

Interactive comment on “Bayesian objective classification of extreme UK daily rainfall for flood risk applications” by M. A. Little et al.

M. A. Little et al.

Received and published: 13 January 2009

We are grateful to the anonymous referees for their time and effort in providing insightful comments. Overall, we are encouraged by the positive responses about importance, scope and the clarity of presentation of the ideas. The referees raised some specific issues and we wish to address these individually below. As referees 1 and 3 appear to mention comments by referee 2, we will address referee 2’s comments first.

N.B. These author comments are split into two files, please see the second comment file as well as this one.

Referee 2

(1) If we have understood correctly, the issues raised by this reviewer surround the ap-

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



propriateness of the k-means clustering technique, as used in this study. The reviewer suggests that use of a latent class model (finite multivariate Bernoulli mixture model) would improve the statistical methodology, because the Bernoulli model would be more appropriate than the Gaussian model for the binary representation of extremes.

This is an interesting technical point; the main thrust of our response is that such technical details should be balanced against the wider context of the ability of the k-means method to capture the main synoptic scale meteorological and seasonal patterns in extreme rainfall, as demonstrated in this paper.

We feel it is probably necessary to refine our discussion of the use of Euclidean distance metric in the k-means algorithm. This distance metric does not logically imply that the vectors are drawn from spherical multivariate Gaussians, instead, if one assumes this distributional model, then the square Euclidean distance metric is implied by maximum likelihood estimation (in other words, spherical Gaussians imply least squares estimation, but not the other way around!) Thus, the only real logical issue in our statistical methodology has to do with the choice of distributional model for the likelihood of each vector (not the choice of distance metric), and its' relationship to the BIC framework. Hence, such choice of distributional models only affects the log likelihood $L(K)$ and hence, potentially, the choice of the optimum number of clusters K . It does not directly affect the estimated cluster centroids.

Thus, although technically correct to state that the spherical Gaussian assumption used to derive the log likelihood expression is inappropriate for binary data, whether this fundamentally invalidates the statistical methodology is unclear. There are many potential weaknesses in any choice of mathematical assumptions: for example, one could also criticise the use of the BIC, because it is only one of a number of methods (including Akaike Information Criterion, Minimum Description Length, Minimum Message Length etc.), and there is thus ambiguity about this choice. Does this ambiguity also imply that our statistical methodology is invalid, even if the BIC approach appears to select a realistic number of precipitation categories? Our approach to resolving such

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

[Discussion Paper](#)



ambiguities is to try the simplest methods first.

Given these points above, whilst we agree with the reviewer that Bernoulli finite mixture models might be more appropriate for the binary representation we use, we do not think that the benefits are as clear cut as perhaps the reviewer suggests. We feel that at least, the choice of distance metric or distributional model for the data would require an extended, side-by-side or systematic study devoted to the technical, mathematical issues about the relative sensitivity of the performance of the clustering to such choices. For example, it would not be enough to test k-means against latent class models; to do justice to all the issues one should also test k-medoids, Hamming-distance k-means and exponential family PCA alongside k-means and the latent class model, for example. This would probably just scratch the surface though, as the literature abounds with comparable techniques.

Such a comparative study may be very valuable, but the technical complexity would obscure the effective hydrological importance of this study, in our opinion. That is why we opted for a simple, well-known technique that was likely to be understood by a wide audience of hydrologists, and indeed, we find this is successfully able to reproduce known meteorological categories.

In our revision of the paper for HESS, we will incorporate some discussion of these interesting issues.

"Page 3041: the meaning of MSLP is missing."

(2) This is actually defined at the bottom of page 3039.

Referee 1

(3) "However, I would fully support the criticism raised by the other referee concerning the method used for clustering 8230;"

If we understand this comment it is a reiteration of the argument of referee 2 about the validity of the Gaussian assumption in the likelihood calculation $L(K)$. We would then

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



point the reviewer towards our responses to referee 2.

(4) "1.1. Parts of Section 3 are somewhat missing the point. The whole paragraphs from the evapotranspiration to the Bergeron-Findeisen theory would be fine in any textbook about atmospheric sciences, but it is not relevant in this context. Please delete this. On the other hand, either in this Section 3, or maybe even better in the introduction, the authors should discuss the state of the art of research on UK precipitation extremes, e.g., Osborn et al, IJOC, 2000, Fowler and Kilsby, IJOC, 2003, Moberg and Jones, IJOC, 2005, Maraun et al, IJOC, 2008. So please delete the mentioned paragraphs and add a discussion of relevant literature."

We agree with the reviewer that a discussion of Bergeron-Findeisen theory may be somewhat elementary, and it is the synoptic spatial character of extremes that is of more importance than the basic mechanism of precipitation production, so we will remove this in the HESS revision.

Regarding the references suggested by the reviewer, this work is important and we are of course well aware of this literature but we felt that it was tangential to the introduction of a new method for spatial typing of extremes we are conducting in this study. By contrast, these studies are concentrated on climate trends in various categories of rainfall amount (Maraun 2008, Osborn 2000), extreme amount indices (Moberg 2005), and in changing parameters of EVT models of extreme rainfall amounts (Fowler 2003). Nonetheless, they do of course overlap with this study so we will include a brief discussion in our HESS revision.

(5) "1.2. The section about the subjective classification is a bit too short, or, one could say, subjective. There are many open questions: -why do the authors introduce this subjective scheme, when there are others available?"

As motivated in the introduction, this work reports an original study aimed at classifying extreme events by spatial layout, meteorological mechanism, and seasonal occurrence on the basis of a novel historical archive, for practical flood risk applications. As cited in

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



the paper, perhaps the closest is reported in Hand et al. (2004), yet this study was (a) based on a restricted number of events (50 events), (b) did not use the new archive data (c) based on meteorological conditions alone ignoring spatial rainfall pattern, and (d) covered events ranging in duration from minutes to days, rather than just on daily rainfall patterns. Thus, to our knowledge, there are no other directly comparable schemes. The only other studies that analyse UK rainfall are climate change studies, and are not based on spatial layout or meteorological mechanism e.g. various categories of rainfall amount (Maraun 2008, Osborn 2000), extreme amount indices (Mober 2005), and in changing parameters of EVT models of extreme rainfall amounts (Fowler 2003).

(6) "-what makes this scheme better, distinct? -does it perform better? Is it evaluated? - what are the criteria for the selection? And how do they enter?"

Since there are no directly comparable studies, we are presenting, for the first time, novel subjective and objective schemes, that are compared in order to provide some indication of the value of both schemes. The main distinction is that our scheme is based purely on the information provided by the British Rainfall archive and not only considers the meteorological mechanisms which caused the rainfall but the spatial distribution of the rainfall. This paper is essentially the evaluation of this scheme. The criteria for selection are the shape, location, magnitude, duration, and information provided in the archive from eye-witness accounts such as the presence of thunder. We will make these points clearer in the HESS revision.

(7) "Without such a discussion, the authors show only 5 nice examples (Figs. 1-5) without justifying the new scheme nor evaluating it. I would urge the authors to expand on this issue, otherwise I would ask to delete the whole subjective scheme. My personal preference would be to delete the whole part about the subjective scheme and publish it in a separate paper (then of course, the objective scheme would have to be compared against an existing scheme)."

Indeed, the focus of the paper is to compare a subjective and objective scheme for

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



classifying UK extreme rainfalls, since this is a novel study with no directly comparable schemes. We will therefore make these points about the justification and evaluation of the subjective scheme more precise in the HESS revision.

(8) "1.3. The authors spend too little time on explaining their methods. They should keep in mind that not all HESS readers might be familiar with cluster algorithms, likelihoods, and Bayesian information criteria. So just writing down the maximum likelihood estimator (and shortly explaining the concept of likelihood) for the cluster variance given a chosen number of clusters, and then introducing the BIC in its general form and the specific form for this case, would help to understand the approach. Some minor comments on this will follow below. Following the recommendations of the other referee, this part will have to be rewritten, and the authors should take my comment as a further recommendation."

We agree that some more technical detail about the statistical methods would be helpful for readers unfamiliar with the statistical concepts, so we will include more detail in the HESS version.

(9) "2. I have got some points related to the slightly overblown language of the paper. I will discuss two points in particular. 2.1. I am wondering, if the point, that the set of possible layouts of extreme rainfall events is exceedingly large, but only a small number of such events have been sampled, is really a manifestation of the curse of dimensionality. The actual idea behind this term is that hypervolumes increase exponentially with dimension. For data analysis, this means the following: When in a one-dimensional setting, N data points are enough for an accurate estimate, you would need N^2 in a two dimensional setting, N^3 in a three dimensional and so on. In the case described here, this does not apply. It is true that there is an infinite number of possible (spatial) shapes of extreme events, and only a small number has been observed. But by attempting to reduce the dimensionality (which is the key point of the paper), the authors imply that the high dimensionality is just a matter of climate noise, and that the effective dimensionality of the system is much lower. This is just the case in any, e.g.,

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



one dimensional system, where you assume that y is a function of x plus some noise - but this is trivial and not an example of the curse of dimensionality. Therefore, I would ask the authors to rather delete this bit."

This is an interesting point. We agree with the reviewer that the curse of dimensionality refers to the exponential increase in hypervolume caused by increasing the degrees of freedom (dimension) of the problem, so that an exponentially expanding number of samples are required to uniformly populate that hypervolume. If we understand the reviewer's argument correctly, it seems that the reviewer has possibly underestimated the broad scope of this idea: it does not only apply to regression problems (i.e. finding the parameters of a functional model with noise terms), but in fact to any general modelling problem where a finite number of samples must be used to populate the increasing hypervolume (Hastie et al., 2001). For example, consider finding the shape of a multivariate distribution, or just finding the modes in that distribution; in fact our clustering technique can be given that latter interpretation, this is described in detail in, for example Hastie et al. (2001) (cited in the paper). Our clustering technique is not a regression problem. Thus we maintain that the curse of dimensionality is the right concept to invoke here, but, to help justify this assertion, we will, in the HESS revision, include a description justifying the introduction of this concept on basis of the argument given here.

(10) "2.2. A similar case is the authors *reductio ad absurdum*. It is a trivial case in statistical modelling that one can describe N data points best with an N parameter model - to the cost of losing all predictive power. Please keep simple things simple."

Of course, we agree with the reviewer that this is a valid point about an N parameter model representing N data points with zero error, at the likely cost of losing generality in the model. But removing any mention of this idea might hinder the accessibility of the description of the techniques to a wider audience unfamiliar with general practices of statistical modelling. Therefore, in the HESS revision, we will revise the description of the concept along the lines proposed here by the reviewer.

S2333

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



(11) "3. How do the authors derive the threshold of 50mm? This is neither linked to quantiles nor to hydrological relevance"

This is actually explained in the paper. We would refer the reviewer to the last paragraph of 3037, here we quote in full:

The events in this record represent one choice of exceedance threshold 8211; 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.

The hydrological importance of such extremely rare events greater than 50mm derives from the resulting flood events cited in detail in the rest of the paper, i.e. please see the discussion and citations in the first few paragraphs of page 3039.

(12) "Is is also not clear if one gauge is representative of the whole grid cell, especially for small scale events (thunderstorms) and small scale orography (e.g. Scottish Highlands). The authors should spent some words on motivating this approach."

We should mention that no grid cell, across the entire country, is represented by one gauge alone. We therefore agree with the reviewer that we should include motivating discussion about this point in the HESS version.

(13) "4. I am not very much convinced by the verification of the objective scheme. First of all, it is based on an unverified subjective scheme (see above), so the verification has not much value."

It is unclear what could be meant by verification in the context of a subjective study, save perhaps for comparison between expert opinion about classification of rainfall patterns. But this is somewhat beside the point: the goal here is to provide a manual reference classification by which to "sanity check" the output from the automated computational scheme. We do not expect the manual scheme to be verifiable in a

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



quantitative sense, but it is based on sound expert observational knowledge encoded in the novel British Rainfall archive. We propose to expand on this point in the HESS version.

(14) "Second, it is very subjective, as the authors themselves state."

It may be that our choice of the term "subjective" is somewhat unfortunate: perhaps "manual" or "expert opinion" might be more appropriate. Whilst the manual scheme is subjective, all schemes, whether based on data computation or manual analysis involve some degree of choice and hence ambiguity: a computational technique must be encoded as an algorithm and this entails mathematical or algorithmic assumptions. The distinction is more one of degree rather than being an absolute one, but we chose these contrasting terms to simplify the discussion. It may be that we will need to be clearer in the HESS revision. For example, whilst the objective scheme will always produce the same quantitative result given the same quantitative rainfall input, it is "blind" to meteorological conditions that are obvious to any trained expert. Similarly, whilst the expert manual classification represents a powerful synthesis of hydrometeorological expertise, historical observational knowledge in the archive, and visual classification from map data, some classification ambiguities will inevitably arise, particularly when the meteorological conditions are indistinct.

(15) "I am in particular concerned by the "Depression" and "East coast" types: According to the observations used by the authors, cluster (e) should be linked to the Depression, and cluster (a) to the East coast type. However, the authors simply swap the two based on the "spatial layout" (page 3046). These two cases are quite clear from the conditional probabilities, so simply swapping them is quite critical. This either suggests that either the cluster algorithm produces unphysical spatial patterns, or that the subjective scheme produces erroneous results. In either case this needs more investigation by the authors. This also suggests a more modest conclusion of the authors, who state that "the objective scheme can be readily interpreted in terms of known meteorological mechanisms" (page 3048)."

S2335

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



This is a good observation that needs some clarification. Firstly, we should make it clear that it is not just the spatial layout that leads us to propose the particular matching of objective to subjective scheme: it is combined reasoning based on the absolute frequency of classes in each scheme, and the seasonal occurrences. We should also make it clear that there is some inevitable ambiguity in meteorological mechanisms; for example, in the duration of one day there can be one or more distinct meteorological causes of extreme rainfall in any one region (thunderstorms sometimes co-occur with mesoscale convective instabilities, for example). Thus, any subjective choice will necessarily have to lump these ambiguities into one category. On such ambiguous days, the subjective and objective schemes may well differ. Thus, a mapping between schemes should not be based on maximising conditional probabilities of occurrences alone, but should take into account other information such as spatial layout, seasonal occurrence, absolute frequencies and any other information available in the rainfall archive. This is indeed the approach we have taken in constructing this mapping: it is not a purely algorithmic process. We will provide a detailed investigation of these issues in the HESS revision.

(16) "page 3034 The sentence "Too little water..." sounds a bit odd. page 3037 The authors should explicitly state that the 257 rainfall events are extreme events page 3043 The authors should explain that $M=132$ stems actually from 11×12 . This whole paragraph could actually be rendered more precisely and shorter at the same time. page 2044 B. I. CriteriON, not Criteria."

We are grateful to the reviewer for pointing out these detailed corrections, which will be made to the HESS version.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 5, 3033, 2008.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

