

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

# Classification procedures in the context of PUB: ways forward?

**R. T. Clarke**

Instituto de Pesquisas Hidraulicas, Universidade Federal do Rio Grande do Sul,  
Porto Alegre, RS, Brazil

Received: 10 January 2011 – Accepted: 11 January 2011 – Published: 21 January 2011

Correspondence to: R. T. Clarke (clarke@iph.ufrgs.br)

Published by Copernicus Publications on behalf of the European Geosciences Union.

**HESSD**

8, 855–867, 2011

**Classification  
procedures in the  
context of PUB: ways  
forward?**

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## Abstract

Limitations of cluster analysis as a procedure for classifying parameters of rainfall-runoff models are discussed, and a procedure is suggested by which such parameters could be estimated, using site characteristics together with a split record test as a measure of performance. It is suggested that geostatistical models may be a possible alternative to procedures based on cluster analysis, and that long-established principles of experimental design (replication, randomization) be used for comparing alternative PUB procedures.

## 1 Introduction

In the context of PUB, “classification” is taken here to refer to a process in which a set  $N$  of river basins, each with a good streamflow record, are put into  $R$  groups (“clusters”) containing  $N_1, N_2 \dots N_R$  basins, which are believed to be similar in terms of variables (“site characteristics”) describing climate, topography, vegetation and geology. The site characteristics of each pair of basins are used to calculate an index of similarity between them; basins with similarities close to each other are assumed to constitute a cluster (or “region”, although the basins need not be contiguous). Basins falling in a given cluster are regarded as “homogeneous”, despite the fact that site characteristics usually vary continuously. Measures of hydrological response (say  $M_1, M_2 \dots M_P$ ) of basins allocated to the  $i$ -th cluster  $G_i$  are calculated, and are combined to give pooled values  $M_1^*, M_2^* \dots M_P^*$  for that cluster. Then given an ungauged basin whose site characteristics of climate, topography, vegetation and geology suggest that it belongs to  $G_i$ ,  $M_1^*, M_2^* \dots M_P^*$  are used as measures of hydrological response for the ungauged basin, possibly with corrections by scaling factors. The measures of hydrological response  $M_1, M_2 \dots M_P$  may be calculated directly from the flow record of each gauged basin (e.g.,  $L$ -moments of the sequence of annual maximum discharges) or indirectly (e.g., as estimates of parameters in a fitted rainfall-runoff model). In the context of

HESSD

8, 855–867, 2011

### Classification procedures in the context of PUB: ways forward?

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

defining regions to estimate hydrological extremes, Chapter 4 of the book by Hosking and Wallis (1997) gives a full account of how cluster analysis should be used, what alternatives exist, and what are their limitations. They conclude that cluster analysis of site characteristics is the most practical method of forming regions from large data sets. They also provide a measure  $H$  of the heterogeneity between the  $R$  clusters, calculated from weighted estimates of  $L$ -moments within each group, and a test of significance based on the size of  $H$ : a region (i.e., the basins within a cluster) is regarded as “acceptably homogeneous” if  $H < 1$ , “possibly heterogeneous” if  $1 \leq H < 2$ , and definitely heterogeneous” if  $H \geq 2$ .

The Hosking-Wallis statistic  $H$  refers only to the specific case where  $L$ -moments are calculated from serially-independent annual extreme values in flow records from the  $N$  sites. In the PUB context, quantities other than  $L$ -moments must be estimated at ungauged sites, a particular example being (say) estimates of the parameters of a rainfall-runoff model fitted to mean daily flows, or quantiles of daily flow. Adaptation of the Hosking-Wallis procedures is then necessary; alternatively, a different approach not based on cluster analysis might be considered, and this paper discusses such issues.

## 2 Limitations of cluster analysis

It is necessary from time to time to take a fresh look at any standard procedure in common use both to remind practitioners of its limitations and to discuss whether research has identified new procedures that might serve as alternatives to the standard.

Cluster analysis is essentially a tool for data exploration: one which, on the basis of similarities between site characteristics, suggests groupings of sites which may or may not have physical interpretation. Where no physical interpretation is possible, the results might suggest that further research, perhaps with the collection of new or different observations, is needed to explain why such clusters have been detected. It is also true to say that not all statisticians hold cluster analysis in high regard: an important review article in the Journal of the Royal Statistical Society by Cormack (1971), although now

**HESSD**

8, 855–867, 2011

---

### Classification procedures in the context of PUB: ways forward?

R. T. Clarke

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



some forty years old, is severely sceptical. The first sentence of his paper said “The availability of computer packages of classification techniques has led to the waste of more valuable scientific time than any other “statistical” innovation (with the possible exception of multiple-regression techniques)” (quoted by Chatfield and Collins, 1981).

5 Stern and Gallagher (2004), although less negative, recommend caution in the use and interpretation of cluster analysis, noting that problems arise for a number of reasons, including (i) the choice of measure adopted as an expression of the similarity between units (basins, in the PUB context); (ii) the occurrence of spurious clusters; (iii) the absence of any optimal clustering method covering all situations; (iv) the fact that different  
10 clustering methods may lead to quite different “solutions”; (v) the groups obtained by cluster analysis are sensitive to the nature of the site characteristics used; (vi) analysis of the results of cluster analysis is subjective. Hosking and Wallis (1997) recognized this element of subjectivity, noting that the output from a cluster analysis need not, and usually should not be final; they note that subjective adjustments can often be found  
15 to improve the physical coherence of the groups and to reduce their heterogeneity as measured by the test statistic  $H$  mentioned above. They note that the kinds of subjective adjustment that are often useful include moving a site from one group to another; deletion of a site or a few sites from the data set; subdivision of a geographical region; breaking up a geographical region by reassigning its sites to other regions; merging the  
20 geographical region with another or others; merging two or more geographical regions for a new cluster analysis; obtaining more data and redoing the cluster analysis.

Thus whilst cluster analysis has been demonstrated (Hosking and Wallis, 1997) to be the preferred tool for delineating homogeneous groups of sites for regionalizing hydrological extremes, it does not necessarily follow that cluster analysis is appropriate  
25 for all other applications in the PUB context.

## HESSD

8, 855–867, 2011

### Classification procedures in the context of PUB: ways forward?

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



### 3 Data analysis for suggesting hypotheses, and testing them

As in any other branch of science, PUB methodology must advance by the formulation of hypotheses set up to explain what has been observed. When a hypothesis has been proposed, an “experiment” is planned which will commonly involve the collection of new observations according to a carefully defined procedure, or which may utilize other relevant data that were not used to formulate the hypothesis being tested; in either case, the analyst seeks to falsify the hypothesis by comparing what is predicted when the hypothesis is adopted, and what is shown by the new (or hitherto unused) observations. If the agreement between hypothesis and prediction is satisfactory according to some criterion specified before the new observations were collected, the proposed hypothesis is retained, until such time as further evidence come along to falsify it. Thus it is important to distinguish between (a) the data that are used to suggest a hypothesis (which might arise, for example, from results of a cluster analysis) and (b) the data used to test the hypothesis. Splitting records when evaluating the performance of rainfall-runoff models, and of testing such models on new data-sets (Clarke, 2008) is evidence of the awareness of a need for such practice.

An analogy can be drawn between the use of cluster analysis as a purely graphical, descriptive procedure for displaying similarities between basins, and the construction of a histogram showing the frequencies of observations falling into different classes. Site characteristics of basins are, for the most part, observations of continuous variables, and cluster analysis using  $M$  site characteristics defines regions of the  $M$ -dimensional space in which the classified objects (basins) are similar, just as a histogram in (say) two-dimensional space defines class-intervals into which data values can be grouped and displayed for pictorial purposes. And just as a histogram is a graphical display of features in data such as position, dispersion and skewness, so cluster analysis with its dendrograms and minimum spanning trees is best considered as a graphical representation of similarities. The many procedures for defining clusters have their analogies

HESSD

8, 855–867, 2011

#### Classification procedures in the context of PUB: ways forward?

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



in rules such as those of Sturges (1926), Scott (1970), and Freedman and Dioco-  
nis (1981) for defining the widths of histogram class-intervals.

Bearing in mind the need to treat cluster analysis with caution, and the need to  
distinguish hypothesis formulation from hypothesis testing, we consider the particu-  
lar case where the  $P$  parameters of a rainfall-runoff model have been estimated from  
good records at each of  $N$  basins, together with the site characteristics of each. It is  
required to obtain values for these parameters for use at an ungauged site with known  
site characteristics, and for which rainfall is also known from a within-basin rain gauge  
network and/or satellite-derived estimates, such as those from TRMM (Huffman et al.,  
2007) and CMORPH (Joyce et al., 2004), giving 3-hourly estimates for grid-squares of  
magnitude  $0.25^\circ \times 0.25^\circ$ ). Then, based on cluster analysis, the following procedure is  
one possible approach.

- *Step 1:* Divide the  $N$  basins, strictly at random, into two groups, equal or ap-  
proximately equal in size, denoted by  $N_+$  and  $N_-$ . The  $N_+$  are used to define the  
clusters (or regions), and the  $N_-$  are used for verification purposes (it is assumed  
that the total  $N$  is sufficiently large for this purpose).
- *Step 2:* Using the site characteristics, denoted by  $X_1, X_2 \dots X_S$ , of the  $N_+$  basins,  
do the cluster analysis to define  $R$  clusters containing  $N_1, N_2 \dots N_R$  basins,  
 $N_1 + N_2 + \dots + N_R = N_+$ .
- *Step 3:* Denoting the  $P$  parameters of the  $j$ -th model in the  $i$ -th cluster by  $\theta_{ij,1},$   
 $\theta_{ij,2}, \dots \theta_{ij,P}$ , calculate their mean values over the  $N_i$  basins in this cluster, de-  
noted by  $\bar{\theta}_{i,1}, \bar{\theta}_{i,2} \dots \bar{\theta}_{i,P}$ . Also, calculate the means of the site characteristics,  
denoted by  $\bar{X}_{i,1}, \bar{X}_{i,2}, \dots \bar{X}_{i,S}$ .
- *Step 4:* Now take the first basin (“Basin One”) in the “validation” group of  $N_-$   
basins. Calculate the Euclidean distance between its site characteristics  $X_1,$   
 $X_2 \dots X_S$ , and each of the mean site characteristics  $\bar{X}_{i,1}, \bar{X}_{i,2} \dots \bar{X}_{i,S}, i = 1 \dots R$ .

**Classification  
procedures in the  
context of PUB: ways  
forward?**

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Thus for the  $i$ -th cluster, the squared distance is

$$d_i^2 = \sum_{j=1}^S \{X_j - X_{i,j}\}^2.$$

– *Step 5:* Identify the cluster, say cluster  $l$ , for which the  $d_i^2$  is a minimum. Thus the site characteristics of Basin One suggest that it should belong to the  $l$ -th cluster. However, as well as identifying which cluster it should belong to, it is also necessary to check that the parameters of Basin One are also closer to the mean parameters  $\bar{\theta}_{l,1}, \bar{\theta}_{l,2} \dots \bar{\theta}_{l,P}$  of this  $l$ -th cluster, than to the mean parameters of any other cluster. Hence: –

– *Step 6:* Calculate, for each of the  $R$  clusters, the squared distance  $\delta_i^2$  given by

$$\delta_i^2 = \sum_{j=1}^P \{\theta_{j,1} - \bar{\theta}_{i,j}\}^2$$

where  $\theta_{1,1}, \theta_{2,1}, \dots, \theta_{P,1}$  are the  $P$  parameters of this first of the  $N_-$  validation basins. Identify the cluster,  $J$  say, for which  $\delta_i^2$  is a minimum;  $\delta_i^2$  is the Euclidian distance between the point in  $P$ -dimensional space defined by the parameters of Basin One and the point defined by the mean parameter values of those basins from the  $N_+$  group that have been shown by the cluster analysis to make up the  $i$ -th cluster.

– *Step 7:* If  $l = J$  (so that the site characteristics of Basin One both suggest that it belongs to the  $l$ -th cluster, and that this is confirmed by the closeness of its  $P$  parameters to this cluster's mean parameter values), record a “Success” for this basin.

## HESSD

8, 855–867, 2011

### Classification procedures in the context of PUB: ways forward?

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

---

## Classification procedures in the context of PUB: ways forward?

R. T. Clarke

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

- *Step 8*: Repeat Steps 4 to 7 for each of the remaining  $N_- - 1$  basins, Basin Two, Basin Three... in the validation group, and record the proportion of successes  $\pi_0 = (\text{number of successes}/N_-)$ . If the clusters are well defined, the value of  $\pi_0$  should be close to one, so the following steps are to establish whether  $\pi_0$  is sufficiently close to one for it to be concluded that the cluster analysis is useful.
- *Step 9*: Redistribute the  $N_+$  basins, at random, amongst the  $R$  clusters, so that there are  $N_1, N_2, \dots, N_R$  (different) basins in them.
- *Step 10*: Cycle through Steps 3 to 8, say  $Q = 1000$  times, recording the proportions of successes  $\pi_1, \pi_2, \pi_3 \dots \pi_Q$ . Arrange these values in ascending order and calculate the quantiles of this sequence. If the observed proportion  $\pi_0$  lies in the upper tail of the frequency distribution of the  $\pi_1, \pi_2, \pi_3 \dots \pi_Q$ , conclude that the clusters defined by the  $N_+$  basins are useful for the purpose of extrapolation to ungauged basins.

Clearly, the procedure defined by the above steps can be repeated by interchanging the two sets of  $N_+$  and  $N_-$  basins, and indeed by using the combined set  $N = N_+ + N_-$ . Furthermore, just as in a cluster analysis there are very many measures of similarity between basins, so there are many possible variants of the above procedure: perhaps using median parameter values instead of mean parameter values in each cluster; or using weights to allow for the fact that some clusters may be quite small, and others much larger; or scaling the parameter values to ensure that contribute more or less equally to cluster definition; or calculating the means of distances, instead of distances between means. It can be argued that each parameter of the rainfall-runoff model could be treated separately, with its own cluster analysis (although this would not take account of the fact that estimates of model parameters calculated from each site's record will be highly correlated).

Exploring these many alternatives and variants could lead to a great deal of probably inconclusive research, so that the question must be asked “are methods based on cluster analysis the only approach to predicting parameters of rainfall-runoff models at



5 ungauged sites?”. At this point, we recall that cluster analysis takes no account of the spatial correlation between quantities recorded or estimated at different sites. For the particular case of estimating an index flood at an ungauged site, linear models based on work by Stedinger and Tasker (1985, 1986) have been discussed by Hosking and Wallis (1997, Chapter 8) which can take spatial correlation into account, and which can allow for spatial trends by using latitude and longitude as site characteristics. The next section discusses the possibility that advances in geostatistical modelling which build upon such models may provide an alternative, if not a full substitute, for conventional approaches through cluster analysis for PUB.

#### 10 4 Geostatistical models for spatial interpolation

15 Geostatistical methods for spatial interpolation have a long history (e.g., Cressie, 1993, and references therein), and in some cases prediction of rainfall-runoff model parameters at ungauged sites can be regarded as an application of spatial interpolation. Just as it is necessary when using cluster analysis to have records from enough sites to allow clusters to be identified, so it is necessary when using geostatistical methods to have enough records (with model parameters estimated from them) to allow spatial correlation structures to be identified: if necessary allowing for anisotropy of correlation. Similarly, a decision is required about whether a separate geostatistical model is to be used to interpolate each rainfall-runoff model parameter separately, or whether a multivariate model such as those discussed by Diggle and Ribeiro (2007, Chapter 3) is to be used. Provided that it is only estimates of the rainfall-model parameters at an ungauged site that are required, and not their variances or confidence regions, a separate geostatistical model for each rainfall-runoff model parameter should be adequate.

25 In this author’s opinion, the geostatistical models developed by Diggle and Ribeiro (2007) and their collaborators (Diggle et al., 1998, 2003) offer an attractive alternative to classification by means of cluster analytical methods for the prediction of rainfall-runoff model parameters at ungauged basins, for the following reasons.

---

**Classification procedures in the context of PUB: ways forward?**

R. T. Clarke

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



First, they recognize explicitly that the hydrological model parameters will exhibit both spatial trend and spatial correlation, both of which are quantified by parametric structures that can be efficiently estimated. These estimation procedures can also be used for selecting those site characteristics which are best for predictive purposes.

Second, geostatistical models for spatial interpolation include not only the familiar linear models, with or without correlated observations, but also generalized linear geostatistical models (GLGMs) discussed by Diggle and Ribeiro (2007). With some notable exceptions (Segond et al., 2006; Yang et al., 2005; Chandler, 2005; Chandler and Wheater, 2002), generalized linear models are not yet part of the familiar tool-kit of practitioners of the hydrological sciences but deserve to be used more widely, where appropriate. Also to the author's knowledge, there have been no reports of the use of GLGMs for spatial interpolation in the PUB context. Powerful software, geoR and geoRglm, is available and can be downloaded free for the analysis of geostatistical data using GLGMs (Ribeiro and Diggle, 2006).

Third, Bayesian methods of model-fitting and spatial interpolation extend the utility of geostatistical models and have the advantage that, contrary to what happens when basins are classified by cluster analysis, any personal subjectivity can be quantified, reported and modified, through the necessary specification of prior distributions for model parameters. Diggle and Ribeiro (2007) give an interesting example, having affinities with PUB, of the use of Bayesian methods with a geostatistical model to map the probabilities, over a fine grid, that Swiss rainfall exceeded 25 mm on 8 May 1986.

## 5 Conclusions

It is clear that many procedures can be suggested by which rainfall-runoff parameters could be estimated for ungauged basins. All will have limitations, particularly where PUB applications require spatial extrapolation and not spatial interpolation, as where the available flow records are from downstream basins, and the problem is to extrapolate to those upstream. Some alternative procedures are simply nested forms of more

## Classification procedures in the context of PUB: ways forward?

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



---

**Classification  
procedures in the  
context of PUB: ways  
forward?**R. T. Clarke

---

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

5 general ones, as where it is required to test whether certain site characteristics contribute significantly to the precision of predictions. Other alternatives are of quite different form, as where predictions given by cluster analysis are compared with predictions by GLGMS. So it is likely that much research effort will be directed to the comparison of  
10 alternatives, and this effort will be applied most efficiently if comparisons are planned according to established principles of statistical experimental design. Whilst very many statistical methods in common use were initially developed by Fisher (1925) and those who followed him at Rothamsted Experimental Station, Fisher's seminal contributions (Fisher, 1935) to experimental design (replication of "treatments" by applying them to  
15 several "experimental units"; allocation of "treatments" to "experimental units" at random) are almost unknown in the water sciences (Clarke, 2008). Yet the comparison of alternative methods for prediction in ungauged basins can be considered as an experimental design in which the "treatments" are the models or procedures to be compared, and the flow records to which they are fitted or applied are the "experimental units".  
20 Measures of model performance become the variates used in Fisher's Analysis of Variance (ANOVA) appropriate to the particular experimental design used; and where the number of reliable flow records is limited, standard experimental designs (e.g., Cochran and Cox, 1950) are available for making most efficient use of them. If the PUB initiative were able to plan and implement procedures for PUB in such a co-ordinated manner, this would be a major contribution to the development of water sciences.

## References

- Chandler, R. E.: On the use of generalized linear models for interpreting climate variability, *Environmetrics*, 16(7), 699–715, 2005.
- Chandler, R. E. and Wheeler, H. S.: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, *Water Resour. Res.*, 38, 1192–1203, ISSN:0043-1397, 2002.
- Chatfield, C. and Collins, A. J.: *Introduction to Multivariate Analysis*, Taylor & Francis Ltd Chapman & Hall/CRC, London, 1981.
- Clarke, R. T.: Issues of experimental design for comparing the performance of hydrologic models, *Water Resour. Res.*, 44, W01409, doi:10.1029/2007WR005927, 2008.
- Cochran, W. G. and Cox, G. M.: *Experimental designs*, ISBN 0-471-54567-8, 1950.
- Cormack, R. M.: A review of classification, *J. R. Stat. Soc. A*, 134, 321–367, 1971.
- Cressie, N.: *Statistics for Spatial Data*, John Wiley & Sons, New York, USA, 1993.
- Diggle, P. J. and Ribeiro Jr., P. J.: *Model-based Geostatistics*, Springer Series in Statistics, New York, 228 pages, 2007.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A.: Model based geostatistics (with discussion), *Appl. Stat.*, 47, 299–350, 1998.
- Diggle, P. J., Ribeiro Jr., P. J., and Christensen, O. F.: *An introduction to model-based geostatistics*, J., Springer, New York, 43–86, 2003.
- Fisher, R. A.: *Statistical Methods for Research Workers*, ISBN 0-05-002170-2, 1925.
- Fisher, R. A.: *The design of experiments*, ISBN 0-02-844690-9, 1935.
- Freedman, D. and Diaconis, P.: On the histogram as a density estimator: L2 theory, *Zeit. Wahr. ver. Geb.*, 57, 453–476, 1981.
- Hosking, J. R. M. and Wallis, J. R.: *Regional Frequency Analysis: An approach based on L-moments*, CUP, 1997.
- Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Hong, Y., Stocker, E., and Wolff, D.: The TRMM Multisatellite Precipitation Analysis (TCMA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, 8, 38–55, 2007.
- Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, *J. Hydrometeorol.*, 5, 487–503, 2004.

# HESSD

8, 855–867, 2011

**Classification  
procedures in the  
context of PUB: ways  
forward?**

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Ribeiro Jr., P. J. and Diggle, P. G.: geoR: Package for Geostatistical Data Analysis, R-NEWS, ISSN 1609-3631, 1(2), 15–18, 2006.

Scott, D. W.: On optimal and data-based histograms, Biometrika, 66, 605–610, 1979.

Segond, M. L., Onof, C., and Wheeler, H. S.: Spatial-temporal disaggregation of daily rainfall from a generalized linear model, J. Hydrol., ISSN 0022-1694, 331, 674–689, 2006.

Stern, R. and Gallagher, J.: Analysing Climatic Data Using DenStat for Windows, Statistical Services Centre, The University of Reading/The UK Met Office, ISBN 0-7049-9829-7, 2004.

Sturges, H.: The choice of a class-interval, J. Am. Stat. Assoc., 21, 65–66, 1926.

Yang, C., Chandler, R. E., Isham, V., and Wheeler, H. S.: Spatial-temporal rainfall simulation using Generalized Linear Models, Water Resour. Res., 41, W11415, doi:10.1029/2004WR003739, 2005.

# HESSD

8, 855–867, 2011

## Classification procedures in the context of PUB: ways forward?

R. T. Clarke

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

