Stochastic Long Term Modelling of a Drainage System with Estimation of Return Period Uncertainty

Thorndahl, Søren Liedtke

Published in:
Conference Proceedings : 11th International Conference on Urban Drainage

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Stochastic long term modelling of a drainage system with estimation of return period uncertainty

S. Thorndahl¹*

¹ Department of Civil Engineering, Aalborg University, Sohngaardsholmsvej 57, 9000 Aalborg Denmark

*Corresponding author, e-mail st@civil.aau.dk

ABSTRACT
Long term prediction of maximum water levels and combined sewer overflow (CSO) in drainage systems are associated with large uncertainties. Especially on rainfall inputs, parameters, and assessment of return periods. This paper proposes a Monte Carlo based methodology for stochastic prediction of both maximum water levels as well as CSO volumes based on operations of the urban drainage model MOUSE (Lindberg and Joergensen 1986) in a single catchment case study. Results show quite a wide confidence interval of the model predictions especially on the large return periods. Traditionally, return periods of drainage system predictions are based on ranking, but this paper proposes a new methodology for the assessment of return periods. Based on statistics of characteristic rainfall parameters and correlation with drainage system predictions, it is possible to predict return periods more reliably, and with smaller confidence bands compared to the traditional methodology.

KEYWORDS
Urban drainage modelling; long term simulation; extreme statistics; uncertainties; Monte Carlo simulation; flooding; combined sewer overflow; return period

INTRODUCTION
The use of simulation models for hindcast prediction of flooding and combined sewer overflow occurrences in urban drainage systems has become an important tool for many consulting engineers and planners. Models, however, are often used with non-catchment specific parameters and are very rarely calibrated except for research use. The rainfall input (historical rain series recordings in a single point) cause model predictions to be uncertain, because the spatial rainfall variability is often ignored. Furthermore, in order to maintain design criteria, defined by the respective authorities, a certain safety is often implemented in parameter values so that predicted return periods are not exceeded within the specified limits.

Rainfall recordings during shorter periods of time entail great uncertainties in estimation of return periods. If a rain gauge has recorded for 10 years and the recordings are used as model input, it is not possible to determine whether the predicted maximum water level in a specific manhole should be assigned to a return period of 10 years, or maybe 50 or 100 years. Willems (2000) recommends that a maximum of 10 % of the total series length is applied as a statistically reliable return period. Other authors, e.g. Arnbjerg-Nielsen et al. (2002) use a maximum value of 25 % of total time series length. In urban drainage the Eurocodes of practice (EN 752-4. 1997) recommends a minimum return period of 10 years of flooding of combined sewers in residential areas. The design criterion is a return period of 2 years of...
occurrence of surcharge (exceedance of pipe top level). A 100 year series is therefore necessary to estimate (with low uncertainty) whether a certain area will be flooded during a ten year period (using the definition of Willems (2000)). In this paper another approach is investigated. Based on deterministic simulation with a 60 year long rain series, a correlation between characteristic rainfall parameters and water levels as well as overflow volumes is defined. The return period of these are then assigned using a fitted return period model of the characteristic rainfall parameters which is based on statistics from a large number of Danish rain gauges.

Besides the return period uncertainties, parameter uncertainty is also included by random sampling and Monte Carlo simulation of the most important parameters. The uncertainties of the different parameters are propagated through the model in order to estimate the uncertainties on the model predictions. Therefore, this paper has two objectives: (1) to do a forward uncertainty analysis on long term simulations of a drainage system, in order to predict surcharge, flooding or combined sewer overflow with different return periods; and (2) to predict return periods in order to incorporate uncertainties using recorded rainfall time series as model input.

Several authors have investigated both input and parameter uncertainties in urban drainage modelling, e.g. Willems and Berlamont (1999), Arnbjerg-Nielsen and Harremoes (1996), Grum and Aalderink (1999), Lei (1996), Thorndahl et al. (2006), Thorndahl and Willems (2008), Thorndahl et al. (2008), but most often modelling a number of either real or synthetic independent rainfall events. Thus, the extreme events and the return period uncertainty are neglected, although these events are the most interesting when considering the loading of the drainage system.

METHODOLOGY, LONG TERM SIMULATIONS

Thorndahl et al. (2008) present an event based uncertainty analysis of a small urban catchment using the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley 1992). Based on corresponding rainfall-runoff measurements from the Frejlev catchment in Denmark (Figure 1), a stochastic calibration of nine independent events is carried out. The GLUE methodology includes a large number of Monte Carlo simulations with parameter values sampled from some selected prior distributions for each parameter. Propagated through a setup of a MOUSE model and weighted against observations using an empirical likelihood measure, a set of posterior distributions for each parameter is derived. In this paper these posterior parameter distributions are applied as a stochastic input to the same setup of the MOUSE model. It would be preferable to do a complete GLUE analysis using long term simulations, but the observation period is limited and therefore model conditioning on historical data is not possible for the whole period. Therefore, a forward uncertainty analysis, without conditioning on observations, is implemented. (Beven 2008). Furthermore, Thorndahl et al (2008) introduce a comparison between two conceptual model setups. In the first, rainfall recordings from a single local rain gauge is used as model input, whereas in the second, an area averaged input from two local gauges is implemented. Using likelihoods as a measure of goodness of fit, it is shown that the model time series predictions are remarkably improved using the more complex rainfall input. However, the extremes, i.e. maximum water levels and combined sewer overflow volumes, does not seem to be significantly affected by the change in model input. Therefore, the model input to this setup will consist of one 20 year rain series recorded in the Svenstrup rain gauge, approx 4 km from the Frejlev catchment. This series includes 780 events with a rain depth larger than 1 mm. As shown on Figure 1,
two local gauges are installed within the catchment, but as these have a shorter running period (only 10 years) the dataset from these is not applied.

Even though recordings from only one rain gauge are applied, some of the uncertainty, related to the assumption of a homogenous rainfall distribution over the catchment, is implemented by the hydrological reduction factor. This factor determines the part of the impervious catchment area which contributes to the runoff. The reduction factor multiplied by the percentage of the impervious area corresponds to a general runoff coefficient. Based on more than 300 independent rainfall-runoff events in the Frejlev catchment, Thorndahl et al. (2006) and Thorndahl et al. (2008) have presented a remarkably small mean reduction factor of approx. 50%. Furthermore, quite a large dispersion of the reduction factor is observed, which is probably due to the fact that spatially uniform rain is applied knowing that a spatial variation is present. The factor is therefore not only used to represent the part of impervious area that actually contributes to the runoff, but also as a measure of the spatial rain variability, however distributed homogenously over the catchment within one event. The other parameters applied in the model setup are presented in Table 1.

The uncertainty analysis is implemented using Direct Sampling Monte Carlo simulations in order to propagate the uncertainty on the parameters through the model. The simulation time of a long term model operation (780 events) of the Frejlev catchment is approx. 12 hours, but using a cluster of computers it has been possible to complete a total of 1000 simulations.

**Table 1.** Parameters applied in the MOUSE setup of the Frejlev catchment. N(μ,σ) corresponds to a normal distribution with mean μ and standard deviation σ. U(x₁,x₂) corresponds to a uniform distribution with lower and upper limit x₁ and x₂ respectively

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyd. reduction factor, φ (-)</td>
<td>N(0.49,0.15)</td>
</tr>
<tr>
<td>Initial loss, i (mm)</td>
<td>U(0,0.8)</td>
</tr>
<tr>
<td>Surf. concentration time, t_c (min)</td>
<td>U(0,20)</td>
</tr>
<tr>
<td>Dry Weather Flow, DWF (l/(PE·day))</td>
<td>U(90,150)</td>
</tr>
<tr>
<td>Manning number, M (m⁻¹/³/s)</td>
<td></td>
</tr>
<tr>
<td>Smooth concrete</td>
<td>N(85,5)</td>
</tr>
<tr>
<td>Normal concrete</td>
<td>N(75,5)</td>
</tr>
<tr>
<td>Rough concrete</td>
<td>N(68,5)</td>
</tr>
<tr>
<td>Plastic</td>
<td>N(80,5)</td>
</tr>
<tr>
<td>Head loss, K_m (-)</td>
<td></td>
</tr>
<tr>
<td>Round edged outlet</td>
<td>U(0,0.5)</td>
</tr>
<tr>
<td>Sharp edged outlet</td>
<td>U(0.25,0.75)</td>
</tr>
</tbody>
</table>
METHODOLOGY, RETURN PERIOD ASSESSEMENT

The traditional way to assess return periods is by ranking e.g. event maximum water levels or CSO volumes. The event with the highest rank is assigned to the simulation period; the event with the second highest rank is assigned to half of the simulation period; the third half of the latter; and so forth. The return period of the events with the highest ranks is therefore quite uncertain. The ranking approach is applied in prediction of maximum water levels and CSO volumes in Figure 6 and 8, respectively.

The proposed approach to predict the return periods is also implemented a posteriori to the long term Monte Carlo simulations. The idea is to find correlations between some characteristic rainfall parameters and the model’s prediction of maximum water levels and combined sewer overflow volumes. It is possible to assess return periods for the characteristic rainfall parameters quite accurately (and with confidence bands), thanks to a large Danish dataset. Knowing the correlation, it is possible to transform the return periods of the characteristic rain parameters to return periods of the model prediction. Using the Frejlev catchment Thorndahl and Willems (2008) showed a linear correlation between the rainfall event peak intensities and the maximum water level in manholes as well as a correlation between the rainfall event depth and duration and the CSO volumes. This is investigated further by doing a deterministic long term simulation based on the mean parameter values from Table 1 and a combined long Danish rain series from Odense. This series is not representative for the Frejlev catchment, it is only used due to its length of approx. 60 years (approx. 5000 rainfall events). Selected results are shown in Figure 2 and 3 for maximum water level and CSO volume relationships respectively.

Studying Figure 2 it is obvious that there is a linear relationship between the predicted water levels $H_{max}$ and in this case the 10 minute averaged event peak intensity $i_{p10}$. The correlation depends on the parameters ($\Theta$) from Table 1 so that:

$$H_{max}(\Theta) \propto i_{p10}$$

With regards to the correlation between the CSO volume ($V_{CSO}$) and the rainfall event depth ($d$), it is obvious that interdependence with the rainfall duration ($dur$) is present (Figure 3). The diagonal line corresponds to the maximum overflow volume:

$$V_{CSO} = F_{imp} \cdot \varphi \cdot d + b = F_{imp} \cdot \varphi \cdot d - F_{imp} \cdot \varphi \cdot (\alpha \cdot dur + d_{min})$$

$F_{imp}$ is the impervious area; $\varphi$ is the hydrological reduction factor; $b$ is ordinate cut-off value, which can be expressed as a function of the rainfall duration (not shown); $\alpha$ is the correlation between the duration and the abscissa cut-off value; and $d_{min}$ is the minimum cut-off value corresponding to the minimum rainfall depth which will cause overflow. Therefore the overflow volume can be expressed as a function of the rainfall depth and duration:

$$V_{CSO}(\Theta, dur) \propto d$$
Using the derived correlations, the return periods of maximum water levels are estimated using the return period of the peak intensities averaged over a period of time corresponding to the transportation time in the sewer system. Upstream in the sewer system this corresponds to the 10 min. averaged peak intensity, further downstream the 20 and 30 min. averaged peak intensities are applied. Correspondingly, the return periods of the CSO volumes are estimated using the return period of the rainfall event depth.

In Spildevandskomiteen (2006) and Madsen and Arnbjerg-Nielsen (2006) 66 Danish rain gauges have been used to estimate key rainfall parameters with return period uncertainty based on measurements from 1979 – 2005. The dataset from these gauges is statistically approximated to a partial duration series model (Madsen and Rosbjerg 1997a; Madsen and Rosbjerg 1997b; Mikkelsen et al. 1998). Applying a generalised Pareto distribution, a fit of the return periods is estimated (Spildevandskomiteen 2006):

\[ \hat{z}_T = z_0 + \hat{\mu} \frac{1 + \hat{\kappa}}{\hat{\kappa}} \left( 1 - \frac{1}{\hat{\lambda} \cdot T} \right)^{\frac{1}{\hat{\kappa}}} \] (4)

\( z_T \) is the value corresponding to the return period \( T \), e.g. 10 minute peak intensity or rainfall event depth, \( z_0 \) is a cutoff level, \( \mu \) is the mean of the exceedings of \( z_0 \), \( \kappa \) is a shape parameter and the number of yearly exceedings, \( \hat{\lambda} \) is estimated by (Spildevandskomiteen 2006):

\[ \hat{\lambda} = \hat{\beta}_0 + \hat{\beta}_1 \cdot YMP \] (5)

\( \beta_0 \) and \( \beta_1 \) are linear regression parameters and \( YMP \) is the regional accumulated yearly precipitation. The variance of the number of yearly exceedings is based on:

\[ \text{var}(\hat{\lambda}) = \text{var}(\hat{\beta}_0) + 2 \cdot YMP \cdot \text{cov}(\hat{\beta}_0, \hat{\beta}_1) + YMP^2 \cdot \text{var}(\hat{\beta}_1) + \sigma_\delta^2 \] (6)
The statistics presented in Spildevandskomiteen (2006) have resulted in a division of the Danish rainfall statistics in two regional zones, one for the Western part of Denmark and one for the Eastern part. Using the accumulated yearly precipitation a local estimate of the return period uncertainty can be calculated. Shown here are observed local values of the 10 minute peak intensity (Figure 4) and the rainfall event depth (Figure 5) along with the Pareto model. It is obvious in both figures that the return periods estimated in the local series do not fit completely within the 95 % confidence interval of the Pareto model. This is due to the relatively short measuring period. However, in the area of 10 % of the series length the data fits quite well within the confidence bands of the Pareto model using the recommendations of Willems (2000).

Based on the above, the return periods predicted with the correlation between the peak intensity and the maximum water level correlation ($i_p-H_{\text{max}}$-correlation) are assessed as presented in this example. If a peak intensity is observed in a specific recorded rainfall event, the mean return period and the 95 % confidence intervals are estimated based on the Pareto model in Figure 4. The return period of the maximum water level for this specific event is then sampled using a triangular distribution applying the mean estimated return period and the variances. For each simulation of the maximum water level a total number of 100 return periods are sampled. As 780 events are included in the Svenstrup rain series and 1000 Monte Carlo runs are simulated, this yields a total of 78 million events with different return periods.

The return periods of overflow volumes based on the depth-duration-CSO volume ($d$-$dur$-$V_{\text{CSO}}$) correlation are estimated by the simulated overflow volume, the hydrological reduction factor, and the observed rainfall event duration. Using eq. 2, a rainfall depth is estimated and the return period for this specific depth is given by the Pareto model of Figure 5. The return periods of the CSO volumes are also sampled with triangular distributions.

RESULTS
The traditional ranking methodology used to estimate return periods of water levels is presented in Figure 6 for the most critical manhole in the system (Manhole T013520, Figure...
1). It is obvious, doing 1000 Monte Carlo simulations, that the dispersion of the predicted water levels for this manhole is quite large. This is primarily due to the standard deviation applied on the hydrological reduction factor, as this is by far the most crucial parameter (Thorndahl et al. 2008). The dispersion is especially large in prediction of water levels between the pipe top and the ground level, which is evident as the storage volume in this area is very small, and small changes in rainfall intensity or depth involves a great change in the water level. In Figure 7 return periods predicted by the $i_p-H_{max}$-correlation methodology are presented. Comparing the two figures, it is clear that the return periods are equal at low values. However, the high return periods of Figure 6 are more uncertain, as the maximum return period is assessed equally to the time series length. As the $i_p-H_{max}$-correlation, using the Pareto model, has no upper limit, the predicted water levels are assigned larger return periods. In order to compare the probability of flooding of the manhole, the predicted water levels exceeding ground level are selected and the cumulative distribution function of the return periods of these events are derived, Figure 10. Again, it is evident that the difference between the two approaches is large when the return period is large, however quite identical low return periods are present. As an example the probability of flooding this manhole every year is 62 % using the return period ranking methodology and 47 % using the $i_p-H_{max}$-correlation and Pareto model. The manhole is the most critical in the system, with very frequent flooding, and actually the drainage system was recently redesigned, so flooding problems no longer occur.

With regards to the prediction of CSO volumes (Figure 8 and 9) the tendency is the same as in prediction of water levels, i.e. the return periods using the ranking methodology are quite uncertain, but the results are similar on the low return periods. The new proposed methodology reduces the confidence interval on the large return periods. For example, the 20 year return period has a 95 % confidence interval corresponding to 3500 m$^3$ and 7500 m$^3$ applying the proposed methodology, however the ranking methodology show a 95 % confidence interval of 1500 m$^3$ and 10500 m$^3$. So in this case the interval is reduced to approx. one third, applying the new approach.

Figure 6. Return periods by ranking of maximum water levels, Manhole T013520. The solid line is the median and the dotted are the 95% confidence interval.

Figure 7. Return periods by intensity-water level correlation, Manhole T013520. The solid line is the median and the dotted are the 95% confidence interval.
Figure 8. CSO volumes and return periods by ranking. The solid line is the median and the dotted are the 95% confidence interval.

Figure 9. CSO volumes and return periods by $i_p-H_{max}$-correlation. The solid line is the median and the dotted are the 95% confidence interval.

Figure 11 presents the maximum water levels corresponding to the 2 year return period as a profile view of the main sewer line. Both methods show that the median and the mean (not shown) maximum water level exceeds the pipe top level in some of the pipe sections, so that the manholes are surcharged. This causes an exceedance of the design criteria recommended in the Eurocodes of practice (EN 752-4. 1997). The two methodologies perform equally in prediction the 2 year return period. It would be preferable to show the 10 year return period as well. This, however, has only occurred in 2 of the 39 pipe sections shown in Figure 11 (applying the $i_p-H_{max}$-correlation methodology). Therefore, it is concluded that the Eurocodes are kept with regards to the criteria of no flooding within a return period of 10 years, when the proposed methodology is applied. However, when applying the ranking methodology the water level corresponding to the 10 year return period is exceeded several times. It is therefore concluded that the large return periods are over-estimated when applying the ranking methodology.

CONCLUSION
This paper has presented a stochastic urban drainage model in which uncertainties of different parameters are sampled by performing Monte Carlo simulations. The parameter distributions are based on posterior distributions from a previous investigation (Thorndahl et al. 2008) in which a conditioning of the model is carried out using the GLUE methodology. A homogeneous spatial rainfall distribution over the catchment in every event is assumed, however the spatial rainfall variability is assessed using a wide sampling interval of the hydrological reduction factor which varies from event to event. It is clear that this might affect the predicted dynamics (i.e. the temporal flow variations) of the system to some extent. However, as this paper only concerns the extreme event statistics, it is the author’s conviction that omission of the spatial rainfall variation is valid - especially as the catchment in question is rather small.

In order to investigate if an adequate number of Monte Carlo simulations is executed, the model prediction statistics are completed using both 500 and 1000 simulations, however with no significant difference of the confidence interval. This concludes that enough simulations are executed.
The predicted maximum water levels and CSO volumes cover quite a wide confidence band, particularly on the large return periods. This confidence band might be narrowed if a conditioning on some observations is implemented, e.g. using the GLUE methodology. However, this would only affect the prediction of maximum water levels and CSO volumes, and not the uncertainty associated with the return periods, unless continuous observations in the drainage system are available for the whole simulation period. Consequently, an approach to handle uncertainties on the return periods is also presented in the paper. By investigating the correlation between characteristic rainfall parameters and maximum water levels as well as overflow volumes, a model for assessing the return periods is assembled. This is based on 26 years of statistics on several Danish rain gauges (Spildevandskomiteen 2006). Using the return periods of the characteristic rainfall parameters it is possible to assess return periods of maximum water levels and CSO volumes applying linear correlations. Comparing the traditional way of assessing return periods by ranking and the proposed model a good agreement between the two approaches is shown where the return periods are low. However, on the large return periods the proposed correlation model shows more realistic results (and smaller confidence bands), as the statistical basis of assessing return periods is much larger.

As the relationship between the rainfall parameters and the drainage system predictions are based on simple linear correlations, obviously some of the dispersion between the two is neglected, which might introduce a small uncertainty. Furthermore, as the approach is only tested on this small catchment which performs quite linearly, it is not possible to conclude whether the approach will work on a larger and more branched drainage system. Finally, it is not tested whether the predicted return periods actually correspond to actual occurrences in the drainage system. Preferably, further investigations will show that this is the case.

The concept of identity between the return periods of characteristic rainfall parameters and prediction of water levels is also proposed in the scientific literature, e.g. using the Chicago design storm (Kiefer and Chu 1957). This, however, includes a number of disadvantages as some of the rainfall dynamics are neglected (due to synthetic storms) and the fact that it is not possible to estimate CSO volumes using this method. The proposed methodology therefore
presents an alternative to return period prediction of maximum water levels and CSO-volumes on small simple catchments.

ACKNOWLEDGEMENT
The author would like to thank colleagues Kjeld Schaarup-Jensen and Jacob Birk Jensen for their contribution to the paper, and the Municipality of Aalborg for use of the Frejlev drainage system.

REFERENCES