the correlations were calculated. Although not independent, the additional realisations do provide more information about the possible forecast outcomes, particularly for fields such as precipitation with high spatial variability. The extent to which they provide this additional information is related to the meteorological situation: it is expected that more additional realisations are required in situations that are less spatially predictable (i.e when the PDF of possible forecast outcomes is most under-sampled).

When the location-dependent agreement scales were used, the correlations highlighted properties of the convection, and of the convective environment. For example, VCs for cloud fraction highlighted the cloud layers, and VCs for horizontal divergence showed the structure of convective cells. VCCs allowed different variables to be related, and confirmed that the ensemble was producing realistic physical structures. For example, higher/lower cloud fractions were associated with stronger/weaker convective cells (as shown in correlations with horizontal divergence) and higher/lower specific humidity. Similar conclusions could be drawn from TCs with surface rain rates. Additionally, TCs with rain rates allowed the length of time over which cells remained over a given location to be assessed. The fact that these relationships can be extracted from a small operational ensemble, highlights the utility of the methods presented in Chapter 7.

5. What is the dependence of vertical correlation structure on the horizontal scales used? How does this relate to the scales obtained from the spatial methods?

Using the six convective cases introduced in Chapter 5 as examples, it was shown that meaningful correlations could be obtained from a twelve member ensemble, if a sensibly chosen neighbourhood was applied to define the augmented ensemble. Correlations calculated at the grid scale were found to be very noisy for all variables due to under sampling from the 12 ensemble members. Correlations calculated over larger neighbourhoods were less noisy and could be related to physical processes. Very large neighbourhoods were found to overly smooth the correlations, resulting in the loss of physically meaningful smaller scale structure. The $S_{ij}^{A(\overline{mm})}$ defined in Chapter 6 were found to be a good choice of neighbourhood size between these extremes.

6. In what ways might these methods be useful in other areas such as data assimila-

tion?

The methods developed in this thesis could be applied to understand ensemble performance in a variety of situations. For example, they could be used to assess the impacts of model changes, upgrades or ensemble membership, or to compare different resolution ensemble forecasts. The spatial approach for analysing convective scale ensembles could also provide useful information in other research areas, such as convective scale data assimilation. In data assimilation, it is necessary to define the multivariate covariance between all model variables at all points in the domain (background error covariance matrix, Section 2.3.2). This covariance matrix quantifies errors in the initial state for the data assimilation, and enables information to be correctly propagated between variables and grid points. Hence, an accurate representation of the background errors is essential for a successful data assimilation scheme. Using the methods presented in Chapter 7 the correlation structures present in convective scale forecasts, and their dependence on horizontal scale, can be investigated, and used to inform schemes to calculate the background error covariance matrix. Other areas that use convective scale forecasts as input, such as hydrological modelling, could also benefit from the methods presented in Chapters 4 and 6 to better quantify spatial uncertainties.

9.3 Areas of further investigation

Direct Extensions

There are several direct extensions to the work of this thesis:

1. The work in Chapter 4 could be extended by calculating the dFSSmean and eFSSmean for a season of cases (as done in Chapter 8 for the location-dependent agreement scales), to allow the domain-wide evaluation of the growth and scale dependence of spatial errors more statistically. Additional seasons could also be considered, both for the FSS and the location-dependent agreement scales, to get more statistically robust results. This would allow the spatial ensemble characteristics to be evaluated more generally, and allow verification of the operational ensemble system. With a larger sample of cases it would also be possible to subset the cases to investigate the regime-dependence of the spatial uncertainty. For ex-

ample, the cases could be split by the dominant forcing mechanism of convection, or the large scale flow regime.

2. More variables could be considered to provide further insight into the ensemble spatial spread. Variables with high spatial variability such as vertical wind speeds, or variables' derivatives, such as the magnitude of temperature gradients, would be of particular interest. The FSS could also be applied at different vertical levels (in a similar way that the agreement scales were used in Section 6.6) to investigate how horizontal spatial differences vary in height.

3. The work of Chapter 7 could be extended by calculating multivariate correlations for more cases and grid points. This would allow a more in-depth statistical study of the different correlation structures, and enable a more robust physical interpretation. A catalogue of the different correlation structures, and their relationship with physical processes, could also be built up. This would be useful as a quick reference when interpreting the model output, and allow the prevalence of individual correlation structures to be assessed. As the new version of the Met UM dynamical core (ENDgame Wood et al., 2014) can have a significant impact on convective precipitation (N. Roberts personal communication, H. Lean personal communication) it would also be interesting to calculate correlations from forecasts using ENDgame and compare and contrast these with correlations from the New Dynamics core (Davies et al. (2005), used for all model runs in this thesis).

Theoretical understanding

Much of the work in this thesis (such as the criterion for assessing forecast similarity in Section 6.3.2 or the method for comparing spatial spread and spatial skill in Section 6.3.4) has been based on physical and practical considerations. Future work should investigate these areas theoretically. Important questions relate to the use of the average of member-member pairs to measure ensemble spread, and member-radar pairs to measure ensemble skill. For example, what are the theoretical implications of comparing averages over a different number of quantities? What are the implications of the average of member-member pairs and the average of member-radar pairs not being independent? Possible methods of investigating these issues could include the use and extension of idealised experiments (such as those presented in Section 6.4 but with increasing complexity), or the use of a very large ensemble to enable a subset of ensemble members to be used for the spread and skill calculations.

There are additional aspects of the FSS and agreement scale calculations that could be further understood theoretically. These include the quantification of edge effects, such as done for simple idealised deterministic precipitation configurations in Skok (2015), and discussed in Nachamkin and Schmidt (2015). Edge effects could also be investigated by comparing results calculated using the full domain with those calculated using a 'buffer zone' around the edge of the domain to remove edge effects.

The methods presented in this thesis have used a neighbourhood approach to quantify spatial differences. An interesting avenue of further investigation would be to compare these methods with others which use alternative approaches, such as those that are features based or use scale decomposition. As the FSS acts like an implicit spatial low pass filter, and the agreement scales act like a location-dependent low pass spatial filter, these methods (at least qualitatively) perform a similar function to wavelet methods. It would be interesting to investigate this link mathematically. The methodology for calculating the agreement scales could also be used with other criteria to determine the spatial similarity of forecasts. These need not be limited to neighbourhood based methods: other approaches such as the features based SAL score (Wernli et al., 2008) could be used to form alternative similarity criteria.

Technique Development

The methods presented in this thesis have focused on quantifying spatial differences and errors in the horizontal plane. These methods could be further developed to investigate timing differences, and differences in the vertical dimension. Duc et al. (2013) have included timing errors in the FSS for deterministic forecasts by expanding the neighbourhood to include the time dimension (i.e. giving a 3 dimensional neighbourhood, with dimensions X, Y and time). Although useful in gaining an overview of the ensemble performance over a time period, their method leaves some questions unanswered. In particular, as timing errors due to advection (rather than initiation or decay) are already inherently included in the spatial approach, is the method of Duc et al. (2013) counting for some errors twice? What are the implications of considering, together in a neighbourhood (with equal weighting), quantities with different units and scaling lengths? It would be interesting to conduct an in-depth study of cases where timing errors were seen in the forecast, using the methods presented in this thesis. If these methods were found to give a useful indication of the timing errors and differences, they could then be compared and contrasted with other measures of the timescales associated with convective events, such as the convective time scale used by Keil et al. (2014).

The agreement scales, and the FSS, could also be extended to investigate spatial differences in the vertical. This could be achieved by modifying the neighbourhood to include the vertical dimension (making it three dimensional). The effect of the different physical scale lengths in the horizontal and vertical would need to be investigated, along with the best way to modify the similarity criteria.

Including observation uncertainty

Throughout this thesis observation uncertainty has been neglected, with radar observations treated as reality. Of course, errors do still exist in the radar observations, and should ideally be accounted for in the FSS and agreement scale comparisons between ensemble members and radar derived rain rates. One possible method of including observation errors would be by adding a random perturbation to the observations before calculating the FSS or agreement scales (as done for example in Roberts and Lean (2008)). However, for this to be useful the magnitude of the observation errors must first be understood. As the error magnitude is likely to depend on the meteorological situation (for example, higher errors may be expected in very heavy precipitation due to attenuation for the radar beam), this is a non-trivial question. Other methods for including observation error could make use of radar-based ensembles, such as those in development at the Met Office (Allison et al., 2015). Using an ensemble of radar-rain rates the spatial uncertainties associated with the radar observations could be assessed by calculating the member-member agreement scales for the radar ensemble. The best way to compare the radar ensemble and forecast ensemble would have to be investigated; for example how best to form pairs of the forecast ensemble members and the radar ensemble members.

The FSS and agreement scales could also be used to compare ensemble forecasts with the analysis, or with station observations. This is another interesting area of investigation, particularly due to the uneven distribution of observing stations. Comparing these results with those obtained from radar comparisons would allow the effect of observational uncertainties (present in all observation types) on the neighbourhood based approaches to be better understood. If they were shown to give useful information for precipitation comparisons, the analysis could also be used to estimate the spatial spread-skill for variables other than precipitation for which gridded observations are not available.

9.4 Summary of key contributions

This thesis has focused on the development of a spatial approach to the analysis of convective scale ensemble systems. Two novel forecast evaluation methods have been presented; one used to summarise the domain-wide ensemble performance, and another which captures the location-dependent behaviour. Through the evaluation of eight detailed case studies, a simple idealised ensemble system, and a summer season of precipitation forecasts, the importance and utility of the spatial approach has been demonstrated. Using the spatial neighbourhood approach, multivariate relationships were also investigated. In particular it has been shown that:

- Convective scale ensemble forecasts should be considered useful at scales over which the forecast is skillful, not at the grid scale. With the development of higher and higher resolution models, this distinction between the grid scale and the skillful scale is becoming increasingly important.

- The spatial agreement of ensemble forecasts can be quantified, using the methods presented in Chapters 4 and 6, to give a meaningful and compact visualisation for operational forecasting. The information provided using these methods is also useful for ensemble verification.

- Once the spatial uncertainty has been quantified, it can be applied to other areas, such as the calculation of multivariate correlations from a small operational ensemble system, to obtain physically meaningful results.

To summarise, this thesis has demonstrated that a more thorough and appropriate understanding of convective scale ensemble forecasts can be gained by using a spatial approach. This information has direct applications for forecasting, and also provides information which can diagnose and verify spatial ensemble characteristics, hence showing where deficiencies lie and where improvements are needed.

References

Allison, L. C., M. Mittermaier, R. Scovell, C. Vernon, K. Norman, and C. Pierce, 2015: Quantifying the effect of radar observational uncertainty on the verification of kilometer-scale NWP precipitation forecasts using the Fractions Skill Score, american Meteorological Society 37th Conference on Radar Meteorology. URL `https://ams.confex.com/ams/37RADAR/webprogram/Paper274960.html`

Ancell, B. and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Monthly Weather Review*, **135**, 4117– 4134.

Ancell, B. C., 2013: Nonlinear characteristics of ensemble perturbation evolution and their application to forecasting high-impact events. *Weather and Forecasting*, **28**, 1353–1365.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230**, 99–111.

Andersson, E., P. Bauer, A. Beljaars, F. Chevallier, E. Hólm, M. Janisková, P. Kållberg, G. Kelly, P. Lopez, A. Mcnally, E. Moreau, A. J. Simmons, J.-N. Thépaut, and A. M. Tompkins, 2005: Assimilation and modeling of the atmospheric hydrological cycle in the ECMWF forecasting system. *Bulletin of the American Meteorological Society*, **86**, 387–402.

Andersson, E., E. Hólm, P. Bauer, A. Beljaars, G. Kelly, A. McNally, A. Simmons, J.-N. Thépaut, and A. Tompkins, 2007: Analysis and forecast impact of the main humidity observing systems. *Quarterly Journal of the Royal Meteorological Society*, **133**, 1473–1485.

Arakawa, A. and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics*, **17**, 173–265.

Baker, L., A. Rudd, S. Migliorini, and R. Bannister, 2014: Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes in Geophysics*, **21**, 19–39.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: description and sensitivities. *Monthly Weather Review*, **139**, 3887–3905.

Bannister, R., S. Migliorini, and M. A. G. Dixon, 2011: Ensemble prediction for nowcasting with a convection- permitting model - II: forecast error statistics. *Tellus A*, **63**, 497–512.

Bannister, R. N., 2008a: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1951–1970.

— 2008b: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1971–1996.

Barlow, R. J., 1989: *STATISTICS A guide to the Use of Statistical Methods in the Physical Sciences*. The Manchester Physics series, John Wiley and Sons.

Bednarczyk, C. N. and B. C. Ancell, 2015: Ensemble sensitivity analysis applied to a southern plains convective event. *Monthly Weather Review*, **143**, 230–249.

Ben Bouallègue, Z. and S. E. Theis, 2014: Spatial techniques applied to precipitation ensemble forecasts: from verification results to probabilistic products. *Meteorological Applications*, **21**, 922–929.

Bennett, L. J., K. A. Browning, A. M. Blyth, D. J. Parker, and P. A. Clark, 2006: A review of the initiation of precipitating convection in the United Kingdom. *Quarterly Journal of the Royal Meteorological Society*, **132**, 1001–1020.

Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Monthly Weather Review*, **128**, 644–667.

Berre, L. and G. Desroziers, 2010: Filtering of background error variances and correlations by local spatial averaging: A review. *Monthly Weather Review*, **138**, 3693–3720.

Best, M., 2005: Canopy, surface and soil hydrology. Unified Model Documentation Paper 25, UK Meteorological Office, Unified Model Version 5.0.

Birch, C. E., J. H. Marsham, D. J. Parker, and C. M. Taylor, 2014: The scale dependence and structure of convergence fields preceding the initiation of deep convection. *Geophysical Research Letters*, **41**, 4769–4776.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly Weather Review*,

129, 420–436.

Bishop, C. H. and D. Hodyss, 2007: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **133**, 2029–2044.

— 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. *Tellus A*, **61**, 84–96.

— 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus A*, **61**, 97–111.

Blyth, A. M., L. J. Bennett, and C. G. Collier, 2015: High-resolution observations of precipitation from cumulonimbus clouds. *Meteorological Applications*, **22**, 75–89.

Bölöni, G., L. Berre, and E. Adamcsek, 2015: Comparison of static mesoscale background-error covariances estimated by three different ensemble data assimilation techniques. *Quarterly Journal of the Royal Meteorological Society*, **141**, 413–425.

Bouttier, F., B. Vié, O. Nuissier, and L. Raynaud, 2012: Impact of stochastic physics in a convection-permitting ensemble. *Monthly Weather Review*, **140**, 3706–3721.

Bowler, N. E., A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Shutts, 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **135**, 767–776.

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **134**, 703–722.

Browning, K. and F. Hill, 1984: Structure and evolution of a mesoscale convective system near the British Isles. *Quarterly Journal of the Royal Meteorological Society*, **110**, 897–913.

Browning, K. A., 1997: The dry intrusion perspective of extra-tropical cyclone development. *Meteorological Applications*, **4**, 317–324.

Browning, K. A., C. J. Morcrette, J. Nicol, A. M. Blyth, L. J. Bennett, B. J. Brooks, J. Marsham, S. D. Mobbs, D. J. Parker, F. Perry, P. A. Clark, S. P. Ballard, M. A. Dixon, R. M. Forbes, H. W. Lean, Z. Li, N. M. Roberts, U. Corsmeier, C. Barthlott, B. Deny, N. Kalthoff, S. Khodayar, M. Kohler, C. Kottmeier, S. Kraut, M. Kunz, J. Lenfant, A. Wieser, J. L. Agnew, D. Bamber, J. McGregor, K. M. Beswick, M. D. Gray, E. Norton, H. M. A. Ricketts, A. Russell, G. Vaughan, A. R. Webb, M. Bitter, T. Feuerle, R. Hankers, H. Schulz, K. E. Bozier, C. G. Collier, F. Davies, C. Gaffard, T. J. Hewison, D. N. Ladd, E. C. Slack, J. Waight, M. Ramatschi, D. P. Wareing, and R. J. Watson, 2007: The convective storm initiation project. *Bulletin of the American Meteorological Society*, **88**, 1939–1955.

Browning, K. A. and N. M. Roberts, 1994: Use of satellite imagery to diagnose events leading to frontal thunderstorms: Part I of a case study. *Meteorological Applications*, **1**, 303–310.

— 1995: Use of satellite imagery to diagnose events leading to frontal thunderstorms: Part II of a case study. *Meteorological Applications*, **2**, 3–9.

Browning, K. A., N. M. Roberts, and C. S. Sim, 1996: A mesoscale vortex diagnosed from combined satellite and model data. *Meteorological Applications*, **3**, 1–4.

Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, **131**, 2394–2416.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, **131**, 1013–1043.

— 2010: Error statistics in data assimilation: Estimation and modelling. *Data Assimilation*, Springer, 93–112.

— 2012: Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation. *Monthly Weather Review*, **140**, 617–636.

Buehner, M. and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **133**, 615–630.

Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Monthly Weather Review*, **125**, 99–119.

Buizza, R., P. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, **133**, 1076–1097.

Burt, S., 2005: Cloudburst upon Hendraburnick down: the Boscastle storm of 16 August 2004. *Weather*, **60**, 219–227.

Bushell, A., A. Scaife, and C. Warner, 2013: Non-orographic (spectral) gravity wave parametrization. Unified Model Documentation Paper 34, UK Meteorological Office, Unified Model Version 8.5.

Caron, J.-F., 2013: Mismatching perturbations at the lateral boundaries in limited-area ensemble forecasting: A case study. *Monthly Weather Review*, **141**, 356–374.

Caseri, A., P. Javelle, M. Ramos, and E. Leblois, 2015: Generating precipitation ensembles for flood alert and risk management. *Journal of Flood Risk Management*, doi:10.1111/jfr3.12203.

Charney, J. G. and N. Phillips, 1953: Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. *Journal of Meteorology*, **10**, 71–99.

Clark, A. J., J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia Jr, M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Weather and Forecasting*, **28**, 387–407.

Clark, A. J., J. S. Kain, D. J. Stensrud, M. Xue, F. Kong, M. C. Coniglio, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S. J. Weiss, and J. Du, 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.

Clark, A. J., S. J. Weiss, J. S. Kain, I. L. Jirak, M. Coniglio, C. J. Melick, C. Siewert, R. A. Sobash, P. T. Marsh, A. R. Dean, X. Xue, F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, J. Du, D. R. Novak, F. E. Barthold, M. J. Bodner, J. J. Levit, C. B. Entwistle, T. L. Jensen, and J. Correia Jr, 2012: An overview of the 2010 hazardous weather testbed experimental forecast program spring experiment. *Bulletin of the American Meteorological Society*, **93**, 55–74.

Clark, P. A., K. A. Browning, R. M. Forbes, C. J. Morcrette, A. M. Blyth, and H. W. Lean, 2014a: The evolution of an MCS over southern England. Part 2: Model simulations and sensitivity to microphysics. *Quarterly Journal of the Royal Meteorological Society*, **140**, 458–479.

Clark, P. A., K. A. Browning, C. J. Morcrette, A. M. Blyth, R. M. Forbes, B. Brooks, and F. Perry, 2014b: The evolution of an MCS over southern England. Part 1: Observations. *Quarterly Journal of the Royal Meteorological Society*, **140**, 439–457.

Cloke, H. and F. Pappenberger, 2009: Ensemble flood forecasting: a review. *Journal of Hydrology*, **375**, 613–626.

Collier, C. G., 1989: *Applications of weather radar systems*. Wiley-Praxis, second edition.

Craig, G. C. and A. Dörnbrack, 2008: Entrainment in cumulus clouds: What resolution is cloud-resolving? *Journal of the Atmospheric Sciences*, **65**, 3978–3988.

Craig, G. C., C. Keil, and D. Leuenberger, 2012: Constraints on the impact of radar rainfall data assimilation on forecasts of cumulus convection. *Quarterly Journal of the Royal*

Meteorological Society, **138**, 340–352.

Craig, G. C. and M. Würsch, 2013: The impact of localization and observation averaging for convective-scale data assimilation in a simple stochastic model. *Quarterly Journal of the Royal Meteorological Society*, **139**, 515–523.

Cullen, M. and A. Brown, 2009: Large eddy simulation of the atmosphere on various scales. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **367**, 2947–2956.

Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood, 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, **131**, 1759–1782.

Dey, S. R., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A spatial view of ensemble spread in convection permitting ensembles. *Monthly Weather Review*, **142**, 4091–4107.

Dey, S. R., N. M. Roberts, R. S. Plant, and S. Migliorini, 2016a: A new method for the characterisation and verification of local spatial predictability for convective scale ensembles. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.2792. URL `http://dx.doi.org/10.1002/qj.2792`

Dey, S. R., R. S. Plant, N. M. Roberts, and S. Migliorini, 2016b: Assessing spatial precipitation uncertainties in a convective-scale ensemble. *Quarterly Journal of the Royal Meteorological Society*, under review at Quarterly Journal of the Royal Meteorological Society.

Dixon, M., Z. Li, H. Lean, N. Roberts, and S. Ballard, 2009: Impact of data assimilation on forecasting convection over the United Kingdom using a high-resolution version of the Met Office Unified Model. *Monthly Weather Review*, **137**, 1562–1584.

Done, J. M., G. C. Craig, S. L. Gray, and P. A. Clark, 2012: Case-to-case variability of predictability of deep convection in a mesoscale model. *Quarterly Journal of the Royal Meteorological Society*, **138**, 638–648.

Dorninger, M., M. Mittermaier, E. Gilleland, E. Ebert, B. Brown, and L. Wilson, 2013: Mesovict: Mesoscale verification inter-comparison over complex terrain. NCAR Technical Note NCAR/TN-505+STR, NCAR. URL `https://opensky.library.ucar.edu/collections/TECH-NOTE-000-000-000-874`

Dow, G. and B. Macpherson, 2013: Benefit of convective-scale data assimilation and observing systems in the UK models. Forecasting Research Technical Report 585, Met Office.

Duc, L., K. Saito, and H. Seko, 2013: Spatial-temporal fractions verification for high-resolution ensemble forecasts. *Tellus A*, **65**, 1–21.

Duda, J. D. and W. A. Gallus Jr, 2013: The impact of large-scale forcing on skill of simulated convective initiation and upscale evolution with convection-allowing grid spacings in the WRF. *Weather and Forecasting*, **28**, 994–1018.

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, **15**, 51–64.

Eckel, F. A. and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, **20**, 328–350.

Edwards, J., J. Manners, J. Thelen, W. Ingram, and P. Hill, 2014: The radiation code. Unified Model Documentation Paper 23, UK Meteorological Office, Unified Model Version 8.6.

Edwards, J. M. and A. Slingo, 1996: Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Quarterly Journal of the Royal Meteorological Society*, **122**, 689–719.

Ehrendorfer, M., 1997: Predicting the uncertainty of numerical weather forecasts: a review. *Meteorologische Zeitschrift-Berlin*, **6**, 147–183.

Eisenstat, S. C., H. C. Elman, and M. H. Schultz, 1983: Variational iterative methods for nonsymmetric systems of linear equations. *SIAM Journal on Numerical Analysis*, **20**, 345–357.

Errico, R. M. and T. Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. *Monthly Weather Review*, **120**, 1644–1660.

Essery, R., M. Best, R. Betts, P. M. Cox, and C. M. Taylor, 2003: Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *Journal of Hydrometeorology*, **4**, 530–543.

EUMETSAT, 2015: European Organisation for the Exploitation of Meteorological Satellites. Accessed 15/05/2015.
URL `http://www.eumetsat.int/website/home/index.html`

Evensen, G., 2009: *Data assimilation: the ensemble Kalman filter*. Springer.

Fairman, J. G., D. M. Schultz, D. J. Kirshbaum, S. L. Gray, and A. I. Barrett, 2015: A radar-based rainfall climatology of Great Britain and Ireland. *Weather*, **70**, 153–158.

Fritsch, J. and C. Chappell, 1980: Numerical prediction of convectively driven mesoscale pressure systems. Part I: Convective parameterization. *Journal of the Atmospheric Sciences*, **37**, 1722–1733.

Gaborit, E., F. Anctil, V. Fortin, and G. Pelletier, 2014: Hydrologic evaluation of spatially disaggregated global ensemble rainfall forecasts. *Hydrological Processes*, **28**, 4682–4693.

Garcies, L. and V. Homar, 2009: Ensemble sensitivities of the real atmosphere: application to mediterranean intense cyclones. *Tellus A*, **61**, 394–406.

Garratt, J. R., 1994: *The atmospheric boundary layer*. Cambridge university press.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125**, 723–757.

Gebhardt, C., S. Theis, Z. B. Bouallgue, A. Rpnack, N. Schuhen, and M. Buchhold, 2012: COSMO-DE-EPS an operational convection permitting ensemble prediction system for the atmosphere. Deutscher Wetterdienst, DWD, Toulouse, International Conference on Ensemble Methods in Geophysical Sciences. URL `http://www.meteo.fr/cic/meetings/2012/ensemble.conference/presentations/session07/5.pdf`

Gebhardt, C., S. Theis, M. Paulat, and Z. Ben Bouallègue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmospheric Research*, **100**, 168–177.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, **24**, 1416–1430.

Golding, B., S. Ballard, K. Mylne, N. Roberts, A. Saulter, C. Wilson, P. Agnew, L. Davis, J. Trice, C. Jones, D. Simonin, Z. Li, C. Pierce, A. Bennett, M. Weeks, and S. Moseley, 2014: Forecasting capabilities for the London 2012 Olympics. *Bulletin of the American Meteorological Society*, **95**, 883–896.

Golding, B., P. Clark, and B. May, 2005: The Boscastle flood: Meteorological analysis of the conditions leading to flooding on 16 August 2004. *Weather*, **60**, 230–235.

Golding, B. W., 1998: Nimrod: a system for generating automated very short range forecasts. *Meteorological Applications*, **5**, 1–16.

Grant, A., 2001: Cloud-base fluxes in the cumulus-capped boundary layer. *Quarterly Journal of the Royal Meteorological Society*, **127**, 407–421.

Gray, M. and C. Marshall, 1998: Mesoscale convective systems over the UK, 1981–97.

Weather, **53**, 388–396.

Gregory, D. and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Monthly Weather Review*, **118**, 1483–1506.

Hacker, J., C. Snyder, S.-Y. HA, and M. Pocernich, 2011: Linear and non-linear response to parameter variations in a mesoscale model. *Tellus A*, **63**, 429–444.

Halliwell, C., 2014: Subgrid turbulence scheme. Unified Model Documentation Paper 28, UK Meteorological Office, Unified Model Version 9.2.

Hally, A., E. Richard, S. Fresnay, and D. Lambert, 2014: Ensemble simulations with perturbed physical parametrizations: Pre-HyMeX case studies. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1900–1916.

Hamill, T. M. and S. J. Colucci, 1998: Evaluation of eta-rsm ensemble probabilistic precipitation forecasts. *Monthly Weather Review*, **126**, 711–724.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, **129**, 2776–2790.

Hand, W. H., N. I. Fox, and C. G. Collier, 2004: A study of twentieth-century extreme rainfall events in the United Kingdom with implications for forecasting. *Meteorological Applications*, **11**, 15–31.

Hanley, K., D. Kirshbaum, N. Roberts, and G. Leoncini, 2013: Sensitivities of a squall line over central Europe in a convective-scale ensemble. *Monthly Weather Review*, **141**, 112–133.

Hanley, K. E., D. J. Kirshbaum, S. E. Belcher, N. M. Roberts, and G. Leoncini, 2011: Ensemble predictability of an isolated mountain thunderstorm in a high-resolution model. *Quarterly Journal of the Royal Meteorological Society*, **137**, 2124–2137.

Hanley, K. E., R. S. Plant, T. H. M. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2015: Mixing length controls on high resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, **141**, 272–284.

Harris, D., E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit, 2001: Multiscale statistical properties of a high-resolution precipitation forecast. *Journal of Hydrometeorology*, **2**, 406–418.

Harrison, D. L., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorological Ap-*

plications, **7**, 135–144.

Harrison, D. L., K. Norman, C. Pierce, and N. Gaussiat, 2012: Radar products for hydrological applications in the UK. *Proceedings of the ICE-Water Management*, **165**, 89–103.

Harrison, D. L., R. W. Scovell, and M. Kitchen, 2009: High-resolution precipitation estimates for hydrological uses. *Proceedings of the ICE-Water Management*, **162**, 125–135.

Harvey, N. and H. Dacre, 2015: Spatial evaluation of volcanic ash forecasts using satellite observations. *Atmospheric Chemistry and Physics Discussions*, **15**, 24727–24749.

Hohenegger, C., D. Lüthi, and C. Schär, 2006: Predictability mysteries in cloud-resolving models. *Monthly Weather Review*, **134**, 2095–2107.

Hohenegger, C. and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bulletin of the American Meteorological Society*, **88**, 1783–1793.

Hollan, M. A. and B. C. Ancell, 2015: Ensemble mean storm-scale performance in the presence of nonlinearity. *Monthly Weather Review*, **143**, 5115–5133.

Holloway, C. E., J. C. Petch, R. J. Beare, P. Bechtold, G. C. Craig, S. H. Derbyshire, L. J. Donner, P. R. Field, S. L. Gray, J. H. Marsham, D. J. Parker, R. S. Plant, N. M. Roberts, D. M. Schultz, A. J. Stirling, and S. J. Woolnough, 2014: Understanding and representing atmospheric convection across scales: recommendations from the meeting held at Dartington Hall, Devon, UK, 28–30 January 2013. *Atmospheric Science Letters*, **15**, 348–353.

Hoskins, B. J., M. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, **111**, 877–946.

Houze, R. A., 2004: Mesoscale convective systems. *Reviews of Geophysics*, **42**, doi:8755-1209/04/2004RG000150.

Houze, R. A. and A. K. Betts, 1981: Convection in GATE. *Reviews of Geophysics*, **19**, 541–576.

Ingleby, N. B., 2001: The statistical structure of forecast errors and its representation in the Met. Office Global 3-D Variational Data Assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, **127**, 209–231.

Ingleby, N. B., A. C. Lorenc, K. Ngan, F. Rawlins, and D. R. Jackson, 2013: Improved variational analyses using a nonlinear humidity control variable. *Quarterly Journal of the Royal Meteorological Society*, **139**, 1875–1887.

Johnson, A. and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Monthly Weather Review*, **140**, 3054–3077.

— 2013: Object-based evaluation of a storm-scale ensemble during the 2009 NOAA Hazardous Weather Testbed Spring Experiment. *Monthly Weather Review*, **141**, 1079–1098.

Johnson, A., X. Wang, M. Xue, F. Kong, G. Zhao, Y. Wang, K. W. Thomas, K. A. Brewster, and J. Gao, 2014: Multiscale characteristics and evolution of perturbations for warm season convection-allowing precipitation forecasts: Dependence on background flow and method of perturbation. *Monthly Weather Review*, **142**, 1053–1073.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press.

Keil, C. and G. C. Craig, 2011: Regime-dependent forecast uncertainty of convective precipitation. *Meteorologische Zeitschrift*, **20**, 145–151.

Keil, C., F. Heinlein, and G. C. Craig, 2014: The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quarterly Journal of the Royal Meteorological Society*, **140**, 480–490.

Kirchgessner, P., L. Nerger, and A. Bunse-Gerstner, 2014: On the choice of an optimal localization radius in ensemble Kalman filter methods. *Monthly Weather Review*, **142**, 2165–2175.

Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2006: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparsion of coarse- and fine-grid experiments. *Monthly Weather Review*, **134**, 807–833.

— 2007: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part II: Storm-scale experiments. *Monthly Weather Review*, **135**, 759–782.

Kühnlein, C., C. Keil, G. C. Craig, and C. Gebhardt, 2014: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1552–1562.

Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Monthly Weather Review*, **136**, 3408–3424.

Leith, C., 1974: Theoretical skill of monte carlo forecasts. *Monthly Weather Review*, **102**, 409–418.

Leon, D. C., J. R. French, S. Lasher-Trapp, A. M. Blyth, S. J. Abel, S. Ballard, A. Barrett, L. J. Bennett, K. Bower, B. Brooks, P. Brown, C. Charlton-Perez, T. Choularton, P. Clark, C. Collier, J. Crosier, Z. Cui, S. Dey, D. Dufton, C. Eagle, M. J. Flynn, M. Gallagher, C. Halliwell, K. Hanley, L. Hawkness-Smith, Y. Huang, G. Kelly, M. Kitchen, A. Korolev, H. Lean, Z. Liu, J. Marsham, D. Moser, J. Nicol, E. G. Norton, D. Plummer, J. Price, H. Ricketts, N. Roberts, P. D. Rosenberg, D. Simonin, J. W. Taylor, R. Warren, P. I. Williams, and G. Young, 2015: The COnvective Precipitation Experiment (COPE): Investigating the origins of heavy precipitation in the southwestern UK. *Bulletin of the American Meteorological Society*, doi:http://dx.doi.org/10.1175/BAMS-D-14-00157.1, Early Online Release.

Leoncini, G., R. Plant, S. Gray, and P. Clark, 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.

Leoncini, G., R. S. Plant, S. L. Gray, and P. A. Clark, 2010: Perturbation growth at the convective scale for CSIP IOP18. *Quarterly Journal of the Royal Meteorological Society*, **136**, 653–670.

Leoncini, G., N. Roberts, and S. R. A. Dey, 2016: On ensemble spread and skill at the convective scale, manuscript to be submitted to Monthly Weather Review.

Leoncini, G., N. Roberts, and B. Golding, 2011: 8th July 2011 Floods in Scotland. *SEPA report*.

Leutbecher, M. and T. N. Palmer, 2008: Ensemble forecasting. *Journal of Computational Physics*, **227**, 3515–3539.

Lewis, M. W. and S. L. Gray, 2010: Categorisation of synoptic environments associated with mesoscale convective systems over the UK. *Atmospheric Research*, **97**, 194–213.

Lock, A., A. Brown, M. Bush, G. Martin, and R. Smith, 2000: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Monthly Weather Review*, **128**, 3187–3199.

Lock, A. and J. Edwards, 2012: The parametrization of boundary layer processes. Unified Model Documentation Paper 24, UK Meteorological Office, Unified Model Version 8.2.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.

Majumdar, S., C. H. Bishop, B. Etherton, and Z. Toth, 2002: Adaptive sampling with the ensemble transform kalman filter. part ii: Field program implementation. *Monthly Weather Review*, **130**, 1356–1369.

Malia, N., 2012: *Summer Severe convective storms in the United Kingdom*. Master's thesis, Departement of Meteorology, University of Reading.

Manners, J., S. Vosper, and N. Roberts, 2012: Radiative transfer over resolved topographic features for high-resolution weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **138**, 720–733.

Marsigli, C., A. Montani, and T. Paccagnella, 2014: Perturbation of initial and boundary conditions for a limited-area ensemble: multi-model versus single-model approach. *Quarterly Journal of the Royal Meteorological Society*, **140**, 197–208.

Martin, J. E., 2006: *Mid-latitude atmospheric dynamics: a first course*. John Wiley & Sons.

Martin, W. J. and M. Xue, 2006: Sensitivity analysis of convection of the 24 May 2002 IHOP case using very large ensembles. *Monthly Weather Review*, **134**, 192–207.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bulletin of the American Meteorological Society*, **83**, 407–430.

McBeath, K., P. Field, and R. Cotton, 2014: Using operational weather radar to assess high-resolution numerical weather prediction over the British Isles for a cold air outbreak case-study. *Quarterly Journal of the Royal Meteorological Society*, **140**, 225–239.

Melhauser, C. and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *Journal of the Atmospheric Sciences*, **69**, 3350–3371.

Met Office, 2013: Met office weather summaries. Accessed 28/04/2015.
URL `http://www.metoffice.gov.uk/climate/uk/summaries/2013/summer`

— 2015: Met office weather summaries. Accessed 01/05/2015.
URL `http://www.metoffice.gov.uk/public/weather/climate`

Migliorini, S., M. Dixon, R. Bannister, and S. Ballard, 2011: Ensemble prediction for nowcasting with a convection-permitting model I: description of the system and the impact of radar-derived surface precipitation rates. *Tellus A*, **63**, 468–496.

Mittermaier, M., E. Gilleland, M. Dorninger, E. Ebert, and L. Wilson, 2013a: Intercomparison of spatial forecast verification methods: a review and new project launch, 13th EMS Annual Meeting and 11th European Conference on Applications of Meteorology (ECAM). URL `http://meetingorganizer.copernicus.org/EMS2013/EMS2013-148.pdf`

Mittermaier, M. and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the Fractions Skill Score. *Weather and*

Forecasting, **25**, 343–354.

Mittermaier, M., N. Roberts, and S. A. Thompson, 2013b: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorological Applications*, **20**, 176–186.

Mittermaier, M. P., 2014: A strategy for verifying near-convection-resolving model forecasts at observing sites. *Weather and Forecasting*, **29**, 185–204.

Morcrette, C., H. Lean, K. Browning, J. Nicol, N. Roberts, P. Clark, A. Russell, and A. Blyth, 2007: Combination of mesoscale and synoptic mechanisms for triggering an isolated thunderstorm: Observational case study of CSIP IOP 1. *Monthly Weather Review*, **135**, 3728–3749.

Muhlbauer, A., E. Berry, J. M. Comstock, and G. G. Mace, 2014: Perturbed physics ensemble simulations of cirrus on the cloud system-resolving scale. *Journal of Geophysical Research: Atmospheres*, **119**, 4709–4735.

Mylne, K., 2013: Scientific framework for the ensemble prediction system for the UKV. MOSAC PAPER 18.6, UK Meteorological Office. URL `http://www.metoffice.gov.uk/media/pdf/q/0/MOSAC_18.6_Mylne.pdf`

Nachamkin, J. E. and J. Schmidt, 2015: Applying a neighborhood fractions sampling approach as a diagnostic tool. *Monthly Weather Review*, **143**, 4736–4749.

NERC Satellite Receiving Station, 2015: Dundee University, Scotland. Accessed 14/04/2015.
URL `http://www.sat.dundee.ac.uk/`

Nogueira, M. and A. Barros, 2014: The nonconvective/convective structural transition in stochastic scaling of atmospheric fields. *Journal of Geophysical Research: Atmospheres*.

Palmer, T. N., 2000: Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics*, **63**, 71.

Pannekoucke, O., L. Berre, and G. Desroziers, 2007: Filtering properties of wavelets for local background-error correlations. *Quarterly Journal of the Royal Meteorological Society*, **133**, 363–379.

Poterjoy, J. and F. Zhang, 2011: Dynamics and structure of forecast error covariance in the core of a developing hurricane. *Journal of the Atmospheric Sciences*, **68**, 1586–1606.

Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quarterly Journal of the Royal Meteorological Society*, **122**, 121–150.

Radhakrishna, B., I. Zawadzki, and F. Fabry, 2012: Predictability of precipitation from continental radar images. Part V: Growth and decay. *Journal of the Atmospheric Sciences*, **69**, 3336–3349.

Raynaud, L., O. Pannekoucke, P. Arbogast, and F. Bouttier, 2015: Application of a Bayesian weighting for short-range lagged ensemble forecasting at the convective scale. *Quarterly Journal of the Royal Meteorological Society*, **141**, 459–468.

Rezacova, D., P. Zacharov, and Z. Sokol, 2009: Uncertainty in the area-related QPF for heavy convective precipitation. *Atmospheric Research*, **93**, 238–246.

Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, **15**, 163–169.

Roberts, N. M., 2000a: A guide to aspects of water vapour imagery interpretation: the significance of dry regions. Forecasting Research Technical Report 109, Met Office.

— 2000b: The relationship between water vapour imagery and thuderstorms. Forecasting Research Technical Report 110, Met Office.

— 2003: The impact of a change to the use of the convection scheme to high resolution simulations of convective events. Forecasting Research Technical Report 407, Met Office.

Roberts, N. M. and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136**, 78–97.

Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Monthly Weather Review*, **142**, 4519–4541.

Russell, A., G. Vaughan, E. Norton, C. Morcrette, K. Browning, and A. Blyth, 2008: Convective inhibition beneath an upper-level PV anomaly. *Quarterly Journal of the Royal Meteorological Society*, **134**, 371–383.

Russell, A., G. Vaughan, E. G. Norton, H. Ricketts, C. J. Morcrette, T. J. Hewison, K. Browning, and A. M. Blyth, 2009: Convection forced by a descending dry layer and low-level moist convergence. *Tellus A*, **61**, 250–263.

Scaife, A., N. Butchart, C. Warner, and R. Swinbank, 2002: Impact of a spectral gravity wave parameterization on the stratosphere in the Met Office Unified Model. *Journal of the Atmospheric Sciences*, **59**, 1473–1489.

Schmetz, J., P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier, 2002: An introduction to Meteosat Second Generation (MSG). *Bulletin of the American Meteorological Society*, **83**, 977–992.

Schumacher, R. S. and A. J. Clark, 2014: Evaluation of ensemble configurations for the analysis and prediction of heavy-rain-producing mesoscale convective systems. *Monthly Weather Review*, **142**, 4108–4138.

Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, and M. C. Coniglio, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Monthly Weather Review*, **137**, 3351–3372.

Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, M. C. Coniglio, and M. S. Wandishin, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, **25**, 263–280.

Simonin, D., S. Ballard, and Z. Li, 2014: Doppler radar radial wind assimilation using an hourly cycling 3D-Var with a 1.5 km resolution version of the Met Office Unified Model for nowcasting. *Quarterly Journal of the Royal Meteorological Society*, **140**, 2298–2314.

Skok, G., 2015: Analysis of Fraction Skill Score properties for a displaced rainband in a rectangular domain. *Meteorological Applications*, **22**, 477–484.

Smagorinsky, J., 1963: General circulation experiments with the primitive equations: I. the basic experiment. *Monthly Weather Review*, **91**, 99–164.

Smith, R., 1990: A scheme for predicting layer clouds and their water content in a general circulation model. *Quarterly Journal of the Royal Meteorological Society*, **116**, 435–460.

Smolarkiewicz, P. K. and L. G. Margolin, 1994: Variational elliptic solver for atmospheric applications. Technical report, Los Alamos National Lab., NM (United States).

Sobash, R. A. and D. J. Stensrud, 2013: The impact of covariance localization for radar data on enKF analyses of a developing MCS: Observing system simulation experiments. *Monthly Weather Review*, **141**, 3691–3709.

Staniforth, A. and J. Côté, 1991: Semi-lagrangian integration schemes for atmospheric models-a review. *Monthly weather review*, **119**, 2206–2223.

Staniforth, A., A. White, J. Thuburn, M. Zerroukat, E. Cordero, T. Davies, and M. Diamantakis, 2006: Joy of U.M.6.3 - model formulation. Unified Model Documentation Paper 15, UK Meteorological Office, Unified Model Version 6.3.

Steiner, M., R. A. Houze Jr, and S. E. Yuter, 1995: Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. *Journal of Applied Meteorology*, **34**, 1978–2007.

Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, **128**, 2077–2107.

Stratman, D. R., M. C. Coniglio, S. E. Koch, and M. Xue, 2013: Use of multiple verification methods to evaluate forecasts of convection from hot-and cold-start convection-allowing models. *Weather and Forecasting*, **28**, 119–138.

Stratton, R., N. Willett, S. Derbyshire, and R. Wong, 2012: Convection scheme. Annex to Unified Model Documentation Paper 29A, UK Meteorological Office, Unified Model Version 8.2.

Sun, J., M. Xue, J. W. Wilson, I. Zawadzki, S. P. Ballard, J. Onvlee-Hooimeyer, P. Joe, D. M. Barker, P.-W. Li, B. Golding, M. Xu, and J. Pinto, 2014: Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bulletin of the American Meteorological Society*, **95**, 409–426.

Surcel, M., I. Zawadzki, and M. K. Yau, 2014: On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Monthly Weather Review*, **142**, 1093–1105.

Tang, Y., H. W. Lean, and J. Bornemann, 2013: The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications*, **20**, 417–426.

Tennant, W., 2015: Improving initial condition perturbations for MOGREPS-UK. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.2524, Early Online Release.

Theis, S., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteorological Applications*, **12**, 257–268.

Torn, R. D. and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Monthly Weather Review*, **136**, 663–677.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, **125**, 3297–3319.

Verrelle, A., D. Ricard, and C. Lac, 2015: Sensitivity of high-resolution idealized simulations of thunderstorms to horizontal resolution and turbulence parametrization.

Quarterly Journal of the Royal Meteorological Society, **141**, 433–448.

Vetra-Carvalho, S., M. Dixon, S. Migliorini, N. K. Nichols, and S. P. Ballard, 2012: Breakdown of hydrostatic balance at convective scales in the forecast errors in the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **138**, 1709–1720.

Vié, B., G. Molinié, O. Nuissier, B. Vincendon, V. Ducrocq, F. Bouttier, and E. Richard, 2012: Hydro-meteorological evaluation of a convection-permitting ensemble prediction system for Mediterranean heavy precipitating events. *Natural Hazards and Earth System Science*, **12**, 2631–2645.

Walser, A., D. Lüthi, and C. Schär, 2004: Predictability of precipitation in a cloud-resolving model. *Monthly Weather Review*, **132**, 560–577.

Walser, A. and C. Schär, 2004: Convection-resolving precipitation forecasting and its predictability in Alpine river catchments. *Journal of Hydrology*, **288**, 57–73.

Wang, Y., M. Bellus, J.-F. Geleyn, X. Ma, W. Tian, and F. Weidle, 2014: A new method for generating initial condition perturbations in a regional ensemble prediction system: Blending. *Monthly Weather Review*, **142**, 2043–2059.

Warren, R. A., 2014: *Quasi-stationary convective systems in the UK*. Ph.D. thesis, Department of Meteorology, University of Reading.

Warren, R. A., D. J. Kirshbaum, R. S. Plant, and H. W. Lean, 2014: A Boscastle-type quasi-stationary convective system over the UK Southwest Peninsula. *Quarterly Journal of the Royal Meteorological Society*, **140**, 240–257.

Webster, S., A. Brown, D. Cameron, and C. Jones, 2003: Improvements to the representation of orography in the Met Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **129**, 1989–2010.

Webster, S. and H. Wells, 2011: Gravity wave drag. Unified Model Documentation Paper 22, UK Meteorological Office, Unified Model Version 8.1.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0-36-h explicit convective forecasts with the WRF-ARW model. *Weather and Forecasting*, **23**, 407–437.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, **125**, 527–548.

Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SAL-A novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, **136**, 4470–4487.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman filter analyses and forecasts of a severe mesoscale convective system using different choices of microphysics schemes. *Monthly Weather Review*, **142**, 3243–3263.

Wilkinson, J., 2011: The large-scale precipitation parametrization scheme. Unified Model Documentation Paper 26, Unified Model Version: 8.0, Met Office.

— 2012: The large-scale precipitation parametrization scheme. Unified Model Documentation Paper 26, UK Meteorological Office, Unified Model Version 8.2.

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences*, volume 59. Academic Press.

— 2011: *Statistical methods in the atmospheric sciences*, volume 100. Academic press.

Wilson, D., 2011: The large-scale cloud scheme and saturated specific humidity. Unified Model Documentation Paper 29, UK Meteorological Office, Unified Model Version 8.0.

Wilson, D., A. Bushell, and C. Mocrette, 2011: The PC2 cloud scheme. Annex to Unified Model Documentation Paper 29A, UK Meteorological Office, Unified Model Version 8.1.

Wilson, D. R. and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **125**, 1607–1636.

Wilson, D. R., A. Bushell, A. M. Kerr-Munslow, J. D. Price, C. J. Morcrette, and A. Bodas-Salcedo, 2008a: PC2: A prognostic cloud fraction and condensation scheme. II: Climate model simulations. *Quarterly Journal of the Royal Meteorological Society*, **134**, 2109–2125.

Wilson, D. R., A. C. Bushell, A. M. Kerr-Munslow, J. D. Price, and C. J. Morcrette, 2008b: PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quarterly Journal of the Royal Meteorological Society*, **134**, 2093–2107.

Wood, N., A. Staniforth, A. White, T. Allen, M. Diamantakis, M. Gross, T. Melvin, C. Smith, S. Vosper, M. Zerroukat, et al., 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, **140**, 1505–1520.

Zacharov, P. and D. Rezacova, 2009: Using the fractions skill score to assess the relationship between an ensemble QPF spread and skill. *Atmospheric Research*, **94**, 684–693.

Zhang, F., 2005: Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. *Monthly Weather Review*, **133**, 2876–2893.

Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the surprise snowstorm of 2425 January 2000. *Monthly Weather Review*, **130**, 1617–1632.

— 2003: Effects of moist convection on mesoscale predictability. *Journal of the Atmospheric Sciences*, **60**, 1173–1185.

Zimmer, M., G. C. Craig, C. Keil, and H. Wernli, 2011: Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophysical Research Letters*, **38**, doi:10.1029/2010GL046199.