Pergamon

Wat. Sci. Tech. Vol. 37, No. 11, pp. 57–64, 1998.
IAWQ

0273–1223/98 $19·00 + 0·00

PII: S0273-1223(98)00316-3

MODELLING THE EMBEDDED RAINFALL PROCESS USING TIPPING BUCKET DATA

Peter Thyregod*, Karsten Arnbjerg-Nielsen**, Henrik Madsen* and Jacob Carstensen**

** Department of Mathematical Modelling, Technical University of Denmark, 2800 Lyngby, Denmark*
*** Department of Environmental Science and Engineering, Technical University of Denmark, 2800 Lyngby, Denmark*

ABSTRACT

A new method for modelling the dynamics of rain measurement processes is suggested. The method takes the discrete nature and autocorrelation of measurements from the tipping bucket rain gauge into consideration. The considered model is a state space model with a Poisson marginal distribution. In the model there is only one parameter, a thinning parameter. The model is tested on 39 rain events. The estimated value for the various rain events is reflecting a subjective classification of rain events into frontal and convective rain. Finally, it is demonstrated how the model can be used for simulation and prediction.

KEYWORDS

Tipping bucket rain gauge, discrete valued state space model.

INTRODUCTION

Rainfall is of paramount interest in todays society. The time scales of relevance in urban areas ranges from minutes to decades. Of major importance in urban hydrology is the variations on a scale of minutes. Urban runoff causes problems with flooding of subcatchments and pollution, the latter both due to combined sewer overflows and due to transient impacts on waste water treatment plants. From an environmental engineering point of view a better understanding of the rainfall process will enable better modelling of the entire water flow in the cities.

Increasing requirements for averting floods as well as increasing consideration of environmental pollution aspects makes minimisation of peak discharges a key issue today. Among engineers many believe that this goal could be achieved through real-time control of urban hydrological systems. These control systems may require short-term rainfall forecasts as input.

Also in requirements from design and analysis of urban drainage systems there is a need for better understanding and description of the rainfall process. Typically long time series of rainfall are taken as input to rainfall-runoff models to simulate long time series of maximum water levels, overflows and impacts on waste water treatment plants and receiving waters. The simulated model output is analysed, and the system performance evaluated. Thus, the results of any hydrological calculations are highly dependent on the rainfall input that is used. However, there is a serious shortage of rain series long enough to produce long-term extreme statistics. Therefore simulated rain series could be used with advantage. However, before starting to simulate 30 years of rain data a good understanding of single rain events is crucial.

Rain data is most often collected by means of a tipping bucket rain gauge. A tipping bucket rain gauge is a discrete sampler counting the number of times a bucket is filled in each sampling time interval. Knowing the volume of the bucket the rain event can somehow be reconstructed from these counts. Thus, when the rain event is sampled in this way, the observed rain event is represented as a time series of counts.

The literature is scarce concerning analysis of single rain events with a temporal resolution suitable for modelling of urban storm drainage. Most of the rain data collection is due to agricultural needs and for prediction of flooding in rivers. Hourly, or even daily, measurements are quite sufficient for these purposes. However, the temporal aggregation of rain data for use in urban hydrology should not exceed five minutes (Schilling, 1991).

In (Arnbjerg-Nielsen, 1996), the modelling of single rain events is considered. The approach used in this work is to consider the waiting times between consecutive tips of the bucket. These waiting times are modelled using traditional statistical models. One approach is using ARIMA models on the logarithm of the waiting times. Another is a full discrete time Markov chain with 26 states representing waiting times. The latter model is used to construct long artificial rain series with properties close to the observed series. The disadvantage of this approach is that it requires a lot of parameters.

THE DATA

The data used for testing the proposed models consist of 39 single rain events, 16 of which are subjectively classified as being convective and the other 23 are classified as being frontal. The rain events cover a time span of 16 years, the first rain event is from 1979 and the last from 1995. They have been selected using either extreme intensities (an average of more than 9 μm/s in ten minutes) or depths over 20 mm as criterion.

The rain data originates from tipping bucket rain gauge 20211 in Aalborg, Denmark. The gauge is part of a large monitoring program initiated by the Danish Committee on Water Pollution Control (Spildevandskomiteen). For a more detailed description of the monitoring program see (Harremöes and Mikkelsen, 1995).

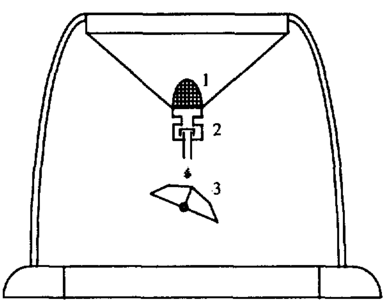

Fig. 1. Tipping bucket rain gauge.

The measuring principles of a tipping bucket rain gauge is illustrated in Figure 1. The rain enters the gauge through the funnel whereafter it goes through a filter (1) into a syphon (2). The purpose of the syphon is to ensure that the water always enter the bucket assembly (3) with the same momentum. When the syphon is full, all the water from the syphon enters the bucket assembly. The water enters one of the two buckets at a time, and when the bucket is full the bucket assembly tips, and the water will now enter the other bucket. The number of tips from the bucket assembly is registered with a sampling frequency of one minute.

In Denmark two types of rainfall are dominating. Convective rain is high intensity rainfall usually of rather short duration, as seen for example during thunderstorms. The 16 convective events selected have durations of between 75 and 564 minutes, and have depths in the interval [9.0; 47.4] mm. As an example a typical convective rain event is shown in Figure 2. The first observations of the events shown are as follows [0 0 1 3 1 3 4 1 2 ⋯].

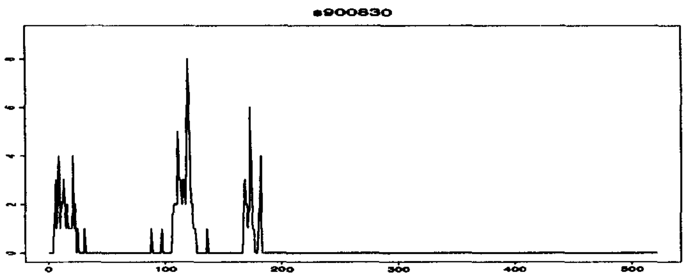

Fig. 2. Convective rain event from August 30th 1990.

The other common type of rain is frontal rain. Frontal rain usually give rise to longer rain events with almost constant rain intensities, as seen for example during the passing of a front. The 23 frontal events selected have durations between 308 and 1460 minutes, and have depths in the interval [20.2; 72.4] mm. An example of a typical frontal rain event is shown in Figure 3. The first observations of the events shown are as follows [0 0 1 0 0 0 ⋯].

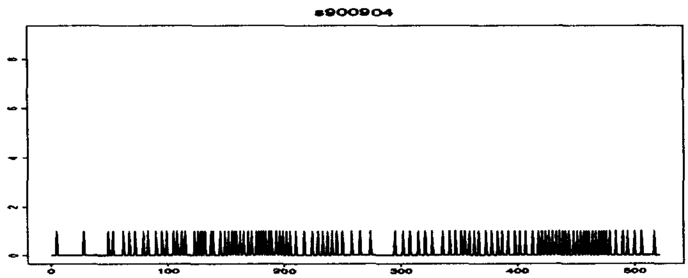

Fig. 3. Frontal rain event from September 4th 1990.

Rain events are rarely pure in the sense that they can uniquely be said to belong to one of the two types. Especially convective storms are likely to exhibit frontal behavior prior to and after the convective region of the storm.

All 39 events have been manually classified as being either frontal or convective by an expert. Those events reflecting both types of behavior have been classified as convective.

THE MODEL

Count data are usually modelled using probability distributions such as the Poisson or negative binomial. When the means of such models are constant, or can be expressed in terms of input variables the estimation can be carried out by using the theory of generalized linear models, see e.g. (McCullagh and Nelder, 1990). The idea behind times series analysis, however, is that the mean is not constant but rather depends on previous observations. In the model considered in this paper the distribution of an observation, y_t, is specified conditional on the mean of an underlying unobserved process, the rain intensity $\{\theta_t\}$. The observations can then be said to consist of two components; one belonging to the embedded process, the true rain intensity, and the other a random error. This error represents both the discretisation and any misspecification of the underlying model.

In (Harvey, 1989) a state space model for count data is presented using the analogy to the random walk plus noise model

$$y_t = \theta_t + \varepsilon_t \tag{1}$$

$$\theta_t = \theta_{t-1} + \eta_t \tag{2}$$

in which the underlying level of the process changes over time according to (2). ε_t and η_t are error terms representing measurement error and system error, respectively. Both have mean zero and constant variance.

The conditional distribution of the observations (y_t) given the level of the underlying process (θ_t) is assumed to follow a Poisson distribution

$$Y_t|\theta_t \in Po(\theta_t) \tag{3}$$

This corresponds to the measurement equation (1). Similarly, the probability distribution of the embedded rain intensity is chosen to be a gamma distribution.

$$\theta_t \in G(\alpha, \beta) \tag{4}$$

In order to formulate the model defined by (3) and (4) as a stochastic process, the time dependence has to be specified. Here it is chosen to specify this dependence in terms of the conditional distribution of θ_t given the previous observations, $\mathbf{Y}_{t-1}$. Let α_{t-1} and β_{t-1} denote the parameters at time $t-1$ computed from the first $t-1$ observations, $\mathbf{Y}_{t-1}$. In other words, conditional on the information at time $t-1$ the distribution of θ_{t-1} is given by

$$\theta_{t-1} \in G(\alpha_{t-1}, \beta_{t-1}) \tag{5}$$

For the random walk plus noise model with normally distributed observations, $\theta_{t-1} \in N(m_{t-1}, p_{t-1})$ implies that $\theta_t \in N(m_{t-1}, p_{t-1} + \sigma_\eta^2)$. Thus, the mean of $\theta_t|\mathbf{Y}_{t-1}$ is the same as that of $\theta_{t-1}|\mathbf{Y}_{t-1}$, but the variance is increased. For the gamma distribution this same effect can be obtained by multiplying α and β with a constant between zero and one. The fundamental difference between this model and the random walk plus noise model is in the model for the latent process, $\{\theta_t\}$.

Derivation of the distribution of θ_t is now done by updating the parameters α and β. By letting

$$\alpha_t = \omega\alpha_{t-1}$$

$$\beta_t = \omega\beta_{t-1} \tag{6}$$

where $0 < \omega < 1$, the following desirable properties hold

$$\mathrm{E}\{\theta_t\} = \mathrm{E}\{\theta_{t-1}\}$$

and

$$\mathrm{var}\{\theta_t\} = \omega^{-1}\mathrm{var}\{\theta_{t-1}\}$$

i.e. exactly the properties of the random walk plus noise model.

Having observed y_t, α_t and β_t can be updated using this new information. The updating equations are given by

$$\alpha_t' = \alpha_t + y_t$$

$$\beta_t' = \beta_t + 1$$

where α_t' and β_t' denote the updated parameter values. These values are then used as the basis for the new prediction using (6).

RESULTS

In most of the frontal events the counts are very rare, indicating a very low minute-to-minute varition. Thus, to have some autocorrelation worth modelling all 39 rain events have been aggregated to the number of tips in 10 minutes. The parameter ω has been estimated for each of the 39 aggregated events using the maximum likelihood method. For a derivation of the maximum likelihood estimator as well as a more detailed description of the model see (Thyregod, 1997). The resulting parameter estimates are shown in Figure 4. The estimate leads to a very distinct separation into convective events and frontal events. The parameter ω can be regarded as a thinning parameter. Thus, ω indicates the strength of the carry-over effect. The events classified as being frontal give rise to higher values of ω, i.e. the frontal events are characterised by having a higher carry-over effect.

These findings are in very good agreement with the common meteorological understanding that convective rain events are innovation processes whereas frontal rain events have more memory. The usual practice for determining whether a rain event is frontal or convective has been to classify the event manually by an expert. The above findings indicate that this instead can be done by inspecting the parameters of the above model.

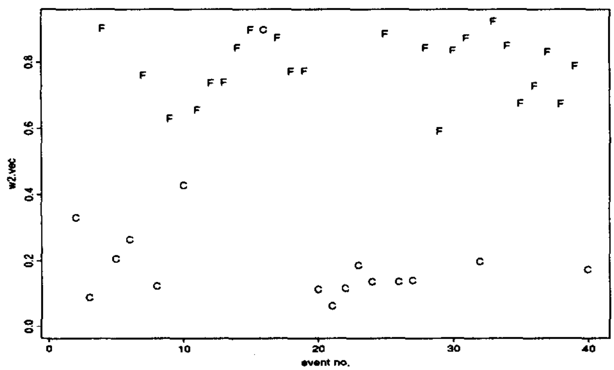

Fig. 4. Estimate of the parameter ω for the 39 events.

One major advantage of the state space approach is that the uncertainty is separated into two contributions, one arising from the measurement process and one arising from the underlying process. When interest is focussed upon predictions or in gaining insight in the rainfall process, it is much easier to work with a smooth process without the measurement error on top.

In Figure 5 the observed process is shown above the underlying process. It is seen that the fast variations have been removed, but the estimated underlying process exhibit a visible variation that corresponds to the picture one would expect intuitively.

A model is estimated for each rain event in order to reflect the properties of this specific event, and hence for comparisons with other events the related parameter estimates are used. This means that it is not possible to carry out an ordinary cross validation procedures.

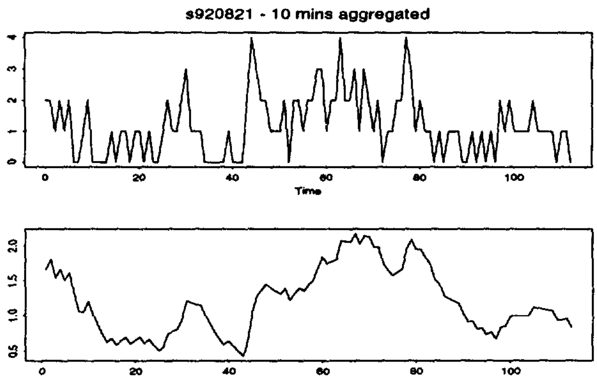

Fig. 5. Original series (top) and estimated latent process (bottom).

APPLICATIONS

The proposed model can be used for making short erm predictions which is useful for instance for on-line control of sewer systems. Furthermore, the model can be used to generate synthetic rain events.

Prediction

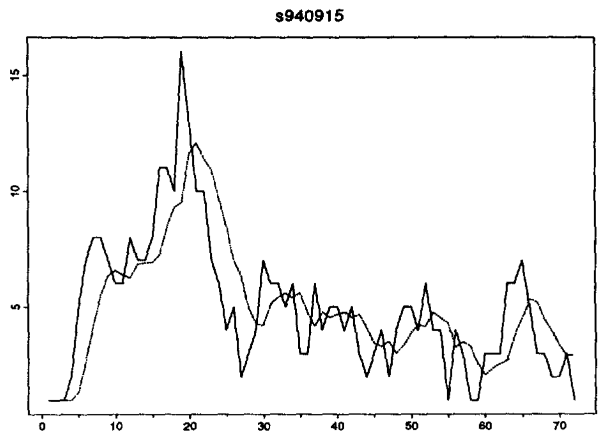

Fig. 6. Poisson-gamma predictions. Predictions: dotted line.

At time τ the mean of the Poisson measurement process will be the level of the underlying process, θ_τ. Hence, the best one-step prediction will be the estimated level of the underlying process. It follows from the updating formulae for the parameters that the predictions can be found as an exponentially weighted average with weights, ω, declining exponentially. Hence, the most recent observations are given the most weight in the predictions.

$$\hat{Y}_{\tau+1|\tau} = \frac{\sum_{j=0}^{\tau-1} \omega^j y_{\tau-j}}{\sum_{j=0}^{\tau-1} \omega^j} \tag{7}$$

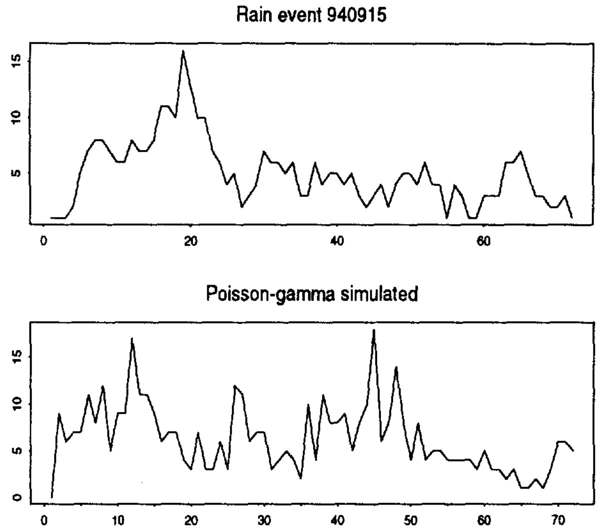

Fig. 7. INAR and Poisson-gamma simulations.

In Figure 6 the series of one-step predictions is shown together with the series of aggregated observations.

Simulation

The way to simulate a rain event using the proposed state space model is by first simulating the level of the underlying process. Conditional on that the number of tips can be simulated as a random Poisson variable with mean equal to the level of the underlying process.

It can be shown that in the transition equation for the dynamics of the embedded rain process the random component originates from a Beta distribution. Thus, the level of the underlying process, θ_t can be simulated by generating random Beta distributed variables with suitable parameters following the updating equations for the parameters.

The parameter used for simulation originate from estimation of the model using one of the aggregated rain events. In Figure 7 the original series is shown above the simulated series. The simulated series seem to have some of the same visual characteristics as the original series.

The model performance in terms of its prediction and simulation capabilities is in this paper judged solely on the basis of graphical comparisons. A more objective method to compare the accuracy of more models could be using the series of prediction errors arising from each model. When making comparisons like that it is essential that the number of parameters for each model is also taken in to consideration.

CONCLUSION

In this paper a new model for the rainfall measurement process have been suggested and its performance is tested on actual rainfall data sampled by a tipping bucket rain gauge.

Compared to some other approaches the model is formulated in the domain most natural, i.e. the time domain for positive counts. Furthermore it holds the Markovian property.

Models with many parameters have the major disadvantage, that they require a very large amount of data for estimation with a reasonable precision. Furthermore in a model with many parameters, there will often be a high degree of substitutability, i.e. the parameter estimates will be highly correlated, and therefore the identification of the individual parameters may be doubtful. Models with few parameters, on the other hand, provide better possibility of interpreting and comparing the estimates of the parameters. Finally models with less parameters are often more robust.

A set of parameters was estimated for each of the available 39 rain events. When estimating a set of parameters based on only one rain event, the parameters will be reflecting properties of this particular rain event. The model leads to parameter estimates that could be closely associated with the distinction between frontal and convective rainfall. This is an important break-through in rainfall modelling. Previously such classification has been done manually, but using the suggested models the estimated parameters can be used in that classification.

In the Poisson-gamma model there is only one parameter, the thinning parameter. For each of the series the thinning parameter was estimated using the ML method. The estimated value for the various rain events reflected in a very distinct way the classification of rain events into frontal and convective rain. Convective rain events gave rise to small values of the thinning parameter which is in agreement with the meteorological known fact, that convective rain is closer to an innovation process.

Finally the model was used for simulation. The simulations resulted in a series with the same visual characteristics as the actual series of observations.

References

Arnbjerg-Nielsen, K. (1996). *Statistical analysis of urban hydrology with special emphasis on rainfall modelling*. PhD thesis, DTU.

Harremöes, P. and Mikkelsen, P. (1995). Properties of extreme point rainfall i: Results from a rain gauge system in denmark. *Atmospheric Research*, 37.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

McCullagh, P. and Nelder, J. (1990). *Generalized linear models*. Chapman and Hall.

Schilling, W. (1991). Rainfall data for urban hydrology: What do we need ? *Atmospheric Research*, 27.

Thyregod, P. (1997). Models for the rainfall measurement process. Master's thesis, IMM/DTU.