Flood Risk Management: Research and Practice – Samuels et al. (eds)

Inter-site dependence in extremes: Unlocking extra information

Duncan W. Reed
DWRconsult, Cholsey, Oxfordshire, UK

ABSTRACT: Where long records have been gathered at a relevant site, and are thought to represent a near-stationary process, their statistical analysis is pivotal to environmental risk estimation. In other situations, pooling data from nearby or similar sites provides a pragmatic way of stretching estimates to the long return period typically required. A well-constructed pooling scheme promotes site-to-site consistency in design values, i.e. estimates of T-year extreme values. This fits well with the "we estimate for all sites" mindset now fashionable. The paper explores the concept that the study of inter-site dependence yields valuable extra information. In this context, dependence is the tendency for extremes at neighbouring sites to occur together. Rank correlation in annual maxima provides a simple measure of pairwise dependence in extreme values at different sites. Results are presented from a number of studies of dependence based on flood data taken from http://www.environment-agency.gov.uk/hiflowsuk/. These include case examples of the Spey, the Yorkshire Ouse, three rivers in south-east Wales and nine urbanised catchments in the London area. Sites are mapped by multi-dimensional scaling so that inter-site distance reflects the degree of independence in their extremes. The paper concludes that dependence analysis can inform a range of applications.

1 INTRODUCTION

Correlation is an association between variables. The word *dependence* carries a similar but less rigid meaning, which is helpful when trying to think more widely about relationships and behaviours.

1.1 *Review*

The study of inter-site dependence in extremes has typically been driven by application. Fricke et al. (1983) assess the vulnerability of a mineral railway to flood-induced failure along its route. Dales & Reed (1989) study inter-site dependence across networks of sites to gain understanding of the clustering in space and time to be expected of extreme rainfalls of a kind to test dam safety. Stewart et al. (1999) present a wider review.

Inter-site dependence in extremes might have rooted more securely in hydrological research had studies continued in *risk appreciation* mould. By exploiting the Dales & Reed model in the derivation and extension of rainfall growth curves, Reed & Stewart (1989) challenged the conventional wisdom (e.g. Hosking & Wallis, 1988) that inter-site dependence in extremes is more problem than opportunity. The ensuing FORGEX method (Reed et al., 1999) remains controversial (e.g. Cox, 2003).

This paper opts for the gentler terrain of reporting studies of inter-site dependence in flooding. The aim is to highlight the scope to unlock extra information.

1.2 *Content*

Inter-site dependence in river flooding is more complicated than inter-site dependence in rainfall extremes. It reflects spatial dependence in rainfall (and snowmelt), spatial dependence in antecedent conditions, spatial variation in catchment properties, and, of course, drainage-path structure.

The measure of dependence used here is the rank correlation, ρ_{mag}, of annual maximum floods. Reed (2002) explains that using annual maxima to assess dependence is not solely a matter of convenience, and advocates the use of consensus measures of dependence. However, ρ_{mag} is a simple and relatively robust measure, e.g. it is largely insensitive to systematic error in rating curves. Robustness is relevant when, as here, many station-pairs are being analysed.

The paper considers:

1. Exploratory analysis of dependence across 516 British catchments;
2. Generalisation of a model for upstream-downstream dependence;
3. The mapping of dependence.

2 EXPLORATORY ANALYSIS

Initial analysis examined 516 British stations for which HiFlows-UK provides complete annual maximum flood data for the 1981 to 2000 water-years. Of the resulting 132870 station-pairs, the 50 having the highest rank correlation were noted. ρ_{mag} values ranged from 0.98 for Stations 8001 & 8006 on the Spey to 0.93 for Stations 37008 & 37011 on the Chelmer. Unsurprisingly, 35 of these highly dependent stations are upstream-downstream pairs. A further nine are on tributaries confluent further down the river system.

Of the six station-pairs whose waters meet only in the ocean, a notable dependence is that between the the Stour at Hammoon (43009) and the Blackwater at Ower (42014). At face value, the catchments are not especially similar: $dist_{\text{SWS}} = 0.89$ units in the Flood Estimation Handbook (FEH) pooling method (Jakob et al. 1999). Picking out exceptional values in this manner tends to exaggeration. The rank correlation, ρ_{mag}, reduces from 0.93 to a less outstanding 0.75 when the full overlapping record (1976 to 2002 water-years) is consulted. The Stour and Blackwater catchments lie respectively west and east of the Hampshire Avon, with centroids 50 km apart. In 12 out of 27 years, their annual maxima lie within a day of each other.

The most *counter-dependent* station-pair is the Mimram at Panshanger (38003) and the Duddon at Duddon Hall (74001), for which $\rho_{\text{mag}} = -0.75$. Consulting the full overlapping record (1967 to 2002 water-years), the effect is less pronounced with $\rho_{\text{mag}} = -0.34$. Neither flood series has a long-term trend or rating change to suggest that the counter-dependence is artifice. It may reflect north-south counter-dependence in climate, in addition to the obvious catchment differences. The contrasting permeability and natural storage of the catchments are conveniently summarised in the baseflow index: with *BFI* values of 0.93 for the Mimram and 0.29 for the Duddon.

Reed (2002) reported an unexpectedly strong dependence ($\rho_{\text{mag}} = 0.93$) between the Upper Lee at Water Hall (38018) and the Turkey Brook at Albany Park (38021), with 16 of 25 annual maxima occurring within a day of each other. Judged from daily mean flow data, the catchment flow regimes are utterly different, with *BFI* values of 0.81 for the Upper Lee and 0.21 for the Turkey Brook. The unusual hydrogeology in this part of Hertfordshire offers an explanation. Reed (2008) notes that: (i) the principal source of floods at Water Hall is typically rainfall on the *lower* part of its topographic catchment, and (ii) karstic influences—both there and on the Turkey Brook–account for the strong dependence.

3 UPSTREAM-DOWNSTREAM DEPENDENCE

By many criteria, British rivers are densely gauged. Yet the 516 stations provide only 504 upstream-downstream pairs, and some of these are of catchments so different in scale (e.g. 8.7 km^2 Severn at Plynlimon and 4325 km^2 Severn at Bewdley) that one seldom analyses them jointly.

The study of upstream-downstream dependence is under-researched given that it is fundamental to judging how satisfactorily, and how extensively, the "river modelling" approach to flood-risk mapping can be applied.

3.1 *A model of upstream-downstream dependence*

A general model was sought for dependence in upstream-downstream floods based on the approach of Reed (2002). The model:

$$\rho_{\text{mag,mod}} = 0.556 + 0.273\, AREA_{\text{rat}} + 0.449\, SAAR_{\text{rat}} - 0.278\, SPRHOST_{\text{rat}} - 0.00485\, ICD \qquad (1)$$

explains 46.1% of the variation in ρ_{mag} for the 504 upstream-downstream station-pairs. $AREA_{\text{rat}}$, $SAAR_{\text{rat}}$ and $SPRHOST_{\text{rat}}$ are ratios of the relevant FEH catchment descriptors at upstream and downstream sites: the ratio in each case being calculated as the smaller value divided by the larger. [*AREA* is drainage area, *SAAR* is average annual rainfall and *SPRHOST* is standard percentage runoff estimated from soil maps.] *ICD* is inter-centroid distance in km between upstream and downstream catchments.

The model represents the greater dependence expected between catchments that almost coincide (i.e. as $AREA_{\text{rat}}$ increases and *ICD* reduces). Indeed, Equation 1 precisely meets the criterion $\rho_{\text{mag}} = 1$ when upstream and downstream sites coincide. [With ρ_{mag}, $AREA_{\text{rat}}$, $SAAR_{\text{rat}}$ and $SPRHOST_{\text{rat}}$ each replaced by their complement, the intercept term in the regression was not significantly different from zero.]

Interpretation of the other terms in Equation 1 is less clear-cut. The sign of the $SAAR_{\text{rat}}$ coefficient is consistent with homogeneity in topography and climate promoting greater dependence. [It should be remembered that dependence in flooding reflects spatial dependence in catchment wetness as well as that in heavy rainfall.] But it is curious that the degree of dependence is modelled to be greater when the soils intervening between the upstream and downstream sites contrast with those of the upstream catchment. Perhaps $SPRHOST_{\text{rat}}$ acts as a surrogate.

3.2 *Application*

Equation 1 provides a tool for estimating dependence between upstream-downstream sites where one or neither is gauged. Consider the Ouse catchment to York, which appears later in Figure 3. As will be discussed, direct analysis indicates that there is strong dependence between the Ure and Ouse ($\rho_{\text{mag}} = 0.89$), and between the Swale and Ouse ($\rho_{\text{mag}} = 0.90$). Were a "river modeling" approach to flood-risk estimation adopted, the more problematic decisions may concern treatment of the moderately dependent Nidd ($\rho_{\text{mag}} = 0.75$) and the weakly dependent Foss ($\rho_{\text{mag,mod}} = 0.45$, when assessed using Equation 1 to a site immediately downstream of the Ouse-Foss confluence in York).

4 DEPENDENCE MAPS

Though instructive and analytically precise, studies of pairwise dependence can be difficult to summarise and digest. An attempt has therefore been made to map inter-site dependence in flood extremes.

4.1 *Mapping technique*

The technique adopted is *multi-dimensional scaling* (e.g. Cox & Cox, 2000). This is the method typically used to map towns according to travel time rather than geographical position. The inter-site distance or *dissimilarity* measure used here is: $\delta = 1 - \rho_{\text{mag}}$. Values of δ range from 0 for fully dependent sites, through 1 for independent sites, to 2 for fully counter-dependent sites.

The aim of *metric scaling* is to seek a configuration of points such that the distance d_{ij} between points i and j equals the dissimilarity δ_{ij}. While a 2-dimensional (2-d) map is typically sought, this will provide an adequate representation only if the underlying dimensionality of the property being mapped is about two. In the approach—often called *classical scaling*—the required dimensionality of the representation is indicated by the number of non-zero eigenvalues of the dissimilarity matrix.

4.2 *Examples of dependence mapping*

Results are presented for four groups of stations (see Table 1).

These case studies have been selected as examples of networks for which the dissimilarity matrix has one or two eigenvalues that are dominant, giving scope for presentation as a 2-d map (see Fig. 1).

The dependence map produced by classical scaling (e.g. Fig. 2b) is compared with one in which stations are positioned at their *catchment centroids* (e.g. Fig. 2a). Station-pairs lying upstream/downstream are indicated by straight lines; thicker lines denote the main stem(s).

Table 1. Case studies for dependence mapping.

Short name	Description	Stations	Water-years
SPEY9	Nine subcatchts of Spey, Northern Scotland	8001–8002, 8004–8010	1953–2002
OUSE7	Seven sub-catchments of Ouse to York	27001, 27007, 27009, 27034, 27053, 27059, 27071	1977–2002
TAFF7	Seven catch$^{\text{ts}}$ in Hyd. Area 57, South Wales	57004, 57005, 57006, 57007, 57008, 57009, 57010	1974–2002
LURB9	Nine urbanised catchments in/ near London	37018, 37019, 38007, 39005, 39012, 39052, 39053, 39093, 39095	1974–2002

Absolute geographical positions play no role in the assessment of ρ_{mag} or in the classical scaling. Some of the dependence maps have therefore been rotated and/or flipped to improve visual correspondence.

4.2.1 *Spey*

Fifty years of data are available for nine subcatchments of the Spey. Their configuration is shown in Figure 2a. It should be noted that each catchment is plotted at its geographic centroid *not* at its outlet. The main stem of the Spey runs from Invertruim to Boat o Brig.

The dependence map is presented in Figure 2b. The Dulnain and Avon are seen to be more closely allied to the main stem of the Spey than suggested by geographic position, while the Tromie is more remote (i.e. less dependent). This may reflect the large export of water from the Tromie for power generation, whereas the Dulnain and Avon are largely natural.

4.2.2 *Ouse at York*

The dependence map for the Ouse (Fig. 3b) is especially compact: reflecting the strong dependence between the Ouse, Swale and Ure discussed in Section 3.2. The Nidd is noticeably less dependent than suggested by geographic position (Fig. 3a), perhaps reflecting its much greater reservoir development. The off-centre position of the Laver is to be expected. This much smaller catchment has distinctive Karstic features.

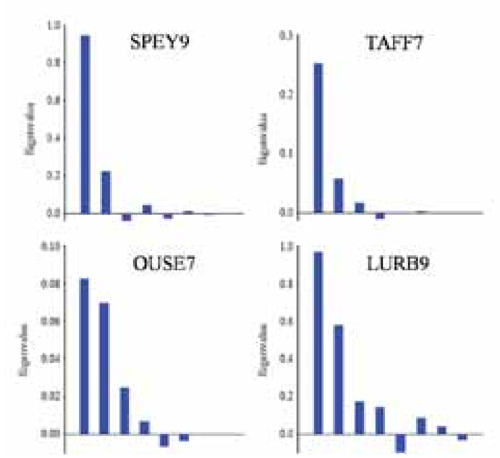

Figure 1. Eigenvalues for case-study networks.

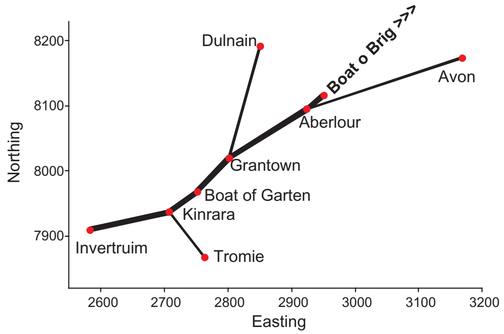

Figure 2a. Subcatchments of the Spey (geographic map).

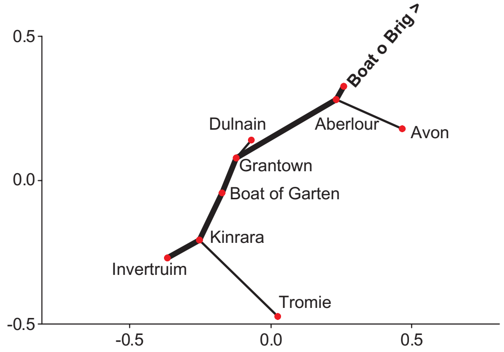

Figure 2b. Subcatchments of the Spey (dependence map).

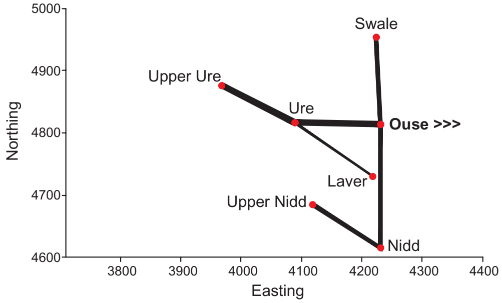

Figure 3a. Subcatchments of the Ouse (geographic map).

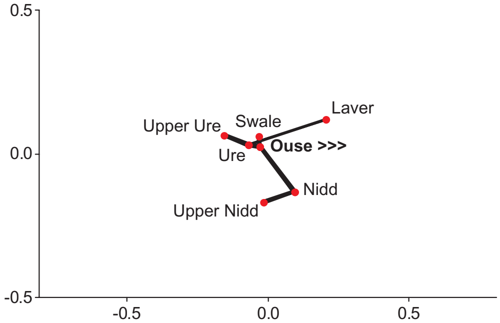

Figure 3b. Subcatchments of the Ouse (dependence map).

4.2.3 *Hydrometric Area 67 (Taff)*

The Taff example (Fig. 4) explores dependence mapping for adjacent as well as nested catchments. The Ely system is confirmed in Figure 4b to be quite different to the Taff. However, the Rhymney is found to be more strongly dependent with the Taff than expected from geographic position alone (Fig. 4a). In this example, it did not prove possible to rotate or flip the dependence map to achieve visual coherence with the geographic map.

A wider study examined dependence across a network of 16 stations in this part of south-east Wales. Although the eigenvalue analysis implied that a 2-d representation might be simplistic, a dependence map (not presented) placed the Taff and Rhymney catchments in a tight grouping with five catchments from the immediate north-west, north, north-east and east. The Ely stations were placed in a looser group with four catchments to the immediate west.

4.2.4 *Heavily urbanised catchments*

The final case-study is of nine heavily urbanised river catchments in/near London (see Fig. 5a). None of these catchments is nested. An appreciation of the

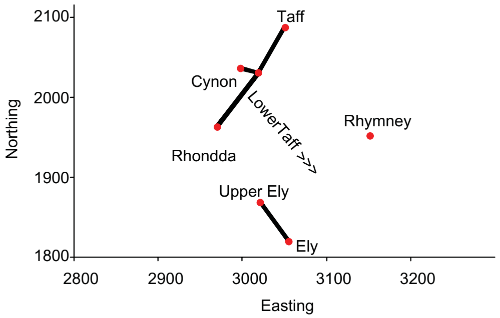

Figure 4a. Catchments in Hydrom. Area 57 (dependence map).

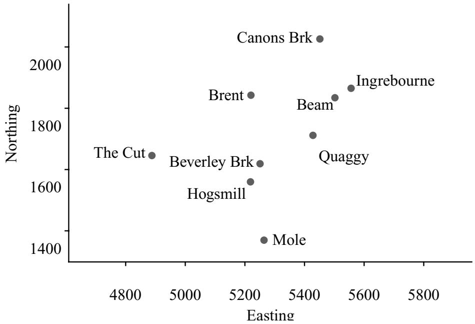

Figure 5a. Urban catchments near London (dependence map).

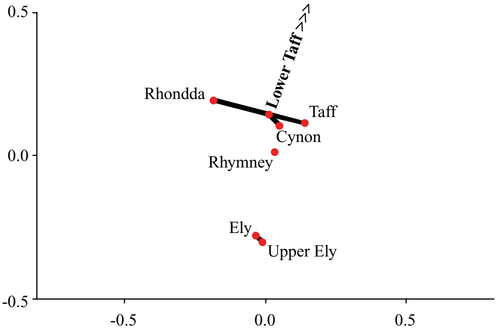

Figure 4b. Catchments in Hydrom. Area 57 (geographic map).

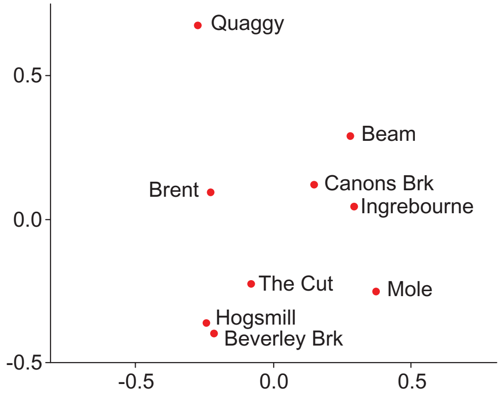

Figure 5b. Urban catchments near London (geographic map).

degree of dependence expected in flooding is crucial to judging the performance of storm-water sewer systems. Typically, many years pass with few problems before major flooding affects many sites within a metropolitan area (Reed, 2007).

From the dependence map (Fig. 5b), the Quaggy is seen to be utterly distinct. Judged by rank correlation of annual maxima, flooding on the Quaggy is independent of that on the nearby Beverley Brook and Hogsmill ($\rho_{\text{mag}} = -0.07$ and -0.05 respectively, based on 29 years of data). The Brent is also seen to be distinctive, presumably through the impact of Brent Reservoir.

A further interesting feature is the extent to which far-flung urbanised catchments—The Cut in Bracknell and the Mole at Horley—are more central in the dependence map.

5 CONCLUDING REMARKS

5.1 *Previously*

Studies of inter-site dependence in extremes have tended to be driven by application: with *risk appreciation* to the fore. Reed (2002) discusses lesser applications such as judging donor suitability (in transfers of hydrological variables from gauged to ungauged sites), reviewing pooling-groups, and selecting stations for flood warning.

5.2 *Bigger aims*

A general model, albeit basic, has been presented of upstream-downstream dependence in British river flood-risk. Such a model could be combined with direct estimates of inter-site dependence (i.e. between gauged sites) to obtain improved estimates of dependence between gauged and ungauged sites. The approach might imitate methods used to transfer estimates of hydrological variables from gauged to ungauged sites.

An approach to general mapping of dependence in flooding has been illustrated. The method works only on relatively small networks of stations and produces a general rather than site-focused map of dependence. However, the mapping technique is relevant to the scale at which some "river modelling" is attempted. Development of a method for producing a *site-focused* map of dependence would nevertheless be very useful.

Particular sites are known to be of prime importance. Sometimes this is a site with long-term or historical flood data: data whose interpretation is pivotal to flood-risk estimation throughout the river system. In other cases, the site is important simply because of the criticality of flooding there.

In the first instance, a site-focused assessment of dependence will help to inform how far up/down the main stem of the river network a particular flood-risk mapping approach might reasonably be applied. Knowledge of the extent to which dependence decays on minor tributaries may help to arrest the excessive subdivision of catchments in "river models" and to promote "continuous simulation" above "design event" approaches.

In the second instance, a site-focused map of dependence will help emergency planners visualise the likely spatial extensiveness of severe flooding.

5.3 *Smaller aims*

The analysis of inter-site dependence provides a way of exploring flood data that is largely free from assumption. By analysing data series collectively, new information is extracted.

The peculiarly strong dependence found between annual maximum floods on the Upper Lee at Water Hall and on the Turkey Brook led (eventually!) to diagnosis of a singular phenomenon. In other circumstances, the analysis of inter-site dependence may draw attention to data entry error and/or to flood peaks that are missing at a particular station.

Aiming big, or aiming small, it seems relevant to find ways of using this extra information.

REFERENCES

Cox, D.R. 2003. Some comments on 10,000 year return period rainfall. Report to Defra. www.defra.gov.uk/environment/water/rs/pdf/coxreport.pdf

Cox, T.F. & Cox, M.A.A. 2000. *Multidimensional scaling*. 2nd edition, Chapman & Hall, 328pp.

Dales, M.Y. & Reed, D.W. 1989. Regional flood and storm hazard assessment. Rep. No. 102, Institute of Hydrology, Wallingford, 159 pp.

Fricke, T.J., Kennedy, M.R. & Wellington, N.B. 1983. The use of rainfall correlation in determining design storms for waterways on a long railway line. *Proc. Hydrol. & Wat. Resour. Symp.*, Hobart, Nov 1983, 5pp.

Hosking, J.R.M. & Wallis, J.R. 1988. The effect of intersite dependence on regional flood frequency analysis. *Wat. Resour. Res.* 24: 588–600.

Jakob, D., Reed, D.W. & Robson, A.J. 1999. Selecting a pooling-group. Chapter 16 in *Statistical procedures for flood frequency estimation*, Vol. 3 of Flood Estimation Handbook, Institute of Hydrology, Wallingford, 153–180.

Reed, D.W. 2002. Why analyse inter-site dependence in river flooding? *Proc. 8*th *BHS National Hydrol. Symp.*, Birmingham, 8–11 September 2002: 169–174.

Reed, D. 2007. Floods: feast and famine. *Proc. WaPUG Autumn Conf. 2007*, Blackpool, Nov 2007, Paper 8.

Reed, D.W. 2008. Eight things you don't want to know about flood risk. *Proc. 43*rd *Flood & Coastal Management Conf.* Manchester, 1–3 Jul 2008, EA/Defra.

Reed, D.W. & Stewart, E.J. 1989. Focus on rainfall growth estimation. *Proc. BHS National Hydrol. Symp.*, Sheffield, Sep 1989: 3.57–3.65.

Reed, D.W., Faulkner, D.S. & Stewart, E.J. 1999. The FORGEX method of rainfall growth estimation, II: Description. *Hydrol. & Earth System Sci.*, 3: 197–203.

Stewart, E.J., Reed, D.W., Faulkner, D.S. & Reynard, N.S. 1999. The FORGEX method of rainfall growth estimation, I: Review of requirement. *Hydrol. & Earth System Sci.*, 3: 187–195.