Bayesian Objective Classification of Extreme UK Daily Rainfall for Flood Risk Applications

Max A. Little[1], Harvey J. E. Rodda[2], and Patrick E. McSharry[1]

[1]{Systems Analysis, Modelling and Prediction Group, University of Oxford, Oxford OX1 3PJ, United Kingdom}

[2]{Hydro-GIS Ltd., 10 Coles Lane, Chalgrove, Oxfordshire OX44 7SY, United Kingdom}

Correspondence to: Max A. Little (littlem@robots.ox.ac.uk)

Abstract

In this study we describe an objective classification scheme for extreme UK daily precipitation to be used in flood risk analysis applications. We create a simplified representation of the spatial layout of extreme events based on a new digital archive of UK rainfall. This simplification allows a Bayesian clustering algorithm to compress these representations down to eight prototypical patterns of extreme falls. 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. 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. We discuss the practical use of this new typing scheme in flood simulations and climate change applications.

1 Introduction

Water is vital to life and humanity has always sought to control it to support an expanding population. Too little water and life becomes impossible, conversely, too much is dangerous and puts communities at risk of flooding. Understanding the motion of water on land is critical therefore, and is a major application of hydrological analyses and physical modelling.

Modelling for flood risk applications involves application of the basic physical processes of hydrology and knowledge of the specific meteorological and geographical circumstances of the region under study.

However, for many reasons, including rapidly changing land usage patterns, complex topographical and flow routing details, and highly nonlinear physical processes (Beven, 2006), accurate and detailed hydrological modelling is an involved undertaking. It therefore becomes necessary, in practice, to focus efforts on conditions under which flooding is most frequent and/or dangerous. Since the primary cause of both fluvial and pluvial flooding is *extreme rainfall*, it therefore becomes important to find a simplified rainfall representation to make extensive hydrological analysis manageable and routine. The temporal and spatial layout of extreme rainfall can determine which rivers flood and by how much.

Taken *prima facie* as a vast collection of weather variable observations of, e.g. air pressure, temperature and moisture, the number of possible atmospheric states is essentially infinite and even a large but finite set of observations, without analysis, offers no real insight into the underlying mechanisms at work (Holton, 2004). *Dynamical meteorology* tries to represent mathematically the atmospheric state at each time instant. The resulting equations can be considered a form of *data compression* (Ockendon, 2003;Cover and Thomas, 2006); equations that could, in principle, be solved to simulate all possible extreme rainfall patterns. However, general solutions of even the relatively simple Navier-Stokes equation are, unfortunately, unknown (Fefferman, 2000). As a result, mechanistic meteorological understanding tends to be *qualitative*, based on illustrative applications of the underlying physical mechanisms verified against observations, rather than the detailed *quantitative* modelling typical of numerical weather forecasting (Barry and Chorley, 2003;Holton, 2004).

Such qualitative descriptions can be used to produce simplified classifications known as *weather types* (Lamb, 1972;Hess and Brezowsky, 1977). As a whole this set of types can be a considerably compressed representation of all possible atmospheric states. This representation can be very useful in hydrological practice, as, for example, are synthetic storm profiles for hydrological design problems (Reed et al., 1999). They have the advantage that they can be chosen on the basis of simple but widely-appreciated mechanistic physical insights.

Nonetheless, qualitative types have an ineliminable *subjective* component because they are based on interpretations of the designer (Bardossy et al., 1994), and to this extent involve assumptions that may not lead to an optimally compressed representation. Therefore, more recent work has applied *data mining* techniques to a representative collection of atmospheric observations (Wilks, 2006;Jones et al., 1993;James, 2007;Bardossy et al., 1994;Lauzon et al., 2006). These are *objective* in that no human input is required once the classification algorithms have been designed (Bardossy et al., 1994); they are subjective to the extent that the mathematics is always based on some assumptions which are unverifiable in principle, and they are limited to compressing information from data observed over a particular spatial region for a specific time interval. That data can never be entirely representative of all possible atmospheric states.

For rainfall therefore, equipped with either subjective or objective types, we can then perform a manageable number of hydrological simulations or flood risk assessments with rainfall of each different type, expecting that this will be representative of the broad range of possible flooding scenarios. An objective typing scheme is preferred though, because the results are readily reproducible without human effort and variability. Nonetheless, it is important to be able to check the validity of an objective method against known meteorological and hydrological knowledge and physical insights. To this aim we produce both a subjective scheme and a novel objective scheme for extreme UK rainfall, based on a new digital archive of extreme rainfall events (Rodda et al., 2008). We use the subjective scheme and a number of notable hydrological events as a check on the validity of the results of the objective scheme.

The structure of the paper is as follows. Sect. 2 describes the data used in the study, and Sect. 3 the methods and their rationale for both the subjective and objective typing schemes presented in this study. Sect. 4 presents the results of applying both schemes, and then discusses and compares the resulting rainfall types in terms of their applicability in hydrological flood risk analysis. Finally, Sect. 5 draws conclusions and directions for future research.

2 Data

The data for this study comes from two sources. The first source is 257 rainfall events taken from the publication British Rainfall (BR) over the years 1866-1968. These years were used because for this period the publication contained a section on extreme rainfall events under the heading “Heavy Falls on Rainfall Days” or “Heavy Falls in 24 Hours”. Within this section all observed 24 hour rainfall depths were listed which exceeded a certain threshold. This was set at 2.5 inches (63.5mm) 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 50mm 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 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 has been compiled into a new digital archive (Rodda, Little et al. 2008).

The second source of information is the rainfall depth observations from the UK Meteorological Office MIDAS surface daily weather observation network. Extreme depths over 50mm were extracted, covering the same time range as the BR archive. The two sources of information (BR 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.

Although all the observations are subjected to manual quality control, there remains the possibility of some error in the total recorded amount; as with all raingauge data there is the potential for other forms of precipitation such as snow or hail to be recorded as rainfall if it melts.

The events in this record represent one choice of exceedance threshold – certainly other thresholds are possible. However, this particular threshold captures very rare events for the UK (more extreme than the 90th percentile for most locations), and many such events have led to dangerous flooding and so that the record is of considerable hydrological importance.

3 Methods

3.1 The observed character of UK rainfall

Rainfall is a complex phenomenon. Two dominant processes: evaporation and vegetative transpiration release water which rises to condense out as clouds at the saturation vapour pressure (Barry and Chorley 2003). Atmospheric water migrates due to wind-driven transport (advection), and turbulent mixing. The current understanding of general atmospheric circulation views the combined effect of differential heating at equator and poles, travelling atmospheric waves, and the rotational Coriolis force as the main causes of prevailing winds which can transport moisture over large distances. Also, the spatial land-sea layout interacts with prevailing winds and seasonal heating shaping rainfall patterns.

Airborne water precipitates out under the right conditions. One of the most important precipitation mechanisms is freezing nuclei (Bergeron-Findeisen theory): tiny ice crystals forming around atmospheric contaminants growing rapidly by deposition or aggregation, gathering enough weight to overcome any updrafts in the cloud. They can then fall and melt into raindrops which may eventually reach the ground. The other mechanism is coalescence; large and therefore heavy raindrops have large terminal velocities, and can therefore 'sweep up' smaller droplets as they fall (Barry and Chorley 2003).

Spatial patterns of rainfall can be identified at a range of scales. Fronts (cyclones, depressions) are large-scale atmospheric wave-like structures, 1,000 to 3,000km long, forming at interfaces between cold and warm air. The low-pressure frontal core, when transported with moist air over land can bring widespread bands of heavy rainfall (frontal rain). Saturated air rising over topographical features can result in localised rainfall (orographic rain). Rainfall can also be convective: differential land heating causing atmospheric instability, generating thunderclouds (cumulus) leading to localised downpours (Barry and Chorley 2003). These physical observations point towards three broad classes of rainfall: frontal, orographic and convective, although some combined and/or sub types of these are normally observed.

In the UK, the prevailing westerly Atlantic wind brings moist air over the exposed, western coastline, which is likely therefore to experience orographic rainfall. Frontal rain can cover the entire country, and the flat, central and southern parts of the country are particularly at risk of convective instabilities during summer.

Characterised by clusters of intense thunderstorm cells within a mesoscale region of stratiform rain which forms a shallow depression, mesoscale convective complexes (MCC) (Collier and Hardaker 1996) are rare in the UK, more often initiated from deep convection over continental interiors (Browning and Hill 1984). In the UK they occur only in summer, often originating from the Bay of Biscay and mostly confined to south and southwest regions. MCCs have been responsible for the UK's largest 24 hour rainfall totals and some of the most devastating floods, such as that in Lynmouth, 1952 (Bleasdale and Douglas 1952). Spatially localised thunderstorms, more scattered and less intense than MCCs and without the defining depression can only cause localized flooding. This usually affects urban areas; an example is the Hampstead Storm of 1975 (Bailey, Carpenter et al. 1981).

Finally, a commonly observed east coast pattern exists: depressions travelling in a westerly direction slow down over land and bring moist, easterly winds from the North Sea. In winter these cause snow, but in summer they can bring prolonged rainfall and notable flooding, for example the Norwich floods of 1912 (Mill 1913).

3.2 Subjective UK Rainfall Typing

Most subjective weather typing concentrates on circulation patterns in mean sea-level pressure (MSLP) fields. The well-established circulation typing scheme of Lamb (1972) identifies 26 types grouped into four broad classes of cyclonic, anticyclonic, stationary and unclassified days. It is very widely used, particularly as a predictor variable in many statistical models of UK rainfall. The Grosswetterlagen (GWL) (Hess and Brezowsky 1977) applies to the whole of the European continent, and is of similar influence to the Lamb scheme.

Rodda et al. (2000) described a UK extreme rainfall classification scheme as part of a flood risk model for insurance purposes. A large number of flooding scenarios were generated based on synthetic rainfall events of each type. The scheme was based on spatial patterns of rainfall associated with 72 historical flood events. Four rainfall types were identified: frontal, thunderstorm, east coast and mesoscale convective complex (MCC). Frontal events encompassed all sub-types of fronts (e.g. warm, cold, occluded, stationary) associated with the westerly track of Atlantic depressions characterised by rainfall over a wide area but heaviest over western parts of Britain.

Hand et al. (2004) studied 50 extreme rainfall events based more on meteorological conditions rather than spatial rainfall patterns. Three primary classes were identified – orographic, frontal and convective, with additional frontal sub types (embedded convection and convective events with frontal forcing). These latter two sub types corresponded closely with the MCC events of Rodda et al. (2000). East coast events were labelled as frontal in this scheme. The orographic label was given to events where general topographic lifting was determined as the dominant mechanism for very extreme rainfall. The convective label was attributed to highly localized and temporally intermittent events, even if the convection was triggered by a previously dissipated frontal system.

Here, we introduce a new subjective typing scheme for UK extreme daily rainfall. The classification was constructed from meteorological interpretations of rainfall amounts, and descriptive text produced by observers of 257 extreme rainfall events in the British Rainfall (BR) publication. Inspired by the Rodda and Hand classifications described above, the five types were: depression, orographic, convective (MCC), thunderstorm, and east coast. The 257 BR events were those for which good quality isohyetal maps were available to make a classification based largely on spatial rainfall patterns. Additional information provided by the text such as thunder, description of the depression path and overall rainfall amounts also assisted in the typing.

3.3 Objective UK rainfall typing

The literature describes a wide range of approaches to objective typing of weather states, but the predominant approach involves the use of *principal components analysis* (PCA) (White et al., 1991). This method compresses an observed weather variable field down to a small set of components which, taken in linear combination, can be used to reconstruct the field to within a prescribed level of accuracy defined by the fraction of variance explained. Typically, PCA will be followed by a *clustering* or *classification* technique which finds a grouping for the observed sets of weather variables which is optimal under some numerical criteria. Examples include *k-means* or *classification and regression trees* (CART), but also *correlation-based* pattern recognition (Kirchofer, 1973), *fuzzy rule-based* methods (Bardossy et al., 1994) and *compositing* (Moses et al., 1987).

For the Columbia River Basin, US, Hughes *et al.* (1993) applied PCA and CART to classify MSLP states; rainfall simulations conditioned on these states were then used to estimate seasonal streamflows and flood frequencies under climate change scenarios. Similar methodology is applied in Schoof *et al.* (2001). Zorita *et al.* (1995) applied PCA/CART to SLP associated with rainfall occurrence and amounts at selected gauging stations. Three weather types were found and used to simulate local rainfall from the output of a GCM. For Sardinia, Benzi *et al.* (1997) classified temperature and precipitation fields using PCA, followed by a clustering method; eleven spatially-distinct precipitation types were identified. Recently, *Kohonen* and *multilayer perceptron neural networks* have also been used to cluster daily precipitation fields (Lauzon et al., 2006).

In our application, we wish to construct a set of types that optimally represent the spatial layout of UK rainfall extremes, where extreme in this context is any daily rainfall amount greater than 50mm at any raingauge, using the same dataset as that used to construct the subjective scheme. However, we face the problem that the set of possible layouts of extreme rainfall events is exceedingly large, but we only have small number of examples of event layouts (because such events are rare). This is an example of the *curse of dimensionality* and in general therefore, if we apply an automated classification technique to the raw, unprocessed extreme rainfall event data, the resulting classification will be highly unstable: a slight perturbation of the data may cause significant changes to the resulting types (Hastie et al., 2001). One approach to stabilizing the classification is to *compress* or simplify the data such that the set of all possible events is more constrained.

Often, PCA is used to accomplish this stabilization step, but the distribution of rainfall amounts and extremes is highly *non-Gaussian*. This is problematic in this context because, under the statistical interpretation, PCA is a *linear-Gaussian* technique so that, if the data is drawn from a multivariate Gaussian, truncation of the full set of principal components has a consistent data compression interpretation in information-theoretic terms (Cover and Thomas, 2006). Furthermore, although the 'linear combination of a small number of bases' model underlying PCA is very simple and hence produces stable results, it is *too stable* for our purposes here because the rare events we are studying often have highly variable layouts, and we wish to maintain a good representation of this variability.

Therefore, we need to find a compressed representation that is both (a) consistent with the non-Gaussian nature of the data, and (b) sufficiently simplified to allow stable clustering. There are many ways in which this could be achieved; largely this will depend on the eventual application and we describe and justify our particular approach next.

In this study we are interested in the broader, UK-wide spatial variability of extremes. The extreme amounts are listed by UK grid reference and day, and we therefore simplify the data by rounding the grid location of each rainguage to whole latitude/longitude values. The gauge locations ranged from longitude 8°W to 2°E, and latitudes 49°N through 60°N, covering the UK by a grid of 11 horizontal by 12 vertical, a total of 132, 1° by 1° cells. Thus, on average across the UK, each grid cell covers an area of approximately 110km vertical by 65km horizontal, or 7,150km^2.

An array element $\mathbf{v}_{tij}$ representing each cell on each event day is set up, where $i = 1, 2 \ldots 11$ and $j = 1, 2 \ldots 12$ are the horizontal and vertical indices respectively. The index $t = 1, 2 \ldots N$ selects the event number, and here $N = 257$. Next, if on any day, a grid cell (i, j) contains at least one gauge with a daily amount of greater than 50mm, that grid cell is indicated as 'extreme' with $\mathbf{v}_{tij} = 1$, otherwise $\mathbf{v}_{tij} = 0$. For the subsequent clustering, we rearrange all events into a set of M-element vectors where $M = 132$, by horizontal column-first stacking. We denote these stacked vectors by $\boldsymbol{u}_t = \mathbf{v}_t$.

Note that the grid cells themselves cover very large areas with respect to the size of typical extreme events over the UK. Therefore, although the stacked vectors omit the sense of spatial neighbourhood between cells, each cell is relatively insensitive to local variations in the position of events that occur on different days. Thus, this simplified extreme indicator representation captures the spatial variability of different events at sufficient spatial resolution for the subsequent analysis to preserve spatial proximity.

Next, *Bayesian k-means* clustering is applied to this array of extreme indicators. This method solves three problems: finding a small set of representative 'cluster centroid' (template) vectors, a unique assignment to one of these clusters for each vector t, and the optimal number of clusters. The first two problems are solved by an alternating optimization algorithm, for details see Hastie *et al.* (2001). The centroids are chosen such that the *within-cluster dissimilarity* between vectors, measured using a distance function $d(\ ,\ )$, is minimized:

$$W(K) = \sum_{k=1}^{K} \sum_{C(t)=k} \sum_{C(t')=k} d(\boldsymbol{u}_t, \boldsymbol{u}_{t'}) \tag{1}$$

Here, K is the chosen number of clusters. The assignment function $C(t)$ maps an event number on to the class to which it is assigned. In this study, we use the standard square Euclidean distance function, which is the sum of squares of the difference between corresponding elements in each vector.

Although minimization of $W(K)$ is possible with the alternating optimization algorithm, this does not allow us to choose the *number* of clusters K. To see this, note that increasing the number of clusters reduces $W(K)$: in the extreme case $K = N$ the number of events, and each event vector is assigned to a single cluster which is its own cluster centroid, so that $W(N) = 0$. This *reductio ad absurdum* can be avoided if the *complexity* of the model is taken into consideration: we want the smallest possible number of clusters that simultaneously minimizes $W(K)$. One approach to doing this is using the *Bayesian Information Criteria*

(BIC), which effectively penalizes the average dissimilarity over all clusters, for using a large number of clusters.

The BIC for most statistical inference problems such as this is often difficult to calculate analytically. However, the assumption in the basic *k*-means algorithm is that the clusters are *spherical multivariate Gaussian distributions* sharing the single variance σ^2 (Note that it is important to distinguish between this spherical Gaussian “shape” of the statistical model for each cluster *in the space of extreme indicator vectors*, from the geospatial shape in physical space of each cluster – these two “spaces” are not in any way related). This makes a tractable approximation to the BIC possible (Pelleg and Moore, 2000). To do this, we must find the maximum likelihood estimate of the variance of the clusters, which is given by:

$$\sigma^2 = \frac{1}{N-K} W(K) \tag{2}$$

(Note that this occurs because the within-cluster distance is equivalent to the sum of squared residuals from each cluster centroid (Hastie et al., 2001), and is thus equivalent to the maximum likelihood estimator of the shared variance). We also need the *log likelihood* $L(K)$ of the data given the Gaussian models:

$$L(K) = \sum_{k=1}^{K} \left[-\frac{N_k}{2}\log(2\pi) - \frac{MN_k}{2}\log(\sigma^2) - \frac{N_k - K}{2} + N_k \log(N_k) - N_k \log(N) \right] \tag{3}$$

Here, N_k is the number of events assigned to cluster *k*. Additionally, the approximate BIC *log prior* $P(K)$ is given by:

$$P(K) = -\frac{K + MK}{2}\log(N) \tag{4}$$

Finally, the *Negative Log Posterior* (NLP), that is essentially, the *information* in the posterior distribution evaluated at *K* is given by $NLP(K) = -P(K) - L(K)$. Obtaining the number of classes that gives the *maximum a-posteriori* probability is equivalent then to minimizing $NLP(K)$.

In the alternating minimization algorithm for $W(K)$, the initial assignments $C(t)$ are chosen at random, and because the *k*-means problem has no guaranteed unique solution, we replicate the optimization from 10 random initial assignments, and take the cluster centroids and assignments that yield the smallest value of $W(K)$ over all replications. Finally, note that the cluster centroids represent mean values of all the indicator variables in each vector assigned to that cluster: therefore the individual vector elements of the centroids will have real values that can lie anywhere between 0 and 1 – a value proportional to the likelihood of an extreme event occurring in each grid cell.

4 Results and discussion

4.1 Subjective classification

Representative examples of the subjective classes obtained are shown in isohyetal maps depicted in Figure 1 through Figure 5, and definitions of each type are described in Table 1. The table also lists the percentage of extreme event days in the BR archive, showing that orographic and depression events are the most common. The extremes predominate in July, August and December.

For the hydrological consequences of the subjective types, the following flood events (Black and Law, 2004) are associated with each type: *depression* (26th April 1908, Thames Valley, Wallingford, Oxfordshire), *orographic* (3rd December 1960, Uckfield, Sussex), *MCC* (10th – 11th July 1968, Bristol and Somerset), *thunderstorm* (10th August 1959, Newquay, Cornwall), *east coast* (20th – 23rd July 1930, Esk and Leven Valleys, near Whitby, North Yorkshire).

4.2 Objective classification

See Figure 6 for a graphical depiction of the automated classifications obtained after applying the *k*-means procedure, and Figure 7 for the results of the cluster number selection process. Eight types are selected as the optimum number of clusters, in contrast to the five types in the subjective scheme. The most common events, Type (a) has extremes stretching across the entire central region of the country, with the most probable location being in the mid-south coast area. The second most common, Type (b), are located in the far south, coastal regions of England. The third-ranked Type (c) events are localised over Wales and the northern English west coast, and the next, Type (d), over the far south west, Devon and Cornwall. Type (e) events are concentrated over the central and east coast areas in the English midlands, whereas Type (f) has very high concentration of extreme rainfall over the far, northwest Scottish coast. The final two types (g) and (h), have extremes concentrated over the north-western English and Scottish coast, and east central England.

We now validate these results against those obtained in the subjective typing scheme. Table 2 gives the conditional probability of subjective type given the specific automated classification type, and allows an assessment of the overall strength and character of association between the different types.

Objective Type (a) is most strongly associated with both east coast and depression subjective types, although it has non-negligible, but much weaker, associations with all the other subjective types as well. From the spatial layout, however, we would be more strongly inclined to associate this with the depression subjective type. Objective Type (b), however, associates most strongly with thunderstorm, east coast and depression subjective types. It does not associate at all with the orographic type.

Type (c) pairs very strongly with the orographic type; the dominant west-coast layout makes the overlap between these two types particularly close. This is very similar to the situation with Types (d), (f) and (h) which can be readily matched with the orographic subjective class. Type (e), being of central layout, matches most closely with depression and, to a lesser extent, thunderstorm types. Finally, objective Type (g) associates almost equally with all but the east coast subjective class.

On the basis of these observations we can propose the following objective-subjective mapping: Type (a) – depression, Types (c), (d), (f) and (h) – orographic, Type (g) – MCC, Type (b) – thunderstorm, Type (e) – east coast. This mapping is itself somewhat subjective but provides a simple way in which the objective results can be verified against the subjective typing for the purposes of interpretation in terms of meteorological mechanisms.

Turning to seasonal aspects of the objective typing, Figure 8 shows the conditional probability of the month given a specific objective classification type. Notable features include the fact that objective Type (a) rarely occurs in winter months, whereas Types (d), (g) and (h) almost never occur during early winter or spring months. Types (b), (e) and (h) have a very strong peak in July, and class (c) appears to have no obvious preference for month of occurrence. Class (f) is more likely to occur in December than any other month.

In terms of hydrological consequences, the 26th April 1908 Thames Valley *depression* event is here classified under Type (a). The subjective *orographic* event on 3rd December 1960, Sussex is classified under Type (h) and the *thunderstorm* 10th August 1959, Cornwall event was associated with Type (d). Types (e) and (a) associate with the *east coast* event on 20th –

23rd July 1930, North Yorkshire, and the subjectively-typed *MCC* event of 10th – 11th July 1968, Bristol and Somerset associates with Type (h).

5 Summary and conclusions

In this paper we demonstrated two different schemes for classifying the pattern of rainfall in the UK, one subjective and one objective, based on a comprehensive new archive of extreme events. We showed that, using a simplified, grid-based encoding of the spatial layout of extremes making up each event, eight objective types were optimal under a Bayesian complexity constraint at minimizing the average within-type dissimilarity. These eight types were strongly associated with similar spatial layouts in the five-class, subjective typing scheme. We also demonstrated that the objective scheme was able to account reliably for the spatial layout and seasonal timing of notable flooding events in the past century. We thus conclude that the objective scheme can be readily interpreted in terms of known meteorological mechanisms, and at the same time is also hydrologically relevant.

In practice then, the objective typing scheme has the obvious advantage over the subjective scheme 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 described above: the objective type for that day is given by the cluster that has the smallest total dissimilarity to the encoding.

In practical hydrological flood studies where design or other rainfall simulations are used, the cluster centroids in Figure 6 can be used to inform the spatial layout and Figure 8 the seasonal timing of extreme rainfall in each type. 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.

Finally, applying the objective typing to all past and future events could be useful in climate studies, for example to assess the extent to which patterns of extremes may be changing in response to global and regional temperature variations.

Acknowledgements

The authors would like to thank the British Atmospheric Data Centre for providing the MIDAS surface observations. The research was support by the UK Natural Environment Research Council grant NE/E002013/1 under the Flood Risk from Extreme Events programme. Dr Patrick McSharry is grateful for support from the UK Royal Academy of Engineering and the UK Engineering and Physical Sciences Research Council.

References

Bardossy, A., Muster, H., Duckstein, L., and Bogardi, I.: Knowledge based classification of circulation patterns for stochastic precipitation modeling, in: Stochastic and statistical methods in hydrology and environmental engineering, edited by: Hipel, K. W., Kluver Academic Publishers, 19-32, 1994.

Barry, R. G., and Chorley, R. J.: Atmosphere, weather, and climate, 8th ed., Routledge, London ; New York, xvi, 421 p. pp., 2003.

Benzi, R., Deidda, R., and Marrocu, M.: Characterization of temperature and precipitation fields over Sardinia with prinicipal component analysis and singular spectrum analysis, International Journal of Climatology, 17, 1231-1262, 1997.

Beven, K.: Searching for the Holy Grail of scientific hydrology: Qt=(S, R, Delta t)A as closure, Hydrol. Earth Syst. Sci., 10, 609-618, 2006.

Black, A. R., and Law, F. M.: Development and utilization of a national web-based chronology of hydrological events, Hydrological Sciences, 49, 237-246, 2004.

Cover, T. M., and Thomas, J. A.: Elements of information theory, 2nd ed., Wiley-Interscience, Hoboken, N.J., xxiii, 748 p. pp., 2006.

Fefferman, C. L.: Existence and Smoothness of the Navier-Stokes Equation, Clay Mathematics Institute, 2000.

Hastie, T., Tibshirani, R., and Friedman, J. H.: The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations, Springer series in statistics, Springer, New York, xvi, 533 p. pp., 2001.

Hess, P., and Brezowsky, H.: Katalog der Grosswetterlagen Europas (1881-1976), in: Berichte des Deutschen Wetterdienst, Bd. 15, Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, 1977.

Holton, J. R.: An introduction to dynamic meteorology, 4th ed., Elsevier Academic Press, Burlington, MA, 2004.

Hughes, J. P., Lettenmaier, D. P., and Guttorp, P.: A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation, Water Resources Research, 29, 3303-3315, 1993.

James, P. M.: An objective classification for Hess and Brezowsky Grosswetterlagen over Europe, Theoretical and Applied Climatology, 88, 17-42, 2007.

Jones, P. D., Hulme, M., and Briffa, K. R.: A comparison of Lamb circulation types with an objective classification scheme, International Journal of Climatology, 13, 655-663, 1993.

Kirchofer, W.: Classification of European 500 mb patterns43, 1973.

Lamb, H. H.: British Isles weather types and a register of daily sequences of circulation patterns, 1861-1971116, 1972.

Lauzon, N., Anctil, F., and Baxter, C. W.: Clustering of heterogeneous precipitation fields for the assessment and possible improvement of lumped neural network models for streamflow forecasts, Hydrol. Earth Syst. Sci., 10, 485-494, 2006.

Moses, T., Kiladis, G. N., Diaz, H. F., and Barry, R. G.: Characteristics and frequency of reversals in mean sea level pressure in the North Atlantic sector and their relationship to long-term temperature trends, Journal of Climatology, 7, 13-30, 1987.

Ockendon, J. R.: Applied partial differential equations, Rev. ed., Oxford University Press, Oxford ; New York, xi, 449 p. pp., 2003.

Pelleg, D., and Moore, A.: X-means: Extending k-means with Efficient Estimation of the Number of Clusters, in: Proceedings of the 7th International Conference on Machine Learning, Morgan Kaufmann, San Francisco, 727-734, 2000.

Reed, D. W., Faulkner, D., Robson, A., Houghton-Carr, H., and Bayliss, A. C.: Flood Estimation Handbook, Centre for Ecology & Hydrology, 1999.

Rodda, H. J. E., Little, M. A., Wood, R. G., MacDougal, N., and McSharry, P. E.: Extreme Rainfalls in the British Isles over the Period 1866-1968, Weather, 2008.

Schoof, J. T., and Pryor, S. C.: Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, International Journal of Climatology, 21, 773-790, 2001.

White, D., Richman, M., and Yarnal, B.: Climate regionalization and rotation of principal components, International Journal of Climatology, 11, 1-25, 1991.

Wilks, D. S.: Statistical methods in the atmospheric sciences, 2nd ed., Academic Press, Amsterdam ; Boston, xvii, 627 p. pp., 2006.

Zorita, E., Hughes, J. P., Lettemaier, D. P., and von Storch, H.: Stochastic Characterization of Regional Circulation Patterns for Climate Model Diagnosis and Estimation of Local Precipitation, Journal of Climate, 8, 1023-1042, 1995.

Table 1: Description of the subjective types of extreme rainfall events.

Subjective Classification	Description	Percentage of Extreme Event Days Classified	Predominant Month
Depression	Rainfall associated with a depression but not showing the features of the other classes. Can occur throughout the year.	27.2%	August
Orographic	Rainfall associated with the normal west-east movement of Atlantic depressions which is enhanced over the mountainous areas and occurs throughout the year.	27.6%	December
Convective (MCC)	Small depressions characterized by intense precipitation from convective cells within a larger area of continuous rain. Occur in summer affect S and SW Britain, falls of over 200mm in 24 hours can be observed.	9.7%	August
Thunderstorm	Isolated occurrences or progressions of convective cells; occur in summer but lacking the synoptic structure of an MCC.	16.0%	July
East Coast	Depressions where the eastwards progression stalls over the UK bringing moist air and rainfall from the North Sea to affect areas of the east coast. Occur in summer and can bring continuous rain for up to 4 days.	19.5%	August

Table 2: Conditional probability of obtaining a subjective classified extreme type (rows) given a specific objective type (columns). The italic entries indicate the proposed objective to subjective mapping (see text).

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Depression	***0.27***	0.22	0.11	0.17	0.52	0.13	0.25	0.08
Orographic	0.10	0.00	***0.82***	***0.57***	0.00	***0.87***	0.17	***0.75***
Convective (MCC)	0.14	0.13	0.00	0.17	0.04	0.00	***0.25***	0.00
Thunderstorm	0.14	***0.38***	0.04	0.09	0.30	0.00	0.25	0.08
East Coast	0.34	0.28	0.04	0.00	***0.13***	0.00	0.08	0.08

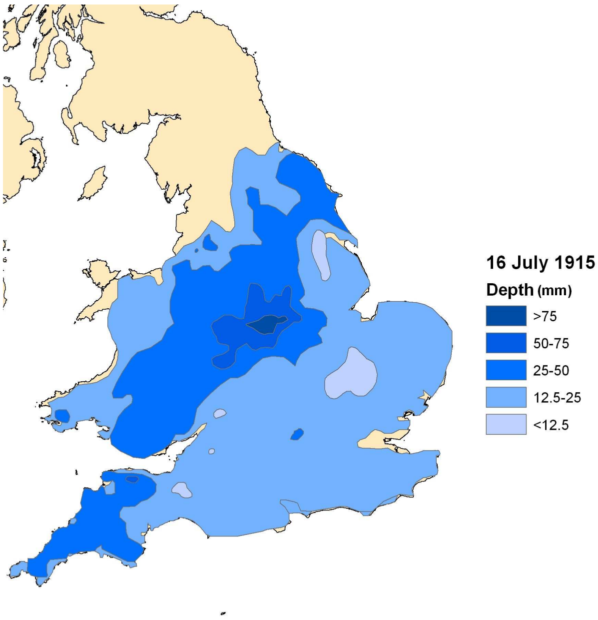

Figure 1: Typical depression extreme rainfall event identified from the British Rainfall archive.

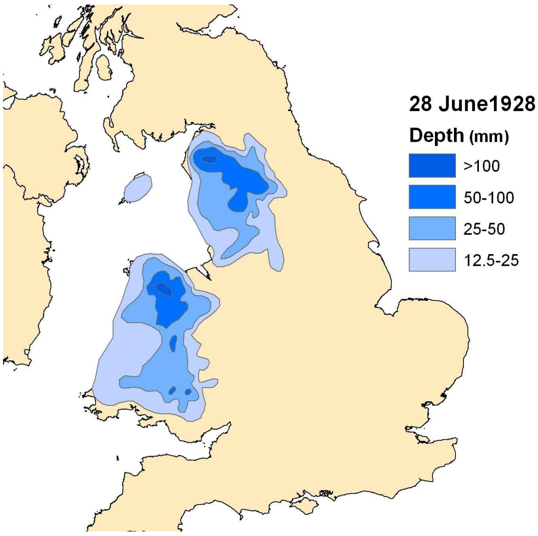

Figure 2: Typical orographic extreme rainfall event from the British Rainfall archive.

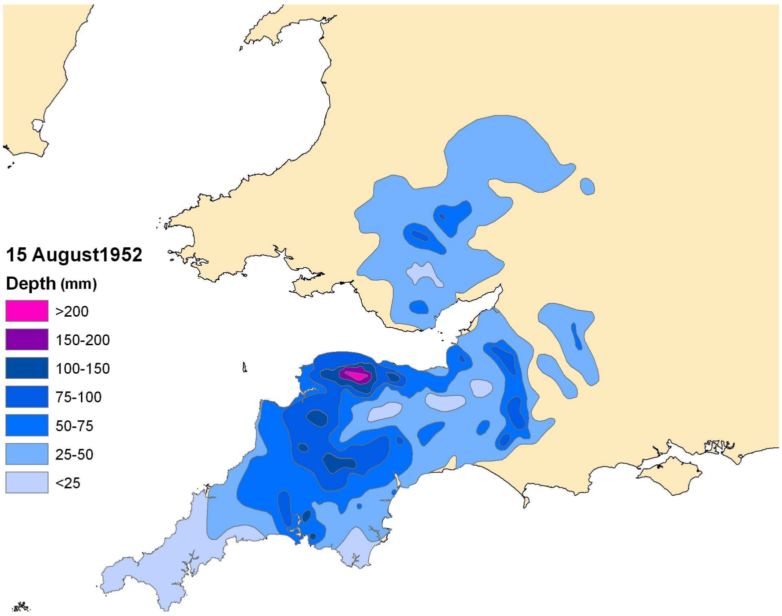

Figure 3: Typical convective extreme rainfall event, also known as a mesoscale convective complex (MCC), identified in the British Rainfall archive.

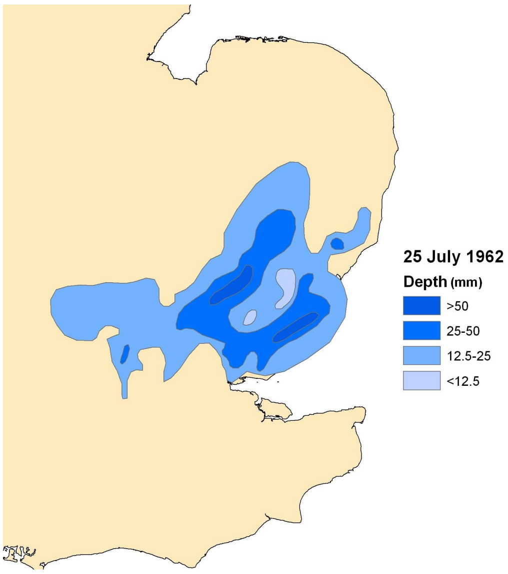

Figure 4: Typical example of a thunderstorm extreme rainfall event from the British Rainfall archive.

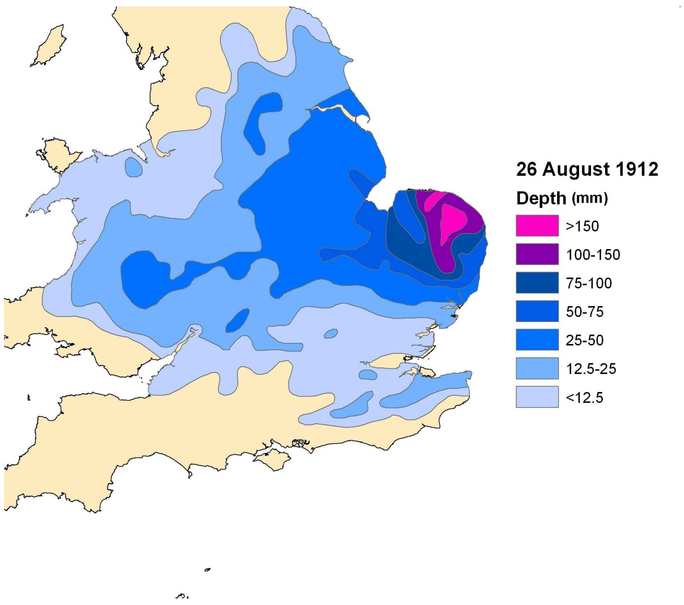

Figure 5: Typical example of an east coast extreme rainfall event from the British Rainfall archive.

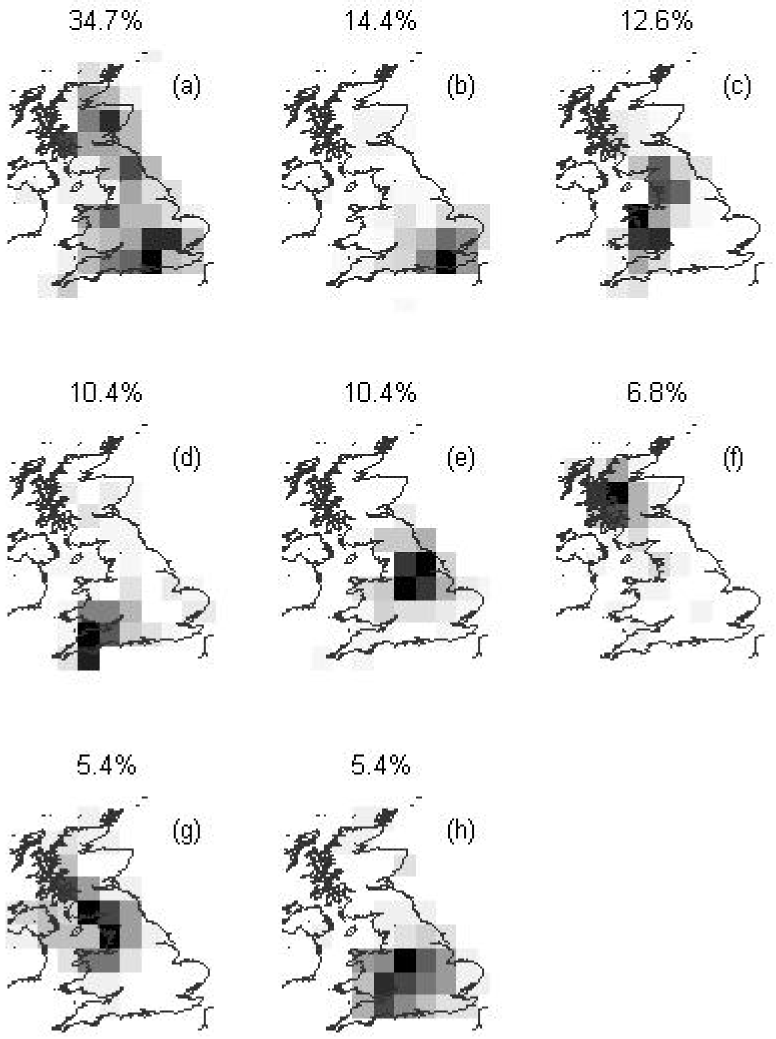

Figure 6: Graphical depiction of the eight objective types (a) – (h). The figure above each type is the percentage of selected days classified in that type. Darker colours indicate higher prevalence of extremes in that 1 degree by 1 degree grid cell.

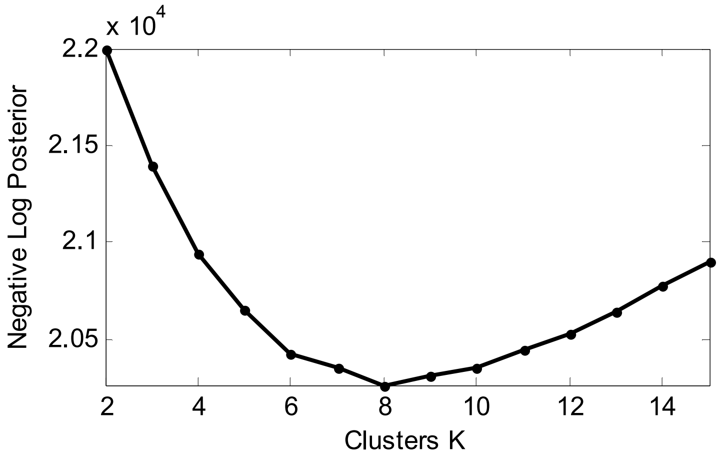

Figure 7: Results of the cluster number selection process in the objective classification. The horizontal axis is the number of clusters, and the vertical axis is the negative log posterior.

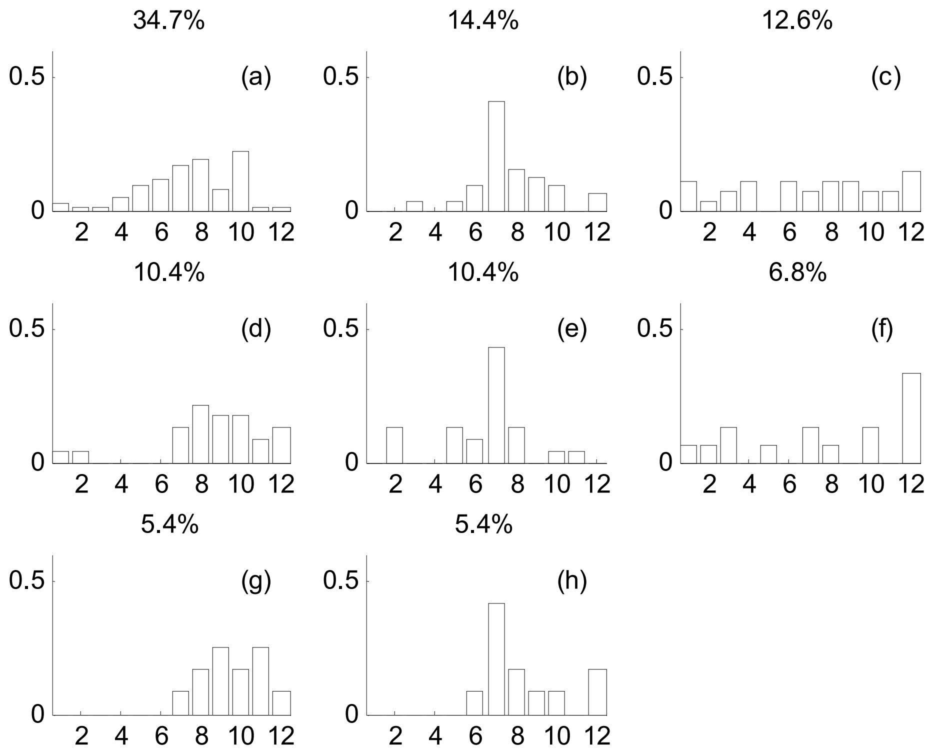

Figure 8: Conditional probability of the month given a specific objective classification type. The horizontal axis is month number, the vertical axis is probability. The figure above each type is the percentage of selected days classified in that type.