퍼지 및 지능시스템학회 논문지 2006, Vol. 16, No. 5, pp. 537-543

# 복수의 임펄스 응답을 이용한 강우-유출 해석

# Rainfall-Runoff Analysis Utilizing Multiple Impulse Responses

유철상*, 박주영**
Chulsang Yoo* and Jooyoung Park**

*Dept. of Civil and Environmental Engineering, Korea University
**Dept. of Control and Instrumentation Engineering, Korea University

## 요 약

최근들어 강우-유출 현상에 관한 비선형 모델링에 관하여 많은 연구가 있어 왔는데, 그 중에서도 신경망을 이용한 결과는 매우 성공적인 것으로 보고되어 왔다. 선형 구조가 갖는 근본적인 한계성으로 인하여, 이 분야에 선형 모델을 활용하는 것은 신경망을 사용하는 경우에 비하여 불리할 것으로 여겨지곤 한다. 하지만 우리는, 선형 모델의 경우 주어진 문제가 갖는 근본적 동특성의 원리를 보다 잘 이해할 수 있도록 해주므로, 적절한 확장 과정을 거치면 선형 임펄스 응답의 개념은 매우 경쟁력 있는 도구가 될 수 있을 것으로 생각한다. 이러한 생각에 따라, 본 논문에서 우리는 복수의 임펄스 응답의 이용을 강우-유출 해석의 문제에 적용하는 방안을 제안한다. 제안된 방법은 복수의 임펄스 응답 모델 사이에 적용되는 단순하고 고정된 스위칭 전략에 기반을 두고 있으며, 각 임펄스 응답은 음이 아닌 성분을 갖도록 하고, 동시에 한개의 봉우리만 갖는 형태를 만족하도록 한다. 우리나라의 특정한 지역의 수문기상학 자료를 대상으로 하여 적용해 본 결과, 제안된 방법은 매우 의미 있는 결과를 제공함을 보여주었다.

**키워드** : 강우-유출 해석, 임펄스 응답, 최적화

## Abstract

There have been many recent studies on the nonlinear rainfall-runoff modeling, where the use of neural networks is shown to be quite successful. Due to fundamental limitation of linear structures, employing linear models has often been considered inferior to the neural network approaches in this area. However, we believe that with an appropriate extension, the concept of linear impulse responses can be a viable tool since it enables us to understand underlying dynamics principles better. In this paper, we propose the use of multiple impulse responses for the problem of rainfall-runoff analysis. The proposed method is based on a simple and fixed strategy for switching among multiple linear impulse-response models, each of which satisfies the constraints of non-negativity and uni-modality. The computational analysis performed for a certain Korean hydrometeorologic data set showed that the proposed method can yield very meaningful results.

**Key Words** : Rainfall-runoff analysis, Impulse response, Optimization

## 1. Introduction

Modeling the rainfall-runoff relationship is one of the most challenging problems in the area of hydrology, where the various factors such as spacial and temporal variability of watershed characteristics and precipitation patterns can be involved. Recently, many successful applications of artificial neural networks have been reported in the water resources literature [1-5]. Due to fundamental limitation of linear structures, studies employing linear models have been relatively rarely attempted in this area. However, we believe that with an appropriate extension, the concept of linear impulse responses can be a viable tool which enables us to understand underlying dynamics. In this paper, we propose the use of multiple linear impulse responses for the problem of rainfall-runoff analysis. Among the works related with the proposed method are the results on the use of competing structures for modeling reported in [6-9]. In [6] and [7], Pawelzik et al. used annealed competing neural networks to segment a non-stationary time series, where non-stationarities are caused by switching dynamics. In [8], Chang et al. extend the results of [6] and [7] toward the use of the competing support vector machines together with a new strategy for adjusting the annealing parameter based on the EM (expectation maximization) algorithm. In [9], the authors of this paper presented a new rainfall-runoff analysis method based on competing linear impulse responses. Although the method of [9] is

접수일자 : 2006년 8월 31일
완료일자 : 2006년 10월 11일
This work was supported by a grant (R01-2003-000-10046-0) from the Basic Research Program of the *Korea Science and Engineering Foundation*. All contributions are gratefully acknowledged.

quite efficient in analyzing the rainfall-runoff data to decompose into multiple linear modes, how to utilize the results for prediction was not clearly addressed in the work. In this paper, we employ a simple fixed switching method together with multiple linear impulse responses satisfying certain constraints in order to find an accurate rainfall-runoff analysis model with reasonable physical interpretations. In the following section, we will explain the proposed method in a step-by-step manner and with mathematical details.

The rest of this paper is organized as follows: In Section 2, the proposed method is presented along with a detailed prcedure for rainfall-runoff analysis. In Section 3, experimental results are reported, and finally in Section 4, concluding remarks are given.

## 2. Proposed Method

Let $r(t)$ and $q(t)$ be the amounts of rainfall and discharge at time $t$, respectively. As is well-known, the amount of discharge is dependent on previous rainfall patterns, thus the rainfall data can be employed as independent input variables for the estimation of discharge value. In general, the main objective of the rainfall-runoff modeling is to find an element $f$ in a certain class of functions which can closely approximate the observed discharge in the following sense:

$$q(t) \cong f(r(t), \cdots, r(t-T+1)) \text{ for each } t. \quad (1)$$

Here, $T$ is the length of the input window, which can be determined using the engineering judgement or the methods such as the rainfall-runoff correlograms [4]. Although the estimation of the discharge $q(t)$ can be made more accurate generally by incorporating $q(t-1)$ as an additional input variable [3], the strategy yields only limited information on how the present river flow results from the rainfall data up to now. Thus in this paper, we stick to the case that independent variables consist of the rainfall data $r(t), \cdots, r(t-T+1)$ only. One of the most simple approach that can be considered along this line would be the case that the function $f$ of (1) is linear. In this case, the equation (1) can be expressed as follows:

$$q(t) \cong f(r(t), \cdots, r(t-T+1)) \cong \sum_{k=0}^{T-1} h(k) r(t-k), \quad (2)$$

where the vector $[h(0), \cdots, h(T-1)]$ shows how the river flow results from the recent rainfall patterns, which is often called the impulse response because $h(n)$ represents the response of the model at time $n$ to a unit impulse input occurring at time $0$. Estimating the amount of river flow based on the model (2) may have several pros and cons. One of the most prominent advantages of the above linear impulse response approach is obviously its simplicity. The method essentially seeks pseudo-inverse based solutions for a typical class of least-squares problems. Another strength of this approach is that a reasonable interpretation can be obtained via the concept of the impulse response. As an index for estimation performance, one can use the MSE (mean squared error), which is defined as follows:

$$E = \frac{1}{M} \sum_{t=M_1}^{M_2} [q(t) - f(r(t), \cdots, r(t-T+1))]^2, \quad (3)$$

where $M = M_1 - M_2$, and $[M_1, M_2]$ is the time window during which the input-output data pair are available for the training, and $T$ is the time-window for the linear impulse response. The simple linear impulse response approach has some obvious disadvantages, especially when the system is not assumed to be linear or when the input data contain excessive noise. Then, some of impulse response parameters $h(0), \cdots, h(T-1)$ could be estimated to be negative (below zero), which may lead negative estimation of output variables. Another possible weakness of this approach is that the resulting impulse response curve is not smooth at all, which is quite contrary to the underlying physical principles as well as our intuition. Due to all these problems, the linear impulse response model has not attracted interests in the area of river flow estimations, and nonlinear alternatives such as MLPs (multi-layer perceptrons) and RBF (radial-basis-function) networks have been attempted instead. However, these neural estimators are not only lack of reasonable interpretation since they are basically black-box type estimators, but also often pathologically resulting in negative flow estimations [3]. Thus here we propose to employ the multiple linear impulse response models together with a simple switching strategy for the rainfall-runoff modeling.

In the proposed method, we consider each $q(t)$ to be best approximated by one of $m$ linear unknown functions, i.e.,

$$q(t) \cong f_{x(t)}(r(t), \cdots, r(t-T+1)) = \sum_{k=0}^{T-1} h_{x(t)}(k) r(t-k), \quad t = T_1, \cdots, T_2, \quad (4)$$

where $f_j$ is the $j$-th linear model which is characterized by the impulse response vector $[h_j(0), \cdots, h_j(T-1)]$, and $x(t) = j$ if the $t$-th flow value $q(t)$ is best approximated by the $j$-th linear model $f_j$. For convenience, we will call $f_j$ the $j$-th mode from now on. In addition to the linearity of $f_j$, we will further assume that each mode has the property of non-negativity and uni-modality, i.e.,

$$h_j(k) \geq 0, \; j = 1, \cdots, m, \; k = 0, \cdots, T-1, \quad (5)$$

and

$$h_j(0) \le \cdots \le h_j(t_{max,j}) \ge \cdots \ge h_j(T-1) \quad (6)$$

for some $t_{max,j} \in \{0, \cdots, T-1\}$. Note that these assumptions are quite reasonable in view of underlying dynamics. Since it is not obvious initially which mode is the most relevant for the description of $q(t)$, one may take a probabilistic approach. Let $p_j(t)$ be the probability that the river flow $q(t)$ is best approximated by the $j$-th mode, i.e.,

$$p_j(t) = \Pr\{x(t) = j\}, \quad j = 1, \cdots, m,\ t = T_1, \cdots, T_2. \quad (7)$$

Since the best choice for $p_j(t)$ cannot be precisely known beforehand, in this paper we depend on the following two steps to seek a simple and efficient estimation:

(Mode Partition Step) First, we need to adjust the $p_j(t)$ under the constaints

$$\sum_{j=1}^{m} p_j(t) = 1,\ p_j(t) \ge 0, \quad j = 1 \cdots, m,\ t = T_1, \cdots, T_2. \quad (8)$$

For this step, we use a simple approach to partition the rainfall events based on the amount of exponentially weighted rainfall accumulation, which is similar to the strategy of the well-known CLS model [12].

(Linear Modes Estimation Step) For each mode $j \in \{1, \cdots, m\}$, we find the best linear model $f_j$ with the $p_i(t)$ fixed. Note that this step can be formulated as the following minimization problem subject to the constraint that each impulse response of $f_j$ is non-negative and uni-modal:

$$\min_{f_j} \sum_{t=T_1}^{T_2} p_j(t)\,(q(t) - f_j(r(t), \cdots, r(t-T+1))^2 \quad (9)$$

In the following, we present the detailed description of the proposed modeling method:

**[Detailed Procedure for the rainfall-runoff modeling]**

Given $m$, $T$, and the rainfall-runoff data, perform the following:

**[1] Partition of modes**: We use a simple and fixed scheme for choosing $p_j(t)$ which satisfies (8). In the first step of our scheme, we compute a set of exponentially weighted sum of rainfalls as follows:

$$R_\lambda(t) = r(t) + \lambda r(t-1) + \lambda^2 r(t-2) + \cdots + \lambda^{T-1} r(t-T+1), \quad (10)$$

where $\lambda \in (0,1)$, and $t = T_1, \cdots, T_2$. Next, the $R_\lambda(t)$ are clustered into $m$ classes, each of which represents the unique linear mode employed for the modeling. A simple partition of the interval $I = [0, \max_{t \in [T_1, T_2]} R_\lambda(t)]$ into sub-intervals $I_1, \cdots, I_m$ is usually good enough for the clustering, and if more systematic approaches are preferred, iterative methods such as k-means clustering [11] could be used instead. Note that the sub-intervals satisfy

$$I_i \cap I_j = \varnothing \text{ for } i \ne j, \text{ and } \cup_{i=1}^{m} I_i = I. \quad (11)$$

In the final step of the initialization, we set for each $t \in T_1, \cdots, T_2$, $p_j(t) = 1$ when $R_\lambda(t) \in I_j$, and $p_i(t) = 0$ for all the other $i \ne j$.

**[2] Optimization to find each mode**: Once the $p_j(t)$ are determined, we can find the best linear approximator $f_j$ for each mode $j \in \{1, \cdots, m\}$ as follows: First, note that when the river flow $q(t)$ is approximated by the $j$-th mode $f_j$, we have

$$q(t) \cong f_j(r(t), \cdots, r(t-T+1)) = \sum_{k=0}^{T-1} h_j(k) r(t-k). \quad (12)$$

Since it is supposed that $q(t)$ is best approximated by the $j$-th mode $f_j$, the problem of finding $f_j$ can be reduced to minimize the following performance index $PI_j$:

$$PI_j(h_j(0), \cdots, h_j(T-1)) = \sum_{t=T_1}^{T_2} p_j(t)\,\{q(t) - (\sum_{k=0}^{T-1} h_j(k) r(t-k))\}^2 \quad (13)$$

Although here in this paper, focus is restricted on the case $p_j(t) \in \{0,1\}$, please note that one may take various strategies utilizing the soft switching with the property $p_j(t) \in [0,1]$ as in [9]. As mentioned before, here it is assumed that each mode has the property of non-negativity and uni-modality. Thus, in order to find the $f_j$, we need to minimize $PI_j$ of (13) under the constraints of (5) and (6). Since the $t_{max,j}$ of (6) is not known beforehand, we first solve a series of the preliminary problems

$$\begin{aligned} \min\ & PI_j(h_j(0), \cdots, h_j(T-1)) \\ \text{s.t.}\ & h_j(k) \ge 0, k = 0, \cdots, T-1 \\ & h_j(0) \le \cdots \le h_j(l) \ge \cdots \ge h_j(T-1), \end{aligned} \quad (14)$$

where the location of $t_{max,j}$ is fixed at $l \in \{0, \cdots, T-1\}$. Then we pick the best result among the $T$ preliminary problems as the solution to $f_j$. Note that this strategy is nothing but to solve a series of QP (quadratic programming) problems, thus for the step, one can utilize any QP solvers, e.g., the function "quadprog" of the optimization toolbox for use with MATLAB [10]. In Table 1, the advantages of the proposed method are highlighted compared to the neural networks approach.

Table 1. Advantages of the proposed method highlighted against the neural network approach.

| Propose method | Neural networks |
|---|---|
| - Interpretation available<br>- Simple<br>- Flow estimates always non-negative | - Black-box approach<br>- Complex<br>- Flow estimates often negative |

## 3. Application examples

### 3.1 Data

The Seolma-Chun experimental basin is located at the upstream part of the Seolma-Chun, which is the first tributary of the Imjin river basin. Seolma-Chun experimental basis is a small basin of size just about 8.5 ㎢ and channel length of 5.8 ㎞. Most parts of the basin are mountainous, and because it is located just behind the DMZ (de-militarized zone) between North and South Korea, most of the basin remains undeveloped. Six rain gauges, two stream gauges, and one meteorological station are being operated within the basin. This study used the data collected at the outlet of the Seolma-Chun experimental basin. The rainfall-runoff data used in this study were collected during the summer of August 19 to September 28, 2003, which is plotted in Figure 1. Nine independent rainfall events were recorded during this period.

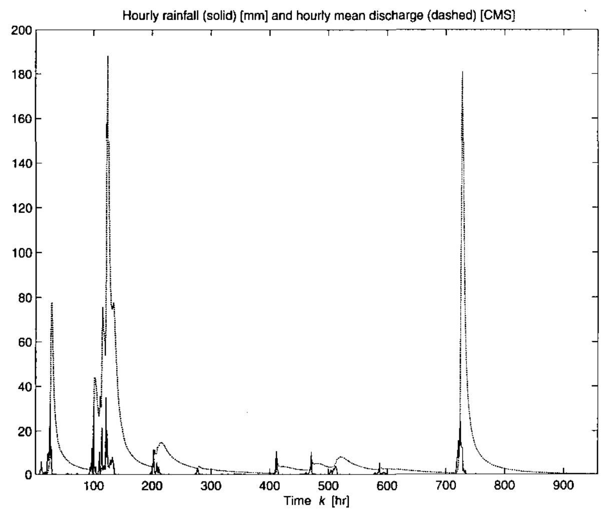


Figure 1. Rainfall (solid) and runoff (dashed) measured at the Seolma-Chun experimental basin.

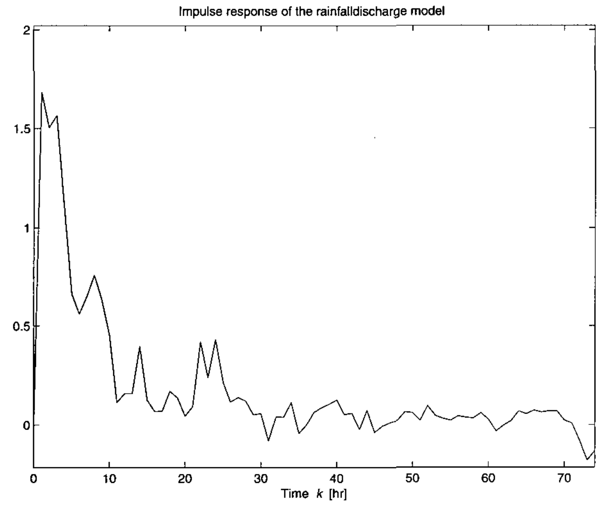


Figure 2. Impulse response of the rainfall-runoff model for the Seolma-Chun experimental basin.

### 3.2 Linear Impulse Response Approach

Applying the simple linear impulse response approach based on equation (2) to the rainfall-runoff data in Figure 1 yields the impulse response of Figure 2 and runoff estimation results of Figure 3. As shown in Figure 3, the simple linear impulse response approach based on equation (2) with $T=75$ resulted in the MSE value of 47.9018. Although the MSE value itself is not bad, the results are unreasonable in the sense that the flow estimations in Figure 3 often tend to be negative. This problem is due to some of the impulse response parameters $h(0), \cdots, h(T-1)$ being negative, which is clearly shown in Figure 2. As mentioned before, another possible weakness of this simple impulse response approach is that the resulting impulse response curve of Figure 2 is not smooth, which is quite contrary to the underlying physical principles as well as our intuition.

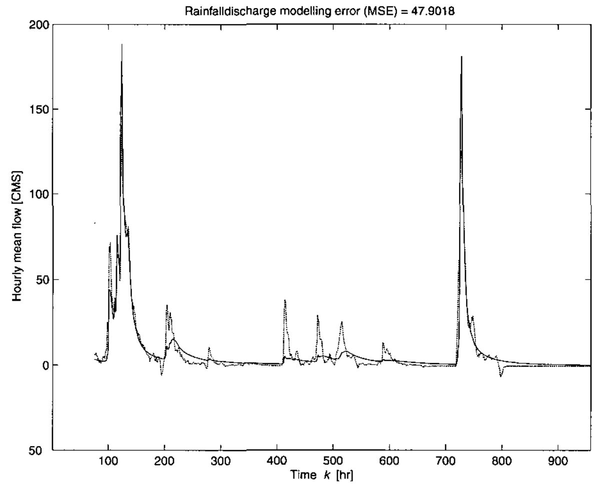


Figure 3. Actual (solid) and estimated (dashed) runoff values for the Seolma-Chun experimental basin resulting from the simple impulse response approach.

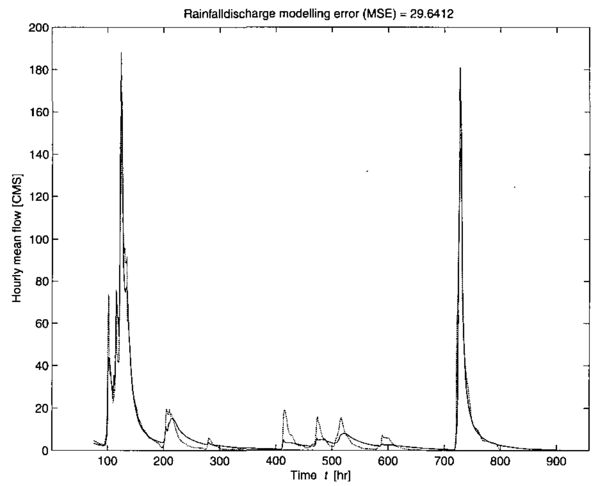


Figure 4. Actual (solid) and estimated (dashed) runoff values for the Seolma-Chun experimental basin resulting from the proposed method (the m=2 case).

### 3.3 Proposed multiple impulse response approach

As an example for illustration of the proposed method, we report the modeling results performed for the rainfall-runoff data set of Figure 1. With $T = 75$, $\lambda = 0.8$, and $m = 2$, we obtained the simulation results of Figure 4. Comparing the figure with Figure 3, we see that the proposed method yields better fitting results than the simple impulse response approach. Also, as indicated at the top of the figure, its error value (MSE=29.6412) is much smaller than the error value (MSE=47.9018) of Figure 3.

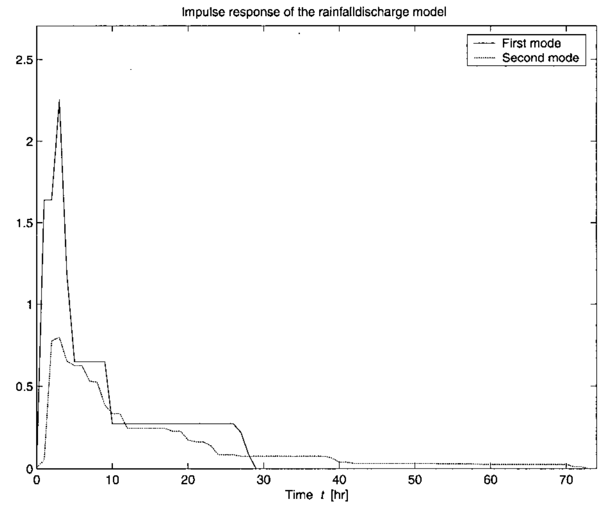


Figure 5. Impulse responses of the rainfall-runoff model for the Seolma-Chun experimental basin resulting from the proposed method (the m=2 case).

Shown in Figure 5 are the two impulse responses found by the proposed method. It is apparent from the figure that the first mode (solid curve) represents a component with quick and reactive response while the second one (dashed curve) is for relatively slow and less reactive response.

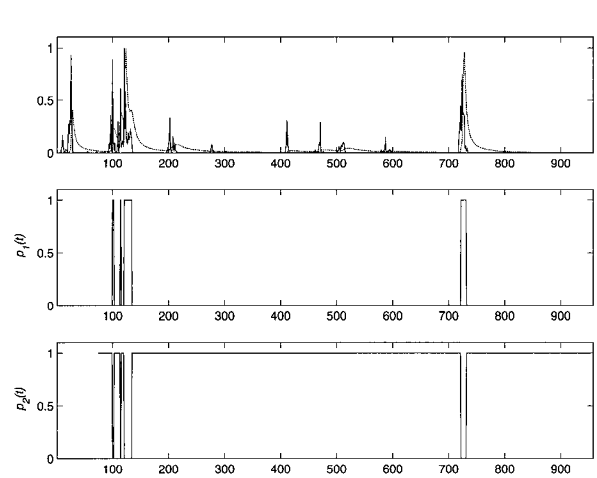


Figure 6. Normalized rainfall and discharge plot (upper) together with the $p_j(t)$ values (lower two plots).

In Figure 6, the $p_j(t)$ are plotted together with normalized rainfall and discharge values in order to show which modes are more responsible for the description of the given data at each time slot. The contents of Figures 4-6 show that the proposed method is capable of yielding reasonable results with rich interpretations. We also applied the proposed method to the same data set with more number of modes ($m = 3$) and obtained the results of Figures 7-9. As shown in the figures, the use of three modes turns out to further reduce the fitting error, and seems to be more effective in describing the given data.

Finally, note that although the proposed method is much simpler than the method taking the strategy of more sophisticated switching rule such as [9], it has an obvious advantage over the sophisticated method that extension toward the one-step prediction problem is more straightforward. Investigation on the ability of the proposed method as the runoff predictor is a topic that needs to be pursued in the future.

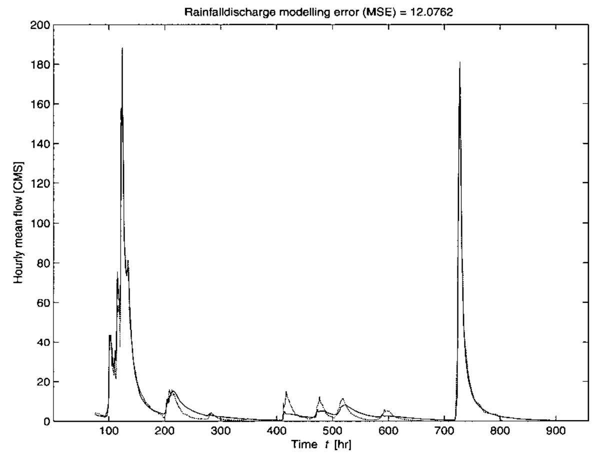


Figure 7. Actual (solid) and estimated (dashed) runoff values for the Seolma-Chun experimental basin resulting from the proposed method (the m=3 case).

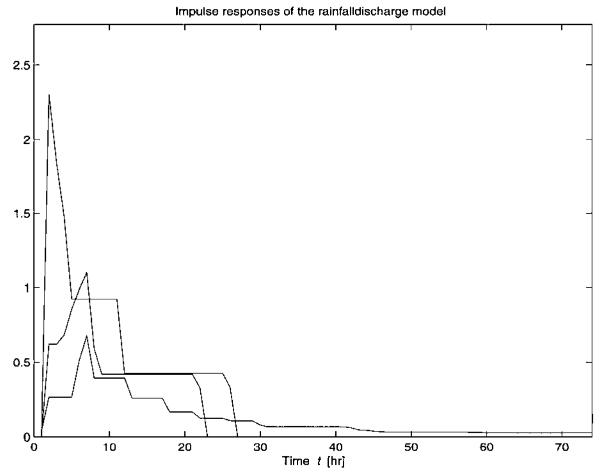


Figure 8. Impulse responses of the rainfall-runoff model for the Seolma-Chun experimental basin resulting from the proposed method (the m=3 case).

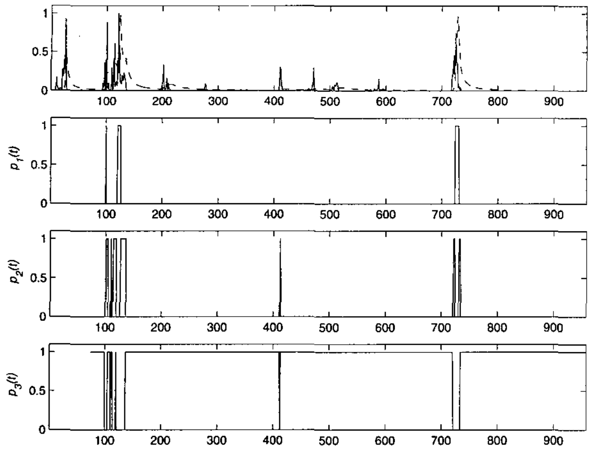

Figure 9. Normalized rainfall and discharge plot (upper) together with the $p_j(t)$ values (lower three plots).

## 4. Concluding Remarks

Many recent studies have successfully used neural networks for nonlinear rainfall-runoff modeling. Due to fundamental limitation of linear structures, linear modeling has recently been considered inferior to neural network approaches in this area. However, we believe that with an appropriate extension, the concept of linear impulse responses can be a viable tool for understanding the underlying dynamics of nonlinear rainfall-runoff processes, and in this paper, we proposed an extension toward the use of multiple linear impulse responses for the rainfall-runoff analysis. The proposed method is based on a simple switching over multiple linear impulse response models, each of which satisfies the constraints of non-negativity and uni-modality. The numerical analyses performed for the rainfall-runoff data in the Seolma-Chun experimental basin showed that the proposed method can yield very meaningful results. Works yet to be done include further investigation on the prediction capability and extensive simulation studies to identify the strengths and weaknesses of the proposed method.

### References


[1] H. K. Cigizoglu, “Estimation, forecasting and extrapolation of river flows by artificial neural networks," *Hydrological Sciences Journal*, Vol. 48, No. 3, pp. 349-360, 2003.

[2] H. K. Cigizoglu, "Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons," *Advances in Water Resources*, Vol. 27, pp. 185-195, 2004.

[3] H. K. Cigizoglu and M. Alp, "Rainfall-runoff modelling using three neural network methods," L. Rutkowski et al. (Eds.): ICAISC 2004, *Lecture Notes in Artificial Intelligence*, Vol. 3070, pp. 166-171. Springer-Verlag, Berlin, 2004.

[4] K. P. Sudheer, A. K. Gosalin, and K. S. Ramasastri, "A data-driven algorithm for constructing artificial neural network rainfall-runoff models," *Hydrological Processes*, Vol. 16, pp. 1325-1330, 2002.

[5] H. R. Maier and G. C. Dandy, "Neural network for the prediction and forecasting of water resources variaboes: a review of modeling issues and applications," *Environmental Modeling and Software*, Vol. 15, pp. 101-124, 2000.

[6] K. Pawelzik, J. Kohlmorgen, and K.-R. Muller, "Annealed competition of experts for a segmentation and classification of switching dynamics," *Neural Computation*, Vol. 8, No. 2, pp. 340-356, 1996.

[7] K.-R. Muller, J. Kohlmorgen, and K. Pawelzik, "Analysis of switching dynamics with competing neural networks," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E78-A, No. 10, pp. 1306-1315, 1995.

[8] M.-W. Chang, C.-J. Lin, and R. C. Weng, "Analysis of switching dynamics with competing support vector machines," *IEEE Transactions on Neural Networks*, Vol. 15, pp. 720-727, 2004.

[9] C. Yoo and J. Park, "Rainfall-runoff analysis based on competing linear impulse responses," Submitted, 2006.

[10] T. F. Coleman, M. A. Branch, and A. Grace, *Optimization Toolbox for Use with MATLAB, User's Guide*, Version 2, Math Works Inc., 1999.

[11] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, Upper Saddle River, NJ, 2000.

[12] E. Todini, "The CLS model," In *Mathematical models of large watershed hydrology*, Singh, V. P., Frevert, D. K., and Meyer, S. P. (eds.), Water Resources Publications, Littleton, Colorade, Chapter 20, pp. 861-886, 2002.


## 저 자 소 개


**Chulsang Yoo** received his BS and MS degrees in civil engineering from Korea University, Korea, in 1987 and 1989, respectively, and his PhD degree in civil engineering from the Texas A&M University, USA, in 1995. He joined Korea University in 2003, where he is currently a Professor in the Department of Civil and Environmental Engineering. His research interests

include hydrology and stochastic processes.

**Jooyoung Park** received his BS degree in electrical engineering from Seoul National University, Korea, in 1983, and his PhD degree in electrical and computer engineering from the University of Texas at Austin, USA, in 1992. He joined Korea University in 1993, where he is currently a Professor in the Department of Control and Instrumentation Engineering. His research interests include neural networks, nonlinear systems, and control.