Generation of spatially consistent rainfall data

Richard Chandler, H.S. Wheater, V.S. Isham, C. Onof, S. Bate,
P.J. Northrop, D.R. Cox and D. Koutsoyiannis

INTRODUCTION

Rainfall is the primary input to river flows. For continuous flow simulation therefore, one possible approach is to develop a continuous simulation model for rainfall, and to route its output through runoff models to obtain simulated flow sequences. Here, we review some recent developments in stochastic rainfall modelling. The aim is to provide rainfall simulations that are appropriate for hydrological application.

The next section deals with the development of models that can be calibrated using radar rainfall data. Such data represent an important source of information regarding the fine-scale structure of spatial rainfall fields. However, recognising their current limitations with respect to accuracy and record length, we subsequently describe some alternative methods that can be used when radar data are scarce, unreliable or absent. A more detailed account of the work is given by Wheater *et al.* (2000).

SPATIAL-TEMPORAL MODELLING USING RADAR DATA

There are several different approaches to the modelling and simulation of precipitation. In this section we focus upon a category of stochastic models, which aim to provide a simplified conceptual representation of the rainfall process in continuous space and time. This is done using probabilistic assumptions that enable rainfall structures to be represented using a few physically interpretable parameters. Such an approach is particularly suited to continuous simulation in the context of flow generation: the models' representation of rainfall in continuous space and time gives complete flexibility of application, and they are cheap to simulate.

Our approach is a spatial analogue of models that have been used to represent the temporal process of rainfall at a single site (e.g. Rodriguez-Iturbe *et al.*, 1987; 1988). The models use a hierarchical representation of rainfall, in which the smallest element is a *rain cell* – a localised area of relatively intense rainfall. It is known that cells tend to cluster within larger-scale structures that we call *storms*. These storms themselves cluster, to form *rain events*. At the spatial scale of UK catchments, these correspond to weather systems. We treat the interior of a rain event as a stationary stochastic process. However, there are many different types of event, and the parameters of this process will typically vary between event types. We therefore construct, for each month of the year, a library of model parameters to reflect the prevailing types of weather system during that month.

Modelling event interiors

In our model, a rain event consists of a collection of storms. Storm *centres* originate in a Poisson process in space (in practice, this means some large region containing the area of interest) and time, and they move with a constant velocity during their lifetimes. Each storm consists of a cluster of elliptically shaped rain cells, scattered about the storm centre in both space and time and moving with the same velocity as the storm centre. Each cell deposits rain at a constant rate (or intensity) over its entire area, and throughout its lifetime. The numbers of cells, their intensities, shapes and sizes, their temporal durations and their displacements about the storm centre, are random variables. Various independence assumptions are made, but are not discussed here. The spatial structure of a storm is shown schematically in Figure 1.

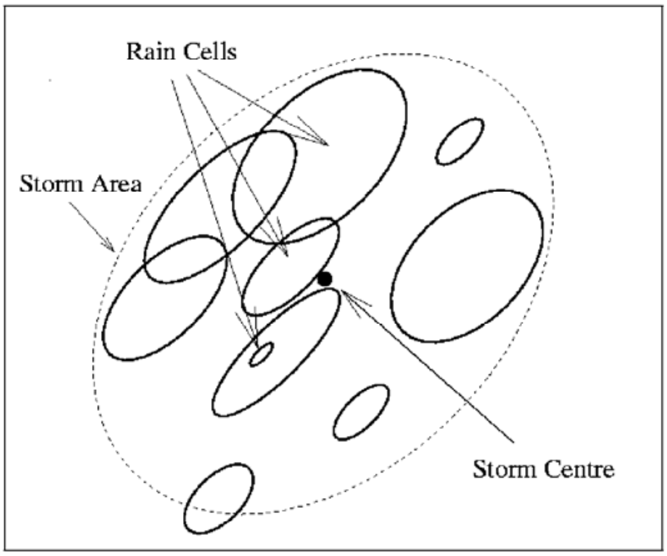

Figure 1. Schematic diagram showing the spatial structure of storms within the spatial-temporal model for event interiors

In the model used here, the spatial distribution of cells about a storm centre is bivariate normal. We therefore refer to the model as the *Gaussian Displacements Spatial-Temporal Model* (GDSTM). By varying the parameters of the displacement distribution, storms with different characteristics can be produced. For example, we can obtain 'scattered showers' if the variance of this distribution is large relative to the mean cell size. Rain bands result from a distribution with elongated elliptical contours. For simplicity, the shapes of cells within a storm are taken to be the same as the elliptical contours of the cell displacement distribution. A full description of the model requires distributions to be specified for all of the random components. Details may be found in Wheater *et al.* (2000), and in Northrop (1998).

The GDSTM has 11 parameters. These are physically interpretable quantities such as the expected cell duration, mean cell intensity and expected number of cells per storm. They can be estimated from a sequence of radar images by selecting some properties of the data, and finding parameter values that reproduce them as closely as possible (usually in a weighted least-squares sense). These properties might include the mean and variance of rainfall intensities, spatial and temporal autocorrelations at different scales, and the probability that an arbitrary region experiences no rainfall during a specified period. Their values, under the model, can be derived from its continuous space-time representation – see Northrop (1998) for details. Numerical methods are required to estimate the parameters from a given set of fitting properties. To ensure the stability of such methods, it is helpful to estimate groups of parameters separately where possible. Section 2.8 of Wheater *et al.*

(2000) contains some discussion of the issues involved. In general, parameter estimates for the GDSTM agree closely with previous studies, with respect to such quantities as cell and storm size and duration.

Validation of the GDSTM can be carried out by comparing observed and theoretical properties not used in the fitting procedure. For use as input to a continuous flow simulation, it may be important to reproduce relatively complex properties of rainfall fields, such as the extent and duration of localised areas of heavy rainfall (which can be quantified using suitable summaries of a rainfall sequence in which all rainfall intensities below some high threshold are set to zero). Such properties cannot be determined analytically, but can be explored by simulating the fitted model.

To illustrate the appearance of realisations from this model, Figure 2 shows an hourly sequence of radar images from the Wardon Hill radar station in south-west England,

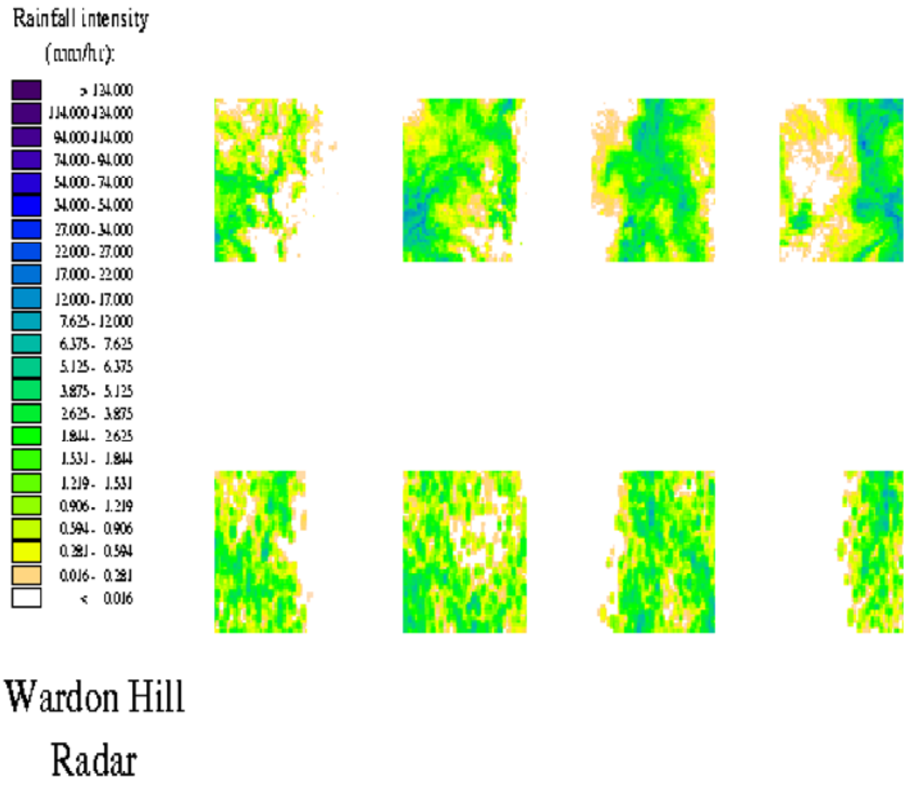

Figure 2. Visual assessment of the GDSTM fitted to the rain event of 6 February 1994, 13:00-14:00. The temporal separation of the images is 1 hour. Top: Observed data. Bottom: Simulation

together with an hourly sequence of images simulated from the GDSTM fitted to this event. The images are at $2\times2km^2$ spatial resolution, and cover an area of $104\times104km^2$. The model parameters were estimated from a 1-hour sequence of 5-minute radar images from the middle of this event. Figure 2 includes time periods outside that used for model fitting, to illustrate the movement of the rain event over the area (see the following Section). The observed and simulated sequences exhibit broadly similar features, although the simulations have a more obvious cellular structure than the data. Quantitative assessments of fit, described in Wheater *et al*., (2000), indicate that the model reproduces many hydrologically significant features of this event.

The GDSTM has been fitted to data from a total of 207 rain events, recorded at Wardon Hill between September 1993 and March 1997. The events have been split by month, to allow for seasonal variations. We therefore have 12 monthly libraries of parameter estimates that can be used to generate different types of events as appropriate in each month of the year.

Modelling event sequences

The GDSTM is stochastically stationary in both space and time, and hence is only appropriate for modelling the interiors of rain events. To obtain realistic continuous simulations of rainfall sequences, we need to model the process by which events enter and leave a catchment. Wheater *et al*. (2000) give the motivation, and supporting data analyses, for the scheme described here. We define two types of event: *Type 1* events are those that extend over the entire area of interest for part of their lifetime, and *Type 2* events are those that may produce a significant amount of rainfall, but do not cover the entire area. Analysis of Wardon Hill radar data indicates that, even at spatial scales as large as 10,000km^2, Type 1 events are responsible for a large proportion of the total rainfall. We model the stochastic process of events and dry periods using a semi-Markov process with 3 states ('dry' periods, *Type 1* events and *Type 2* events). The durations of sojourns in each state are independent random variables, modelled using Weibull distributions. For simulation purposes, events of any duration can be generated by simulating the GDSTM within a quadrilateral, of appropriate size, that moves across the catchment. The random geometries of these quadrilaterals are determined by analysing the edges of observed rain events. For the simulation in Figure 2, the quadrilateral is an approximately rectangular region, aligned in a North-South direction and moving from west to east: the leading (eastern) and trailing (western) edges appear in the first and final images respectively.

Seasonality is an important feature of rainfall sequences. It can be incorporated into the process of event arrivals, by fitting separate Weibull sojourn distributions for each month of the year. To reproduce seasonal variations in event types, in any month we generate events using GDSTM parameter sets sampled randomly from the corresponding library of parameter estimates. The overall algorithm for continuous simulation is as follows: (i) generate the time until the next event, by sampling from the current month's Weibull distribution for dry period lengths (ii) determine the event type, according to the appropriate frequencies for the month (iii) determine the event duration by sampling from the appropriate duration distribution (iv) select a GDSTM parameter set from the current month's library (v) simulate the GDSTM within a moving quadrilateral, to give the correct duration (vi) return to step (i).

The performance of this scheme has been investigated by comparing properties of simulations with those of observed radar sequences. Comparisons have been done for data at time scales of 15 minutes, 1 hour and 1 day, and at spatial scales ranging from 2×2km^2 up to 104×104km^2. Various properties have been investigated, ranging from basic statistical measures (means, variances, autocorrelation functions and proportions of an area that are dry) to analyses of extreme behaviour. In general, the scheme reproduces a wide range of features of observed rainfall fields. Again, details may be found in Wheater *et al.* (2000).

ALTERNATIVES, FOR APPLICATION IN THE SHORT-TERM

At present, the continuous simulation scheme described above has two drawbacks. The first is its reliance on radar data, which are limited both in terms of availability and quality. Realistically, it may be 5-10 years before adequate data are available to support the scheme operationally in the UK. A simplified version of the model, which may be calibrated using sub-daily data from a raingauge network, is described in Section 5.1 of Wheater *et al.* (2000). However, it has been found that even dense gauge networks do not provide sufficient spatial information to identify features on the scale of rain cells. Moreover, the availability of sub-daily gauge data is also limited in the UK at present.

The second drawback is that, except for seasonally varying parameters, the scheme is stationary in both space and time. It may therefore be inappropriate, in its present form, for use in regions with a lot of topographic variability, or for the generation of long rainfall sequences when long-term changes in climate are suspected. It is anticipated that this may be overcome by further model development. However, at present it may be appropriate to consider alternative methods, which are suitable for application in the short term.

Generalized Linear Models

Although radar and sub-daily raingauge data are relatively scarce in the UK at present, long *daily* records are more abundant. These can be used to study non-stationarities in rainfall sequences, and to develop daily rainfall simulation models. In principle, the output from such models can be downscaled to any desired resolution using some appropriate method. Hourly data may be adequate for most flow simulation purposes.

Many techniques are available for generating daily rainfall sequences. Here, one of the primary concerns is the incorporation of non-stationarities. A powerful and flexible technique, which allows non-stationarities to be quantified and incorporated, is that of Generalized Linear Models (GLMs). These are standard in the statistical literature, and were introduced into the study of daily rainfall by Coe and Stern (1982). The basic idea, which is an extension of linear regression, is to use the values of various predictors to forecast a probability distribution for the amount of rainfall at a site on a given day. In fact, this distribution is specified in two parts – we model the probability of rainfall occurrence separately from the amount of rain if non-zero. Previous days' rainfall amounts can be used as predictors, to account for autocorrelation. Other predictors might include quantities representing regional variation (such as site altitude), seasonal variability, long-term trends and 'external' climatological factors such as the North Atlantic Oscillation (NAO). Models can be specified in such a way that the effect of a predictor depends on the values of others – for example, it is known that the NAO affects European climate predominantly in the winter months. Models can be fitted, and compared, using Maximum Likelihood; they can be tested using a variety of simple but informative checks. By defining suitable dependence structures between sites, it is possible to build a multivariate GLM that allows simulation of non-stationary daily rainfall sequences over a network of sites.

The GLM methodology has been applied to many rainfall datasets from the UK and elsewhere. It has been found to be extremely useful for interpreting historical records, particularly when we want to study changes in the climate of an area. Moreover, providing models have been specified carefully, GLM simulations can reproduce a variety of features (including extreme behaviour) of observed rainfall sequences. Chapter 4 of Wheater *et al.* (2000) contains technical details and a selection of results. Other applications include: a flooding study in Ireland, where GLM outputs were downscaled to hourly resolution for input to continuous flow simulations (OPW 1998); and an investigation into drought in the Yangtze River in China (Yang, 2001).

Spatial-temporal rainfall disaggregation

The GDSTM can provide stationary sequences in continuous space and time, whereas the GLM can provide non-stationary sequences at a daily timescale. It is natural to try and combine these two techniques, so as to obtain the advantages of both. Various methods of doing this have been explored. Perhaps the most promising to date uses a method due to Koutsoyiannis (2001); this is described in Section 5.3 of Wheater *et al.* (2000). The method is appropriate for use when one or more sub-daily time series are available, along with several daily time series at a network of sites. The sub-daily time series may be either observations, or simulated data from a single-site version of the GDSTM, or data produced by disaggregating a simulated daily series. The daily series may be observational, or may have been generated by simulating a GLM. The basic idea is to generate hourly rainfall sequences at the sites with daily data, that respect both the fixed daily totals and the spatial structure of sub-daily rainfall fields. This spatial structure can be deduced from a simplified version of the GDSTM, fitted to the available data. The procedure produces hourly sequences with the correct means, variances, coefficients of skewness and proportions of dry intervals. Promising preliminary results are reported in Wheater *et al.* (2000).

SUMMARY

Continuous simulation of rainfall fields, incorporating fine-scale spatial and temporal structure, can only be achieved if supported by adequate radar data. Models such as the GDSTM appear to offer the potential for generating realistic scenarios, once adequate data become available (although further work is required to address issues of non-stationarity). In the meantime, other methods using downscaled daily sequences may be more feasible. Such daily sequences may be generated using GLMs. Although these methods cannot describe the full spatial structure of rainfall fields (they are essentially multi-site models), they are extremely powerful and have a lot of potential for hydrological application. An ongoing research project, sponsored by DEFRA, is carrying out a detailed investigation of some of the methods reported here, with a view towards making recommendations for practitioners. A report, using case studies to illustrate the techniques, is at an advanced stage of preparation.

REFERENCES

Coe, R. and Stern, R. (1982). Fitting models to daily rainfall. *Journal of Applied Meteorology*, 21, 1024-1031.

Koutsoyiannis, D. (2001). Coupling stochastic models of different time scales. *Water Resources Research*, 37, 379-392.

Northrop, P. J. (1998). A clustered spatial-temporal model of rainfall. *Proc. Roy. Soc. London*, A454, 1875-1888.

OPW (1998). *An investigation of the flooding problems in the Gort-Ardrahan area of South Galway, by Southern Water Global and Jennings O'Donovan and partners*. Office of Public Works, Dublin.

Rodriguez-Iturbe, I., Cox, D. and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proc. Royal Society London*, A410, 269-288.

Rodriguez-Iturbe, I., Cox, D. and Isham, V. (1988). A point process model for rainfall: further developments. *Proc. R. Soc. London*, A417, 283-298.

Wheater, H.S., Isham, V.S., Onof, C., Chandler, R.E., Northrop, P.J., Guiblin, P., Bate, S.M., Cox, D.R. and Koutsoyiannis, D. (2000). *Generation of spatially consistent rainfall data*. Report to the Ministry of Agriculture, Fisheries and Food (2 volumes). Also available as Research Report No. 204, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT (http://www.ucl.ac.uk/Stats/abstracts.html).

Yang, C. (2001). *Observed changes and simulative predictions of climate extremes in China*. PhD thesis, Institute of Atmospheric Physics, Beijing. (In Chinese.)