A FILTERED RENEWAL PROCESS AS A MODEL FOR A RIVER FLOW

MARIO LEFEBVRE

Received 13 July 2004

Various models, based on a filtered Poisson process, are used for the flow of a river. The aim is to forecast the next peak value of the flow, given that another peak was observed not too long ago. The most realistic model is the one when the time between the successive peaks does *not* have an exponential distribution, as is often assumed. An application to the Delaware River, in the USA, is presented.

1. Introduction

In [2] (see also [5]), a filtered Poisson process was used to forecast the various peaks of rivers. Let $\{N(t),\ t \geq 0\}$ be a homogeneous Poisson process and let $X(t)$ be the river flow at time t. It was assumed, in the previous references, that

$$X(t) = \sum_{n=1}^{N(t)} Y_n e^{-(t-\tau_n)/c}, \tag{1.1}$$

where the random variables τ_n are the arrival times of the Poisson events, Y_n is the magnitude of the signal that occurred at time τ_n, and c is a constant which characterizes the river system. The authors also assumed that the random variables Y_n have an exponential distribution. The stochastic process $\{X(t),\ t \geq 0\}$ defined by (1.1) is indeed a particular case of what is known as filtered Poisson processes. This type of stochastic process has been used to model various phenomena; see [3, page 144]. In civil engineering, filtered Poisson processes have served as models for stochastic rainfall [6] and seismic hazards [4], in particular.

Going back to the $\{X(t),\ t \geq 0\}$ process, the authors considered the limiting stationary version of this stochastic process, from which they defined a Markov chain $\{X_n,\ n = 1,2,\dots\}$ for the sequence of peak river discharges. Using various estimators based on the transition probability function of the X_n's, they obtained the forecasted value of the "next flood," given the value of the "last flood" observed.

The model set up in [2] worked only relatively well, mainly because the correlation coefficient between the successive peaks is rather weak, in general. It is well known that

Mathematical Problems in Engineering 2005:1 (2005) 49–59
DOI: 10.1155/MPE.2005.49

trying to predict the next peak value of a river flow is a very difficult task. However, we believe that we can at least improve the results obtained so far by rendering the model more realistic. Indeed, many mathematical assumptions made in the formulation of the model are often not realistic at all or are only used to make the model tractable or for lack of better alternatives.

There are two main criticisms that one can state with regard to the $\{X(t),\ t \geq 0\}$ process above. First, it assumes that an event that occurs at time τ_n has an immediate maximum effect and that this effect decreases with time. In practice, a more or less steep increase of the river flow is almost always observed before it begins to decrease. Therefore, the choice of an exponential function as a "response function" can surely be criticized.

Second, the main assumption in the model above is that the time between two consecutive peaks has an exponential distribution, so that events occur according to a Poisson process. Again, in practice this will almost surely be false. At least, this assumption should be checked by performing a statistical test to make sure that the exponential distribution is indeed a good model for the random variables representing the times between the events. We will see, in Section 3, that in the case of the Delaware River this assumption is clearly false.

In Section 2, the notion of a filtered renewal process will be introduced. We will see how the next peak flow value could be forecasted, based on the most recent peak observed. An application to the Delaware River will then be presented in Section 3. Finally, a few conclusions will be drawn in Section 4.

2. A filtered renewal process

A homogeneous Poisson process is characterized by the fact that the times $T_1, T_2, \ldots$ between the consecutive events are independent and have the same exponential distribution with constant parameter λ. It is a particular renewal process, namely, a counting process for which the times $T_1, T_2, \ldots$ are independent but can follow any common distribution, be it discrete or continuous.

Now, the probability density function of an exponential distribution is strictly decreasing from zero. However, in practice, the distribution of the times between the various flow peaks is *not* strictly decreasing. Rather, it generally increases toward a maximum value and then is strictly decreasing until infinity. Because of that, assuming that the sequence of occurrence times of peak flow values forms a Poisson process is not appropriate.

In the case of the Delaware River, we will see in Section 3 that the distribution of the T_k's is well approximated by a Rayleigh distribution, that is,

$$f_{T_k}(t) \simeq \frac{t}{\alpha^2} e^{-t^2/2\alpha^2} \quad \text{for } t \geq 0, \tag{2.1}$$

where α is an unknown parameter that must be estimated.

The general model that we propose for a river flow is of the form

$$X(t) = \sum_{n=1}^{N(t)} w(Y_n, t - \tau_n), \tag{2.2}$$

where $w(\cdot,\cdot)$ is the response function and $\{N(t),\ t \geq 0\}$ is a renewal process whose distribution of the times T_k's between the events must be determined.

Remark 2.1. Although the distribution of the T_k's is generally not an exponential distribution, it is sometimes possible to apply a function to the T_k's that transforms them into exponential random variables. This amounts to working on a different time scale. For instance, in the case of the Rayleigh distribution with parameter $\alpha = 1/\sqrt{2}$ above, if we define

$$T_k^* = (\lambda T_k)^{1/2}, \tag{2.3}$$

where $\lambda > 0$, then we easily find that T_k^* has an exponential distribution with parameter λ. Thus, by taking the square root of all the time variables, we can assume that $\{N(t),\ t \geq 0\}$ is indeed a Poisson process.

Next, by choosing a response function of the form

$$w(Y_n, t-\tau_n) = Y_n e^{-(t-\tau_n)/c}, \tag{2.4}$$

we neglect the fact that there is almost always a period during which the river flow increases before decreasing again; that is, the flow increase is not instantaneous. To obtain this feature, we can consider a response function given by

$$w(Y_n, t-\tau_n) = Y_n (t-\tau_n)^k e^{-(t-\tau_n)/c}, \tag{2.5}$$

where k is a positive constant that must be estimated. Of course, it will be more difficult to forecast the future peak values from the more general model. In the case of the basic model, we have

$$X(t+\delta) = \sum_{n=1}^{N(t+\delta)} Y_n e^{-(t+\delta-\tau_n)/c} \tag{2.6}$$

for any $\delta \geq 0$, so that

$$\begin{aligned} X(t+\delta) &= e^{-\delta/c}\left\{\sum_{n=1}^{N(t)} Y_n e^{-(t-\tau_n)/c} + \sum_{n=N(t^+)}^{N(t+\delta)} Y_n e^{-(t-\tau_n)/c}\right\} \\ &= e^{-\delta/c}\left\{X(t) + \sum_{n=N(t^+)}^{N(t+\delta)} Y_n e^{-(t-\tau_n)/c}\right\}, \end{aligned} \tag{2.7}$$

where t^+ is equal to $t+\epsilon$, with $\epsilon > 0$ as small as we want. If $\{N(t),\ t \geq 0\}$ is a Poisson process, we may write

$$X(t+\delta) = e^{-\delta/c}\{X(t) + X^*(\delta)\}, \tag{2.8}$$

where $X^*(\delta)$ has the same distribution as $X(\delta)$. It follows that

$$E[X(t+\delta)|X(t)] = e^{-\delta/c}\{X(t) + E[X(\delta)]\} = e^{-\delta/c}X(t) + \frac{\lambda c}{\mu}(1 - e^{-\delta/c}) \tag{2.9}$$

if X_n has an exponential distribution with parameter μ for all n (see [3]). With the more general response function (2.5), we cannot obtain such a simple formula.

To be able to use the model defined by

$$X(t) = \sum_{n=1}^{N(t)} Y_n (t - \tau_n)^k e^{-(t-\tau_n)/c} \tag{2.10}$$

to forecast peak flow values, we first need to estimate the unknown parameters k and c. This will be done as follows: let

$$g(t) = t^k e^{-t/c}. \tag{2.11}$$

This function attains its maximum value at $t_{\max} = kc$. Therefore, we can estimate the value of the product kc by computing the mean time taken by the river flow to reach a peak from the preceding minimum.

Next, if the time between the consecutive peaks is assumed to be large enough, we can neglect the effect of the signals $Y_1, \ldots, Y_{N(t)-1}$, retaining only $Y_{N(t)}$, and write that

$$X(t+\delta) \simeq Y_{N(t)} \left(t + \delta - \tau_{N(t)}\right)^k e^{-(t+\delta-\tau_{N(t)})/c}. \tag{2.12}$$

Remark 2.2. In the next section, we will see that with the value of k estimated from the data, the function $g(t)$ is indeed such that we can neglect all the events prior to the one that occurred at time $\tau_{N(t)}$.

We deduce from (2.12) that

$$\begin{aligned} \frac{X(t+\delta)}{X(t)} &\simeq \frac{Y_{N(t)} \left(t + \delta - \tau_{N(t)}\right)^k e^{-(t+\delta-\tau_{N(t)})/c}}{Y_{N(t)} \left(t - \tau_{N(t)}\right)^k e^{-(t-\tau_{N(t)})/c}} \\ &= e^{-\delta/c} \left\{ 1 + \frac{\delta}{t - \tau_{N(t)}} \right\}^k. \end{aligned} \tag{2.13}$$

This formula is valid for values of t and $t+\delta$ between two consecutive peaks. If we choose t to be the time at which the most recent peak was observed, then we may write

$$\frac{X(t+\delta)}{X(t)} \simeq e^{-\delta/c} \left\{ 1 + \frac{\delta}{kc} \right\}^k. \tag{2.14}$$

Since kc can be estimated, we solve for k in (2.14) and obtain that

$$k \simeq \ln \frac{\left(X(t+\delta)/X(t)\right)}{\ln\left(1 + \delta/kc\right) - \delta/kc}. \tag{2.15}$$

Finally, to estimate k (and hence c from the estimated value of kc), we will compute the mean value obtained for k if $t+\delta$ is the time at which the minimum following the last recorded peak was observed. That is, we compute the ratio of the minimum over the preceding maximum for the observations, and we calculate the mean value of the expression in the right-hand member of (2.15).

Our aim, in the next section, will therefore be to forecast the next peak flow value, based on the preceding peak, once we have observed that the river flow has started to increase again. It is an objective less futile than trying to predict the next peak flow based on the most recent one.

In theory, we must also find a model for the distribution of the signals Y_n, $n = 1, 2, \ldots$. However, in practice only the mean value of the previous signals will be used to forecast the next peak flow, so that the actual distribution of the Y_n's is not needed.

To forecast the value of the next peak flow, we will use the following estimator:

$$\widehat{\text{Peak}}_1 = \text{Max}\left(\bar{N}_I + N_D\right)^{\hat{k}} e^{-(\bar{N}_I + N_D)/\hat{c}} + \bar{I}, \tag{2.16}$$

where Max is the value of the most recent peak flow, $\bar{N}_I$ is the average number of days taken by the river to go from a minimum to a maximum flow, N_D is the number of days between Max and the following minimum flow, and $\bar{I}$ is the average difference between the various peaks and the preceding minima. Finally, $\hat{k}$ and $\hat{c}$ are the point estimates of k and c. To obtain $\hat{k}$, $\hat{c}$, $\bar{N}_I$, and $\bar{I}$, we will use part of the available data, and we will then apply our estimator to the rest of the data.

We will compare the results obtained with $\widehat{\text{Peak}}_1$ to the corresponding ones when the constant k is taken equal to zero, so that we neglect the time taken by the river to reach a peak. Based on this model, a simple estimator of the next peak flow is

$$\widehat{\text{Peak}}_2 = \text{Min} + \bar{I}, \tag{2.17}$$

where Min is the minimum flow that has just been observed (so that this estimator, as the previous one, is produced one day after the minimum was observed), and $\bar{I}$ has been defined above. Notice that the $\widehat{\text{Peak}}_1$ estimator does not make use of the value of Min (or the value of the flow one day after the minimum), but rather it only uses the variable N_D representing the number of days elapsed between the maximum and minimum flows.

We will also consider the estimators $\widehat{\text{Peak}}_3$ and $\widehat{\text{Peak}}_4$ obtained by using linear regression, with one and many response variables, respectively. This technique has given excellent results for short-term forecasting in other works (see [1]).

As a criterion to assess the quality of the various estimators considered in the paper, we will use the correlation coefficient. As mentioned previously, the reason why the problem of forecasting peak river flows is so difficult is that the correlation between consecutive peaks is small. If we can obtain forecasts that are relatively highly correlated with the actual observed peaks, we will be satisfied.

3. An application to the Delaware River

To test our model on real data, we have chosen the Delaware River, which is an important river whose flow values are freely available on the WWW (see http://nwis.waterdata.usgs.gov and http://pa.water.usgs.gov). More precisely, we have used the data for the years 1993–2002 at the Montague Station, NJ (no. 01438500). During this time period (until September of 2002, actually), there have been 91 peak flow values greater than or

Table 3.1. Goodness-of-fit test of a gamma distribution for the random variable T.

j	$[0,5]$	$(5,10]$	$(10,15]$	$(15,20]$	$(20,\infty)$
n_j	6	26	15	8	6
p_j	0.107	0.345	0.293	0.155	0.100
m_j	6.53	21.05	17.87	9.46	6.10

equal to 10000 ft^3/s, of which 61 were followed by another peak (greater than or equal to 10000 ft^3/s) in a short-enough interval.

Our objective will be to first find a model for the flow of the Delaware River. Next, we will use the data from the years 1993–1997 to estimate the various parameters and quantities in the model, and then we will forecast the 33 peak flows that were preceded by another peak a few days beforehand during the 1998–2002 time period. We will thus compare the estimators $\widehat{\text{Peak}}_1, \ldots, \widehat{\text{Peak}}_4$.

3.1. Model fitting. The first step in fitting a model of the type defined by (2.2) to the data is to find an approximate distribution for the times T_n's between the successive events. If we denote by T the general random variable representing the time between two events, we find, using the 61 data points, that the average value of T and its standard deviation are given by, respectively,

$$\bar{t} \simeq 11.689, \qquad s_T \simeq 6.125. \tag{3.1}$$

We immediately notice that an exponential distribution is *not* an appropriate model for these data, since if T has an exponential distribution, then we know that

$$E[T] = \text{STD}[T], \tag{3.2}$$

which is clearly not the case here. Therefore, we must conclude that we should not consider a filtered Poisson process as a model for the flow of the Delaware River. Indeed, we will try to fit a gamma distribution to the data. Knowing that $E[T] = \alpha/\lambda$ and $\text{STD}[T] = \sqrt{\alpha}/\lambda$ if T has a gamma distribution with parameters α and λ, we deduce from (3.1) that

$$\hat{\alpha} \simeq 3.64, \qquad \hat{\lambda} \simeq 0.31. \tag{3.3}$$

Applying a chi-square goodness-of-fit test, we find that we can accept this model with a large p-value (of approximately 0.39). The test is summarized in Table 3.1.

In this table, n_j is the number of peak flow values observed in interval j (in days), p_j is the probability of having an observation in interval j if the model is correct, and m_j is the expected number of observations in interval j (again, if the model is correct). We obtain a D^2 statistic equal to approximately 1.895, which is compared to the quantiles of a chi-square distribution with two degrees of freedom.

Remark 3.1. When we apply a chi-square goodness-of-fit test, we assume that the data are *independent* observations of the random variable, which is not exactly true here. We could use only a subset of the data that are (almost) uncorrelated. However, since

Table 3.2. Goodness-of-fit test of a Rayleigh distribution for the random variable T.

j	$[0,10)$	$[10,15)$	$[15,20)$	$[20,\infty)$
n_j	27	19	8	7
p_j	0.437	0.288	0.174	0.101
m_j	26.66	17.57	10.61	6.16

the number of observations and the correlation coefficient of the consecutive observations are not large, we prefer to keep all the data points.

Although the model $T \sim \text{Gamma}\,(\alpha = 3.64; \lambda = 0.31)$ is surely acceptable, we also try to fit a Rayleigh distribution to the data. That is, we look for a random variable with density function

$$f_T(t) = \frac{t}{\alpha^2} e^{-t^2/2\alpha^2} \quad \text{for } t \geq 0, \tag{3.4}$$

where α is an unknown parameter. We have

$$E[T] \overset{(*)}{=} \left(\frac{\pi}{2}\right)^{1/2} \alpha, \qquad \text{STD}[T] \overset{(**)}{=} \left(2 - \frac{\pi}{2}\right)^{1/2} \alpha. \tag{3.5}$$

From (3.1), we find that (*) implies that $\alpha \simeq 9.3265$, while (**) yields $\alpha \simeq 9.3492$. Hence, the model seems very good. Using Table 3.2, a chi-square goodness-of-fit test is performed with $\alpha = 9.33$. This time, we obtain $D^2 \simeq 1.026$, which corresponds to a p-value of approximately 0.60. This is therefore the model that we will use for the random variable T.

Remark 3.2. Actually, if we use the same intervals as in Table 3.1, we obtain that the p-value is approximately 0.20, which is less good than with the gamma distribution. However, when the number of observations is not very large, the choice of the intervals plays a big role in the conclusion of the test. At any rate, there is the same number of degrees of freedom in both tests, since we only had one parameter to estimate in the case of the Rayleigh distribution. Furthermore, there is another reason to prefer the Rayleigh distribution, as will be seen below.

Next, we find that the mean value of the time between a minimum flow and the following maximum is approximately 4.044 days. Hence, we set

$$kc \simeq 4. \tag{3.6}$$

Then, we write that (see (2.15))

$$k \simeq \ln \frac{(X(t+\delta)/X(t))}{\ln(1+\delta/4) - \delta/4}, \tag{3.7}$$

where δ is the number of days between the maximum at time t and the following minimum. We obtain that the mean value of k is approximately 3.85, so that we have $c \simeq 2.16$.

Therefore, the model that we propose for the flow of the Delaware River at time t is

$$X(t) = \sum_{n=1}^{N(t)} Y_n \left(t - \tau_n\right)^{1.85} e^{-(t-\tau_n)/2.16}. \tag{3.8}$$

Remark 3.3. Let

$$g_1(t) = t^{1.85} e^{-t/2.16}. \tag{3.9}$$

In obtaining the formula for the parameter k, we assumed that (2.12) is valid. This implies that we can neglect the events that occurred before time $\tau_{N(t)}$. Because the mean value of the time elapsed since the previous event occurred, when we are at a maximum (resp., minimum) flow, is approximately $11.7 + 4 = 15.7$ (resp., $2 \times 11.7 = 23.4$) days, we can indeed neglect the events before time $\tau_{N(t)}$. To see why this is true, notice that the maximum value of $g_1(t)$ is approximately 2 (attained around $t = 4$), while $g_1(15.7) \simeq 0.11$ (resp., $g_1(23.4) \simeq 0.0067$). Thus, a signal of a given size that occurred 15.7 (resp., 23.4) days ago is almost 20 (resp., 303) times less important than one of the same size that produced the current peak flow.

Now, because we prefer to use the Rayleigh distribution as a model for the random variable T, and since the square root of T has an exponential distribution (see Section 2), we should take the square root of all the time variables before estimating the parameters k and c in the model. This is advisable, because the transformed process is then a filtered Poisson process, for which many exact and explicit results are known (see [3]), which is not the case when $\{N(t),\ t \geq 0\}$ is a general filtered renewal process.

Proceeding as mentioned above, we find that ($kc \simeq 1.93$)

$$k \simeq 1.983, \qquad c \simeq 0.973, \tag{3.10}$$

so that

$$X(t) = \sum_{n=1}^{N(t)} Y_n \left(t - \tau_n\right)^{1.983} e^{-(t-\tau_n)/0.973}, \tag{3.11}$$

in which t is measured in square roots of days.

To complete this work, we must find a model for the distribution of the Y_n random variables. It is often assumed that an appropriate distribution for these random variables is an exponential one. For the years 1993–2002, we find that the mean value and the standard deviation of the signals are

$$\bar{y} \simeq 15468, \qquad s_Y \simeq 15202. \tag{3.12}$$

This tends to confirm that an exponential distribution could be adequate for the Y_n's. Performing a chi-square goodness-of-fit test, we obtain $D^2 \simeq 3.545$, which corresponds to a p-value of around 0.17. Thus, though not an exceptional fit, the exponential distribution is indeed acceptable for the magnitude of the signals. Moreover, as already mentioned above, only the mean value of the observed signals will be used to forecast the peak flows, so that the actual distribution of the Y_n's is not needed.

3.2. Forecasting. Based on the model fitted in the preceding subsection, the value of the forecasted peak flow that will follow the current minimum flow is given by

$$\widehat{\text{Peak}}_1 = \text{Max}\,(1.93 + N_D)^{1.983} e^{-(1.93+N_D)/0.973} + 15468, \tag{3.13}$$

where 15468 is the mean difference between the peaks and the preceding minima during the years 1993–2002 and the variable N_D is measured in square roots of days. Using this predictor, we find that the correlation coefficient between the observed and forecasted peaks for these years is $r \simeq 0.489$.

Remark 3.4. With the model (3.8), we obtain that $r \simeq 0.475$. Therefore, taking the square root of the time variables slightly improves the results.

Now, the correlation coefficient between the consecutive pairs of flows is actually approximately 0.416. Therefore, we can conclude that the model (3.11) has enabled us to improve the forecasts of peak flow values. However, the usefulness of the model is more apparent when we really forecast the peak flows rather than finding a model once the observations have been gathered.

That is, if we first estimate the parameters k and c in the model by using the data from the years 1993–1997, we obtain that

$$X(t) = \sum_{n=1}^{N(t)} Y_n (t - \tau_n)^{2.178} e^{-(t-\tau_n)/0.926}, \tag{3.14}$$

from which we deduce that the estimator $\widehat{\text{Peak}}_1$ becomes

$$\widehat{\text{Peak}}_1 \simeq \text{Max}\,(2.02 + N_D)^{2.178} e^{-(2.178+N_D)/0.926} + 18163 \tag{3.15}$$

(because $\widehat{kc} \simeq 2.02$ and $\bar{I} \simeq 18163$ for these years).

The correlation coefficient between the forecasted and observed peak flows obtained for the years 1998–2002 is $r \simeq 0.347$. This result is more impressive when we compute the value of r between the observed peak flows during this time period; indeed, we obtain that $r \simeq -0.206$ (whereas $r \simeq 0.516$ for the years 1993–1997). Thus, the filtered renewal process model has been able to transform a negative correlation coefficient into a relatively high (and positive) correlation coefficient.

To get a better idea of the value of the estimator $\widehat{\text{Peak}}_1$ defined in (3.15), we will compute r when we use the estimator (see (2.17))

$$\widehat{\text{Peak}}_2 = \text{Min} + 18163. \tag{3.16}$$

We obtain that $r \simeq 0.084$. Remember that this estimator is the one that corresponds to a filtered Poisson process with a simple response function of the form $e^{-(t-\tau_n)/c}$ (i.e., $k = 0$).

Next, if we try to forecast the peak flows observed during the years 1998–2002 by making use of the linear regression equation obtained from the data for the 1993–1997 time period, we first compute the regression equation

$$\widehat{\text{Peak}}_3 \simeq 0.320\,\text{Max} + 16111. \tag{3.17}$$

The correlation coefficient between the peaks forecasted by this regression equation and the observed peaks in 1998–2002 is $r \simeq -0.206$, as should be. We can at least improve the results by computing the regression equation

$$\widehat{\text{Peak}}_4 = 11977 + 0.241\,\text{Max} + 0.283\,\text{Min} + 0.283\,\text{Min}_{+1}, \tag{3.18}$$

where Min_{+1} is the value of the river flow observed one day after the current minimum flow (i.e., the most recent data point available when we want to forecast the oncoming peak). Although this predictor uses more variables and the most recent data point obtained, the correlation coefficient computed between the forecasted and observed peaks in 1998–2002 is only $r \simeq 0.053$.

We have considered other estimators of the future peak flow values, whether based on the previous models or not, and we were not able to *beat* the correlation coefficient $r \simeq 0.347$ obtained with $\widehat{\text{Peak}}_1$.

4. Conclusions

In this paper, we have developed a filtered renewal process for the flow of a river. The estimation of the parameters in the model is based on the observed maxima and minima of the river flow. The model is intended to be used to forecast the oncoming peak flow when we notice that the flow has begun to increase from a minimum value.

For simplicity and/or tractability reasons, many authors have used over-simplifying assumptions in this type of models. We have seen in the application to the Delaware River that the fact of assuming that the response function $g(t)$ has the form $e^{-(t-\tau_N)/c}$ leads to a poor predictor of the future peaks. Similarly, the assumption that the river flow can be modeled as a filtered Poisson process implies that the time between the various signals is exponentially distributed, which is clearly wrong in most applications.

We have also seen that linear regression, which has been found to provide very good results when the aim is to forecast the river flow over a short period of time, cannot compete with the filtered renewal process when it comes to forecasting peak flows, and this even if we use more variables and the most recent data point in the regression model.

The data set that we used to compare the various predictors considered in this paper is rather special. Indeed, the correlation coefficient between the consecutive peaks during the first five years (1993–1997) is relatively high and positive (0.516), while it is small and negative (-0.206) for the last five years (1998–2002). However, we feel that it is in such a challenging situation that the quality of an estimator can be established.

Trying to forecast peak flow values is surely a difficult task. To be more worthwhile, this task should only be attempted when the correlation coefficient between the peaks is high enough. In practice, this implies that the time between the consecutive peaks should not be too large. Moreover, we have decided to forecast the oncoming peak when we have observed that the river flow has just moved from a minimum value.

One way of rendering the filtered renewal process even more realistic would be to choose a response function that is not deterministic. In reality, the river flow does not decrease in a perfectly regular way from a maximum value. Rather, the descent is more

or less smooth. Such behavior could be obtained with a response function which is a random variable having a given distribution.

Finally, another subject on which more work is needed is a method to estimate the parameters in the filtered renewal process when we cannot assume that (2.14) holds, that is, when we cannot neglect all the signals that occurred before the most recent one.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] M. Lefebvre, *Short-term hydrological forecasts using linear regression*, Rev. Sci. Eau **16** (2003), no. 2, 255–265.

[2] M. Lefebvre, J. Ribeiro, J. Rousselle, O. Seidou, and N. Lauzon, *Probabilistic prediction of peak flood discharges*, Proc. 9th International Conference on Applications of Statistics and Probability in Civil Engineering Vol. 1 (Calif, 2003) (A. Der Kiureghian, S. Madanat, and J. M. Pestana, eds.), Millpress, Rotterdam, 2003, pp. 867–871.

[3] E. Parzen, *Stochastic Processes*, Holden-Day Series in Probability and Statistics, Holden-Day, California, 1962.

[4] S. Rahman and M. Grigoriu, *Markov model for seismic reliability analysis of degrading structures*, J. Struct. Eng. **119** (1993), no. 6, 1844–1865.

[5] J. Ribeiro-Corréa, *Étude de quelques problèmes reliés à l'estimation des débits de crue*, Ph.D. thesis, Department of Civil Engineering, École Polytechnique de Montréal, Montréal, 1994.

[6] J. Yoon and M. L. Kavvas, *Probabilistic solution to stochastic overland flow equation*, J. Hydrol. Eng. **8** (2003), no. 2, 54–63.

Mario Lefebvre: Département de Mathématiques et de Génie Industriel, École Polytechnique, C.P. 6079, Succursale Centre Ville, Montréal, Québec, Canada H3C 3A7

E-mail address: mlefebvre@polymtl.ca