

An assessment of kriging-based rain-gauge–radar merging techniques

Sharon A. Jewell* and Nicolas Gaussiat
Met Office, Exeter, UK

*Correspondence to: S. A. Jewell, Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, UK.
E-mail: sharon.jewell@metoffice.gov.uk

This article is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

Networks of rain-gauges provide an accurate but highly localized measure of rainfall, with limited coverage and resolution, whereas radars provide rain-rate and accumulation estimates over wide areas at high spatial and temporal resolution but low accuracy. When quantifying rainfall accumulations for applications such as flood forecasting, combining the two sets of data can be beneficial for producing a high-resolution merged product with a lower error than the gauge-only or the radar-only product. In this study, three kriging methods (kriging with external drift (KED), kriging with radar-based error correction (KRE) and ordinary kriging (OK)) as well as a multiquadric (MQ) scheme have been used to merge radar and gauge data. The results were cross-validated with true rainfall readings at the surface for a number of different meteorological events covering England and Wales. Overall, all the merging schemes trialled produced a merged product that was superior to the individual radar or gauge data. The KED was the best performing method across all rainfall thresholds and meteorological conditions, with the use of a parametric variogram best suited to short (15 min) accumulation periods and a non-parametric variogram preferable for hourly accumulations. The study also shows that the merged product deteriorates when the gauge-network density is reduced; exceptions to this are made in situations where spatially isolated rainfall is observed by the radar and where the gauges that are found least likely to represent the rainfall are carefully removed prior to merging. The latter process was found to improve the quality of the merged product, regardless of the meteorological conditions, providing a viable method for producing a KED-based merging scheme that is applicable to all meteorological conditions.

Key Words: gauge–radar merging; kriging; radar; rain-gauges

Received 17 March 2014; Revised 18 December 2014; Accepted 20 January 2015; Published online in Wiley Online Library
17 March 2015

1. Introduction

Accurate, real-time rainfall measurements are critical inputs to any flood forecasting scheme. A flood forecast model requires rainfall measurements across an entire domain to be combined with a surface run-off model to provide the end-user with reliable, up-to-date information on the expected water level for decision-making processes (Moore, 2002; Bell *et al.*, 2007; Cole and Moore, 2008). Rain-gauge networks can provide an accurate but very localized measure of rainfall, with limited coverage and spatial resolution. Conversely, radars provide high-resolution observations of the precipitation field, but the quantitative precipitation estimation (QPE) at the ground is riddled with uncertainties. The QPE errors result from various factors such as the calibration of the radar, attenuation from intense precipitation and hail as well as the wet radome, the regional variations of the vertical reflectivity profile and of the $Z - R$ relationship. Combining independent observations of the rainfall from radar and rain-gauges enables high-resolution merged products to be

produced that are of greater precision and accuracy than the original gauge or radar datasets separately.

The first method for combining meteorological radar data with rain-gauge readings was proposed by Brandes (1975) and used a calibration factor derived from the ratio of radar and rain-gauge readings to correct the entire rainfall field derived from the radar. However, such a technique lacks any sensitivity to localized variations in the rainfall due to spatially isolated rainfall events that are not representative of the wider area and fails to adequately compensate for localized biases induced by beam blockages or attenuation. In contrast, the use of geospatial interpolation methods to create gridded merged products in real-time using weighting factors to produce localized corrections to radar data has the potential to account for localized variations in the nature of the rainfall and the measurement conditions, but at considerably higher computational cost. The computational time taken to produce an operational product is an important consideration, as for flood-forecasting applications in the United Kingdom during periods of intense rainfall a merged 15 min accumulation product is required every 15 min. For the non-optimized codes

used for this study it takes 5–15 min for one kriging merging scheme to run for a single accumulation period (depending on the merging scheme and the total rainfall). The consequence of this is that a careful balance has to be made between the number of rainfall events tested (in this case, four different distinct rainfall events) and the computational time required when testing the performance of multiple merging schemes.

Gauge–radar merging methods have been the subject of many reviewed articles highlighting the key features of different interpolation schemes (Sinclair and Pegram, 2005; Schiemann *et al.*, 2011). Kitzmiller and co-workers (Kitzmiller *et al.*, 2013) provide an overview of multisensor quantitative precipitation-estimate techniques used by the National Weather Service (NWS) for use in hydrological operations, where the refinement of the radar data is mainly limited to mean-field and local bias adjustments. The use of more sophisticated schemes is discussed by Goudenhoofdt and Delobbe (2009). They applied a number of geospatial methods to daily rainfall accumulations over a 4 year period in the Walloon region of Belgium, and the results highlight the added benefit of using a gridded interpolated merged product over a simple bias correction, with the best performing product being a form of kriging (Wackernagel, 2010). Kriging is a general term for a number of geostatistical techniques that can be used to interpolate the value of a random field (in this case the gauge network) at a location where the true value is unknown. The key feature of kriging is that the weighting factors used to calculate the rainfall at a point take into account not only the distance from neighbouring gauges, but also how far apart neighbouring gauges are from each other. The process assumes that points that are within a predetermined distance of each other have a certain degree of spatial correlation (as would generally be expected in a rainfall field), whereas points that are separated by extensive distances are statistically independent. This spatial correlation is incorporated into the merging algorithm through the use of a variogram, as discussed in section 3.

In the absence of radar data, ordinary kriging (OK) has been shown to produce an interpolated rainfall field from a network of rain-gauges that is superior to one generated from simpler techniques such as inverse distance schemes (Nour *et al.*, 2006; Belo-Pereira *et al.*, 2011). However, it is when two independent, synchronous datasets such as rain-gauges and radar measurements are used that the true benefit of kriging becomes apparent. Numerous kriging techniques have been developed to merge the gauge and radar data and they can be broadly split into two categories. The first category uses the datasets to produce two separate interpolated fields (one based on the gauge data and the other on the radar) with the difference between the two results used to modify the original radar field; examples of this method include kriging with radar-based error correction (KRE) (Ehret *et al.*, 2008) and conditional merging (Sinclair and Pegram, 2005). The second category uses a more rigorous approach by combining the information from the two datasets prior to the generation of any interpolated field. This effectively constrains the merged value by requiring the application of the weights to the radar values at the gauge locations to produce a radar value at the ungauged point consistent with the measured radar value. Approaches such as these include kriging with external drift (KED) (Haberlandt, 2007; Velasco-Forero *et al.*, 2009) and co-kriging (CK) (Schuurmans *et al.*, 2007; Sideris *et al.*, 2014).

The study presented here explores the results of merging historical gauge and radar data collected from across England and Wales using OK, KED and multiquadric (MQ) schemes. The specific aim of this work was to determine the optimal scheme to use to produce a merged 15 min accumulation product and a 60 min accumulation product in real-time every 15 min for flood forecasting purposes. The quality and detection efficiency of the schemes are examined for four different periods that exhibited a variety of different meteorological conditions. In addition, for each scheme the influence of parameters such as gauge density, accumulation time and degree of rainfall correlation are systematically assessed.

Table 1. A summary of the characteristics of the four different meteorological events examined during this study.

	Slow moving	Fast moving
Stratiform	1–6 October 2010 Slow and erratically moving weather fronts. Widespread heavy rain across the country	12–16 January 2011 Widespread rain and heavy orographically enhanced rainfall. Front lingered in the south
Spatially uncorrelated	15–18 July 2011 Deep depression tracked across the country bringing heavy localized showers	3–8 August 2011 Scattered showers and localized thunderstorms following a belt of rain

The article is organized as follows: first, the meteorological conditions over the periods chosen for intercomparing the merging schemes are described and the gauge network and cross-validation scheme are introduced. Next, each of the merging schemes used in the study are presented and their performances compared against one another on 60 min accumulations. The same analysis is then conducted with 15 min accumulation merged products, derived over the same period, and combined to form 60 min accumulations. Finally, the resilience of each of the kriging scheme to a reduction in the gauge network density is assessed, and the sensitivity of the results to the degree of correlation of the rainfall field is shown.

2. Meteorological data and assessment methods

2.1. Meteorological conditions

The gauge–radar merging study was performed using canned rain-gauge data collected from a total of 1064 Met Office and Environment Agency gauges during specified periods between 1 October 2010 and 31 August 2011. The radar data for the same period was obtained from the UK Met Office radar archive. Four different periods, described below, were selected for the test, which covered a range of meteorological conditions. These events can be broadly split into categories as shown in Table 1.

2.1.1. 1–6 October 2010

The period was defined by a low pressure west of Ireland and high pressure over the Baltic. Atlantic fronts made slow and erratic progress across the United Kingdom. Rain fell widely and often heavily on the first 2 days. There was further rain and showers daily until the 6th.

2.1.2. 12–16 January 2011

A mild southwesterly airflow covered the British Isles between the 12th and 16th bringing widespread rain; there were some heavy orographically enhanced falls in upland western districts. On the 16th a cold front brought a drop in the temperature, but the front lingered near southern England with widespread rainfall.

2.1.3. 15–18 July 2011

An unusually deep depression for the season tracked southeastwards across Scotland. Heavy rain fell over most parts of the British Isles during this period.

2.1.4. 3–8 August 2011

Scattered showers and local thunderstorms occurred on the 3rd, notably in the east Midlands and Yorkshire. A belt of rain spread northeastwards across the country on the 4th. This was followed by a depression on the 6th, bringing heavy rain to Scotland

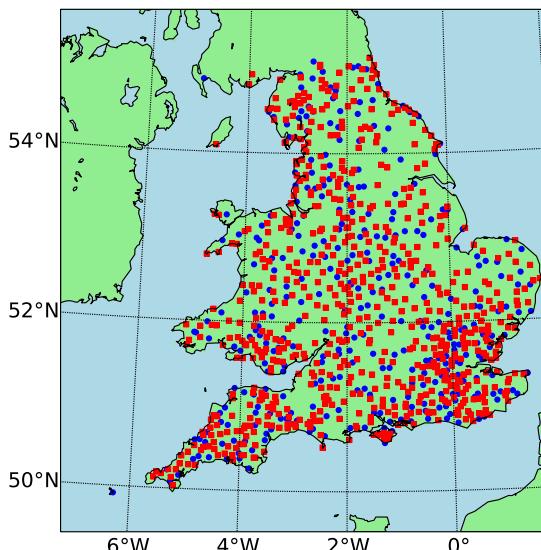


Figure 1. Distribution of merging (red squares) and reference (blue dots) gauges used on the gauge–radar merging study. One thousand and sixty-four gauges are available in total.

and northern England. Heavy showers and local thunderstorms affected England and Wales on the 7th.

2.2. Rain gauge network

The 1064 rain-gauges used were distributed across England and Wales as shown in Figure 1. All of the gauges used were tipping bucket rain-gauges (TBRs) quantized at 0.2 mm accumulations and recorded time-of-tip data. For each gauge, 15 min accumulation totals commencing at hh:00, hh:15, hh:30 and hh:45 were available for the period from September 2010 to September 2011. The accumulation data available comprised gauge location, gauge accumulation and a data-quality flag which read either 0 or 1. This flag indicated whether a gauge reading was suspicious (e.g. due to a suspected blocked gauge or an unexpectedly high reading), but any further diagnostic information relating to the flag was not available. Although rain-gauges are regarded as reliable instruments for providing a ‘ground truth’ value of the rainfall at a specific point, they can still occasionally be subject to random and systematic errors. These can arise from meteorological factors such as wind redirecting the rainfall or solid precipitation blocking the gauge, as well as mechanical failures such as double-tips (in the case of tipping bucket gauges) and telemetry systems failing to report a reading (Larson and Peck, 1974; Sevruk, 1989; Ciach, 2003). To reduce errors being introduced into the merging scheme, standard quality control procedures were used to check for erroneous gauge readings, and the suspected gauges were rejected. In addition, to ensure that a fair comparison could be made between the rainfall periods, only meteorological events involving non-solid precipitation were used.

In order for the results from the merging to be verified, the gauge network was randomly split into two groups. Two-thirds of the gauges (711 in total) were assigned as merging gauges (denoted by red squares in Figure 1) and the remaining 353 (blue circles) were used as reference gauges to provide a ‘true’ rainfall reading for comparison with the gauge–radar merged values. The same merging and cross-validation gauges were used for each of the merging schemes tested, apart from the gauge density study, which is detailed separately.

2.3. United Kingdom radar network

The Met Office operates 15 C-band (5 cm wavelength) weather radars covering the whole of the United Kingdom. Reflectivity scans are taken from four or more elevations every 5 min with high-resolution reflectivity volume data collected out to a range

of 255 km from the radar site in real-time. All quality control procedures to flag errors, such as ground clutter, anaprop, bright band correction and attenuation due to intense rainfall conditions, are carried out centrally on the highest resolution polar format data prior to the calculation of the QPEs. For each single-site QPE a single adjustment factor is calculated from the ratio of radar and gauge values in the region and applied to the single-site rainfall prior to compositing. The radar composite (multiradar QPEs) are generated directly from the processed polar formats from individual radars and the spatial averaging and reprojection is performed in a single step. The radar composites provide a measure of the instantaneous rainfall rate across the United Kingdom on a 1 km resolution grid every 5 min. Multiple composites for the required time period are then combined to produce an estimate of the rainfall during the required accumulation period.

It is important to note that all of the kriging schemes presented in this study follow the approach of Goudenhoofdt and Delobbe (2009), Schiemann *et al.* (2011) and Velasco-Forero *et al.* (2009) and use the rainfall data in their original form rather than transforming them into a Gaussian space. The kriging process is more suited to a transformation of the data to a Gaussian distribution (Erdin *et al.*, 2012), however, it was found that for the current merging code a Gaussian transformation tended to introduce an increased number of numerical instabilities in the final output.

2.4. Cross-validation and skill-score statistics

The performance of a merging scheme during a meteorological event was assessed by calculating the skill-score statistics for reference gauges measuring ‘true’ rainfall above a given threshold. Unless otherwise stated, a low rainfall was defined as $<2 \text{ mm h}^{-1}$ and high rainfall as $>6 \text{ mm h}^{-1}$. In all cases the equivalent threshold is applied to the rain-gauge accumulation at the cross-validation locations. As the TBR gauges used for recording the rainfall are quantized at an accumulation of 0.2 mm, this is the minimum rainfall accumulation used in the cross-validation statistics, and gauges registering no rainfall are not included in the cross-validation process.

2.4.1. Continuous statistics

The merged data quality was assessed by continuous statistics. For each reference gauge reading above the minimum threshold, the magnitude of the difference between the merged and true rainfall value (or radar and true rainfall values in the case of radar-only data) was calculated. These values were then combined over every accumulation period of every day in the merging scheme to calculate the root mean square error (RMSE) and mean average error (MAE) of the scheme.

2.4.2. Binary statistics

The detection efficiency for each merging scheme was assessed using binary statistics. For a chosen rainfall threshold the binary techniques assign the merged and measured accumulations at each reference gauge as being in one of four categories. The categories are as follows: correctly predicted accumulations above the threshold (‘hits’), incorrectly predicted accumulations below the threshold (‘misses’), merged values incorrectly giving rainfall above the threshold (‘false alarms’) and merged and true rainfall both below the threshold (‘correct rejections’).

The critical success index (CSI) of each scheme (which has a maximum value of 1) was calculated for each meteorological event by dividing the number of hits by the total number of hits, false alarms and misses for the event. In comparison, the frequency bias (fBIAS) is the ratio of the number of merged events above the threshold (‘hits’ plus ‘false alarms’) to the true number of events above the threshold (‘hits’ plus ‘misses’). This provides a measure of how biased the scheme is, with a score >1 indicating

a tendency to produce false alarms and a score <1 suggesting that the scheme tends to underestimate the precipitation.

3. Overview of rain-gauge–radar merging methods

A number of different mathematical approaches to generating gauge and radar merged products have been proposed and implemented in recent years with varying degrees of success. These geospatial interpolation techniques range from statistical methods based on applying weighting functions to individual gauges to fitting mathematical functions to the entire gauge network. This study focuses on a number of techniques that have shown the greatest potential for use in operational gauge–radar merging. An overview of each technique is given here, with references to more detailed descriptions of the merging process.

3.1. Multiquadric

The multiquadric surface-fitting technique is an analytical method that represents an irregular surface through the summation of a series of quadric surfaces centred on discrete points (Moore *et al.*, 1994; Balascio, 2001). In the context of gauge–radar merging, the result is a weighted rainfall field generated from the rain-gauge and radar measurements obtained at each gauge location. It provides a simple, direct and computationally efficient method of combining the two sources of data. In this study, the process follows the approach of Moore *et al.* (1994).

The process first calculates the ratio between the rain-rate reading at gauge location i (R_g^i) and the radar rain rate (R_r^i) at the co-located radar pixel (i.e. the calibration factor) at n gauge locations as follows:

$$z_i = \frac{R_g^i + \varepsilon_g}{R_r^i + \varepsilon_r} \quad (1)$$

where ε_g and ε_r are small constant values (in this case 0.3 mm h^{-1}) to ensure that the calibration factor is defined for zero rainfall. The multiquadric calibration surface $s(\mathbf{x})$ is defined as a weighted sum of n distance functions ($g(\mathbf{x} - \mathbf{x}_j)$) centred on each gauge, where a_j are parameters of the surface:

$$s(\mathbf{x}) = a_0 + \sum_{j=1}^n a_j \times g(\mathbf{x} - \mathbf{x}_j) \quad (2)$$

Moore *et al.* (1994) used a simple Euclidean distance function

$$g(\mathbf{x}) = \sqrt{(|\mathbf{x} - \mathbf{x}_j|)^2 + (|\mathbf{y} - \mathbf{y}_j|)^2 + \alpha^2}, \quad (3)$$

but in this study we use $\alpha = 0$ which corresponds to using a surface of n right-sided cones centred at each of the n gauge locations.

The a_j surface parameters can be solved as follows from Eq. (2)

$$s(\mathbf{x}_i) = a_0 + \sum_{j=1}^n a_j \times g(\mathbf{x}_i - \mathbf{x}_j) = z_i, \quad (4)$$

which can be expressed in matrix form as

$$\mathbf{G}\mathbf{a} + \mathbf{a}_0\mathbf{1} = \mathbf{z}, \quad (5)$$

where \mathbf{G} is an $n \times n$ matrix with the (i, j) th element given by $G_{ij} = g(\mathbf{x}_i - \mathbf{x}_j)$ for $i \neq j$, $\mathbf{1}$ is a unit vector of order n and \mathbf{z} is the vector of calibration factors. To prevent discontinuities at $i = j$ a negative term K (set to -18 km here) is used. An additional constraint of $\mathbf{a}^T \mathbf{1} = 0$ is imposed to prevent anomalies forming in the surfaces at large distances. This is combined with Eq. (5) and the matrix is solved to find \mathbf{z} , which is a mathematical function describing the rain-gauge–radar calibration factor over the entire field. This is then multiplied by the radar field to produce the final merged product. The process is then repeated for the next accumulation period using the gauge and radar data from the next polling time.

3.2. Kriging

Kriging is a general term for a number of geostatistical techniques that can be used to interpolate the value of a random field (in this case, the gauge network) at a location where the true value is unknown (Wackernagel, 2010). Kriging methods use a linear least-squares estimation algorithm to evaluate the rainfall value at a non-rain-gauge grid point by using observations at nearby locations. The position of the gauges in the network relative to each other is included in the algorithm through a variogram that relates the correlation between two gauges to their separation. The output of the kriging algorithm produces a map of the rainfall distribution across the entire field spaced at a resolution higher than that of the original random field. Where the location of a pixel exceeds a maximum distance from the nearest available rain-gauge (in this study, 60 km), the merged result is replaced with the original radar data for that pixel to ensure that a complete rainfall map for the United Kingdom is produced.

In all cases, the rainfall data are used in ‘real-space’ (i.e. no transformation is applied to the rainfall field). Pixel locations registering no rainfall in the radar field are set to 0 in the merged field and gauges registering 0 are excluded from the merging dataset. In cases where there are no gauges available within the nearest-neighbour radius of an ungauged pixel (either due to errors in the gauge data or the rainfall accumulation being below the 0.2 mm quantization threshold of the gauges) the merged rainfall defaults to the radar value and the kriging process is therefore bypassed. No further optimization has been made to the code, which runs smoothly and produces consistent images from one accumulation period to the next.

3.2.1. Variogram generation

The information regarding the correlation between gauges within the neighbourhood is contained within a variogram that relates the difference between two gauge measurements. The variogram can either be *parametric* or *non-parametric*; both of these approaches have been tested in this study, and in both cases the variogram is recalculated for each time step, as each accumulation period is treated independently from the previous ones.

3.2.1.1. Parametric variogram

A parametric variogram is produced by calculating half the square of the difference between the rain-gauge readings at two different gauge locations against the distance (lag) between them for every unique pair of gauges in the network. The data are binned and a function, $\gamma(x)$ (usually spherical, Gaussian or exponential), is fitted to the graph of difference *versus* lag. For n gauges $\frac{n(n-1)}{2}$ points are produced and so the generation of $\gamma(x)$ from a large number of data points can be computationally demanding. The merged field produced depends strongly on the suitability of the function used to represent the variogram and the choice of parameter values (Schiemann *et al.*, 2011). In this study a spherical function is used to produce the parametric variogram. This model is particularly suited to the gauge–radar study, as the gauge correlation flattens out at large separation distances and when implementing the function a maximum separation can be selected to allow gauges at large separations to be ignored.

Figure 2 shows examples of the parametric variograms used during this study for a 15 min accumulation during a stratiform and a convective event corresponding to the radar images shown later in Figure 8. The scattered data-points show the dissimilarity (γ^2) between the rainfall accumulations z_g for each unique pair of gauges located at points (\mathbf{x}) and $(\mathbf{x} + \mathbf{h})$ with respect to their separation $|\mathbf{h}|$ where:

$$\gamma^2 = \frac{1}{2n} \sum_{\alpha=1}^n (z_g(\mathbf{x}_\alpha + \mathbf{h}) - z_g(\mathbf{x}_\alpha))^2 \quad (6)$$

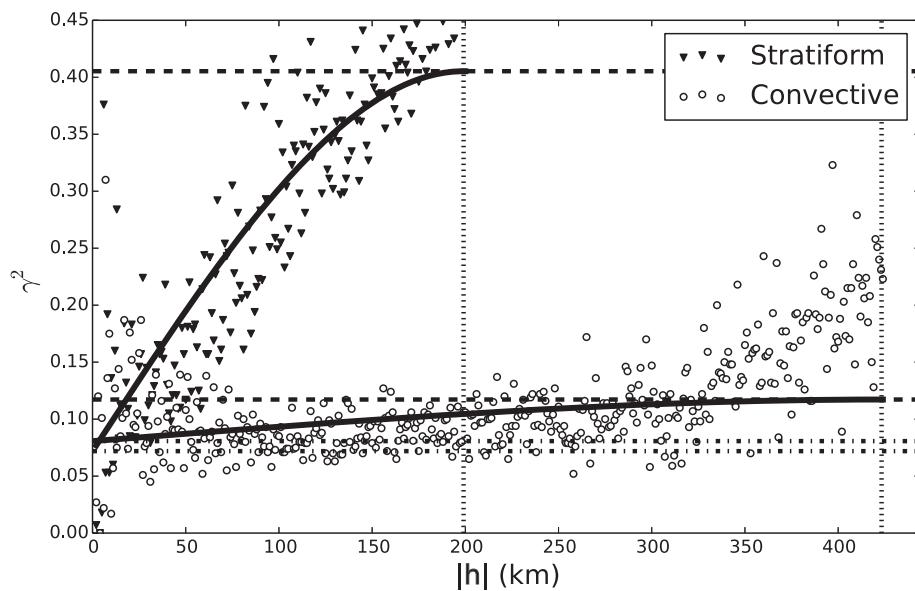


Figure 2. Examples of typical parametric variograms generated using gauge data for a 15 min accumulation period during a stratiform and a convective event. The symbols indicate the data points used from the gauge location and accumulation data and the solid lines show the best fit to the data using a spherical function.

and n is the number of non-zero gauges available. The solid line shows the result of fitting a spherical function of the form

$$\gamma^2 = (\text{sill} - \text{nugget}) \times \left[\frac{3}{2} \left(\frac{|\mathbf{h}|}{\text{range}} \right) - \frac{1}{2} \left(\frac{|\mathbf{h}|}{\text{range}} \right)^3 \right] + \text{nugget} \quad (7)$$

to the data using a least-squares fitting method where the sill, nugget and range are determined by the fitting procedure. The sill is the upper boundary of the variogram function, the nugget represents any discontinuity at zero separation and the range is the separation over which the covariance is valid. These values are indicated by the dashed, dash-dot and dotted lines respectively in Figure 2. It is clear from this figure that, although the distance over which the stratiform variogram is valid is half that of the convective case, the dissimilarity is much smaller at short separations and is much more sensitive to the separation distance than in the convective case.

3.2.1.2. Non-parametric variogram

As the name suggests, a non-parametric variogram does not require the user to make any prior assumptions about the correlation of the data. The Wiener–Khinchin theorem (Reif, 1956) demonstrates that the magnitude of the discrete Fourier transform (DFT) of the standardized observations (in this case the radar field) is the spectral representation of the correlation function (Velasco-Forero *et al.*, 2009). The DFT of the entire two-dimensional radar field is computed, multiplied by its complex conjugate and the inverse DFT is then calculated and rescaled to range from 0 to 1. The output is independent of any user-defined parameters and can be computed using a standard library. However, in cases when the radar is underperforming (e.g. in situations of strong attenuation) errors in the variogram will compound errors in the merged gauge–radar product. In addition, the process of computing the DFT means that the covariance can be calculated only for gauges separated by a distance of less than half that of the domain (although this is not a significant issue when calculating the merged product across geographically large domains, as used in this study).

3.2.2. Ordinary kriging

Ordinary kriging is one of the most widely used kriging techniques in geostatistics and it allows an unknown value at a point to be estimated by using a variogram compiled from data in the

surrounding neighbourhood. Ordinary kriging is not an actual merging technique, as it only has one set of data as input (in this case, the rain-gauge data). In this study it is referred to as gauge-only (GO) kriging and is used to generate an interpolated rainfall field on a regularly spaced grid at a spatial resolution much greater than the original gauge network (in this case, equal to the 1 km grid used for the UK radar data). Generally, it provides a reference field to compare the gauge–radar merged products against.

In the OK process a parametric variogram, $\gamma(\mathbf{h})$, is created from the available rain-gauge readings in the rainfall field. The value at an unobserved point is calculated by summing the weighted contributions from surrounding gauges

$$Z(x_0) = \sum_{\alpha=1}^n w_\alpha Z(x_\alpha), \quad (8)$$

with the condition that

$$\sum_{\alpha=1}^n w_\alpha = 1. \quad (9)$$

Minimizing the estimation variance with this constraint through the use of a Lagrange multiplier (μ_1) produces the ordinary kriging system:

$$\sum_{\beta=1}^n w_\beta \gamma(x_\alpha - x_\beta) + \mu_1 = \gamma(x_\alpha - x_0) \text{ for } \alpha = 1, \dots, n. \quad (10)$$

These conditions can be formalized in a matrix as follows:

$$\begin{bmatrix} \gamma(x_1 - x_1) & \dots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu_1 \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{bmatrix}. \quad (11)$$

The matrix is solved to obtain the n weighting factors and these are then used in Eq. (8) to determine the merged rainfall value at x_0 .

3.2.3. Kriging with radar-based error correction (KRE)

The KRE process uses the difference between the OK gauge data and the radar data from the same period to refine the final

merged output field. The rain-gauge data are Kriged using the OK method to produce a rain-gauge field at grid-points that coincide with the radar data grid. The radar data is then Kriged with the same variogram as the gauges (and hence the same weighting factors generated from Eq. (11)) but the radar data values at the n rain-gauge locations are used to generate the rainfall field from Eq. (8). This produces an interpolated field of rainfall values that follows the mean field of the original radar data. At each grid point the ratio of the observed ($Z_R(x_0)$) and ($Z_{R(\text{Kriged})}(x_0)$) radar values is calculated. A logarithmic transformation method is used to enable the size of the difference in relation to the rainfall accumulation to be treated fairly across all accumulations. The procedure used to calculate the deviation field $c(x_0)$ was

$$c(x_0) = \exp \left(\arctan \left[\ln \frac{Z_R(x_0)}{Z_{R(\text{Kriged})}(x_0)} \right] \right), \quad (12)$$

and this ensures that the value of the deviation field at each rain-gauge point is equal to 1. Finally, the deviation field $c(x_0)$ is multiplied by the Kriged rain-gauge field to produce the merged product. The result is a rainfall field on a regularly spaced grid that follows the near-field of the rain-gauge interpolation but maintains the spatial variability of the rainfall that is contained within the rainfall data (Ehret *et al.*, 2008).

This is the approach followed in this study, but an alternative form of KRE, referred to as ‘conditional merging’ has been successfully used, where the natural log of the gauge and radar data are calculated prior to merging and the fields are merged by addition (Sinclair and Pegram, 2005). This effectively removes any problems associated with zero rainfall (which may be encountered in other kriging methods) and has been shown to be successful for merging daily rainfall accumulations with radar data.

3.2.4. Kriging with external drift

Kriging with external drift uses the radar data during the merging process to add additional constraints to the weighting factors used at each gauge point. Whereas OK and KRE methods assume that the global mean and covariance across the whole of the field is constant (known as first-order stationarity), KED assumes non-stationarity, with the local mean field varying across the field. In effect, KED partitions the data into a deterministic trend (or drift) in the mean field of the radar data and a residual noise component from the gauge information. In the context of the gauge–radar merging scheme, the local mean-field rainfall is inferred from the radar data and small-scale variations are incorporated by interpolating residual variations. This is achieved by using the radar data to apply an additional constraint on the value of the weighting factor calculated for the rain-gauges rather than as a second, independent comparison factor (as in KRE).

In total, three constraints determine the KED scheme. First, as in the OK scheme, the sum of the kriging weights must be equal to 1 (Eq. (9)). Second, the sum of the weighting factor multiplied by the radar value for each gauge must be equal to the measured radar value at the interpolated grid point:

$$\sum_{a=1}^n w_a \cdot Z_R(x_\alpha) = Z_R(x_0). \quad (13)$$

Finally, as in the OK scheme, the estimation covariance is minimized. As the scheme now has the addition of the radar information a second Lagrange multiplier (μ_2) is required, giving the final constraint as:

$$\begin{aligned} \sum_{\beta=1}^n w_\alpha \gamma(x_\alpha - x_\beta) + \mu_1 + \mu_2 \cdot Z_R(x_\alpha) \\ = \gamma(x_\alpha - x_0) \text{ for } a = 1, \dots, n. \end{aligned} \quad (14)$$

The radar data therefore provide the *external drift* term and it is important that the measurements are highly linearly correlated to

the initial (rain-gauge) data. The removal of the (non-physical) assumption that the global mean rainfall field is constant is a significant advantage of the KED scheme over other OK-based kriging schemes. However, this use of a ‘local mean-field’ is at the expense of the scheme’s sensitivity to regions of rainfall not centred on a gauge, and this is likely to be poorly represented in the final merged output.

The conditions in Eqs (9), (13) and (14) can be combined in matrix form and solved to enable the weighting factors to be determined:

$$\begin{bmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 & 1 \\ 1 & \cdots & 1 & 0 & 0 \\ Z_R(x_1) & \cdots & Z_R(x_n) & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_n \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \\ Z_R(x_0) \end{bmatrix}. \quad (15)$$

The weighting factors are then applied to the nearest neighbour gauge values (Eq. (8)) and the merged value at the pixel location is obtained.

In this study, the KED scheme has been used to determine the influence that the choice of variogram has on the final merged product. Merging has been performed using the same sets of gauge and radar data, but in one case the variogram is approximated using a spherical function (KED_s) and in the other, a non-parametric variogram produced from the Fourier transform of the radar data is used instead (KED_n).

3.3. Radar only

The performance of the radar-only data will be used as a benchmark against which the performance of the merged data product is assessed. Although this rainfall measurement is not a truly merged product, the radar data have been automatically bias-adjusted prior to use through the application of a calibration coefficient generated by comparing the raw radar data with gauge measurements at ~ 300 gauge sites across the United Kingdom.

3.4. Summary of assumptions of the merging schemes tested

In order to allow the selected schemes to be compared and contrasted with regard to their suitability for merging, Table 2 compares and contrasts a number of key assumptions and physical reasonings behind their application to gauge and radar merging. It should be noted that the MQ and kriging schemes are based on rather different error models and physical reasoning. The MQ scheme is a physically deterministic process that uses a kernel function (in this case a right cone) to reproduce the rainfall surface. Unless careful constraints are placed on the weighting factors, this approach can lead to a bias in the final results (Balascio, 2001). In contrast, the kriging-based schemes are stochastic processes utilizing a randomized probability distribution, where the kriging weights are adjusted at each grid-point to minimize the variance of the estimation errors at that location (Moulin *et al.*, 2009). Such schemes are better able to separate the systematic bias and residual errors than the MQ process (Pulkkinen *et al.*, 2014), but tend to be reliant on a greater amount of rainfall data to characterize the stochastic nature of the rainfall (Balascio, 2001).

4. Performance of schemes at 60 min accumulations

Examples of the gauge–radar merged product for a single 60 min accumulation period are shown in Figure 3. A qualitative analysis of the images shows that the KED and MQ schemes are most effective at maintaining the fine details of the radar data. In order to quantify how successful each scheme is over each

Table 2. Summary of the key assumptions and physical reasonings behind each of the gauge-radar merging schemes tested.

Feature	MQ	GO	KRE	KEDs	KEDn
Data sources required	Gauge and radar	Gauge data only	Gauge and radar		
Variogram choice and limitations	Does not technically use a variogram BUT the use of the conical functions can be regarded as a simple limited variogram choice (see Shaw (1983))	Uses a parametric variogram from the gauge data – approximated to a spherical function			Uses a non-parametric variogram generated from the two-dimensional Fourier transform of the radar data covering the whole of the merging region
		Quality is limited by the network density			
Performance at rain-gauge locations	Accumulation can vary within fixed boundaries	Set to gauge accumulation value			Limited by the radar coverage but no approximations required
Distribution of input data	No preference	Optimized for a Gaussian data distribution			
Mean field characteristics	Global mean field applied across the domain			Local mean-field inferred from the radar data	

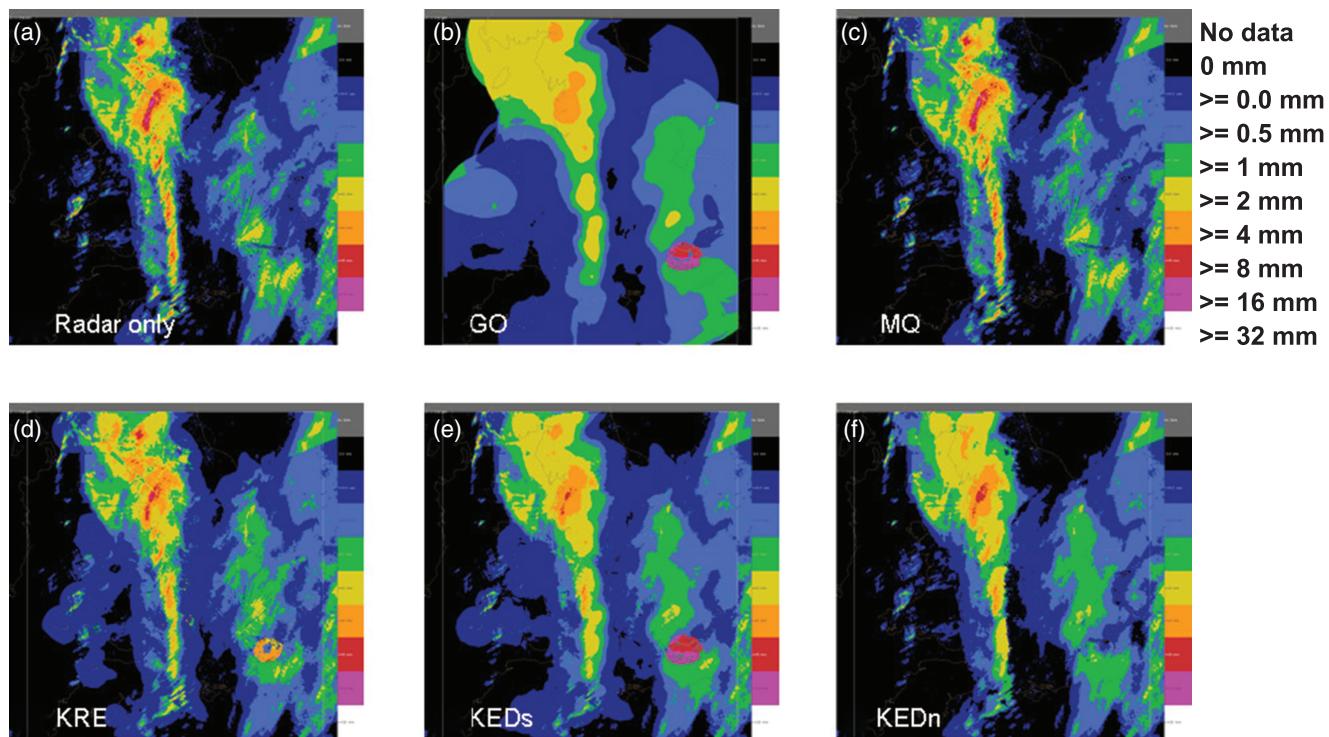


Figure 3. Examples of the merged products produced for a single 60 min accumulation period on 6 October 2010 commencing at 0300 UTC. The results highlight the differences in spatial information contained in the merged products. The schemes are (a) original radar data, (b) gauge only (ordinary) kriging, (c) multiquadric, (d) kriging with radar error correction, (e) kriging with external drift (parametric) and (f) kriging with external drift (non-parametric).

meteorological event, however, the merged result at each time-step must be cross-validated, as described in section 2.4.

The cross-validation results obtained from the merged products using a 60 min accumulation time are shown in Figure 4. The results compare the continuous and binary statistics calculated for low (2 mm h^{-1}) and high (6 mm h^{-1}) rainfall intensities for each of the four meteorological conditions given in Table 1. Overall, the KEDn scheme outperforms the other schemes in terms of both data quality and detection efficiency in all but the highest rainfall thresholds occurring during the most uncorrelated conditions (where the MQ has a slight advantage). The KEDs is consistently the next best-performing scheme and has the advantage that it can be run in less than half the time of the KEDn scheme, making it more computationally efficient. The largest difference between the two KED schemes occurs during the lowest rainfall intensity threshold, which suggests that

the additional information available from the use of the non-parametric variogram improves the performance of the scheme. In general, the radar and gauge schemes are the two worst performing schemes across all events, although the radar-only data are closer in quality and efficiency to the KED scheme during uncorrelated rainfall, whereas the gauge-only data are better during the stratiform events. These observations are consistent with the results of the 24 h accumulation study performed by Goudenhoofdt and Delobbe (2009).

In general, the KED always adds value over the non-merged schemes (OK and radar), with the other schemes (MQ and KRE) also having a benefit over the non-merged product in the vast majority of cases. The MQ scheme has the additional benefit that it is the quickest to run of all the schemes and is a suitable alternative in low rainfall conditions.

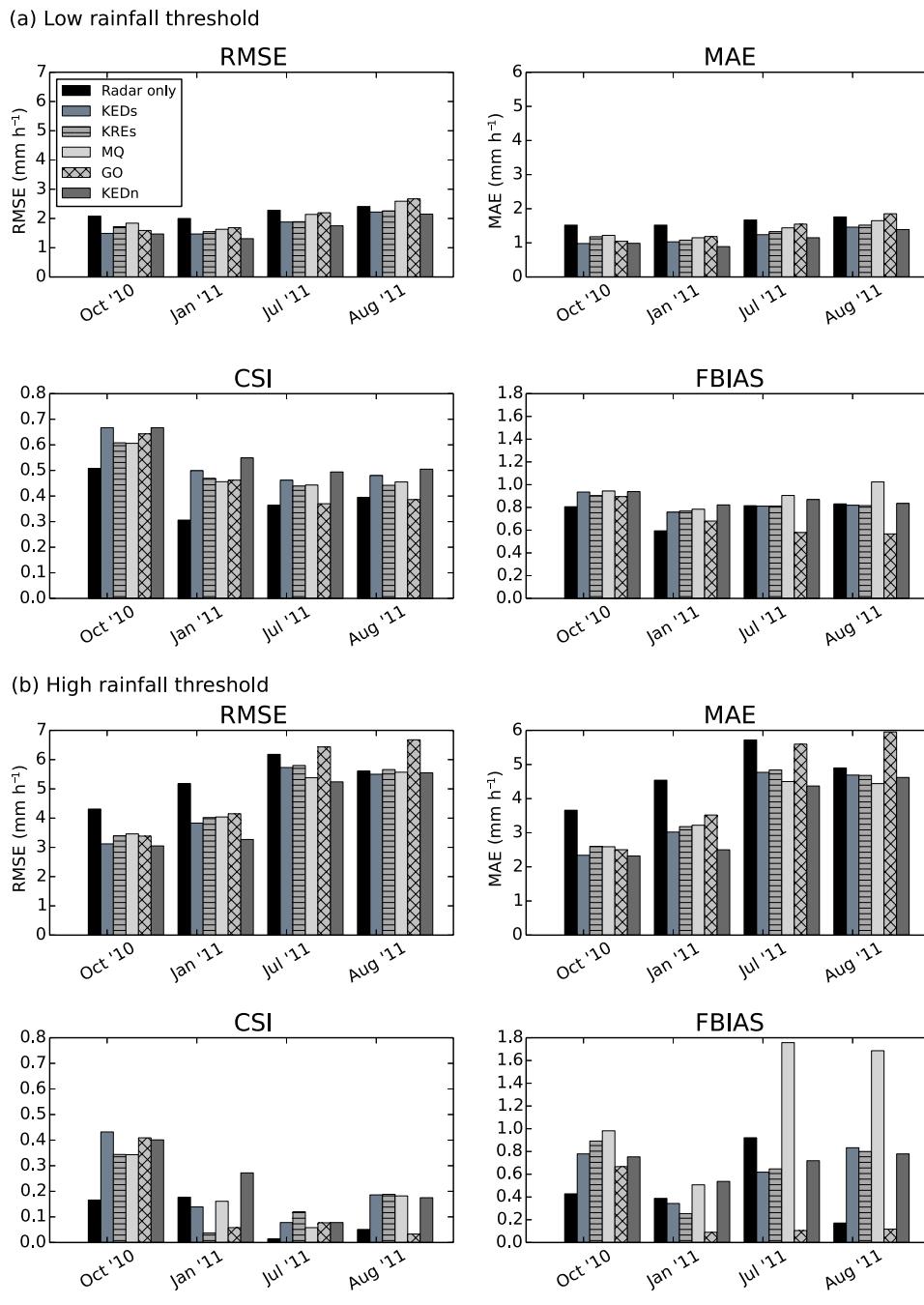


Figure 4. Comparison of merging scheme performances at 60 min accumulation for (a) 2 mm h^{-1} rainfall threshold and (b) 6 mm h^{-1} rainfall threshold.

5. Comparison of 15 and 60 min accumulation periods

The use of the merged gauge–radar product for flood forecasting requires it to be produced for subhourly accumulations in cases of fast-changing meteorological conditions. The best-performing schemes at 1 h accumulations (KEDs, KEDn and MQ) were rerun using 15 min gauge and radar accumulations over the same four meteorological events to enable the effect of accumulation time on the merging schemes to be examined. In general, the benefits of merging gauge and rainfall data over short (subhourly) time-scales are limited due to the differences in representativity between gauge and radar observations, which increase as the accumulation time decreases. For a 15 min accumulation in particular, the radar accumulation is derived from just three instantaneous rainfall measurements within the accumulation window, whereas the gauge data are producing an integrated (but quantized) measurement over the same period. Similarly, the spatial coverage of the two observation schemes differs significantly, with a radar pixel representing 1 km^2 whereas a gauge has an operational area of 50 cm^2 (Gires *et al.* (2014)). Therefore, because of the large spatial resolution differences between the

radar and the gauges, the shorter the accumulation period the more likely very localized showers will be seen by the radar (which integrates over a large volume) and missed by the nearest gauges.

To ensure a fair comparison between the hourly and subhourly merged products, the sum of four 15 min merged outputs on one hand and the corresponding hourly rainfall merged product on the other hand were used for producing the skill-score statistics. The result is shown in Figure 5, with the solid bars representing the results of the 15 min merge product and the lighter, hatched colours indicating the result of the corresponding hourly accumulation product. The results show that the 60 min product tends to outperform the 15 min accumulations, with the KEDs scheme outperforming KEDn during some of the rainfall situations examined. This suggests that a 15 min accumulation of radar data may not be sufficient to generate a reliable variogram during low-intensity stratiform conditions. This is most likely due to the 15 min accumulation being formed from only three 5-min radar composites and therefore having a larger signal-to-noise ratio than that associated with a longer accumulation period.

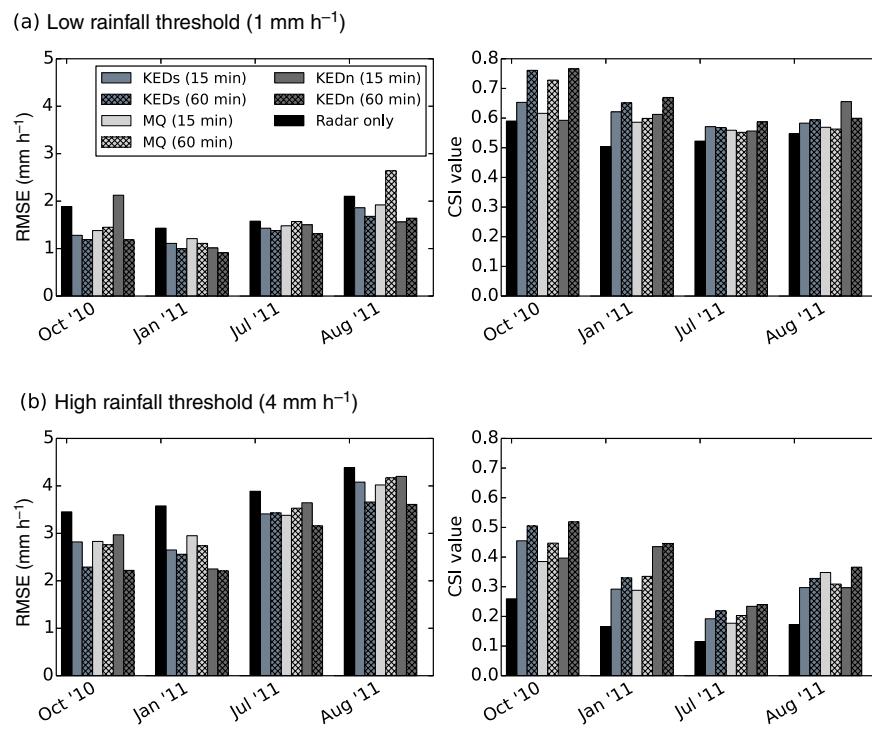


Figure 5. Comparison of 60 min merged products (solid bars) with 4×15 min merged products (hatched bars) for a selection of rainfall conditions and thresholds, for rainfall thresholds of (a) 1 mm h^{-1} and (b) 4 mm h^{-1} . The left-hand panels indicate the RMSE scores in mm h^{-1} and the right-hand panels are for the CSI scores (no units).

One approach to incorporate additional temporal information into the merging scheme is to use a spatial and temporal co-kriging technique Sideris *et al.* (2014). This uses temporal data from a previous time-step as a secondary co-Kriged variable as well as the current spatial correlation information. Due to the increased computational time required by this scheme, it has not been possible to use this approach in the current study.

5.1. Merging gauge and rainfall data with 15 min accumulation times

For certain hydrological applications a 15 min rainfall accumulation every 15 min is desirable. Comparing the 4×15 min accumulations in Figure 5, and also examining the binary and continuous statistics (data not shown), shows that there is still an advantage in merging 15 min gauge and radar accumulations, as the statistics consistently outperform both the gauge and radar-only results. The KED scheme has the best all-round performance, with the MQ having a slight edge during high-intensity uncorrelated rainfall. The KEDs and KEDn produce comparable results in high-intensity rainfall conditions, however, in uncorrelated conditions, due to the low number of gauges available for the parametric variogram, the KEDn performs best. Conversely, in stratiform conditions the KEDs is the better scheme and also has the advantage of taking less time to run. Such a scheme therefore may be advantageous for uncompromising real-time applications, such as flood warnings, at a very short lead time.

A simple yet effective method for generating subhourly accumulations from hourly rainfall measurements has been proposed recently by Sideris and co-workers at Meteo-Swiss (Sideris *et al.*, 2014), which has been derived from the disaggregation of daily to hourly rainfall accumulations by Paulat *et al.* (2008). The scheme involves producing an hourly merged product and then disaggregating the total rainfall into subhourly accumulations in proportion to the rainfall totals measured by the 5 min radar scans. Such a scheme has the potential to enable the improved variogram performance associated with longer accumulation periods to enhance the quality of a subhourly merged product.

6. Effect of gauge density on the merged product

The density of gauges within a network is an issue that requires a balance between the ability of the gauges to accurately represent a local area and the infrastructure requirements and cost for regularly polling and maintaining them. In addition, gauges cannot easily be deployed in mountainous or remote locations. Once a gauge network has been established, as described in the previous section, gauges may be unavailable due to communications or mechanical issues and so the effective gauge density will temporarily decrease.

The effect of gauge density on the performance of the KEDs, MQ and gauge-only schemes was tested. These schemes were chosen due to their relatively quick running time, enabling each scheme to be run multiple times within a suitable time-frame. The test was centred on a particularly dense and homogeneously distributed subset of gauges located in Southeast England, which were part of the original UK-wide network used in previous sections. The data from the October 2010 period were used for the stratiform conditions and August 2011 was used for the uncorrelated conditions. The 123 gauges in the selected area were split into two sets using a random number generator. The first set, containing 61 gauges, was the merging set and the remaining 62 gauges were used as reference gauges. A typical distribution of these gauges and a histogram showing the nearest-neighbour distributions is given in Figure 6(a). The merging scheme (KEDs, MQ, GO) was run using the 61 merging gauges, and the binary and continuous statistics were calculated. The density of the merging gauges was then reduced to 80% of the original (49 gauges in total) by using an algorithm to remove the 12 gauges that produced the greatest increase in the median separation. The merging scheme was re-run, the statistics were calculated and the process was repeated until a density of 20% was reached. The entire process was then repeated a further nine times with a new randomly generated set of initial 61 merging gauges each time. This was undertaken in order to randomize the gauges, which were removed with respect to the merging gauges to prevent any bias arising. The average results were then calculated for each density.

The RMSE and CSI statistics for low and high thresholds as a function of density (denoted by area per gauge) are shown

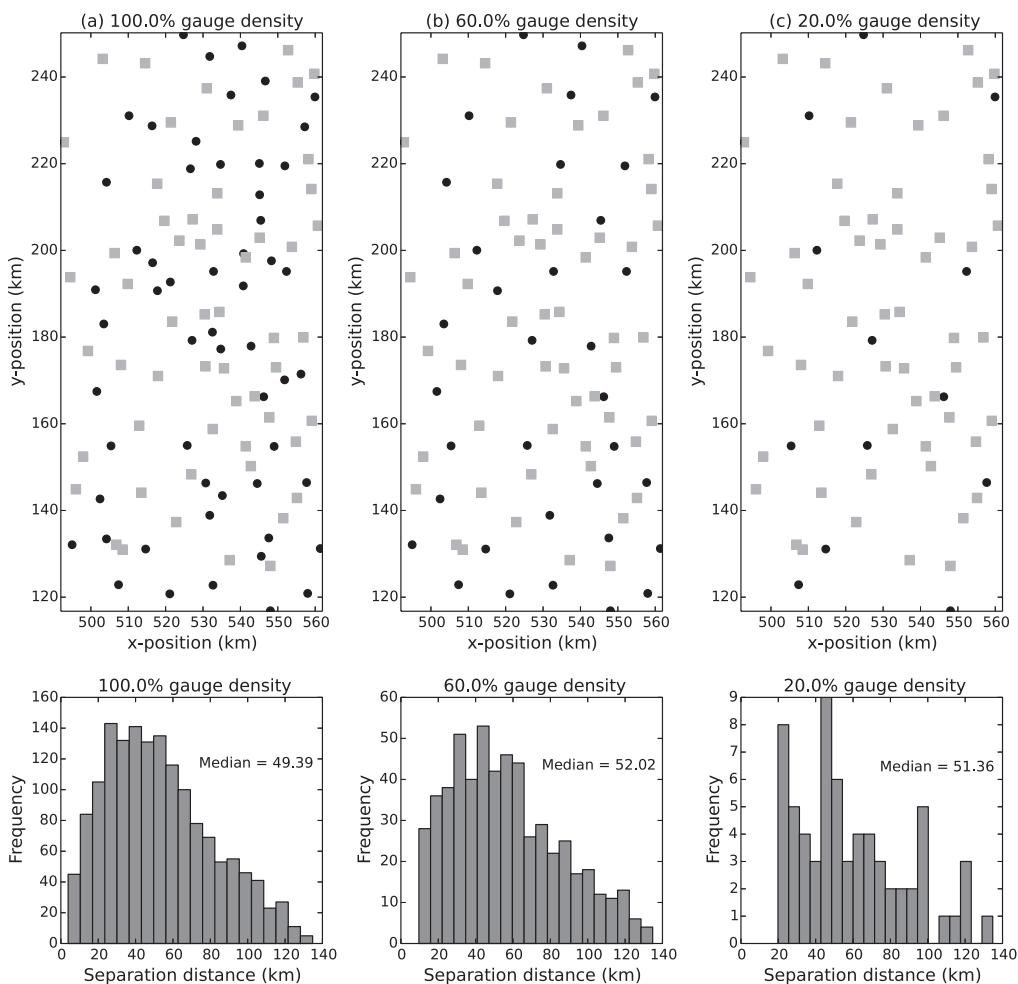


Figure 6. Illustration of the systematic reduction in gauge density and the distribution of the merging gauge separations for (a) all available merging gauges in the region, (b) 60% of the available merging gauges and (c) 20% of the merging gauges. In the upper row, merging gauges are represented by black circles and reference gauges by grey squares (note that the number and locations of the reference gauges is the same in all images).

in Figure 7. The performance of the schemes during convective conditions are shown as dashed lines and the stratiform cases are solid lines. The results show that all of the schemes have a clear dependence on the gauge density, with this being most significant at high rainfall thresholds. As discussed in section 4 the KED scheme has the best continuous and binary statistics and the GO scheme the worst. The two schemes exhibit a noticeable difference in sensitivity to the gauge density, however, particularly in stratiform conditions, with the error associated with the KEDs scheme starting low and then rapidly increasing, whereas the GO scheme starts high at high density but shows a much slower increase in error with reduced gauge density. The results for the MQ scheme are mixed because although the scheme worked well in stratiform conditions at high densities, it failed to produce meaningful results at the lowest density used. This was also the case when uncorrelated rainfall data were used, with the scheme running successfully only when the maximum density of gauges was available. In all but the lowest density case during uncorrelated conditions, the merged product still produced a superior result to the radar-only or GO products. This indicates that even when gauges are temporarily unavailable due to communications problems, etc., or the gauge network is low in some areas, it is still preferable to merge the remaining gauges (down to a specified minimum density). These results also show that the KED is the most robust of the methods tested to changes in gauge density.

7. Influence of rainfall correlation on merged products

The previous sections on merging have routinely highlighted how the meteorological conditions affect the quality and detection efficiency of the merged product. In all cases, the results obtained

during the uncorrelated rainfall conditions in July and August are inferior to those obtained in the stratiform periods of October 2010 and January 2011. For example, the continuous statistics given in Figure 4 show that the RMSE and MAE scores for the KEDn during the uncorrelated events in August is double that of the stratiform event in October.

One approach to producing a universal merging scheme that can be run in all meteorological conditions is to identify uncorrelated regions of rainfall prior to the merging process and treating the background (stratiform) and uncorrelated regions separately. A successful example of this is the ANTILOPE merging scheme developed by MeteoFrance (Laurantin, 2008). This scheme provides an hourly rainfall analysis, merging rainfall and rain-gauge data with a spatial resolution of 1 km. The scheme identifies convective ‘cells’ (groupings of radar pixels) from each 5 min radar accumulation (using an analysis of the surrounding rainfall intensities, which is analogous to that presented in section 7.1 below) and calculating a convective/stratiform ratio to identify small-scale rainfall events. These data are temporarily removed from the merging scheme while the pseudo-stratiform background is merged with data from the rain-gauge field using block kriging. Information from the convective pixels is then recombined with the merged field in the final stage of processing, in order to maintain the schemes’ sensitivity to the small-scale events.

A similar approach has been used here to investigate the effect of rainfall correlation on the performance of the merged products of the KEDn, KEDs and MQ. in the current study, however, the central correlation value (CCV – see below) at each gauge location for each 15 min period was calculated and those gauges above a given CCV threshold were removed from the merging and cross-validation scheme, the CCV threshold defining the ‘level of

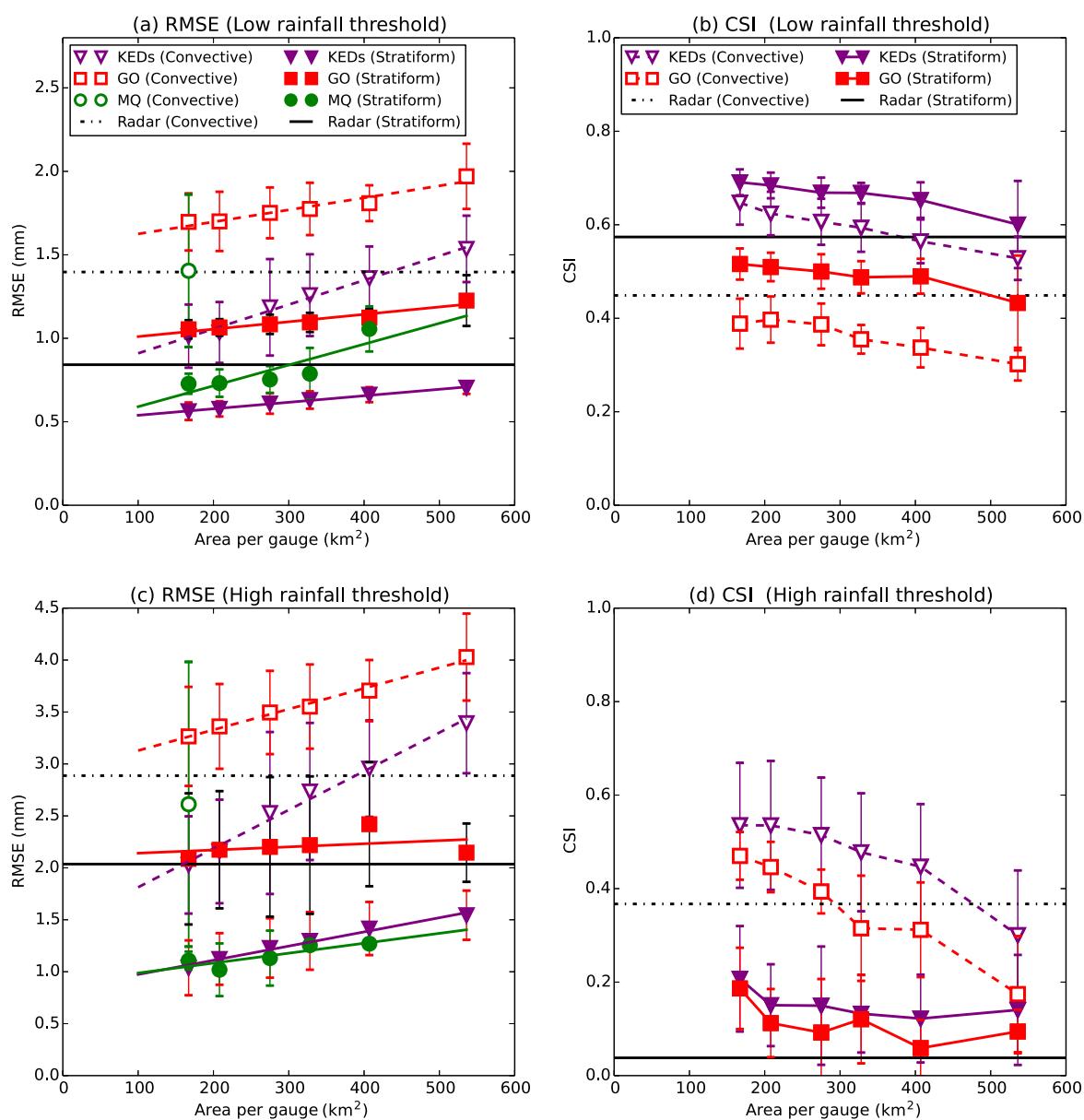


Figure 7. Effect of gauge density on the data quality and detection efficiency of the merged product for low thresholds (a,b: 0.8 mm) and high thresholds (c,d: 3.2 mm).

correlation' of the rainfall. The merging schemes were then run as before for the October 2010 and August 2011 periods and in the case of the KED schemes, for large regions lacking in gauge readings the merged product is automatically replaced with the radar readings. The skill scores were then calculated as before to enable the performance of the merging schemes to be assessed against the level of uncorrelated rainfall included in the merging process.

7.1. The CCV measurements

The CCV is a measure of how well correlated the radar pixel at a point is with the surrounding field. The process for calculating the CCV is performed sequentially for pixel measurement locations (x_0). To reduce the time to generate the CCV values, the measurement locations were based on a regular $10 \times 10 \text{ km}^2$ grid superimposed on the radar grid.

- (1) The radar rain-rate value at x_0 in the first 5 min rainfall radar composite of the accumulation period is recorded.
- (2) Concentric circles centred on x_0 with diameters of 3–13 km (in 1 km steps) were generated.
- (3) The number of valid pixels (i.e. free of any error flags) with their central point located within each circle are counted and the rain rate at each pixel is recorded.

Steps 1–3 are repeated for the additional 5 min composites required to form the accumulation period and the total rainfall at x_0 ($Zr(x_0)$) in Eqs (16) and (17) and at each of the surrounding pixels (Zr_i) is calculated.

The distribution of pixel rainfall accumulation values within each circle is produced, with the values ordered from lowest to highest readings and the 25th (Q_{25}), 50th (Q_{50}), 75th (Q_{75}) and 95th (Q_{95}) percentiles are calculated. In addition, the median absolute deviation of all radar accumulations (Zr_i) in the array are calculated:

$$\text{MedianAbsDev} = \frac{1}{N} \sum_{i=1}^n |Zr_i - Q_{50}|. \quad (16)$$

Finally, the CCV value for the gauge location at the given accumulation period is calculated using the method of Kondragunta and Shrestha (2006) where:

$$\text{If MedianAbsDev} = 0, \text{CCV} = 0. \quad (17)$$

$$\text{If } Q_{25} \neq Q_{75} \text{ then } \text{CCV} = \left| \frac{Zr(x_0) - Q_{50}}{Q_{75} - Q_{25}} \right| \quad (18)$$

$$\text{Else } \text{CCV} = \left| \frac{Zr(x_0) - Q_{50}}{\text{MedianAbsDev}} \right|. \quad (19)$$

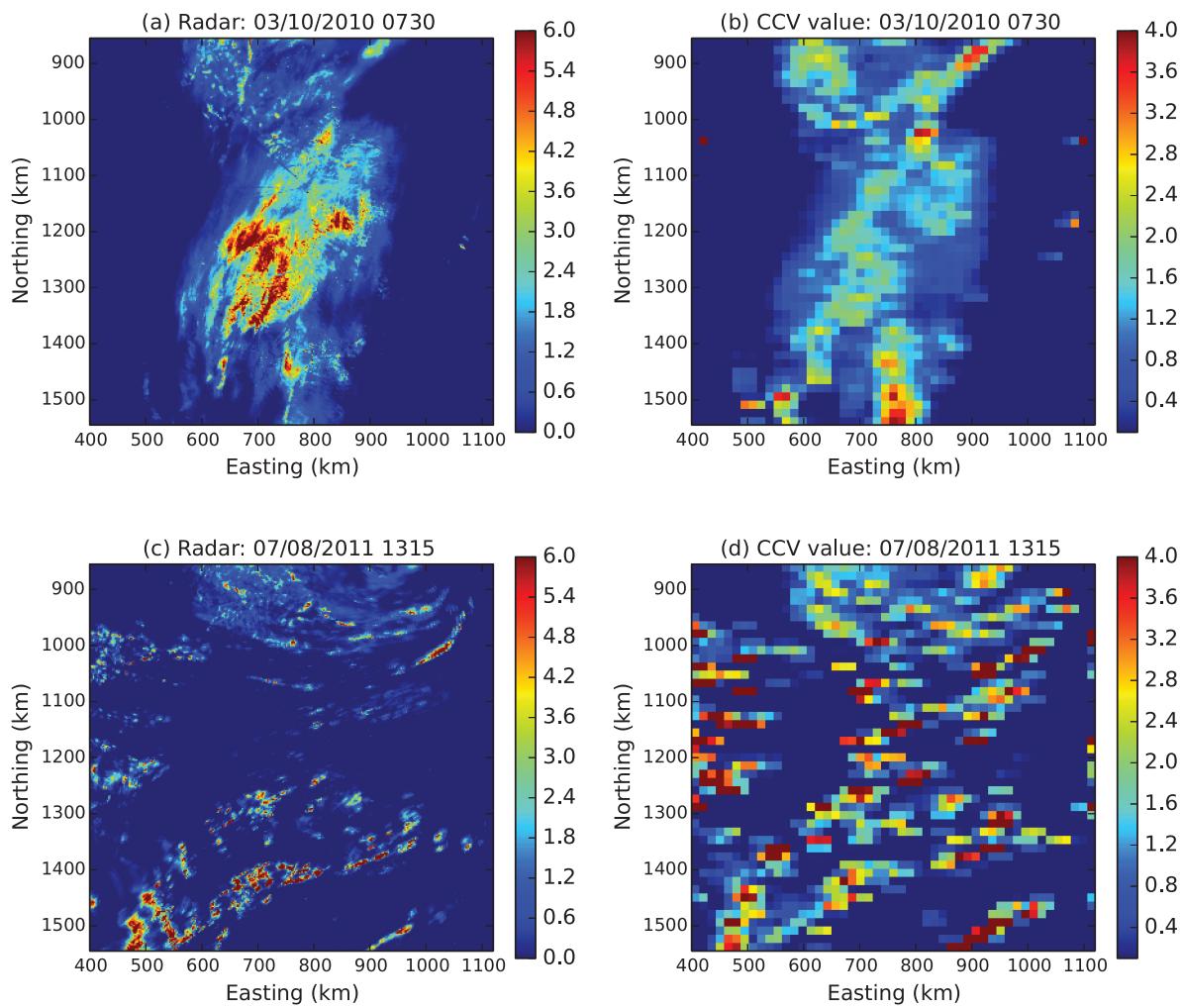


Figure 8. Fifteen minute radar accumulation (converted to mm h^{-1}) and illustration of the CCV value distribution during (a,b) stratiform and (c,d) convective rainfall events.

This process is then repeated from step 1 for each of the remaining grid points and then repeated for all further time-steps. Figure 8(b,d) shows the result of plotting the value of the CCV value at each point (x_0) along with the original radar data (Figure 8(a,c)) for two different 15 min accumulation periods covering England and Wales. Low rainfall correlation results in a higher CCV value, with CCV values above 1 being considered as uncorrelated.

The first case is for a generally stratiform rainfall event during October 2010. There is some amount of uncorrelated rainfall embedded within the regions of particularly high-intensity rainfall but correlation values tend to smoothly vary over the image. In contrast, the second case is for a convective rainfall event during August 2011. In this case the CCV values are much higher than in the stratiform case and are widely scattered all over the region, coinciding with the locations of thunderstorms and convective rainfall events occurring during the accumulation period.

7.2. Thresholded CCV merging

The CCV values at each of the 1064 gauge locations were calculated for the first 15 min accumulation of each 1 h period and the result was used to determine which gauges were used for merging with the radar data for this period. Separate lists were compiled for each merging period for CCV values below 0.5, 1, 2, 3, 4, 5, 10 and 100 (i.e. all gauges) to ensure that the gauges used during the merging process represented the required minimum level of correlation. In addition, gauges with a CCV value of 0 were also removed. This is because any gauges on the edge of a highly localized convective rainfall event that quickly moves out of the sampling region over the 15 min would register a high degree of

rainfall, but would have a median absolute deviation of 0 and therefore, from Eq. (19), a CCV value of 0. This would imply that the rainfall is perfectly correlated when, in fact, it is extremely uncorrelated. The merging schemes (KEDs, KEDn and MQ as well as radar only) were run as before using the depleted lists and cross-validation was performed using only the reference gauges that also satisfied the threshold criteria.

Figure 9 shows the results of plotting the continuous and binary statistics as a function of the mean CCV value for any gauges registering rainfall (i.e. a rainfall threshold of 0.2 mm). The October and August periods have been plotted separately to allow the CCV threshold method to be compared across the two meteorological conditions. The final point located on the far right-hand side of each graph corresponds to the result produced when no thresholds are imposed and all 711 merging gauges shown in Figure 1 were used.

The results clearly show that as the mean CCV value decreases (through the removal of higher CCV gauges) the quality of the merged data improves (Figure 9(a,b)). In addition, the error associated with the August data decreases more quickly with decreasing mean CCV value than in the October case and begins to approach that of the stratiform measurements. However, this effect cannot be attributed to the CCV effect alone; removing the high CCV gauges causes the density of merging gauges to decrease and also the mean rainfall value to decline. This will therefore influence the range of rainfall accumulations measured, which will therefore influence the final measured skill scores. For the August data, when the CCV threshold is reduced to 0.25 the error begins to increase again due to the effect of the significantly decreased gauge density. Note that the density is expected to decrease as we remove high CCV gauges, but the decrease is not

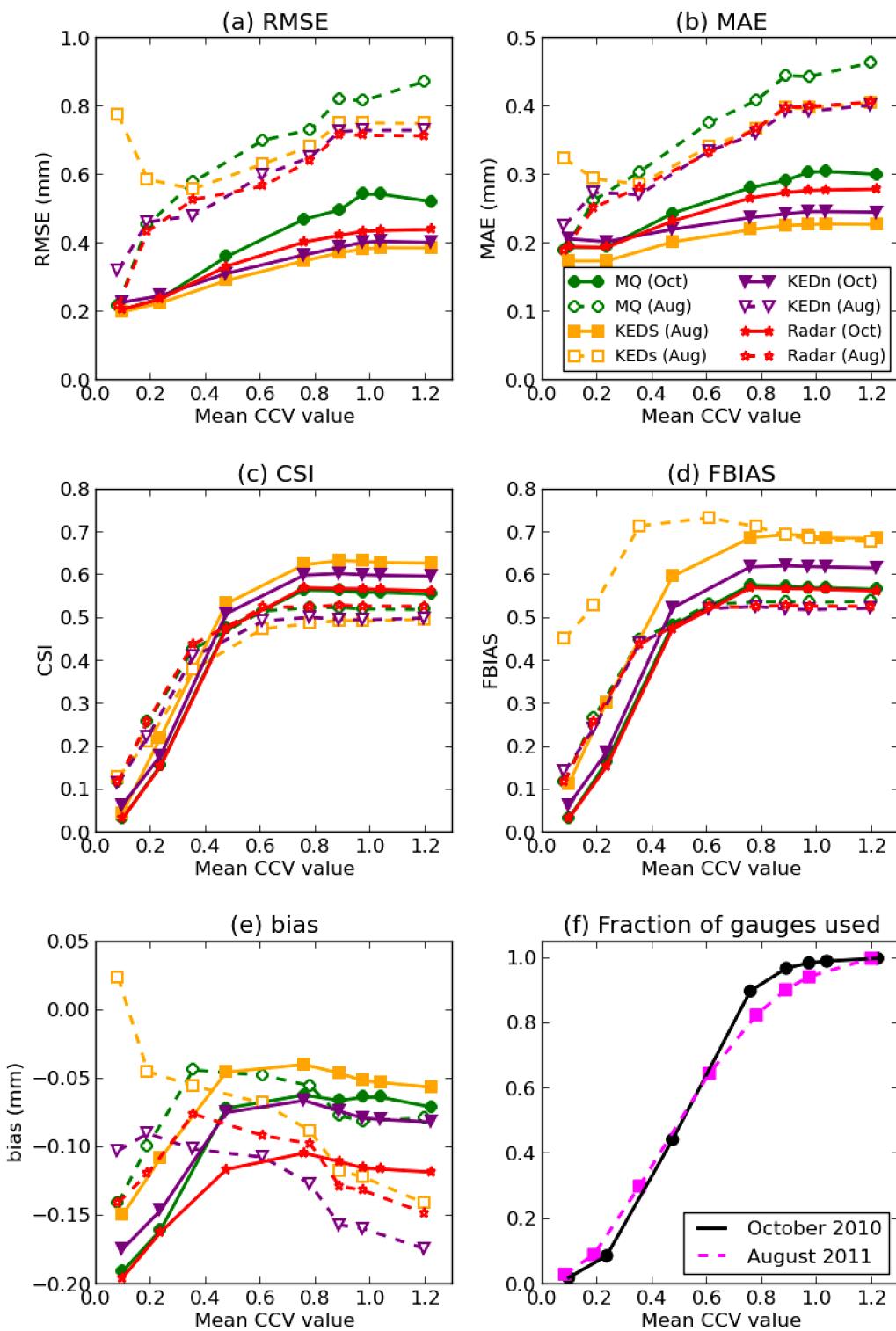


Figure 9. Influence of mean CCV value on (a) RMSE, (b) MAE, (c) CSI, (d) frequency bias, and (e) bias for the merged product for stratiform (Oct) and convective (Aug) conditions for a rainfall threshold of 0.2 mm. (f) The fraction of total available gauges that were used for each mean CCV value calculation.

necessary monotonic because the removal of the high CCV gauges is not constrained by a reduction of the median separation.

Despite the data quality appearing to steadily improve with decreasing CCV levels, the detection efficiency results show a different response. In both datasets the CSI and FBIAS scores stay relatively constant until the mean CCV level reaches around 0.5 (derived from a CCV threshold of 1.0 in both cases), at which point it then goes into a steep decline. It is likely that beyond this point the density of the gauge network is insufficient to support meaningful results. Figure 9(f) shows that at this point the number of readings used is roughly half the number that are actually available in both August and October, yet the merging scheme is still performing well. Although the magnitude of the continuous statistics is higher for the August dataset than the

October set (most probably due to higher rainfall rates), the behaviour of the two datasets is very similar, with the decline in data-quality commencing at approximately the same CCV values.

The use of a CCV threshold below 1 for selecting merging gauges is not a practical option due to the severe drop in available gauges (which suggests that rainfall that is extremely well correlated is unlikely to produce a significant level of rainfall). The results also show that, as the CCV level (and therefore gauge density) continues to decrease, the RMSE and MAE converge towards the value of the bias (0.2 mm). The exception is the August KEDs rainfall, which sees an increase in the errors at very low CCV values, which is accompanied by a reduction in the bias.

The performance of the merging schemes plotted as a function of the CCV level threshold display a common response of improvement in data quality (reduction of error while maintaining detection efficiency) when the CCV threshold is reduced down to 0.5. A universal scheme could therefore be to use the KED (either KEDs or KEDn) with a CCV threshold of 1 without further consideration of the meteorological conditions.

8. Conclusions

This study has provided a broad overview of a number of different gauge–radar merging schemes, directly comparing their performances through the use of canned data from different meteorological conditions. The outcome of the study has shown that kriging with external drift is overwhelmingly the preferred method for combining gauge and radar data, with the choice of a parametric (KEDs) or non-parametric (KEDn) variogram depending on the accumulation time used. Based on the analysis of the 60 min versus 4 × 15 min analysis, for 15 min accumulation periods (with slow-changing conditions) the KEDs performs best, but for fast-changing conditions the KEDn is preferred. For 60 min accumulations KEDn consistently outperforms all the other schemes. The density of the gauge network is a parameter that strongly influences the merging performance. The KED is the least sensitive to the median separation of the merging gauges. The sensitivity of the merging schemes to the gauge-network density is linked to the separation scale of the precipitation. To examine this in detail, the level of correlation of the rainfall at each gauge site was quantified using the local CCV. Excluding gauges with a CCV above a threshold of 0.5 from the KED merging scheme produces an improvement in the data quality while maintaining the detection efficiency at an adequate level. In addition, the improvement in the merged product is equally applicable to stratiform and uncorrelated meteorological conditions, removing the need to use different merging schemes for different meteorological conditions. This result indicates that careful consideration of the quality and representativeness of the meteorological data should be considered in conjunction with refinements to mathematical techniques when developing gauge–radar merging products.

Acknowledgements

This work was funded through a joint Met Office and Environment Agency project. The authors would like to thank Dr Alan Hewitt of the Met Office for assistance with the merging code development and data analysis.

References

- Balascio CC. 2001. Multiquadric equations and optimal areal rainfall estimation. *J. Hydrol. Eng.* **6**: 498–505.
- Bell VA, Kay AL, Jones RG, Moore RJ. 2007. Development of a high resolution grid-based river flow model for use with a regional climate model output. *Hydrol. Earth Syst. Sci.* **11**: 532–549.
- Belo-Pereira M, Dutra E, Viterbo P. 2011. Evaluation of global precipitation data sets over the Iberian Peninsula. *J. Geophys. Res.: Atmos.* **116**: D20101, doi: 10.1029/2010JD015481.
- Brandes EA. 1975. Optimizing rainfall estimates with the aid of radar. *J. Appl. Meteorol.* **14**: 1339–1345.
- Ciach GJ. 2003. Local random errors in tipping-bucket rain gauge measurements. *J. Atmos. Oceanic Technol.* **20**: 752–759.
- Cole SJ, Moore RJ. 2008. Hydrological modelling using raingauge- and radar-based estimations of areal rainfall. *J. Hydrol.* **358**: 158–181.
- Ehret U, Gotzinger J, Bardossy A, Pegram G. 2008. Radar based flood forecasting in small catchments, exemplified by the Goldersbach catchment, Germany. *Int. J. River Basin Manage.* **6**: 7.
- Erdin R, Frei C, Künsch HR. 2012. Data transformation and uncertainty in geostatistical combination of radar and rain gauges. *J. Hydrometeorol.* **13**: 1332–1346.
- Gires A, Tchiguirinskaia I, Schertzer D, Schellart A, Berne A, Lovejoy S. 2014. Influence of small scale rainfall variability on standard comparison tools between radar and raingauge data. *Atmos. Res.* **138**: 125–138.
- Goudenhoofdt E, Delobbe L. 2009. Evaluation of radar–gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.* **13**: 9.
- Haberlandt U. 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrol.* **332**: 144–157.
- Kitzmüller D, Miller D, Fulton R, Ding F. 2013. Radar and multisensor precipitation estimation techniques in National Weather Service hydrologic operations. *J. Hydrol. Eng.* **18**: 10.
- Kondragunta CR, Shrestha K. 2006. ‘Automated real-time operational rain-gauge quality control tools in NWS hydrologic operations’. In *20th AMS Conference on Hydrology*, Atlanta, GA.
- Larson LW, Peck EL. 1974. Accuracy of precipitation measurements for hydrologic modelling. *Water Resour. Res.* **10**: 857–863.
- Laurantin O. 2008. ‘ANTILOPE: Hourly rainfall analysis merging radar and rain-gauge data’. In *Proceedings of the International Symposium on Weather Radar and Hydrology*, Grenoble, France.
- Moore RJ. 2002. Aspects of uncertainty, reliability and risk in flood forecasting systems incorporating weather radar. In *Risk, Reliability, Uncertainty, and Robustness of Water Resources Systems*, Bogardi ZJ, Kundzewicz ZW. (eds.). Cambridge University Press: Cambridge, UK.
- Moore RJ, May BC, Jone DA, Black KB. 1994. Local calibration of weather radar over London. In *Advances in Radar Hydrology*, Almeida-Teixa ME, Fanetti R, Moore R, Silva VM. (eds.). In *Proc. Int. Workshop*, Lisbon, Portugal, 11–13 November 1991, European Commission, Report EHR 14334 EN, 186–195.
- Moulin L, Gaume E, Obled C. 2009. Uncertainties on mean areal precipitation: Assessment and impact on streamflow simulations. *Hydrol. Earth Syst. Sci.* **13**: 99–114.
- Nour MH, Smit DW, El-Din MG. 2006. Estimation of missing precipitation records integrating surface interpolation techniques and spatio-temporal association rules. *Water Sci. Technol.* **53**: 101–110.
- Paulat M, Frei C, Hagen M, Wernli H. 2008. A gridded dataset of hourly precipitation in Germany: Its construction, climatology and application. *Meteorol. Z.* **17**: 719–732.
- Pulkkinen S, Koistinen J, Kuitunen T. 2014. ‘Gauge–radar adjustment by using multivariate kernel regression and spatiotemporal kriging’. In *Extended Abstract, ERAD2014: The 9th European Conference on Radar in Meteorology and Hydrology*, Garmisch-Partenkirchen, Germany.
- Reif F. 1956. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill: New York, NY.
- Schiemann R, Erdin R, Willi M, Frei C, Berenguer M, Sempere-Torres D. 2011. Geostatistical radar–raingauge combination with nonparametric correlograms: Methodological considerations and application in Switzerland. *Hydrol. Earth Syst. Sci.* **15**: 22.
- Schuurmans JM, Bierkens MFP, Pebesma EJ, Uijlenhoet R. 2007. Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar. *J. Hydrometeorol.* **8**: 1204–1224.
- Sevruk B. 1989. ‘Reliability of precipitation measurement’. In *International Workshop on Precipitation Measurement*, Geneva, Switzerland.
- Shaw E. 1983. *Hydrology in Practice*. Van Nostrand Reinhold: Wokingham, UK.
- Sideris IV, Gabella M, Erdin R, Germann U. 2014. Real-time radar–raingauge merging using spatio-temporal co-kriging with external drift in the Alpine terrain of Switzerland. *Q. J. R. Meteorol. Soc.* **140**: 1097–1111.
- Sinclair S, Pegram G. 2005. Combining radar and rain gauge rainfall estimates using conditional merging. *Atmos. Sci. Lett.* **6**: 4.
- Velasco-Forero CA, Sempere-Torres D, Cassiraga E, Gomez-Hernandez JJ. 2009. A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Adv. Water Resour.* **32**: 17.
- Wackernagel H. 2010. *Multivariate Geostatistics: An Introduction with Applications* (3rd edn). Springer: Berlin and Heidelberg, Germany.