# Quantifying flood risk of extreme events using density forecasts based on a new digital archive and weather ensemble predictions

Patrick E. McSharry,[a]* Max A. Little,[b] Harvey J. E. Rodda[c] and John Rodda[c]

[a]*Smith School of Enterprise and the Environment, University of Oxford, UK*
[b]*Media Laboratory, Massachusetts Institute of Technology, USA*
[c]*Hydro-GIS, UK*

*Correspondence to: P. E. McSharry, Smith School of Enterprise and the Environment, University of Oxford, 75 George Street, Oxford OX1 2BQ, UK. E-mail: patrick.mcsharry@smithschool.ox.ac.uk

**Non-coastal flood events in the UK are usually associated with extreme rainfall and can last from minutes to weeks. Efficient management and mitigation of flood risk requires accurate and reliable precipitation forecasts as inputs to flood risk models. We constructed an archive of *British Rainfall* data from 1866 to the present day to improve our understanding of historical extreme rainfall events. The relationship between record rainfall and flooding is nonlinear and uncertain, implying that probabilistic forecasts of rainfall are required. We developed an objective classification scheme of extreme rainfall events consisting of eight types, analysed extreme rainfall events and produced probabilistic forecasts by combining statistical techniques with the outputs of ensemble predictions from a numerical weather predictions model. Copyright © 2013 Royal Meteorological Society**

*Key Words:* daily rainfalls; precipitation; risk analysis; risk management; forecasting; extremes; climate change; climate variability

*Received 20 October 2011; Revised 4 February 2013; Accepted 7 February 2013; Published online in Wiley Online Library*

*Citation:* McSharry PE, Little MA, Rodda HJE, Rodda J. 2013. Quantifying flood risk of extreme events using density forecasts based on a new digital archive and weather ensemble predictions. *Q. J. R. Meteorol. Soc.* **139**: 328–333. DOI:10.1002/qj.2136

## 1. Introduction

Flood events, generated by extreme rainfall, cause greater destruction than any other type of hazard in the United Kingdom. The summer floods of 2007 cost the economy £3.2 billion (of which approximately £3 billion were insured losses) and resulted in the deaths of 13 people (EA, 2007). Approximately two-thirds of these losses were experienced by households and businesses. Furthermore, significant impacts on national security were caused by the closure of a water treatment plant and the threat to a generating station.

A wide range of practical applications rely on information about the likelihood of extreme rainfall events (Rodda, 1967, 1972; Hand *et al.*, 2004). This information can be obtained by investigating the relationship between magnitude, frequency and distribution of intense rainfall (Coles, 2001). Specific applications include the design of drainage schemes, reservoir spillways and other hydraulic structures, and studies of landslides. With increasing recognition of the adverse effects of global warming, accurate knowledge about the evolution of rainfall extremes in space and time is proving relevant for studies about the impacts of climate change (New *et al.*, 2001; Osborn and Hulme, 2002). Rainfall forecasts are used to quantify the risk of floods and droughts and have commercial relevance for pricing flood and crop insurance, weather derivatives and other weather-dependent commodities (Campbell and Diebold, 2005; Taylor and Buizza, 2006).

Effective reduction and management of risk associated with flooding requires collaboration across a range of governmental agencies and private sector organizations. Similarly the identification of and quantification of risk require multidisciplinary research. The Flood Risk from

Table 1. Example extreme rainfall observations from the British Rainfall Digital Archive.

| Date | County | Gauge name | Rainfall (in) | Rainfall (mm) | % |
|---|---|---|---|---|---|
| 11/11/1929 | Glamorgan | Glyncorrwg Colliery | 4.93 | 125.2 | 5.8 |
| 11/11/1929 | Glamorgan | Ystradyfodwg (Tvntyla Hospital) | 3.09 | 78.5 | 4.4 |
| 11/11/1929 | Glamorgan | Treharris Service Res. | 3.60 | 91.4 | 5.6 |
| 11/11/1929 | Glamorgan | Mountain Ash (Darranlas Reservoir) | 3.89 | 98.8 | 5.0 |
| 11/11/1929 | Glamorgan | Rhondda (Castell Nos Res.) | 4.72 | 119.9 | 5.1 |
| 11/11/1929 | Glamorgan | Rhondda (Lluest Wen Filter) | 6.20 | 157.5 | 6.0 |
| 11/11/1929 | Glamorgan | Rhondda (LluestWen Res.) | 8.31 | 211.1 | 7.2 |
| 11/11/1929 | Glamorgan | Treherbert (Tynywaun) | 5.25 | 133.4 | 5.0 |
| 11/11/1929 | Glamorgan | Treorchy Pumping Station | 3.63 | 92.2 | 4.7 |
| 11/11/1929 | Camarthen | Liyn-y-fan Fach (Nant Coch) | 3.02 | 76.7 | 3.4 |
| 11/11/1929 | Camarthen | Llyn-y-fan Fach (Troed-y-foel) | 2.55 | 64.8 | 3.5 |
| 11/11/1929 | Brecon | Ystradgynlais (Yniscedwyn House) | 3.52 | 89.4 | 5.1 |
| 11/11/1929 | Brecon | Swansea W.W. (Nantyrwydd) | 3.32 | 84.3 | 4.0 |

Extreme Events (FREE) programme funded by NERC brought about such collaboration and has yielded a number of insights that have important practical implications.

Our main contribution was to create a new digital archive, which then allowed us to uncover the patterns in extreme rainfall over the past century and a half. Using state of the art classification, machine learning, forecasting and time-series-analysis techniques we explored this archive in conjunction with existing databases. The motivation for the project was to apply novel statistical techniques to the available time series in order to improve our understanding of extreme rainfall and to help quantify future risks of flooding. Access to accurate techniques for modelling extreme rainfall is of substantial interest to those involved in hydrological studies within the environmental management, civil engineering and insurance sectors.

In this article we present the outcomes of the research project and highlight the importance of the results for environmental consultancies, the insurance industry and policymakers. Section 2 describes the source of our digital archive, *British Rainfall*, and summarizes the statistics of this long-term record. Section 3 presents the results of the classification analysis. Section 4 demonstrates the changes in rainfall extremes exhibited by comparing the return period analyses of the periods 1911 to 1960 and 1961 to 2006. Section 5 discusses the advantages of probabilistic rainfall forecasting and the benefits of combining ensemble weather predictions and statistical time-series approaches. Section 6 concludes and summarizes the highlights.

## 2. British extreme rainfall observations

The British Rainfall Digital Archive (BRDA) was constructed using data from *British Rainfall*, an annual publication of rainfall observations started by George J. Symons in 1860 under the title Symons's British Rainfall. The publication contained a section on extreme rainfall events under the heading 'Heavy falls on rainfall days' or 'Heavy falls in 24 hours' during the period 1866–1968. Within this section all observed 24 h rainfall depths were listed that exceeded a certain threshold. This was set at 2.5 in (63.5 mm) or 7.5% of the annual total at the specific gauge up to 1961. For the editions from 1961 to 1968 the threshold was set at 50 mm or 4% of the annual total. Descriptive text from observers was included in the chapter, which provided a range of information such as an overview of the synoptic meteorology, a description


**Figure 1.** Photograph of flooding from the British Rainfall Digital Archive.

of the characteristics and intensity of the rainfall, and accounts of resulting flooding and damage. For the most extreme and interesting events isohyetal maps, estimates of rainfall over specific areas, and photographs were included. All of this information was digitized to form the BRDA. The BRDA provides 24 h rainfall depths at up to 6000 locations with records ranging from 1866 to 1968 (Rodda *et al.*, 2009). In total over 28 000 rainfall depths are listed (Table 1 gives an example), over 1000 pages of text, and over 250 rainfall maps and photographs (Figure 1) are included. Unfortunately *British Rainfall* was ceased in 1969 so accounts of later extreme rainfalls, such as the Hampstead storm in August 1975 and the Calderdale storm in May 1989 (Institute of Hydrology, 1999) were not included. Publications investigating the conditions leading to extreme rainfall events in the UK include Thielen and Gadian (1996) and Golding *et al.* (2005).

The digital archive has been made freely and publicly available through the British Atmospheric Data Centre (BADC, www.badc.rl.ac.uk). Further information about the *British Rainfall* (BR) archive and how to access the data is available from Hydro-GIS (www.hydro-gis.co.uk). This database is an invaluable resource for the fields of meteorology, climatology and hydrology. It can be used to support decision-making and facilitate improved management of flood risk. The BRDA appeals

Table 2. Record rainfall extremes in the United Kingdom over 200 mm in 24 h, based on data from *British Rainfall*, 1866–1968 and other sources including the Met Office (2011), Environment Agency, BBC and MeteoGroup.

| Date | Depth (mm) | Depth (in) | Location | County |
|---|---|---|---|---|
| 18/07/1955 | 279.4 | 11.0 | Martinstown (The Chantry) | Dorset |
| 19/11/2009 | 253 | 9.96 | Seathwaite Farm | Cumbria |
| 28/06/1917 | 242.8 | 9.6 | Bruton (Sexey's School) | Somerset |
| 18/07/1955 | 241.3 | 9.5 | Upwey (Friar Waddon) | Dorset |
| 18/08/1924 | 238.8 | 9.4 | Cannington (Brymore) | Somerset |
| 17/01/1974 | 238.4 | 9.4 | Loch Sloy Main Adit | Strathclyde |
| 19/11/2009 | 229.2 | 9.02 | Honister | Cumbria |
| 15/08/1952 | 228.6 | 9.0 | Longstone Barrow | Devon |
| 18/07/1955 | 228.6 | 9.0 | Upwey (Higher Well) | Dorset |
| 19/11/2009 | 228.2 | 8.89 | High Snab Farm | Cumbria |
| 22/11/1908 | 217.9 | 8.6 | Snowdon (Llyn Llydaw Copper Mill) | Caernarvonshire |
| 28/06/1917 | 215.4 | 8.5 | Bruton (King's School) | Somerset |
| 28/06/1917 | 213.1 | 8.4 | Aisholt (Timberscombe) | Somerset |
| 11/11/1929 | 211.1 | 8.3 | Rhondda (Lluest-wen Reservoir) | Glamorgan |
| 18/07/1955 | 211.1 | 8.3 | Upwey (Elwell) | Dorset |
| 11/10/1916 | 208.3 | 8.2 | Loch Quoich (Kinlochquoich) | Inverness-shire |
| 12/11/1897 | 204.0 | 8.0 | Seathwaite | Cumberland |
| 08/06/1957 | 203.2 | 8.0 | Camelford (Roughtor View) | Cornwall |
| 28/06/1917 | 200.7 | 7.9 | Bruton (Pitcombe Vicarage) | Somerset |
| 18/07/1955 | 200.7 | 7.9 | Wynford House | Dorset |
| 16/08/2004 | 200.7 | 7.9 | Otterham, near Boscastle | Cornwall |

to environmental consultants, in particular where they are required to submit a flood risk assessment, surface water management plan or sustainable drainage design including evidence of historical flooding as part of a planning application. A link to the digital archive using geographical information system (GIS) software was developed for the Thames Region of the Environment Agency as a pilot study. This enabled any user to undertake a spatial search to access data in the archive for particular gauge locations.

The BRDA has the advantage of illustrating the particular weather mechanisms associated with many of the extreme events. For example, the extreme rainfall pattern, which resulted in an estimated £2.0 billion of damage due to surface water flooding on 24–25 June 2007, was very similar to events listed within the BRDA for 7 June 1910 and 10 July 1968. Such information in the BRDA can provide ideal scenarios for flood forecasting and emergency planning operations. In addition, retrospective studies can be of great benefit to the insurance industry by estimating the financial losses from the current insured portfolio if an extreme historical event such as those listed above were to occur today.

The *British Rainfall* archive contained a total of 16 observations of 200 mm or more of rainfall in 24 h (Table 2). Nine of the observations over 200 mm are from two events: 18 July 1955 and 28 June 1917. Five further measurements in excess of 200 mm in 24 h have been observed in the UK in the years since 1968.

Based on data collected from the BRDA and other sources (Table 2), there have been a total of 21 cases of rainfall exceeding 200 mm in 24 h. The most extreme event was 279.4 mm measured in Martinstown, Dorset, in 1955. The 24 h observation of 314 mm of rain at Seathwaite Farm, Cumbria in November 2009 was not included since this was over a moving 24 h period rather than the 09:00–09:00 standard adopted by *British Rainfall*.

Rodda *et al.* (2009) found that the distribution of all extreme rainfall observations on a monthly basis shows a much greater frequency in the latter half of the year (July to December), with a maximum in July (16.6%) and a minimum in April (1.6%).

## 3. Classifying extreme rainfall patterns

An objective classification scheme for extreme UK daily precipitation is needed for flood risk analysis applications (Lamb, 1972). Record rainfalls caused devastating floods in Cumbria in 2009 and Lynmouth in 1952, but the rainfall patterns were fundamentally different. Detailed information about the pattern of rainfall that caused the flooding is required. We developed an automated system for discovering the most likely spatio-temporal pattern in extreme rainfalls using Bayesian machine learning techniques.

This was achieved using a simplified representation of the spatial structure of extreme events based on the new digital archive of UK rainfall (BRDA), using daily rainfall totals as described above. This approach enabled us to employ a Bayesian clustering algorithm to compress the representations down to eight prototypical patterns of extreme rainfalls. These patterns are then verified against a five-class, manual, subjective typing scheme, produced independently using known meteorological mechanisms, isohyetal maps and additional descriptive text from the archive. This manual classification incorporated all of the mapped archive events, and covered the most extreme events as listed in Table 2. The classification and causal factors of these events were assessed based on the meteorological processes known and demonstrated to have occurred at the time. Compared against the manual scheme, the new objective scheme can reproduce the known meteorological conditions, both in terms of spatial layout and seasonal timing, and is shown to be of hydrological relevance when matched to several notable flooding events in the past century. Furthermore, it is computationally simple and straightforward to apply in classifying future extreme rainfall events.

 

The BRDA provided us with a total of 257 rainfall events over the years 1866–1968 for which maps and more detailed written descriptions were available (as described in the previous section). A second source of information was derived from the UK Meteorological Office MIDAS surface daily weather observation network. This allowed us to extract extreme depths over 50 mm, covering the same time period as BRDA. The two sources of information (BRDA and MIDAS) were merged into one dataset of extreme rainfall depths for the 257 events identified in the BR archive, covering the years 1866–1968 and the whole of the UK.

There are considerable practical benefits arising from the availability of an objective approach to classifying rainfall mechanisms. The objective typing scheme has the obvious advantage over the subjective scheme in that it can be used to automatically type all future events without additional manual effort. Extreme rainfall on any given day can be encoded using the simplified representation. The objective type for that day is given by the cluster that has the smallest total dissimilarity to the encoding.

In the case of hydrological flood studies, where design or other rainfall simulations are utilized, the cluster centroids (Figure 2) can be used to inform the spatial layout and the seasonal timing of extreme rainfall in each type (Figure 3). The most intense rainfall in the simulation can be located in the grid cells with the highest values and timed according to the month with the largest probability. In addition, the objective typing could be applied to all past and future events in climate studies, in order to assess the extent to which patterns of extremes may be changing in response to global and regional temperature variations. Further details of the eight prototypical patterns of extreme rainfalls are given in Little *et al.* (2008).

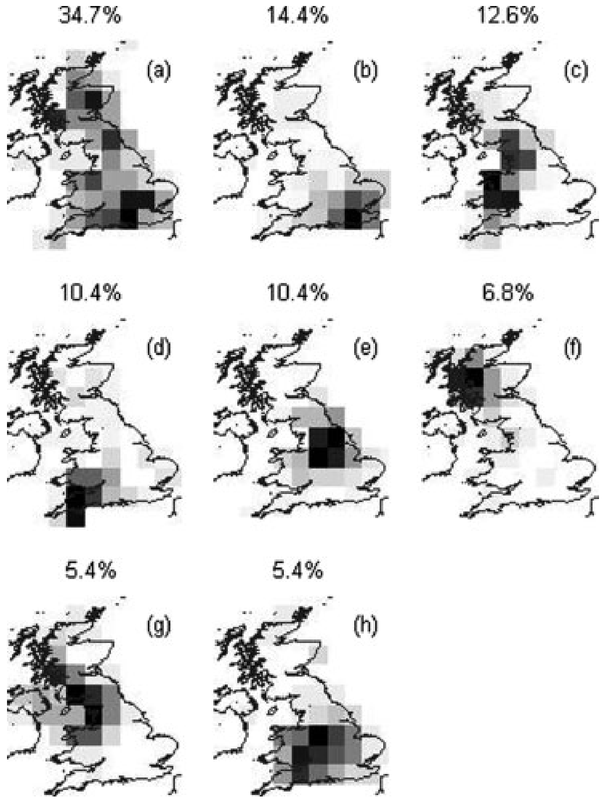


**Figure 2.** (a)–(h) Graphical illustration of the eight objective types. Each map shows the percentage of selected days on a $1^\circ \times 1^\circ$ grid classified in that type, with darker colours indicating a higher prevalence of extremes.

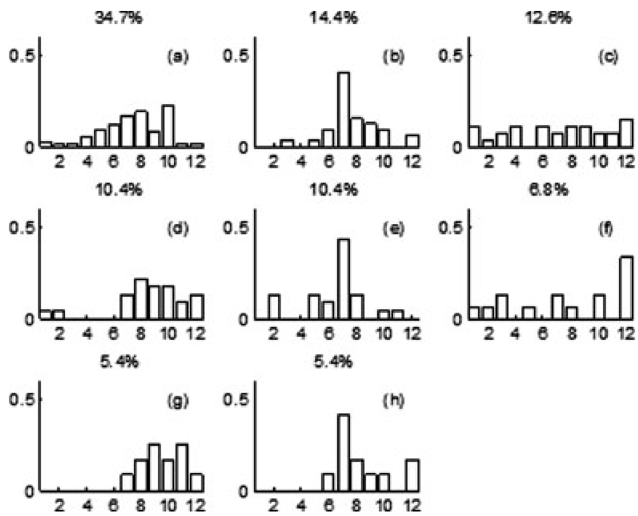


**Figure 3.** Seasonal variation for each of the objective types. The bar chart reflects the conditional probability of the month given a specific objective classification type.

## 4. Quantifying the risk of extreme events

Knowledge of the likelihood of extreme events is necessary for planning and designing in order to withstand the effects of flooding and develop policy for risk reduction. Return period plots provide the probability of an event of a specific magnitude.

Regulation in Europe, known as Solvency II (EU, 2009), requests that all insurance organizations estimate the financial loss corresponding to a one in a hundred year event and demonstrate that they have sufficient capital to withstand such an event. Extreme value theory offers a means of establishing the magnitudes of events corresponding to particular return periods (Coles, 2001). Climate variability has the potential to modify the dynamical processes giving rise to extremes, thereby complicating the use of extreme value theory, which assumes a stationary process.

During the 1960s, a study was undertaken of the magnitude, frequency and distribution of intense rainfall over the UK, employing data from more than 120 daily read rain-gauges covering the period 1911 to 1960. Using the same methodology, we updated that study utilizing data for the period 1961 to 2006 for the same gauges, or from those nearby (Rodda *et al.*, 2010). The purpose of updating the study was to investigate the changes in the patterns of extreme rainfalls for the two periods.

The results indicated that increases of up to 20% have occurred in the northwest of the UK and in parts of East Anglia. There have also been changes in other areas, including decreases of the same magnitude over central England.

## 5. Forecasting rainfall

Weather forecasts are inherently uncertain due to errors in the measure of the current weather conditions, misspecification of the model structure and inaccurate parameter estimates. The magnitude of the forecast error usually increases with the size of the forecast horizon. A point forecast represents the best guess or expected value for a particular weather variable. Although simple to convey, point forecasts fail to communicate the level of uncertainty in the forecast. This misrepresentation of the uncertainty, typically an underestimation of the uncertainty due to the lack of confidence intervals, has serious implications for quantifying flood risk, planning and risk reduction. In contrast we have promoted the construction of probabilistic forecasts, which provide a transparent means of communicating the expected level of rainfall and the associated uncertainty. This was achieved by developing density forecasts, which combine the outputs of physically based numerical weather predictions (NWP), known as ensemble forecasts (Buizza *et al.*, 1998) with statistical models. The resulting forecasts were shown to have considerable advantages in terms of their performance under probabilistic forecast evaluation metrics such as calibration and continuous ranked probability score (Little *et al.*, 2009). We acknowledge that ensemble forecasts have limitations for very short lead times, where high spatial and temporal resolution is required. One solution is to integrate radar, satellite and rain-gauge data with the NWP forecasts (Golding, 2000).

In particular we produced viable statistical, probabilistic UK daily rainfall forecasts based on the output of an existing ensemble prediction system (Little *et al.*, 2009). We investigated the spatio-temporal correlations in UK daily rainfall amounts over the Thames Valley and constructed statistical, Markov chain generalized linear models (Markov GLM) of rainfall. We compared point and density forecasts of total rainfall amounts, and forecasts of probability of rainfall occurrence. Our candidate models included benchmarks such as persistence and statistical climatology, and a range of statistical models such as Markov chain, unconditional gamma and exponential mixture models, and density forecasts from GLM regression post-processed NCEP numerical ensembles with a resolution of $1^{\circ}$. The continuous ranked probability score, measuring the qualities of both sharpness and reliability (Pinson *et al.*, 2010), was used to quantify the performance of the probabilistic forecasts generated by the models (Figure 4).

A subset of 10 gauges in the Thames Valley region was selected as a compromise between minimizing the number of missing observations, economic relevance (Heathrow) and hydrological interest (Brize Norton). The probability integral transform was used to quantify the calibration for one-day-ahead forecasts. Forecast performances of competing models were compared using the mean absolute error, Brier score and continuous ranked probability score. The Markov GLMs and GLM processed ensembles produced skilful one-day ahead and short-term point forecasts. Diagnostic tests demonstrated that all models are well calibrated, but GLMs perform best under the continuous ranked probability score (Figure 4). For lead times of greater than one day, no models were better than the GLM processed ensembles at forecasting occurrence probability. Of all models, the ensembles are best able to account for the serial correlations in rainfall amounts. In conclusion, we recommend GLMs for future site-specific density forecasting. Investigations explain this conclusion in terms of the interaction between the autocorrelation properties of the data and the structure of the models tested. The use of ensembles with resolution higher than $1^{\circ}$ would probably improve the raw forecast capability at short forecast ranges.

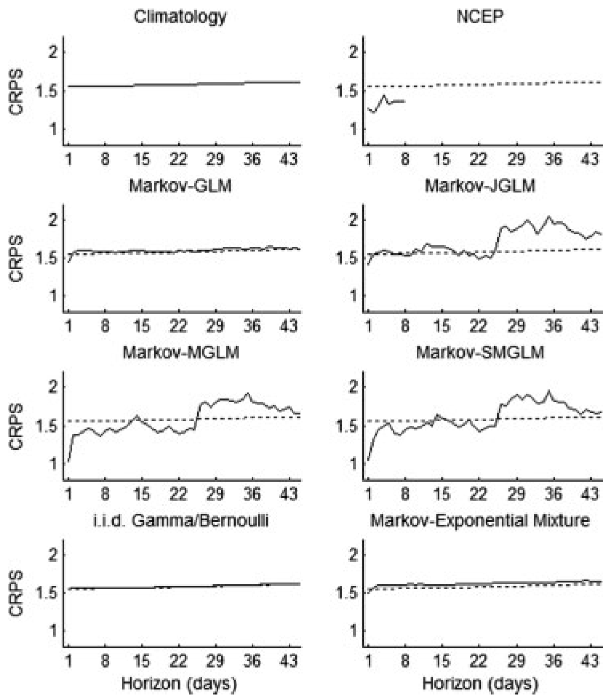


**Figure 4.** Continuous ranked probability score (CRPS) versus forecast horizon for each model, averaged over all 10 gauges. The dashed line indicates the climatology CRPS for comparison.

## 6. Conclusions

The research presented here has emphasized the need for improved access to historical and real-time meteorological data. Our project will facilitate academic research and scientists working in environmental consultancies by providing a rich source of information concerning extreme rainfall events. The BRDA has renewed recognition amongst the community regarding the need to digitize as much of the remaining archival rainfall data as possible.

The investigation of probabilistic rainfall forecasting has proven the practical viability of statistically based probabilistic forecasts that use state-of-the-art ensemble predictions as inputs. Oxford University's technology transfer organization, Isis Innovation, is supporting the commercialization of the resulting intellectual property from the database and associated quantitative models.

The implications for current operational model output statistics are that probabilistic forecast outputs from statistical methods can substantially improve ensemble forecast accuracy, and that only modest computational hardware is required.

The automated detection of rainfall patterns has demonstrated that underlying the spatial pattern of extreme UK rainfall events, there exists a simple clustering of the events that allows classification according to their spatio-temporal occurrence. The new classification scheme has implications across broad areas in hydrology, in both research and practice, justifying the concentration of simulation effort on the most probable extreme rainfall event locations and seasons.

## References

Buizza R, Petroliagis T, Palmer T, Barkmeijer J, Hamrud M, Hollingsworth A, Simmons A, Wedi N. 1998. Impact of model resolution and ensemble size on the performance of an Ensemble Prediction System. *Q. J. R. Meteorol. Soc.* **124**: 1935–1960.

Campbell SD, Diebold FX. 2005. Weather forecasting for weather derivatives. *J. Am. Stat. Assoc.* **100**: 6–16.

Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer: London.

EA. 2010. *The Costs of the Summer 2007 Floods in England*. Environment Agency: Bristol, UK.

EU. 2009. *Directive 2009/138/EC of the European Parliament and of the Council*. Official Journal of the European Union: Brussels.

Golding B. 2000. Quantitative precipitation forecasting in the UK. *J. Hydrol.* **239**: 286–305.

Golding B, Clark P, May B. 2005. The Boscastle flood: meteorological analysis of the conditions leading to flooding on 16 August 2004. *Weather* **60**: 230–235.

Hand WH, Fox NI, Collier CG. 2004. A study of twentieth-century extreme rainfall events in the United Kingdom with implications for forecasting. *Meteorol. Appl.* **11**: 15–31.

Institute of Hydrology. 1999. *Flood Estimation Handbook*. Wallingford Hydrosolutions: Wallingford, UK.

Lamb HH. 1972. *British Isles Weather Types and a Register of Daily Sequence of Circulation Patterns, 1861–1971*. *Geophysical Memoir*, 116, HMSO: London.

Little MA, Rodda HJE, McSharry PE. 2008. Bayesian objective classification of extreme UK daily rainfall for flood risk applications. *Hydrol. Earth Syst. Sci. Disc.* **5**: 3033–3060.

Little MA, McSharry PE, Taylor JW. 2009. Generalised linear models for site-specific density forecasting of UK daily rainfall. *Mon. Weath. Rev.* **137**: 1031–1047.

Met Office. 2011. Meteorological Office, Exeter, UK. http://www.metoffice.gov.uk/about-us/who/how/case-studies/cumbria-floods

New M, Todd M, Hulme M, Jones P. 2001. Precipitation measurements and trends in the twentieth century: review. *Int. J. Climatol.* **21**(15): 1899–1922.

Osborn TJ, Hulme M. 2002. Evidence for trends in heavy rainfall events over the UK. *Phil. Trans. Roy. Soc, Lond. A* **360**(1796): 1313–1325.

Pinson P, McSharry P, Madsen H. 2010. Reliability diagrams for nonparametric density forecasts of continuous variables: accounting for serial correlation. *Q. J. R. Meteorol. Soc.* **136**(646): 77–90.

Rodda JC. 1967. A country-wide study of intense rainfall for the United Kingdom. *J. Hydrol.* **5**: 58–69.

Rodda JC. 1972. A study of magnitude, frequency and distribution of intense rainfall in the United Kingdom. *British Rainfall* **1966**: 204–215.

Rodda HJE, Little MA, Wood RG, MacDougall N, McSharry PE. 2009. A digital archive of extreme rainfalls in the British Isles from 1866 to 1968 based on 'British Rainfall'. *Weather* **64**(3): 71–75.

Rodda JC, Little MA, Rodda HJE, McSharry PE. 2010. A comparative study of the magnitude, frequency and distribution of intense rainfall in the United Kingdom. *Int. J. Climatol.* **30**: 1776–1783.

Taylor JW, Buizza R. 2006. Density Forecasting for Weather Derivative Pricing. *Int. J. Forecast.* **22**: 29–42.

Thielen J, Gadian A. 1996. Influence of different wind directions in relation to topography on the outbreak of convection in northern England. *Ann. Geophys.* **14**: 1078–1087.

 