Journal of Hydrology (2007) **347**, 358–370

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

A stochastic model for daily rainfall disaggregation into fine time scale for a large region

Yeboah Gyasi-Agyei *, S.M. Parvez Bin Mahbub

Centre for Railway Engineering, Faculty of Sciences, Engineering & Health, Central Queensland University, Rockhampton QLD 4702, Australia

Received 19 March 2007; received in revised form 7 August 2007; accepted 13 September 2007

KEYWORDS
Stochastic;
Rainfall;
Disaggregation;
Regionalisation;
Parameter uncertainty;
Capping

Summary A robust model for disaggregation of daily rainfall data at a point within a large region to any fine timescale of choice is presented. Limited fine timescale data are required to calibrate only three parameters for the regional model, to establish monthly variation of simulation timescale lag-1 autocorrelations, and also to establish a scaling law between the simulation timescale and the 24-h aggregation levels. Site specific parameters are obtained using the 24-h statistics to disaggregate a long record of daily data by repetition and proportional adjusting techniques with capping. An Australia-wide data set has been used as a case study to illustrate the capability of the model. It has been demonstrated that the disaggregation model predicts very well the gross statistics (including extreme values) of rainfall time series down to 6-min timescale. The possibility of linking the disaggregation model to daily, or global circulation, models that can capture the inter-annual variability of the rainfall process for simulation beyond the number of years of record is being explored.

Introduction

Fine timescale (sub-hourly) rainfall data are limited worldwide largely because of the high cost and low reliability of monitoring such data. Where fine timescale data exist, the generally short record length undermines their direct use for water resources projects. However, continuous rainfall–runoff simulation is increasingly becoming the preferred option over the use of isolated events data for hydrological systems design and performance analysis. A logical way of addressing the scarcity of the fine timescale data is by stochastic daily rainfall disaggregation. Long records of daily rainfall are available worldwide because of the low cost and ease of collection.

Stochastic models attempting to address this problem can generally be divided into two groups. The first group of models (e.g. Bo et al., 1994; Glasbey et al., 1995; Gyasi-Agyei, 1999, 2005; Cowpertwait, 2006; Segond et al., 2006) are based on point process theory of the Bartlett–Lewis and Neyman–Scott rectangular pulses proposed by

* Corresponding author. Tel.: +61 408 309977; fax: +61 7 49306984.
E-mail address: y.gyasi-agyei@cqu.edu.au (Y. Gyasi-Agyei).

0022-1694/$ - see front matter
doi:10.1016/j.jhydrol.2007.09.047

Rodriguez-Iturbe et al. (1987, 1988). Scale invariance theory of cascade, fractal and multifractal approaches fall into the second group of models for disaggregation (e.g. Ormsbee, 1989; Olsson, 1998; Olsson and Berndtsson, 1998; Güntner et al., 2001; Molnar and Burlando, 2005).

The motivation for this research is the availability of the Australian SILO Data Drill facility for extracting long-term (1889 to date) continuous daily rainfall data from an archive of interpolated rainfall (http://www.nrm.qld.gov.au/silo/datadrill/, February, 2007). Another source of motivation is the desire to model runoff and erosion on railway embankment steep slopes at a very fine timescale. Comparing SILO Data Drill output to observed rainfall data at daily timescale for the same time frame, Gyasi-Agyei (2005) observed that there is not much difference between the two data sets. Hence the facility is a valuable asset for generating long-term daily rainfall for sites within Australia. One advantage is that annual trends exhibiting climate change are preserved by the facility generated daily data. However, the full usefulness of the facility will be very much realised if linked to a fine timescale disaggregation model. Current sub-daily timescale stochastic models are unable to preserve the day-to-day rainfall variability if they are not used to disaggregate daily rainfall data. The need to generate sub-daily time series fully consistent with the observed daily totals while preserving the stochastic structure of multiple sub-daily time scales cannot be overemphasised. It is not sufficient for a model to only reproduce the statistics such as mean, variance, autocorrelations and the probability of an interval being dry (referred to as dry probability) at several time scales. The generated rainfall series should also reflect annual trends, in particular for sites that exhibit a clear climate change.

This paper is focussed on the extension of the point process disaggregation model presented by Gyasi-Agyei (1999, 2005) with the following aims:

(a) generalisation for any fine timescale at a point;
(b) a simplified approach to determine the parameters of the regional model for a very large region such as the Australian continent;
(c) evaluation of the uncertainty of the calibrated parameters;
(d) enhancement of the capping procedure;
(e) testing of an Australia-wide parameter set with 9 years continuous rainfall data of 6-min timescale at a point.

A full description of the model is first presented. This is followed by Australia-wide model parameter identification. The enhanced capping methodology is then presented followed by the repetition techniques and proportional adjusting procedure. Evaluation of the model is given in the penultimate section. The last section provides conclusions from the paper.

The regional disaggregation model

The regional disaggregation model is a hybrid composed of a binary chain model (wet and dry sequence) and an autocorrelated jitter (intensity process) model. To use the model for daily rainfall disaggregation, limited fine timescale data are required to calibrate the binary chain model parameters for the region, with one of the parameters linked to the 24-h aggregation level dry probability. Similarly, a regional model is required to downscale the first and second order moments of the 24-h timescale to the simulation timescale (Gyasi-Agyei, 1999, 2005). In this way, the 24-h aggregation level properties are preserved. Site-specific parameters of the model are estimated from the regional parameter models and the 24-h aggregation level properties. To allow for seasonality, the analysis is carried out on a monthly basis.

The hybrid model

The hybrid model at a point $\{\overline{Y_t}(h)\}$ is a product of the binary chain model $\{Y_t(h)\}$ and an autocorrelated jitter model $\{A_t(h)\}$ expressed as

$$\{\overline{Y_t}(h)\} = \{A_t(h)\}\{Y_t(h)\} \tag{1}$$

where h is the simulation timescale. The jitter process is assumed to be lognormal, $A_t(h) = \exp[Z_t(h)]$, where $\{Z_t(h)\}$ is a stationary Gaussian process modelled by a first-order autoregressive model given as

$$Z_t(h) = \mu_Z(h) + \rho_Z(h)[Z_{t-1}(h) - \mu_Z(h)] + E_t \tag{2}$$

The random process $\{E_t\}$ is normally distributed, with a mean value of zero and variance $\sigma_E^2(h) = [1 - \rho_Z(h)^2]\sigma_Z^2(h)$ with the lag-1 autocorrelation condition $|\rho_Z(h)| < 1$. Setting the jitter model mean value to 1, that is making $\mu_{\overline{Y}}(h) = \mu_Y(h)$, implies $2\mu_Z(h) = -\sigma_Z^2(h)$. The variance $\sigma_Z^2(h)$ and the lag-1 autocovariance $c_Z(h)$ of $\{Z_t(h)\}$ are given, respectively, as (Gyasi-Agyei and Willgoose, 1997)

$$\sigma_Z^2(h) = \ln\left[\frac{\sigma_{\overline{Y}}^2(h) + \mu_{\overline{Y}}^2(h)}{\sigma_Y^2(h) + \mu_{\overline{Y}}^2(h)}\right] \tag{3}$$

and

$$c_Z(h) = \ln\left[\frac{c_{\overline{Y}}(h) + \mu_{\overline{Y}}^2(h)}{c_Y(h) + \mu_{\overline{Y}}^2(h)}\right] \tag{4}$$

The binary chain model

The binary chain model at a point generates a string of two numbers: $Y_i = 0$ for a dry period and a constant value $Y_i = w(h)$ for a wet period, where Y_i is the cumulative rainfall depth over time interval i of duration h. The moments of the binary chain model are given by (Gyasi-Agyei and Willgoose, 1999)

$$w(h) = \frac{\mu_{\overline{Y}}(h)}{1 - P(h)} \tag{5}$$

the variance $\sigma_Y^2(h)$ as

$$\sigma_Y^2(h) = [\mu_{\overline{Y}}(h)]^2 \frac{P(h)}{[1 - P(h)]} \tag{6}$$

and the lag-1 autocovariance $c_Y(h)$ as

$$c_Y(h) = [\mu_{\overline{Y}}(h)]^2 \left\{ P_{00}(h) - 2P_{01}(h)\left[\frac{P(h)}{1 - P(h)}\right] + P_{11}(h)\left[\frac{P(h)}{1 - P(h)}\right]^2 \right\} \tag{7}$$