Derived distributions for stormwater detention basins design

Dimensionnement des bassins d’orage par l’utilisation de lois dérivées

Leviandier Thierry, Payraudeau Sylvain.

Centre d’Ecologie végétale et d’hydrologie. Unité mixte Université Louis Pasteur- ENGEES 1 quai Koch. 67 Strasbourg

thierry.leviandier@engees.u-strasbg.fr
sylvain.payraudeau@engees.u-strasbg.fr

RESUME

Les bassins de retenue d’eaux pluviales sont destinés à protéger les zones urbaines contre un débordement ou le dépassement d’un certain débit aval. Ce risque est la combinaison d’un aléa naturel et d’un contrôle hydraulique. La méthodologie appropriée, qui est d’utiliser un modèle stochastique de pluie et de simuler le fonctionnement du bassin, peut être complexe dans un contexte d’optimisation et trop gourmande en données. Nous donnons à l’utilisateur final un moyen de court-circuiter cette étape par la technique des « distributions dérivées » qui déduit la distribution de fréquence des sorties de celle des entrées (y compris dépendance temporelle) et des paramètres du système. Nous proposons une nouvelle méthode, approchée mais générique, fondée sur les probabilités conditionnelles. L’application au dimensionnement d’un bassin répondant à l’acceptation d’un risque donné se résume alors en la résolution d’une équation algébrique représentant la distribution dérivée. La méthode fait appel à des statistiques pluviométriques usuelles et à un paramètre spécifique, supposé régional, déduit d’une analyse d’épisodes pluvieux.

ABSTRACT

The aim of storm basins is to protect urban areas against some predefined risk of exceeding a given value of downstream runoff, or a risk of overflow for a bounded storage capacity. This risk results from the combination of a natural hazard and hydraulic properties. The proper way to address this issue is to use a stochastic rainfall model, but it may require unavailable data and be cumbersome to use in the framework of an optimization procedure. We give the end user a way to by-pass this step, by the technique of “derived distributions”. The purpose of such a method is to calculate the parameters of the probability density function (pdf) of outputs as a function of the pdf of inputs and of the parameters of the dynamic deterministic system between inputs and outputs. We propose to apply a new way of designing derived distribution, based on conditional probabilities, which is approximate but generic. For application to dimensioning of basins, the determination of the value parameter(s) corresponding to an acceptable risk simply consists in solving an algebraic equation representing the derived model. The methodology needs usual rainfall statistics and a specific parameter inferred from analysis of storms, or supposed to have a regional value.

KEYWORDS

Derived distribution, Metamodel, Stormwater detention.
1 INTRODUCTION

Design of stormwater detention with a given risk of overflow of the storage capacity is clearly related to the statistical properties of rainfall. As the storage is intrinsically cumulative, a large number of storms must be used to encompass the variability of hyetographs. If on-site data are available, simulation of storm detention basins with different geometrical or hydraulic parameters on observed hyetographs enables to find an acceptable or optimal solution (Phillips, 1995, Herrmann and Schmida, 1999, Vaes and Berlamont, 2001). But generally, data are not available, and only statistics of rainfall can be inferred from near by and regional meteorological information. We want to provide a method usable in such a context, but consider that we can use models and data for designing the method, that the end user will not have to use. In this case the derived probability distribution approach (Chen and Adams, 2005) is more convenient.

2 METHOD

2.1 Two variants of the hydraulic behaviour

It is generally assumed that the basin is emptying while storing (otherwise, the problem would be simple, but the efficiency dubious) either with a constant discharge (with a pump) or in a passive way by gravity. In the latter case, the discharge is function of the current level of storage, and the basin performs mitigation of flood. In this paper we assume a linear relationship (only an approximation from a physical point of view). We do not simulate what happens when the storage or hydraulic capacities are saturated, but let run unrealistic simulations and count such events.

2.2 Two ways of introducing rainfall information

The variation of the level of the basin can be simulated on real event observed in the past or on a set of events drawn from a stochastic model. When the process is simple, there is no reason to chose a unique reference event, and using a large set is preferable. The second method can be subdivided into Monte-Carlo simulation or numerical discretization of the probability space, less common but used in the present study.

2.3 The volumes design method

This method (Ministères, 1977, quoted in Chocat, 1997) is used to design dry stormwater detention basins. This method is based on two main hypotheses. First, the basin discharge is supposed to be constant and is usually expressed in mm per hour or litre per second per hectare. Secondly, the transfer of the rainfall to the detention basin is supposed to be instantaneous without lag time. These hypothesis constrains the application of this method to small urban catchment, less than some dozen hectares, without other upstream detention basins.

This method requires the availability of about ten years of rainfall data with short time step and can be applied with three steps.

First, each rainfall event has to be differentiated in the rainfall series on the basis of dry period threshold with the next event. Secondly, for each event the cumulated hyetograph is performed and compared to the cumulated basin discharge to calculate the maximum level in the detention basin during this event. Applied to all the rainfall events of the series, the annual maximum levels in the detention basin are extracted.
and fitted to a well-known probability density function (pdf). The choice of the pdf depends on the goodness of fit between observed data and statistical model used. The Gumbel law, frequently fitted to extreme values, usually allows a good fit with the maximum annual levels of detention basin. Exponential distributions may be preferred for data sampled as peaks over thresholds, what may be convenient to analyse severe storms in the same year. Thirdly, the statistical model, i.e. the Gumbel law for example, could be used to calculate the maximum level in the detention basin required to avoid failure of basin according to a return period, 10-years for example. This maximum level associated to a return period allows to design the detention basin by taking into account the upstream area collected.

2.4 Insight into the stochastic process with conditional probabilities

Usually, a stochastic model is built with independent random variables and exact relationships between dependant and independent variables. On the other hand, multivariate probability distribution functions, more convenient for explicit calculation, have conditional statistical relationships but these relationships do not represent easily the dynamics and causality between variables. We use simultaneously one model of each type, in order to combine their advantages. The multivariate model (we will call it the conditional model) is defined by three properties:

- the independent variable (R Rainfall) is exponentially distributed;
- the conditional variable (output of the system) under condition that Rainfall is greater than some value, is exponentially distributed;
- the median value of the conditional variable is linear with respect to the reduced independent variate.

With notation Q always taken for the output variable, which is a discharge in the case of linear outflow, but represents the maximum storage reached during the storm in the case of constant discharge, the equations are:

for the conditional distribution: \( q_R = q_0 + \lambda u q_R + \mu u_R \)

with \( \text{Prob}(Q > q_R | R > r) = e^{-u q_R} \)

\( \text{Prob}(R > r) = e^{-u R} \)

and parameters \( q_0, \lambda, \mu \)

and by analytical integration, for the cdf of \( q \):

\( (\lambda - \mu) e^{-u} = \mu e^{-u \lambda} + (\lambda - 2 \mu) e^{-u \mu} \)

with \( \lambda u_0 = \mu u_0 = q - q_0 \)

\( \text{Prob}(Q > q) = e^{-u^*} \)

we use the notation with * for this model, so that (2) is true by definition, whereas (4) is only an approximation if \( Q \) belongs to another model. Variables named \( u \) are taken for variables representing frequency, whether or not they are strictly speaking reduced variates, what is true when the random variable follows an exponential distribution.

The stochastic model is defined as the CECP (Contrast enhancing clustering process, Leviandier et al, 2000) model as input of a simple deterministic behaviour of the basin, according to the hypotheses of § 2.1. The hypotheses of CECP are:
the total amount of Rainfall is exponentially distributed;

real hyetographs are successively separated in two unequal intervals of constant intensity, such that the rainfall depth of each interval is kept equal to its true value during this span of time, and such that intensities and duration of intervals are linked by a « splitting equation »;

the stochastic process is given by the probability distribution function of the time of separation, at each order (order of successive separation), which is a power law;

In opposition with the early paper, in which the exponent was the same at each order (some scale invariance), the power law is downgraded to a uniform distribution at order greater than one, and an exponent $g_1$ different of 1, characteristic of the station, is used only at first order.

There is no theoretical equivalence between these models - « derived distributions » should be understood in an extended meaning -, so that we have to check empirically if one can be considered as an approximation of the other, that is if the conditional probabilities generated by the stochastic model satisfy the properties of the multivariate model. To be useful, the approximation must hold for a range of parameters. In other words, the models, despite the differences in their concepts, should have the same parameters. The approximation must also be validated for large return periods, and the models should have the same asymptotic behaviour. It must be noticed that the CECP model, though it provides no theoretical analytical expression of asymptotic behaviour, is very convenient to calculate conditional distribution, due to the separation of the shape of the hyetograph (given by the times of separation) from the total rainfall.

**Approximation of the pdf of output variables as functions of the parameters of the system, through conditional probabilities**

The conditional asymptotic probabilities of variables of interest were calculated numerically, running the stochastic model for different values of the parameters of the system. We call parameters of the system, the parameters of the stochastic model of rainfall and the parameters of the detention basin. Linear regressions were calculated between parameters of conditional asymptotic probabilities (that is $q_0$, $\lambda$, $\mu$ of equation (1) and parameters of the system.

If the models were identical, the only thing to do would be to apply equation (2).

Due to imperfect linearity of conditional probabilities, especially in low return periods, it is necessary to fit a correction formula between the two models:

We use the notation $\text{Prob}(Q > q) = e^{-u}$ for the stochastic model (5)

A possible simple correction is a linear relationship between reduced variates

$u = \alpha u^* + \beta$ (6)

Another one is to consider that Equation (2) is exactly satisfied with $\text{Prob}(Q > q)$ on the left hand side, but that equation (3) contributing to the right hand side, defines $q^*$, and to seek a transformation of $q^*$ into $q$. A relationship between logarithms was tried:

$\ln(q) = \gamma \ln(q^*) + \delta$ (7)

Linear regressions were also calculated between $\alpha$, $\beta$ and the parameters of the system.

1110 NOVATECH 2007
Design of geometrical or hydraulic parameters by an Invert method

For an application, the end user is supposed to know the characteristics of rainfall. The problem is to choose a pair (storage capacity, emptying facility), with often a constraint on one of these items, with a given frequency of overflow. He has not always the resources to run a model and may prefer a formula (Loganathan et al., 1985) or a graphical method (McEnroe, 1992, Ministères, 1977).

For instance, in the case of linear outflow, he wants the outflow not to exceed a value \( q \), with a probability \( F \).

Let us write \( f(\lambda, \mu, q_0, q) \) the right hand side of equation (2):

If the transformation (6) has been used, we have to eliminate \( u^* \) between

\[
(\lambda - \mu) e^{-u^*} = f(\lambda, \mu, q_0, q)
\]

and \( -\ln(F) = u = \alpha u^* + \beta \)

which consists in finding the root of \( \ln(F) + \beta - \alpha (\ln f) - \ln(\lambda - \mu) \)

for the only variable which is the basin discharge parameter \( b \) appearing in the linear regressions above-mentioned, the numerical values being given in next section.

If the transformation (7) has been used, we have to eliminate \( q^* \) and find the root of

\[
(\lambda - \mu) F - f(\lambda, \mu, q_0, q^*(q, \gamma, \delta))
\]

For a given station, the results can be presented in a two dimension graph, with a curve for each frequency in axis \((b, q)\) or a curve for each \( b \) in axis \((\text{frequency}, q)\).

3 RESULTS

3.1 Relevance of linear conditional dependance

Figure 1 shows that the conditional distribution of \( Q \) (represented here for 10 years)

\[ \text{Fig 1. Left: conditional discharge with exponential reduced variate.} \]
\[ \text{Rigt: median conditional discharge for Rainfall reduced variate return period rainfall, and different durations) is asymptotic exponential and that the median conditional value of } Q \text{ is linear with the rainfall reduced variate.} \]
3.2 Validation of the approximations by the conditional model

The model must be validated on a range of statistical parameters encountered on observed data. Several datasets were used with extraction of rainfall during 2 hours exceeding a threshold of 10 mm in Northern France, or 20 mm in Mediterranean regions). This threshold gives from 2 to 3 events per year. Under this conditions the scale parameter of the exponential distribution is never far from the half of the position parameter, and the exponent $g_1$ varies from 0.3 near Montpellier to 0.6 in Nancy.

The second step is to check the assumption that the equation (6) or (7) is able to transform the pdf calculated by the theoretical conditional model to the pdf calculated by numerical integration of the CECP model. Though deep reasons are unknown, the linear correlation between the two values is very strong, with a coefficient $R^2$ always greater than 0.99 and often greater than 0.999.

The next step is to find relationships between parameters of the system and parameters of the conditional model. Linear regressions were calculated for an arbitrary value $p_0 = 10$ mm in the distribution of rainfall $p = p_0 + u g_1$.

$b$ is the depth of water, in mm, that can be filled out at constant rate during the storm.

\[
q_0 = 11.5 - 0.36 b - 1.624 g_1^{-1} + 1.80 g \quad R^2 = 0.981
\]

\[
\lambda = -6.61 - 0.096 b + 3.517 g_1^{-1} + 1.89 g \quad R^2 = 0.993
\]

\[
\mu = -1.13 - 0.248 b + 0.536 g_1^{-1} + 2.12 g \quad R^2 = 0.982
\]

\[
\alpha = 1.333 + 0.0228 b + 0.937 g_1^{-1} - 0.149 g \quad R^2 = 0.934
\]

\[
\beta = 0.324 - 0.0386 b + 1.503 g_1^{-1} + 0.220 g \quad R^2 = 0.905
\]

and for the linear outflow model, $c$ representing the coefficient in the linear equation:

\[
q_0 = 10.6 - 0.053 c - 0.497 g_1^{-1} + 1.972 g \quad R^2 = 1.000
\]

\[
\lambda = -3.72 + 0.1512 c + 3.561 g_1^{-1} + 1.875 g \quad R^2 = 0.998
\]

\[
\mu = -0.091 - 0.033 c + 0.105 g_1^{-1} + 0.970 g \quad R^2 = 0.999
\]

\[
\gamma = 1.820 + 0.0133 c + 0.737 g_1^{-1} - 0.150 g \quad R^2 = 0.978
\]

\[
\delta = -0.927 - 0.0614 c - 2.051 g_1^{-1} + 0.455 g \quad R^2 = 0.975
\]

the most efficient correction was found to be on $u$ in the first case (constant outflow) and on $\ln(q)$ in the second (linear outflow).
The final validation is given by the graph of the two calculations of $u$, after transformation (Fig 2.).

The potential upper limit of efficiency is given by the efficiency of equation (3) for one set of parameters, which is very high. The errors are due to the assimilation of functional relationships to linear relationships. It must be noticed that estimation of parameters by linear regression is not optimal to bring the distribution from conditional model closest to the distribution from stochastic model.

### 3.2 Validation with observed hyetographs

An example is given for a Mediterranean station (Mauguio Airport, near Montpellier). With the choice of 2 hours and a threshold of 20 mm in the modelling approach, we miss only one long event (1 hours) which is the worst in its year, yet only for small basin discharges. However, the derived method underestimates maximum volumes for low basin discharges and low return periods. For this reason, if results in this range of values are expected, it is necessary to take the envelop of the derived method’s results for different durations (using the same equations, in a dimensionless form). In all cases the model’s results are below the direct method’s results.
4 CONCLUSION AND PROSPECTS

A methodology was presented to design stormwater detention basins with a given probability of overflow. It provides a very simple tool for the end user, the relationship between the free parameter of the device and the accepted risk being embedded in an algebraic formula.

Three parameters are used for the description of rainfall stochastic properties. Two of them are common since they are the first two moments of the distribution of rainfall observed in a constant duration. The third one characterizes intensity-duration-frequency relationship, but in a formulation which is specific of the method. It should be considered as regional when there are no data to analyze hyetographs.

The accuracy of the method could be improved by refining the relationships of the parameters of conditional distributions with parameters of the system. However, to keep the method generic and not too much dependent on some details of the system is preferable. In that sense, relaxing the constraint of analytical solution, we are closer to a metamodel than to a derived distribution. Though the conditional probabilities were calculated with a particular stochastic model, we believe they have some intrinsic features. The underestimation could be due to the fact that the model was used in separating the events in only three sub-events, or more probably to the assumption that the storage is empty at the beginning of events, what can be corrected. The simple hydraulic behaviour (in power 0 or 1) assumed in this paper brings some simplification, but ongoing researches let think possible generalisation to upstream hydrological rainfall-runoff transformation with nonlinearities.

As the method takes into account the time variation of rainfall intensity, generalisation to sediment and pollutant modelling is also possible.

REFERENCES