

Research Article

An Application of Filtered Renewal Processes in Hydrology

Mario Lefebvre and Fatima Bensalma

Département de Mathématiques et de Génie Industriel, École Polytechnique, C.P. 6079, Succursale Centre-ville, Montréal, QC, Canada H3C 3A7

Correspondence should be addressed to Mario Lefebvre; mlefebvre@polymtl.ca

Received 21 January 2014; Revised 12 April 2014; Accepted 13 April 2014; Published 5 May 2014

Academic Editor: George S. Dulikravich

Copyright © 2014 M. Lefebvre and F. Bensalma. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Filtered renewal processes are used to forecast daily river flows. For these processes, contrary to filtered Poisson processes, the time between consecutive events is not necessarily exponentially distributed, which is more realistic. The model is applied to obtain one- and two-day-ahead forecasts of the flows of the Delaware and Hudson Rivers, both located in the United States. Better results are obtained than with filtered Poisson processes, which are often used to model river flows.

1. Introduction

Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ . A filtered Poisson (sometimes called *shot noise*) process is a continuous-time stochastic process $\{X(t), t \geq 0\}$ defined by

$$X(t) = \sum_{n=1}^{N(t)} w(t, \tau_n, Y_n) \quad (X(t) = 0 \text{ if } N(t) = 0), \quad (1)$$

in which the random variables τ_1, τ_2, \dots denote the arrival times of the events of the Poisson process and Y_1, Y_2, \dots are assumed to be independent and identically distributed random variables that are also independent of $\{N(t), t \geq 0\}$. The function $w(\cdot, \cdot, \cdot)$ is called the response function.

In many applications, the response function is chosen of the form

$$w(t, \tau_n, Y_n) = Y_n e^{-(t-\tau_n)/c}, \quad (2)$$

where c is a parameter that must be estimated. It then gives the value at time t of an event of magnitude Y_n of the Poisson process that occurred at time τ_n . Moreover, the random variables Y_1, Y_2, \dots are generally assumed to be exponentially distributed with parameter μ . With the above response function, the filtered Poisson process behaves as in Figure 1. Actually, this behavior depends on the form of the response function, but not on the distribution of the time between the events. Therefore, the same behavior would be

observed in the case when $\{N(t), t \geq 0\}$ is a renewal process. Remember that a Poisson process is a particular renewal process.

To model the daily flows of rivers given their most recent observed values, conceptual and physical models based on the filtered Poisson process have been widely used successfully for many decades; see, for example, Weiss [1], Kelman [2], Koch [3], and Konecny [4]. Filtered Poisson processes are still being used to model various phenomena in civil engineering; see Yin et al. [5] and Miyamoto et al. [6].

Now, especially in hydrological applications, the form of the response function in (2) is taken for granted rather than being justified by the observations of the variable of interest. Actually, it is generally simply an assumption made to obtain a mathematically tractable model.

Similarly, the actual distribution of the Y_n 's is not investigated. However, if one is only interested in forecasting the value of $X(t+\delta)$, based on the values of the process up to time t , this does not really cause a problem because the forecast is normally based on the mean of the Y_n 's, and the distribution itself is not needed.

Finally, and even more importantly, the assumption that $\{N(t), t \geq 0\}$ is indeed a Poisson process is not tested. Again, this is a simplifying assumption, because one can then make use of the nice properties of the Poisson process, notably the fact that it has independent and stationary increments.

Next, notice that, with the response function in (2), the effect on the value of the process of an event that occurred at

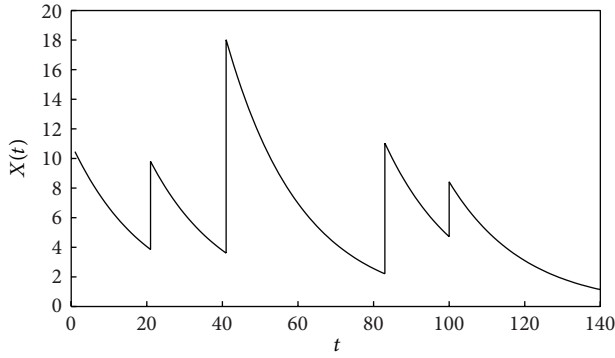


FIGURE 1: Example of a trajectory of a filtered Poisson process defined by (1) and (2).

time τ_n is maximum at that time instant and $\{X(t), t \geq 0\}$ has discontinuities at the arrival times of the events of the Poisson process. In the case of the application in hydrology that consists in modeling the flow of rivers, if one looks at real hydrographs, one sees that there is a more or less extended period of time during which the flow increases from a minimum to a maximum. Even if river flows data are generally available on a daily basis, one does not observe very sudden increases of the flow, followed by an exponential decrease. Rather, it takes on average two to three days for the flow to reach a maximum, and the decrease is more or less rapid.

To obtain a more realistic model for the variations of river flows, by taking into account the periods of flow increase, Lefebvre and Guilbault [7] used the following response function:

$$w(t, \tau_n, Y_n) = Y_n(t - \tau_n)^d e^{-(t - \tau_n)/c}, \quad (3)$$

where d is a positive parameter. With this response function, the river flow begins to increase immediately after time τ_n , but the maximum is reached after dc time units (i.e., generally days) and then it starts to decrease. Their work was improved by the authors, who provided a method to estimate the parameter d . They found that, for the applications that they considered, d was in the interval (0,1).

With a Poisson process being a continuous-time Markov chain, the time that the process spends in a given state is exponentially distributed, which is a strictly decreasing density function. Let T_1, T_2, \dots be the times between two successive events, that is, the interarrival times. In the case of river flows, the interarrival time is generally a random variable with a density function that is increasing at first. Hence, the assumption that $\{N(t), t \geq 0\}$ is a Poisson process is at least doubtful.

In Lefebvre [8], the author tested the hypothesis that the T_n 's are exponentially distributed. He found that, for the Delaware River, it was not accurate. Instead, it was shown that both a gamma and a Rayleigh distribution fitted the data much better. The objective was to forecast the peak flows of the Delaware River, which is a difficult task due to the very weak correlation between the successive peaks. To try to forecast these peak flows, the author used a filtered renewal

process as a model for the river flow. An advantage in using a Rayleigh distribution for the interarrival time is that the filtered renewal process can be transformed into a filtered Poisson process by working on a different time scale.

Poisson and renewal processes have been used in hydrology to model low flows (see, for instance, Loaiciga and Leipnik [9] and Yagouti et al. [10]) as well as various physical phenomena such as traffic noise (see Marcus [11]). Andersen [12, 13] gave new theoretical results on filtered renewal processes (see also the references therein).

The aim of the present paper is to show that we can obtain more accurate short-term forecasts of river flows by making use of filtered renewal processes rather than filtered Poisson processes. To do so, we will first need to develop a formula to forecast the flow at time $X(t + \delta)$, given the history of the process up to time t . Because we cannot appeal to the memoryless property of the exponential distribution, the task of finding an estimator for $X(t + \delta)$ is more difficult. The values of the river flow before time t must be taken into account, contrary to the case of filtered Poisson process, for which the forecasts depend only on $X(t)$.

Moreover, because we want to clearly see the improvement obtained by using a more realistic distribution for the interarrival time, we will take the response function defined in (2). Indeed, as mentioned above, filtered Poisson processes having this response function are commonly used in hydrology, and the forecasting formulas are then easy to derive and implement. Therefore, it will be easier to judge the quality of the model that we propose.

In the next section, filtered renewal processes will be formally defined and the formulas needed to forecast the river flows will be developed. Then, in Section 3, the results will be applied to forecast the flows of the Delaware and Hudson Rivers. Finally, a few concluding remarks will be made in Section 4.

2. River Flows Modeled as a Filtered Renewal Process

Let $X(t)$ denote the flow of a certain river at time t . We assume that

$$X(t) = \sum_{n=1}^{N(t)} Y_n e^{-(t - \tau_n)/c} \quad (X(t) = 0 \text{ if } N(t) = 0), \quad (4)$$

in which $\{N(t), t \geq 0\}$ is a renewal process. The τ_n 's are the arrival times of the events of the process $\{N(t), t \geq 0\}$, and the times $T_n = \tau_n - \tau_{n-1}$ between the successive renewals are independent and identically distributed random variables. Finally, the (positive) random variables Y_1, Y_2, \dots are independent and identically distributed (and independent of $\{N(t), t \geq 0\}$).

Thus, the process is characterized by the sudden increases of the river flow caused by the events that occur at times τ_1, τ_2, \dots . A renewal cycle is defined as the time T_n between two consecutive peaks. For the interarrival time, the following distributions (in addition to the exponential distribution) will be considered in Section 3: lognormal, gamma, Weibull, and Rayleigh.

Next, in order to be able to forecast the river flow at time $t + \delta$, we must first estimate the parameter c that appears in the response function. If there are no signals between t and $t + 1$, then $X(t + 1)$ is given by

$$X(t + 1) = e^{-1/c} X(t). \quad (5)$$

Hence, to estimate the parameter c , we can consider all the days where the flow decreased in the data set and calculate the ratio $X(t + 1)/X(t)$. The point estimate of c is then derived from the arithmetic mean \bar{R} of this ratio:

$$\hat{c} = -\frac{1}{\ln \bar{R}}. \quad (6)$$

In the case of the random variables Y_n , which constitute a random sample of a parent random variable Y , their exact distribution is not needed. Indeed, only the expected value of Y will be used to forecast the value of $X(t + \delta)$. Therefore, we simply have to compute the mean increase of the river flow during the calibration period of the model.

We will now derive formulas to estimate the river flow one and two days in advance. These formulas can be used with any distribution for the interarrival time T , which denotes the parent variable of the T_n 's.

2.1. Forecasting the Flow at Time $t + 1$. We will first try to forecast the river flow at time $t + 1$, given all the information available up to t . To do so, we assume that there will be at most one event in the interval $(t, t + 1]$. At any rate, as mentioned above, in practice flow values are generally available on a daily basis. Therefore, it is not really possible to determine whether more than one event occurred in $(t, t + 1]$. When it does happen, these events can be considered as a single large event.

Apart from the case when the interarrival time T has an exponential distribution, we must take the history of the process into account to calculate the probability that there will indeed be an event in $(t, t + 1]$. Let

$$U_i = \tau_{N(t)+i} - \tau_{N(t)+i-1} \quad \text{for } i = 1, 2, \dots \quad (7)$$

That is, U_i is the time elapsed between the $(i - 1)$ st and the i th event after the most recent one that occurred before or at time t . The random variables T and U_1 are identically distributed. We must compute

$$p_k := P[U_1 \leq k + 1 \mid U_1 > k], \quad (8)$$

where k is the (integer) number of days since the most recent signal was observed. We have

$$p_k = \frac{\int_k^{k+1} f_{U_1}(u) du}{\int_k^{\infty} f_{U_1}(u) du}. \quad (9)$$

With the help of a mathematical software, we can obtain numerical values of this probability for any density function f_{U_1} and any value of k .

Next, if T is exponentially distributed, it is well known that, given that there is a signal (i.e., an event) between t and

$t + 1$, the conditional density of T is uniform on the interval $(t, t + 1]$, so that the signal in question occurred on average at time $t + 1/2$. In the general case, we must compute the conditional density function of U_1 in the interval $(k, k + 1]$, given that $U_1 > k$.

If there is indeed a signal between t and $t + 1$, the forecast of the river flow at time $t + 1$ would be

$$Z_k := X(t) e^{-1/\hat{c}} + E[Y] \frac{1}{P[U_1 > k]} \times \int_k^{k+1} e^{-(k+1-u)/\hat{c}} f_{U_1}(u) du. \quad (10)$$

Therefore, the forecast of the river flow at time $t + 1$, given the entire history of the process up to t , is given by

$$\begin{aligned} \widehat{X}(t + 1) &= X(t) e^{-1/\hat{c}} (1 - p_k) + Z_k p_k \\ &= X(t) e^{-1/\hat{c}} + p_k E[Y] \frac{1}{P[U_1 > k]} \\ &\quad \times e^{-(k+1)/\hat{c}} \int_k^{k+1} e^{u/\hat{c}} f_{U_1}(u) du. \end{aligned} \quad (11)$$

2.2. Forecasting the Flow at Time $t + 2$. Let $U_{1,k} = U_1 \mid \{U_1 > k\}$. To forecast the river flow at time $t + 2$, given all the information available up to time t , we need the distribution of the random variable $S_k := U_{1,k} + U_2$, where U_2 is defined in (7).

We now assume that the probability that there will be more than two signals in the interval $(t, t + 2]$ is negligible (irrespective of the value of k). The density function of S_k is given by the convolution of the density functions of $U_{1,k}$ and U_2 :

$$f_{S_k}(s + k) = \int_k^{\infty} f_{U_{1,k}}(u) f_{U_2}(s + k - u) du. \quad (12)$$

Since U_2 is a nonnegative random variable, we can write that

$$f_{S_k}(s + k) = \int_k^{s+k} f_{U_{1,k}}(u) f_{U_2}(s + k - u) du \quad \text{for } s > 0. \quad (13)$$

In the case when T has an exponential distribution with parameter λ , we know that $(S_k - k)$ has a gamma distribution with parameters $\alpha = 2$ and λ . In the other cases, we must generally evaluate the above integral numerically.

The value of $X(t+2)$ is given by

$$X(t+2) = \begin{cases} e^{-2/c} X(t) & \text{if } U_{1,k} > k+2, \\ e^{-2/c} X(t) + Y_1^* e^{-(k+2-U_{1,k})/c} & \text{if } U_{1,k} \leq k+2 \\ & \text{and } U_2 > 2 - (U_{1,k} - k), \\ e^{-2/c} X(t) + Y_1^* e^{-(k+2-U_{1,k})/c} & \text{if } U_{1,k} + U_2 \leq k+2 \\ + Y_2^* e^{-(2-U_2)/c} & (\text{and } U_3 > 2 - (U_{1,k} - k) - U_2), \end{cases} \quad (14)$$

where Y_i^* denotes the size of the i th signal in the interval $(t, t+2]$. As mentioned above, we assume that

$$P[U_3 > 2 - (U_{1,k} - k) - U_2] \simeq 1. \quad (15)$$

We must first compute

$$p_{1,k} := P[U_{1,k} \in (k, k+2)]. \quad (16)$$

When U_1 has an exponential distribution with parameter λ ,

$$p_{1,k} = P[U_1 \in (0, 2]] = 1 - e^{-2\lambda}. \quad (17)$$

In general, we have

$$p_{1,k} = \frac{\int_k^{k+2} f_{U_1}(u) du}{\int_k^\infty f_{U_1}(u) du}. \quad (18)$$

Next,

$$\begin{aligned} p_{2,k} &:= P[U_{1,k} + U_2 \in (k, k+2)] \equiv P[S_k \in (k, k+2)] \\ &= \int_0^2 f_{S_k}(s+k) ds \\ &= \int_0^2 \left[\int_k^{s+k} f_{U_{1,k}}(u) f_{U_2}(s+k-u) du \right] ds. \end{aligned} \quad (19)$$

Alternatively, using the fact that $U_{1,k}$ and U_2 (≥ 0) are independent random variables, we may write that

$$\begin{aligned} p_{2,k} &:= P[U_{1,k} + U_2 \in (k, k+2)] \\ &= \int_k^{k+2} P[U_{1,k} + U_2 \in (k, k+2) | U_{1,k} = u] f_{U_{1,k}}(u) du \\ &= \int_k^{k+2} P[U_2 \in (k-u, k+2-u)] f_{U_{1,k}}(u) du \\ &= \int_k^{k+2} \left[\int_0^{k+2-u} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du. \end{aligned} \quad (20)$$

Finally, by independence,

$$\begin{aligned} &P[\{U_{1,k} \leq k+2\} \cap \{U_2 > 2 - (U_{1,k} - k)\}] \\ &= \int_k^{k+2} P[U_2 \in (k+2-u, \infty)] f_{U_{1,k}}(u) du \\ &= \int_k^{k+2} \left[\int_{k+2-u}^\infty f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du \\ &= \int_k^{k+2} \left[1 - \int_0^{k+2-u} f_{U_2}(v) dv \right] f_{U_{1,k}}(u) du \\ &= p_{1,k} - p_{2,k}. \end{aligned} \quad (21)$$

The forecast of the river flow at time $t+2$, given the entire history of the process up to time t , is thus given by

$$\begin{aligned} \widehat{X}(t+2) &= X(t) e^{-2/\bar{c}} (1 - p_{1,k}) \\ &\quad + \left\{ X(t) e^{-2/\bar{c}} + E[Y] \int_k^{k+2} e^{-(k+2-u)/\bar{c}} f_{U_{1,k}}(u) du \right\} \\ &\quad \times (p_{1,k} - p_{2,k}) \\ &\quad + \left\{ X(t) e^{-2/\bar{c}} + E[Y] \int_k^{k+2} e^{-(k+2-u)/\bar{c}} f_{U_{1,k}}(u) du \right. \\ &\quad \left. + E[Y] \int_k^{k+2} \left[\int_0^{k+2-u} e^{-(2-v)/\bar{c}} f_{U_2}(v) dv \right] \right. \\ &\quad \left. \times f_{U_{1,k}}(u) du \right\} p_{2,k} \\ &= X(t) e^{-2/\bar{c}} + \left\{ E[Y] \int_k^{k+2} e^{-(k+2-u)/\bar{c}} f_{U_{1,k}}(u) du \right\} p_{1,k} \\ &\quad + \left\{ E[Y] \int_k^{k+2} \left[\int_0^{k+2-u} e^{-(2-v)/\bar{c}} f_{U_2}(v) dv \right] \right. \\ &\quad \left. \times f_{U_{1,k}}(u) du \right\} p_{2,k} \end{aligned}$$

$$\begin{aligned}
&= X(t) e^{-2/\hat{c}} + \left\{ E[Y] e^{-(k+2)/\hat{c}} \int_k^{k+2} e^{u/\hat{c}} f_{U_{1,k}}(u) du \right\} p_{1,k} \\
&+ \left\{ E[Y] \int_k^{k+2} \left[e^{-2/\hat{c}} \int_0^{k+2-u} e^{v/\hat{c}} f_{U_2}(v) dv \right] \right. \\
&\quad \left. \times f_{U_{1,k}}(u) du \right\} p_{2,k}.
\end{aligned} \tag{22}$$

3. Forecasting the Flows of the Delaware and Hudson Rivers

To assess the quality of the forecasts obtained with the renewal filtered process, we need to apply the formulas developed in the previous section to real-life data. As in [7], we chose the Delaware and Hudson Rivers, located in the United States, at the Montague (01438500) and the North Creek (01315500) gage stations, respectively. The flow values are available on the Internet at the following address: <http://nwis.waterdata.usgs.gov>. To calibrate the model, we used the data from October 2008 to September 2009.

Remarks. (i) When one is dealing with real-life data, things are not as simple as the mathematical models suggest. Looking at the actual hydrograph of the Delaware River, we observe a number of weak peaks that occur while the flow is decreasing and for which the increasing period lasts only one day before the flow resumes its decline. Similarly, there are sometimes small increases of the flow value occurring just after a minimum was observed that last only one day. In building the data set, there is always a subjective part. We decided to neglect these weak peaks, thus considering only the peaks that appear quite clearly in the hydrograph.

(ii) Because the river flow is observed on a daily basis, we must discretize the set of possible values taken by the interarrival time T . We computed the number k of days elapsed between consecutive peaks in the data set. In the case of the first observed peak, k is the number of days elapsed since the beginning of the calibration period. Then, k is obtained by subtracting the arrival times of the consecutive peaks. Notice that k is greater than or equal to 1.

We first consider the Delaware River. For this river, we found that the average value \bar{T} of the interarrival times in the data set is equal to 6.8889 days. Moreover, the standard deviation of the observations is given by $s_T = 3.8977$ days. Based on these values, one may at once conclude that T is very unlikely to be exponentially distributed. Indeed, remember that the mean and the standard deviation of an exponential distribution with parameter λ are both equal to $1/\lambda$.

Next, the value of the point estimate of the (unitless) parameter c (see (6)) is $\hat{c} = 9.2592$, and the average value of the magnitude Y of the signals is given by $\bar{Y} = 984.6121$ cubic feet per second.

For the distribution of the random variable T , we tested the following models:

- (i) an exponential distribution with parameter λ ,

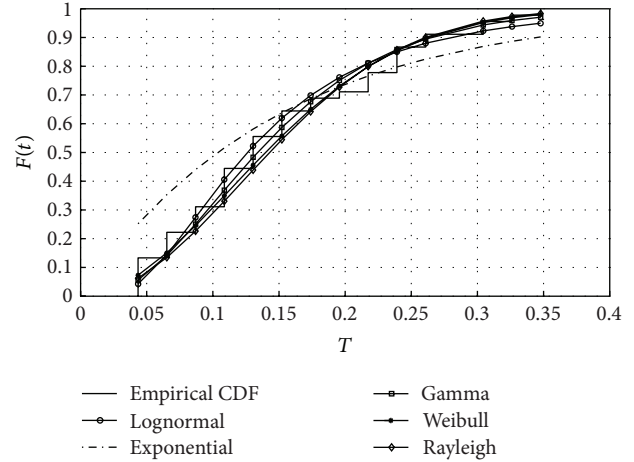


FIGURE 2: Empirical and fitted distribution functions for the random variable T (Delaware River).

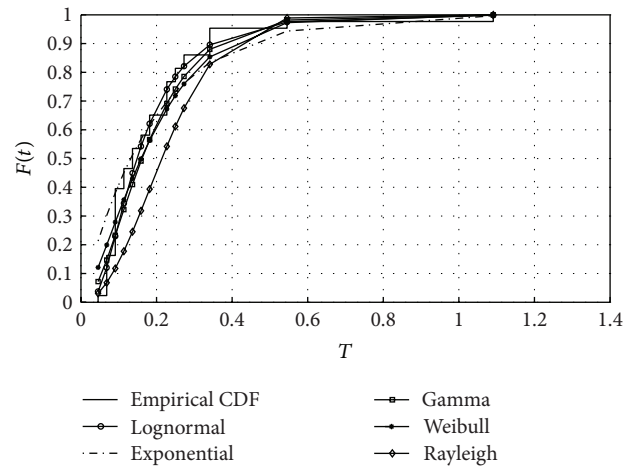


FIGURE 3: Empirical and fitted distribution functions for the random variable T (Hudson River).

- (ii) a lognormal distribution with parameters μ and σ ,
- (iii) a gamma distribution with shape parameter α and scale parameter λ ,
- (iv) a Weibull distribution with shape parameter λ and scale parameter κ ,
- (v) a Rayleigh distribution with parameter σ .

Table 1 summarizes the results of the chi-square goodness-of-fit tests for each distribution and gives the respective parameter estimates. We find that, as expected, the exponential distribution is rejected, whereas the other distributions are all accepted, according to the P -values of the tests. The empirical cumulative distribution function as well as the distribution functions of the various models considered above is shown in Figure 2. We clearly see that the exponential distribution does not fit the data nearly as well as the other four models.

Now, in the case of the filtered Poisson process (1) with the response function defined in (2), to estimate the river flow at

TABLE 1: Goodness-of-fit tests for the distribution of the random variable T (Delaware River).

Distribution	Exponential	Lognormal	Gamma	Weibull	Rayleigh
Parameters	1/6.8889	1.7574, 0.6187	3.0544, 2.2554	7.7974, 1.8975	5.5817
P-value	0.0344	0.7887	0.7723	0.6927	0.7008

TABLE 2: Criteria for comparing the forecasting models (Delaware River).

		Filtered renewal process				Filtered Poisson process
		Lognormal	Gamma	Weibull	Rayleigh	Exponential
$X(t+1)$	Mape	12.58%	12.49%	12.55%	12.56%	12.76%
	Nash	0.8288	0.8290	0.8287	0.8287	0.8356
	Pc	0.1674	0.1664	0.1663	0.1669	0.2522
	Corr	0.9188	0.9186	0.9180	0.9188	0.9187
$X(t+2)$	Mape	20.36%	20.33%	20.34%	20.34%	21.83%
	Nash	0.5447	0.5445	0.5445	0.5445	0.5573
	Pc	0.2930	0.2893	0.2869	0.2863	0.3344
	Corr	0.7735	0.7732	0.7731	0.7730	0.7734

time $t+1$ we compute the expected value of $X(t+1)$ given all the past observations. Because of the memoryless property of the exponential distribution, we find that this conditional expectation only depends on the most recent value of the flow, that is, $X(t)$. We obtain (see [7]) that

$$E[X(t+1) | X(t)] = e^{-1/c} X(t) + E[X(1)], \quad (23)$$

where

$$E[X(1)] = \frac{\lambda c}{\mu} (1 - e^{-1/c}). \quad (24)$$

Similarly, to estimate the river flow at time $t+2$, we only need to calculate

$$\begin{aligned} E[X(t+2) | X(t)] &= E \left[\sum_{n:0 \leq \tau_n \leq t+2} Y_n e^{-(t+2-\tau_n)/c} \mid X(t) \right] \\ &= e^{-2/c} E \left[\sum_{n:0 \leq \tau_n \leq t} Y_n e^{-(t-\tau_n)/c} \mid X(t) \right] \\ &\quad + E \left[\sum_{n:t < \tau_n \leq t+2} Y_n e^{-(t+2-\tau_n)/c} \mid X(t) \right] \\ &= e^{-2/c} X(t) + E[X(2)], \end{aligned} \quad (25)$$

where

$$E[X(2)] = \frac{\lambda c}{\mu} (1 - e^{-2/c}). \quad (26)$$

Moreover, we have that

$$\lim_{t \rightarrow \infty} E[X(t)] = \frac{\lambda c}{\mu}, \quad (27)$$

$$\lim_{t \rightarrow \infty} \text{Var}[X(t)] = \frac{\lambda c}{\mu^2}, \quad (28)$$

$$\lim_{t \rightarrow \infty} \rho_{X(t), X(t+\delta)} = e^{-\delta/c}, \quad (29)$$

where $\rho_{X(t), X(t+\delta)}$ denotes the correlation coefficient of $X(t)$ and $X(t+\delta)$.

By making use of these three equations, we can estimate the parameter λ of the Poisson process, the parameter μ of the assumed exponential distribution of the Y_n 's, and the parameter c in the response function. Actually, we do not need point estimates of λ and μ to forecast the flow. From (27), we deduce that we can simply estimate the ratio $\lambda c / \mu$ that appears in both (24) and (26) by the mean value of the observed data when the process is in steady state. Finally, to estimate the parameter c , we first compute the empirical correlation coefficient r_1 between the flow values at times t and $t+1$ and then we set $\hat{c} = -1 / \ln(r_1)$ (see (29)). With the value of r_1 being close to 1, the previous formula is well defined.

Remark. If one is only interested in forecasting the river flow at time $t+\delta$, one may estimate c by computing the empirical value of $\rho_{X(t), X(t+\delta)}$ and making use of (29).

We are now ready to compare the forecasts derived from both the filtered Poisson process and the filtered renewal process with the various distributions considered for the interarrival time T . We used the formulas given above to forecast the flow of the Delaware River for the 89-day period from February to April 2010.

To assess the accuracy of the forecasts, we consider four criteria commonly used in hydrology: the *mean absolute percentage error* (Mape), the *Nash criterion* (Nash), the *peak criterion* (Pc), and the *correlation coefficient* (Corr) between the observed and forecasted values.

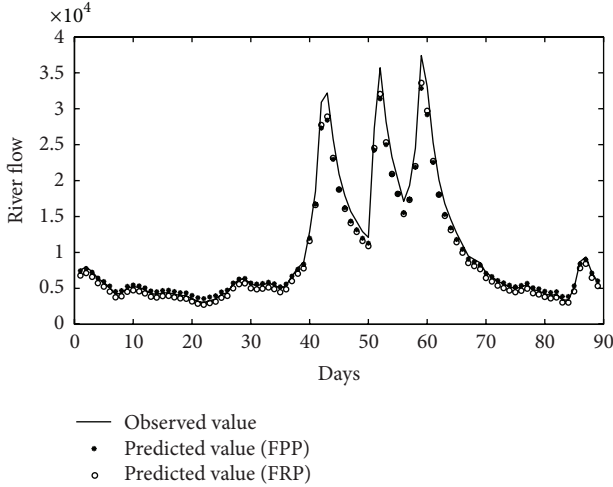
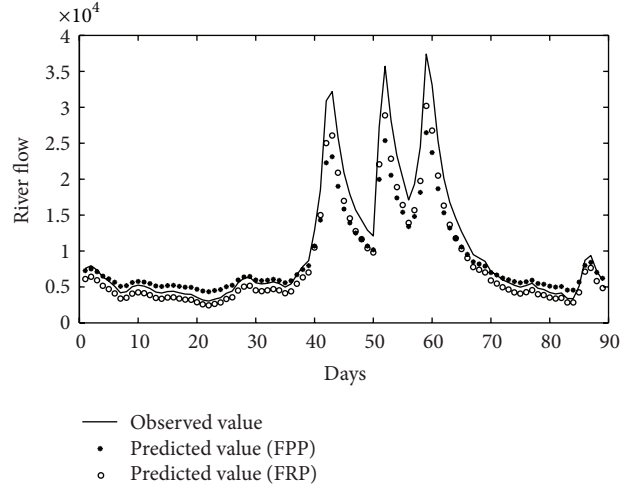
The Nash criterion is based on the mean square error and is used to assess the predictive power of hydrological models. It evaluates the quality of the forecasts from the differences

TABLE 3: Goodness-of-fit tests for the distribution of the random variable T (Hudson River).

Distribution	Exponential	Lognormal	Gamma	Weibull	Rayleigh
Parameters	1/8.3488	1.8754, 0.6625	9.2134, 1.3399	2.1785, 3.8324	7.9978
P -value	0.0013	0.6983	0.3203	0.2532	0.000139

TABLE 4: Criteria for comparing the forecasting models (Hudson River).

		Filtered renewal process			Filtered Poisson process
		Lognormal	Gamma	Weibull	Exponential
$X(t+1)$	Mape	12.33%	12.14%	12.33%	15.58%
	Nash	0.8240	0.8243	0.8240	0.8360
	Pc	0.2875	0.2895	0.2872	0.4818
	Corr	0.9146	0.9146	0.9146	0.9146
$X(t+2)$	Mape	20.84%	21.00%	21.15%	32.78%
	Nash	0.5841	0.5838	0.5835	0.6222
	Pc	0.4558	0.4533	0.4518	0.5575
	Corr	0.7908	0.7909	0.7910	0.7910

FIGURE 4: Observed and forecasted values of the flow of the Delaware River at time $t+1$.FIGURE 5: Observed and forecasted values of the flow of the Delaware River at time $t+2$.

between the expected and observed daily values. It is equal to 1 in case of a perfect fit. For its part, the peak criterion is used to measure the quality of the forecasts during the critical peak period. The closer to 0 it is, the better the forecasts are.

Table 2 displays the results obtained with the two competing models. Looking at these results, we notice that the Mape, Nash, and Corr criteria yield practically the same values for the filtered renewal and filtered Poisson processes. However, the filtered renewal process did much better than the filtered Poisson process when we consider the peak criterion. This criterion is especially important because accurate forecasts are really needed during the peak period.

We now turn to the Hudson River. The value of the point estimate of the constant c for this river is $\hat{c} = 9.1172$ and the average magnitude of the signals is $\bar{Y} = 229.2533$, which is much smaller than in the case of the Delaware River.

Proceeding as above, we performed chi-square goodness-of-fit tests for the various distributions considered. The results

are presented in Table 3 (see also Figure 3). This time, we see that both the exponential and Rayleigh distributions must be rejected, while the lognormal distribution clearly provides the best fit to the data.

The values of the four criteria used to compare the models are shown in Table 4. We discarded the Rayleigh distribution, because it yielded a very bad fit to the data. As in the case of the Delaware River, the values of the Nash and Corr criteria are quite similar, while the important peak criterion and the Mape criterion are much smaller for the filtered renewal process than for the filtered Poisson process.

Finally, we present the observed and forecasted values of the flows of the Delaware and Hudson Rivers at times $t+1$ and $t+2$ in Figures 4, 5, 6, and 7, based on the filtered Poisson process (FPP) and the filtered renewal process (FRP) with a lognormal distribution for the interarrival time T . Notice that the filtered renewal process is generally better than the filtered

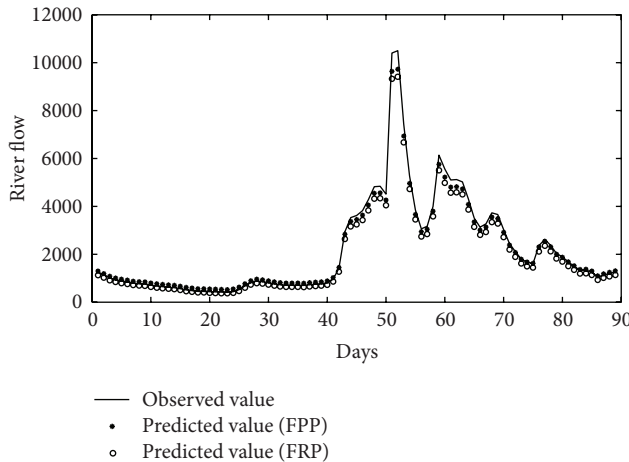


FIGURE 6: Observed and forecasted values of the flow of the Hudson River at time $t + 1$.

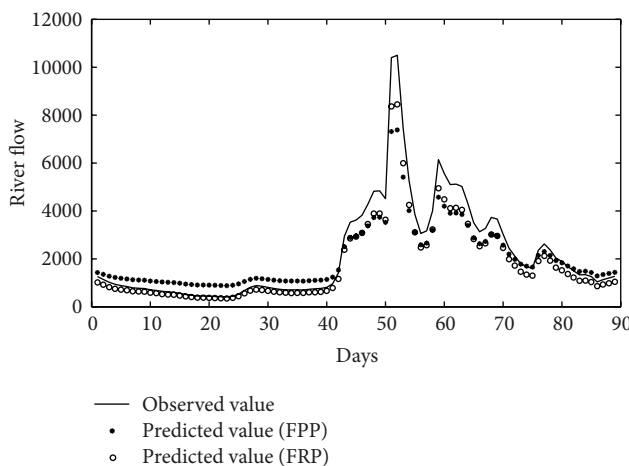


FIGURE 7: Observed and forecasted values of the flow of the Hudson River at time $t + 2$.

Poisson process at forecasting high flow values, which is our main concern.

4. Conclusion

In this paper, we were able to significantly improve the short-term hydrological forecasts produced by filtered Poisson processes by using a more realistic distribution for the inter-arrival time of the signals, thus working with filtered renewal processes. As we saw in the previous section, the filtered Poisson process, despite its lack of realism, is able to yield reasonable short-term forecasts. One possible explanation is the fact that the forecasting formulas derived from this model are very easy to implement. Also, contrary to the filtered renewal process, it does not require subjective decisions to estimate its various parameters. However, the filtered Poisson process was not able to produce forecasts as accurate as those derived from the filtered renewal process during the important peak period.

In addition to being used to forecast flow values, the models considered in this paper can also give us an estimate of the probability that the river flow will exceed a certain threshold in the next few days. This threshold may be a value that corresponds to a flow above which the risk of flooding is very high.

Finally, we could use the response function in (3) to further improve the forecasts. To do so, we would have to be able to estimate the parameter d in that response function and then obtain formulas that generalize the ones in (11) and (22), which is probably not an easy task.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors want to thank the anonymous reviewers for the constructive comments.

References

- [1] G. Weiss, "Shot noise models for the generation of synthetic streamflow data," *Water Resources Research*, vol. 13, no. 1, pp. 101–108, 1977.
- [2] J. Kelman, "A stochastic model for daily streamflow," *Journal of Hydrology*, vol. 47, no. 3-4, pp. 235–249, 1980.
- [3] R. W. Koch, "A stochastic streamflow model based on physical principles," *Water Resources Research*, vol. 21, no. 4, pp. 545–553, 1985.
- [4] F. Konecny, "On the shot-noise streamflow model and its applications," *Stochastic Hydrology and Hydraulics*, vol. 6, no. 4, pp. 289–303, 1992.
- [5] Y.-J. Yin, Y. Li, and W. M. Bulleit, "Stochastic modeling of snow loads using a filtered Poisson process," *Journal of Cold Regions Engineering*, vol. 25, no. 1, pp. 16–36, 2010.
- [6] H. Miyamoto, J. Morioka, K. Kanda et al., "A stochastic model for tree vegetation dynamics in river courses with interaction by discharge fluctuation impacts," *Journal of Japan Society of Civil Engineers B1: Hydraulic Engineering*, vol. 67, no. 4, pp. 1405–1410, 2012.
- [7] M. Lefebvre and J.-L. Guibault, "Using filtered Poisson processes to model a river flow," *Applied Mathematical Modelling*, vol. 32, no. 12, pp. 2792–2805, 2008.
- [8] M. Lefebvre, "A filtered renewal process as a model for a river flow," *Mathematical Problems in Engineering*, vol. 2005, no. 1, pp. 49–59, 2005.
- [9] H. A. Loaiciga and R. B. Leipnik, "Stochastic renewal model of low-flow streamflow sequences," *Stochastic Hydrology and Hydraulics*, vol. 10, no. 1, pp. 65–85, 1996.
- [10] A. Yagouti, I. Abi-Zeid, T. B. M. J. Ouarda, and B. Bobée, "Revue de processus ponctuels et synthèse de tests statistiques pour le choix d'un type de processus," *Journal of Water Science*, vol. 14, no. 3, pp. 323–361, 2001.
- [11] A. H. Marcus, "Traffic noise as a filtered Markov renewal process," *Journal of Applied Probability*, vol. 10, no. 2, pp. 377–386, 1973.

- [12] P. K. Andersen, "Filtered renewal processes with a two-sided impact function," *Journal of Applied Probability*, vol. 16, no. 4, pp. 813–821, 1979.
- [13] P. K. Andersen, "A note on filtered Markov renewal processes," *Journal of Applied Probability*, vol. 18, no. 3, pp. 752–756, 1981.

