Short-term Rainfall Prediction Using a Multifractal Model

by

Yi-Ju Chou

B.S. Department of Agricultural Engineering
National Taiwan University, 1999

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of requirement for the degree of

MASTER OF ENGINEERING IN CIVIL AND ENVIRONMENTAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2003

Signature of Author:____ _____

Yi-Ju Chou
Department of Civil and Environmental Engineering
May 9, 2003

Certified by:_____ _____

Daniele Veneziano
Professor, Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by:_____ _____

Oral Buyukozturk
Chairman, Departmental Committee on Graduate Studies

Short-term Rainfall Prediction Using a Multifractal Model

Yi-Ju Chou

Submitted to the Department of Civil and Environmental Engineering
On May 9, 2003 in partial fulfillment of requirement for the degree of
Master of Engineering in Civil and Environmental Engineering

Abstract

This study develops a method to predict multifractal measure of temporal rainfall intensity by using Kalman filter, and gives some examples of prediction for generated rainfall. The model for the rainfall generation proposed here is established using a continuous-time, discrete-scale lognormal cascade (CLC) with AR(1) process for each component. This model allows us to simulate rainfall field with the property of the multifractality, which indicates the invariance for scaling of rainfall measure. Through the observation from the synthetic rainfall simulated by this model, Kalman filter is used as the tool for short-term rainfall prediction. We compare different results of predictions made under different simulations and discuss the extensions of this study, prediction for the wet/dry process while looking at real rainfall and issues about space-time rainfall modeling.

Keywords: Multifractality, Bayesian estimation, Kalman filter

Thesis Supervisor: Dr. Daniele Veneziano
Title: Professor, Department of Civil and Environmental Engineering

Acknowledgements

I would like to express my gratitude to my thesis advisor Professor Daniele Veneziano for his guidance and encouragement. Professor Daniele Veneziano is always available for questions and very dedicated to me on solving problems.

I would like to thank many of my friends for their warm and true friendship: Dave, Jongsok, Matt, Jr. Shin, Su-Shin, Tzu- Yang, Meng-Hsien and James.

.

Finally, I want to dedicate this thesis to my family. Their support and encouragement motivate my study at MIT.

Table of Contents

List of Figures

Chapter 1. Introduction

This study develops a method to predict multifractal measure of temporal rainfall intensity by using Kalman filter, and gives some examples of prediction for generated rainfall. The model for the rainfall generation proposed here is established using a continuous-time, discrete-scale lognormal cascade (CLC) with AR(1) process for each component. This model allows us to simulate rainfall field with the property of the multifractality, which indicates the invariance for scaling of rainfall measure. Through the observation from the synthetic rainfall simulated by this model, Kalman filter is used as the tool for short-term rainfall prediction. In Kalman filter, we first use Bayesian estimation as the regression approach in the estimation step to obtain the conditional distribution of each parameter based on current observation and prior guess from past data and then, in the prediction step, prediction is made with the correlation of each parameter. In Chapter 2, we review briefly some of the scaling properties of multifractal measure and methods of discrete random cascade construction. Then, we simulate rainfall by using a CLC process based on the concept of discrete cascade and illustrate the simulation results with different given conditions. In Chapter 3, we first review the algorithm of Bayesian estimation and the prediction step in Kalman filter. Then we apply

Kalman filter into our synthetic rainfall process and compare different results of prediction made with different simulation conditions. In Chapter 4, besides making conclusions of the work in the previous chapters, we discuss the extensions of this study, which is mainly dealing with the wet/dry process while looking at data of real rainfall and issues about space-time rainfall modeling.

1.1 Multifractal Measure of Rainfall

Past research on stochastic modeling of precipitation has addressed mainly two classes of models: cluster based and fractal/multifractal models. The cluster-based models, which have been exploited by LeCam models [*LeCam, 1961*], represent rainfall through the superposition of pulses with clustering in space and time. An early review of these models is made in *Waymire and Gupta* [1981]. LeCam model reflects the physical organization of rainfall into rain-bands, meso-scale precipitation areas, convective cells, etc., but they are not scale invariant. Following the second line of research, more recent papers deal with multifractal models based on random cascades [*Deidda et al. 1999; Deidda, 2000; Gupta and Waymire, 1990, 1993; Ladoy et al., 1993; Lovejoy and Scherzer, 1995; Tissier et al. 1993*]. Multifractalality, which was initially applied to modeling velocity fluctuations in turbulent flows [*Benzi et al., 1984*], implies that the rainfall

process looks statistically the same at small and large scales, except some simple transformations [*Gupta and Waymire, 1990; Veneziano, 1999*]. The formalism of multifeactalality allows a robust statistical control over any moment of a given distribution of measures if some kind of similarity holds over a range of scales. This theory represents very powerful approach to nonlinear phenomena such as turbulence velocity fluctuation or intermittency of precipitation.

1.2 Multifractal Forecasting

Although the literature on precipitation forecasting is vast, here we review only methods developed for multifractal processes and describe the approach we proposed for rainfall application. There are two main approaches to multifractal prediction: one is the filter white noise (FWN) approach proposed by *Marsan et al.* [1996], and another one is the state-space (SS) approach developed by *Calvet and Fisher* [2001] in the context of financial time series.

Marsan et al. [1996] have developed a method of this type to predict the log of universal multifractal rainfall fields in space and time. The main contribution is the derivation of a causal FWN representation for universal multifractal processes in space and time, and the extension to dynamic scaling. This method is theoretical optimal and

computationally attractive, but has two limitations: (1) the method is approximate because it does not account for the fact that rainfall observations are "dressed" not "bare" quantities; (2) since the past noise cannot be recovered exactly, the method is not well suited to handle short duration data; (3) the method is also not well suited for non-universal multifractal processes (in particular process with lacunarity), nonstationary processes, and cases when information other than that past precipitation measurement is available.

Calvet and Fisher [2001] have developed a forecasting method for what they call Poisson Multifractal Measures (PMM). A PMM is a variant of the discrete cascade, with irregular tile sizes defined by a Poisson point process. The PMM model is attractive for prediction because its value depends on a hidden continuous-time Markov process. The hidden Markov property and the Poisson structure of the PMM model make it conceptual easy (under simplifying assumptions) to find the distribution of the future state and future value of the process given the current state. High computation demand is the main drawback of the method.

In this study, a simple example of random process continuous in time is the continuous lognormal cascade (CLC) that will be described in Chapter 3. Like PMM, the CLC has a latent Markov state, which at time t comprises all the log multipliers $W'_k(t)$. Also, one can

approximate the latent state by the set of multipliers up to k and absorb the rest of the variability of the log measure, the dressing quantity, into an independent fluctuation factor Z. If the factor Z is approximated as having itself lognormal distribution, one can use the Kalman filter to update the mean and covariance matrix of the log state vector at a series of discrete times. This results in a significant simplification of the Calvet-Fisher updating procedure. This simplification is critical if one extends the prediction problem to space-time processes, for which the latent state varies with spatial locations.

Chapter 2. Multi-fractal Rainfall Simulation

2.1 Scaling Properties of Multifractal Measure

Let us consider a positive temporal rainfall rate $P(t)$ defined on $t \in [0,T]$, and describe the mass within the time interval between t and $t+r$ as

$$m_t(r) \equiv \int^{+r} P(t)dt \tag{1}$$

The multifractal analysis in time can be performed by investigating the scaling behavior of Eq. (1) for different time scale r and $P(t)$ is said to display multifractality if the following scaling law holds:

$$\left\langle [m_t(r)]^q \right\rangle \sim r^{H(q)} \tag{2}$$

where the scaling exponent $H(q)$ is a nonlinear function of moments q, and also referred to as multifractal exponent of $P(t)$. The $\langle\ \rangle$ stands for ensemble average or an average of samples with different starting time t. It follows that we must also require stationarity, that is, the properties in Eq. (2) depend only on r.

In this rainfall field, we can also write average rainfall intensity as

$$\begin{aligned} R_t(r) &= \frac{1}{r}\int^{+r} P(t)dt \\ &= \frac{m_t(r)}{r} \end{aligned} \tag{3}$$

and it follows that if Eq.(2) holds, one obtains

$$\left\langle [R_t(r)]^q \right\rangle \sim r^{H(q)-q} \tag{4}$$

Hereafter we will use the notation $K(q) = H(q) - q$. In multifractal measurement, the moment scaling function $K(q)$ determines many of the scaling properties corresponding to different scale r and different moment q which will be described in the following sections.

2.2 Concept of Discrete Random Cascade Construction

A simple way to build up a multifractal rainfall field is to use a cascade model based on random multiplier [*Gupta and Waymire, 1993; Over and Gupta, 1994, 1996*]. The mathematical construction of random cascade begins with given rainfall intensity, say R_0, distributed uniformly over a bounded region of physical interest $L = [0,T]$. Then we divided L into b subintervals, each with length of T/b and b is called branching number. Let $L_1(t)$, $t = 1,2,\ldots b$, denote this partition of L into these subintervals. Now, distribute the rainfall intensity into each of these b subintervals as $R_0 W_{b,1}(t)$. The $W_{b,i}(t)$'s are mutually independent random variables with identical probability distribution, that is, are iid as W_b. In addition, it is stipulated that $E[W_{b,i}] = 1$, which means that the ensemble average of the rainfall intensity R_0 is conserved after this redistribution. At the next step of the cascade construction, each of the subintervals is further divided into b subintervals

with length T/b^2 and each of these subintervals is denoted by $L_2(t)$, in which $t = 1,2,\ldots b^2$. The rainfall intensity in each of these subintervals is redistributed as $R_0 W_{b,1}(t) W_{b,2}(t)$, where $W_{b,2}(t)$'s are independent of $W_{b,1}(t)$'s and R_0 and are iid with W_b. This construction is shown schematically in Fig. 1(a) and (b) for b = 2.

The cascade construction now continues iteratively, such that at the nth generation level, the length L is divided into b^n subintervals with length L/b^n and we denote each of these subintervals by $L_n(t)$, where $t = 1,2,\ldots b^n$. Hence, the rainfall intensity in each of these subintervals is denoted by

$$R(t) = R_0 W_{b,1}(t) W_{b,2}(t) \cdots W_{b,n}(t)$$

$$= R_0 \prod_{i=1}^{n} W_{b,i}(t) \tag{5}$$

As shown on Fig. 1, the discrete cascade construction with branching number, b = 2 and T = 1024, is started from i = 0 to i = 10. At i = 0, the rainfall intensity was distributed uniformly over time, and with increase of i, it becomes more fluctuated.

Now, let us continue the cascade construction beyond level n. At each step $k > n$, we average the measure density inside an n-block, obtaining a new random variable $R_{n,k}(t)$. If $R_{n,d}(t) = \lim_{k \to \infty} R_{n,k}(t)$ exists, then $R_{n,d}(t)$ is called dressed measure density at level n, and at this finite level n, the constant measure density inside each n-block is called bare measure density, denoted by $R_{n,b}(t)$.

Let Z be the average dressed density in the unit block. By cascade construction, the dressed measure density is related to the bare density and to Z as

$$R_{n,d}(t) \overset{dist}{=} R_{n,b}(t)Z \tag{6}$$

where $R_{n,b}(t)$ and Z are independent. Since $R_{n,b}(t)$ is the measure density inside each unit block at the level n the cascade construction terminated, hereafter, we denote $n_{,b}$ by n and $n_{,d}$ by n_d. Based on Eq. (5), we can write

$$\begin{aligned} R_{n_d}(t) &= R_0 \prod_{i=1}^{n+n_d} W_{b,i}(t) \\ &= R_0 \prod_{i=1}^{n} W_{b,i}(t)Z \end{aligned} \tag{7}$$

where Z can be treated as an error term in this measurement up to nth level due to the dressed measure density.

2.3 Continuous Log-normal Cascade

Eq. (7) is the model using discrete random cascade, which consists of R_0= rainfall measured in the total temporal length the multifractal theory applies, $W_{b,i}(t)$= scaling factor in different time scale, and Z = the error term due to dressing. Instead of partitioning into each discrete subinterval by multiplying $W_{b,i}$ as Eq. (7), in our synthetic rainfall simulation, we generate $W_{b,i}(t)$ in each time step, with a given statistical distribution for each i, and thus obtain the rainfall intensity $R(t)$ by multiplying

all these $W_{b,i}(t)$ components with R_0 and Z. Although this approach is mainly based on the concept of discrete random cascade, the difference is that in our model, synthetic rainfall can be generated continuously at any time in time series, instead of discreteness in time series for discrete cascade.

Now, we assume $W_{b,i}(t)$ to be lognormal distributed and rewrite Eq. (7) as

$$R(t) = R_0 \exp[\sum_{i=1}^{n} W'_{b,i}(t) + Z'] \tag{8}$$

where $W'_{b,i}(t)$ and Z' are natural logarithm of $W_{b,i}(t)$ and Z, respectively, and $W_{b,i}(t)$ has Gaussian distribution. Then, normalizing $R(t)$ by dividing by R_0, Eq. (8) is rewritten as

$$R'(t) = \sum_{i=1}^{n} W'_{b,i}(t) + Z' \tag{9}$$

where $R'(t) = \ln\left[\frac{R(t)}{R_0}\right]$.

Since the scaling factor $W_{b,i}(t)$ is a log-normal generator process, for a cascade with linear multiplicity b, the moment scaling function $K(q)$ of $R(t)$ is

$$K(q) = \log_b E[W_b{}^q]$$

$$= C_1(q^2 - q) \tag{10}$$

and the radial spectral density function of $\ln[R(t)]$ is

$$S_{\ln R(t)}(\omega) = 2C_1\omega^{-1} \tag{11}$$

For log-normal distribution, the mean value and variance of $W_{b,i}(t)$ and $W'_{b,i}(t)$ satisfy

$$E(W_{b,i}) = \exp[E(W'_{b,i}) + \frac{1}{2}Var(W'_{b,i})] \tag{12}$$

$$Var(W_{b,i}) = \exp[2E(W'_{b,i}) + Var(W'_{b,i})]\{\exp[Var(W'_{b,i})] - 1\} \tag{13}$$

Because $E(W_{b,i})$ is equal to 1, we have that $E(W'_{b,i}) = -\frac{1}{2}Var(W'_{b,i})$ from Eq. (12), and $Var(W_{b,i}) = \exp[Var(W'_{b,i})] - 1$ from Eq. (13). Then, according to Eq.(10), and $Var[X] = E[X^2] - E^2[X]$, we have

$$Var[W_{b,i}] = E[{W_{b,i}}^2] - E^2[W_{b,i}]$$

$$= b^{2C_1} - 1 \tag{14}$$

and, from Eq. (13) and (14),

$$Var[W'_{b,i}] = 2C_1 \ln(b) \tag{15}$$

Based on Eq.(8) and (9), the strategy to simulate $R(t)$ is to first simulate $Y(\tau) = R(\tau b^{-n_d})$ for a given level of dressing n_d using

$$Y(\tau) = R_0 \exp[\sum_{i=1}^{n+n_d} W'_{b,i}(\tau b^{-n_d})] \tag{16}$$

which is done by following steps:

1. Simulate $X_i(\tau) = W'_{b,i}(\tau b^{-n_d}) - E[W'_{b,i}]$, by using

$$X_i(0) = 0;$$
$$X_i(\tau) = \phi_i X_i(\tau - 1) + \eta_i(\tau) \tag{17}$$

where $\eta_i(\tau)$ is identical independent distributed variable with normal distribution

of zero mean and variance $(1-\phi_i^2)\sigma_{W'_{b,i}}{}^2$. The AR(1) coefficient ϕ_i is the correlation of $W'_{b,i}$ at time lag 1. Under Gauss-Markov conditions the correlation function of $W'_{b,i}$ satisfies $\phi_i = \exp[-\frac{b^{i-1}}{t_0}]$, in which the correlation distance t_0 is a given positive constant. Fig.2 shows ϕ_i for different *i*, with t_0 equal to $2^8 = 256$, and the correlation at time lag 1 decreased as *i* increases.

2. Then, from Eq.(12), and $E(W'_{b,i}) = -\frac{1}{2}Var(W'_{b,i})$, we can get

$$\begin{aligned} Y(\tau) &= R_0 \exp[\sum_{i=1}^{n+n_d} W'_{b,i}(\tau b^{-n_d})] \\ &= R_0 \exp\{[\sum_{i=1}^{n+n_d} X_i] + (n+n_d)\times E[W'_{b,i}]\} \\ &= R_0 \exp\{[\sum_{i=1}^{n+n_d} X_i] - \frac{1}{2}(n+n_d)\times Var[W'_{b,i}]\} \end{aligned} \tag{18}$$

where $Var[W'_{b,i}]$ is obtained by Eq. (15). As shown on Fig. 3, $Y(\tau)$ was simulated by using Eq. (18), which also shows different simulation of different *n* with $C_1 = 0.2$, $n_d = 3$, $R_0 = 1$ and simulation τ equals to 8192.

3. By averaging $Y(\tau)$ over consecutive b^{n_d} blocks, we can get *R*(*t*) as

$$R(t) = \frac{1}{b^{n_d}} \sum_{\tau=(t-1)b^{n_d}+1}^{tb^{n_d}} Y(\tau) \tag{19}$$

Using this continuous lognormal cascade (CLC), we generate synthetic rainfall *R*(*t*) as shown on Fig. 4, in which all the parameters used are the same as that in Fig. 2 except *n* =

9. In Fig. 2(a), Eq. (18) is used to generate $Y(\tau)$, while in (b) $R(t)$ is obtained by averaging over b^{n_d} blocks.

Chapter 3. Proposed Approach of Short-term Prediction and Application

In this study, we use Kalman filter as a tool to make short-term rainfall prediction. The Kalman filter comprises two steps, one is estimation step, which updates information of each variable based on observation at each time step by estimating the conditional distribution of each variable from current observation and past data; the other one is prediction step, which predicts the value of each variable based on the given correlation and the obtained value from estimation step. Here, we first review the theoretical methodology in each step of Kalman filter, and then apply it to our proposed rainfall model.

3.1 Estimation Step

Bayesian estimation is used to update information by calculating the conditional distribution of each variable based on the current observation and a prior distribution obtained from past data. An important problem of Bayesian estimation is that of calculating the posterior distribution of a vector, $\underline{X}$, with normal prior distribution, observed through a linear equation

$$\underline{Z} = \underline{z}_0 + \underline{H}\,\underline{X} + \underline{\varepsilon} \tag{20}$$

in which $\underline{z}_0$ is a given k vector; $\underline{H}$ is a given observation matrix, and $\underline{\varepsilon}$ is a random noise vector with normal distribution and zero mean. While applying Eq. (20) into the observation of our synthetic rainfall, we need to calculate the posterior distribution of $\underline{X}$ in each time step, so here we denote $\underline{X}$ and $\underline{Z}$ by $\underline{X}(t)$ and $\underline{Z}(t)$, respectively. Since in rainfall observation, $\underline{Z}(t)$ is a scalar (the detail will be described in following sections), and so are $\underline{z}_0$ and $\underline{\varepsilon}$, and we rewrite Eq. (20) as

$$Z(t) = z_0 + \underline{H}\,\underline{X}(t) + \varepsilon \tag{21}$$

Let $\underline{X}(t)$ and ε have prior joint normal distribution, with $\underline{X}(t) \sim N[\underline{m}(t|t-1); \underline{\Sigma}(t|t-1)]$ in which $(t|t-1)$ means being based on historical data collected up to time step $t-1$; $\varepsilon \sim N(0;\theta)$, and the covariance between $\underline{X}(t)$ and ε is zero. The posteriori density of $\underline{X}(t)$, $f_{\underline{X}(t)|Z(t)}[\underline{x}(t), z(t)]$, satisfies Bayes' equation

$$f_{\underline{X}(t)|Z(t)}[\underline{x}(t), z(t)] = f_{\underline{X}(t)}[\underline{x}(t)] \frac{f_{Z(t)|\underline{X}(t)}[z(t), \underline{x}(t)]}{f_{Z(t)}[z(t)]} \tag{22}$$

From Eq. (21), we know that in Eq. (22) $[Z(t)|\underline{X}(t)] \sim N[z_0 + \underline{H}\,\underline{X}(t); \underline{\theta}]$ and $Z(t) \sim N[z_0 + \underline{H}\,\underline{m}(t|t-1); \underline{H}\,\underline{\Sigma}(t|t-1)\underline{H}^T + \theta]$. After some non-trivial algebra, it is found from Eq. (21) that $[\underline{X}(t)|Z(t) = z(t)]$ has normal distribution with mean

$$\underline{m}(t|t) = \underline{\Sigma}(t|t-1)[\underline{H}^T \theta^{-1}(z(t) - z_0) + \underline{\Sigma}(t|t-1)^{-1}\underline{m}(t|t-1)]$$

$$= \underline{m}(t|t-1) + \underline{\Sigma}(t|t-1)\underline{H}^T \left[\underline{H}\underline{\Sigma}(t|t-1)\underline{H}^T + \theta\right]^{-1} \left(z(t) - z_0 - \underline{H}\underline{m}(t|t-1)\right) \tag{23}$$

and covariance matrix

$$\underline{\Sigma}(t|t) = [\underline{\Sigma}(t|t-1)^{-1} + \underline{H}^T \theta^{-1} \underline{H}]^{-1}$$

$$= \underline{\Sigma}(t|t-1) - \underline{\Sigma}(t|t-1)\underline{H}^T \left[\underline{H}\underline{\Sigma}(t|t-1)\underline{H} + \theta\right]^{-1} \underline{H}\underline{\Sigma}(t|t-1) \tag{24}$$

Through Eq. (21), (23) and (24), one can update the mean vector and covariance matrix of $\underline{X}(t)$ based on prior distribution and observation at current time step.

3.2 Prediction Step

To predict rainfall at the time Δ after the current time t, we introduce a parameter $\rho_i(\Delta)$, i = 1,2,...n, which represents the correlation coefficient of each element in $\underline{X}(t)$ between the prediction time $t + \Delta$ and the current time t, written as

$$\rho_i(\Delta) = \exp\left(-\frac{b^{i-1}\Delta}{t_0}\right) \tag{25}$$

where we recall that b is the branching number in the random cascade and t_0 is the correlation length. Thus, we can write a correlation vector as

$$\underline{\rho}(\Delta) = \begin{pmatrix} \rho_1(\Delta) \\ \rho_2(\Delta) \\ \vdots \\ \rho_n(\Delta) \end{pmatrix} \tag{26}$$

and also

$$sqr[\underline{\rho}(\Delta)] = \begin{pmatrix} \rho_1^{\,2}(\Delta) \\ \rho_2^{\,2}(\Delta) \\ \vdots \\ \rho_n^{\,2}(\Delta) \end{pmatrix} \tag{27}$$

Different values of $\rho_i(\Delta)$ with different i and different time lag Δ are shown on Fig. 5, from which we can see that the W component of rainfall model with lower i has higher correlation.

Using Eq. (26) and (27), we can obtain the predicted mean vector $\underline{m}(t+\Delta|t)$ and covariance matrix $\underline{\Sigma}(t+\Delta|t)$ as

$$\underline{m}(t+\Delta|t) = \underline{m}(t) + diag[\underline{\rho}(\Delta)][\underline{m}(t|t) - \underline{m}(t)] \tag{28}$$

and

$$\underline{\Sigma}(t+\Delta|t) = diag\{sqr[\underline{\rho}(\Delta)]\}\underline{\Sigma}(t|t) + diag\{\underline{\Sigma}(t)[I - sqr[\underline{\rho}(\Delta)]]\} \tag{29}$$

Eq. (28) and (29) are used to make prediction of mean vector and correlation matrix after Δ based on the observation at current time, historical information, and correlation coefficient. As the case of $\Delta = 1$, $\underline{m}(t+\Delta|t)$ and $\underline{\Sigma}(t+\Delta|t)$ are equal to $\underline{m}(t+1|t)$ and $\underline{\Sigma}(t+1|t)$ which are used as the parameters of prior distribution of $\underline{X}(t)$ in each prediction step.

3.3 Application of Kalman filter

In our rainfall simulation, the value of C_1 and b in Eq. (15) has been set to 0.2 and 2, respectively. Thus the variance and mean of $W'_{b,i}(t)$ from Eq. (15) are 0.28 and -0.14, respectively (recall that $E(W'_{b,i}) = -\frac{1}{2}Var(W'_{b,i})$). In the estimation step, we first take the natural logarithm, *R'(t)*, of observed rainfall, *R(t)*. The Bayesian estimator Eq. (21) at each time step is obtained by using parameters as follows

1. Observed log average rainfall intensity $R'(t)$ is used as $Z(t)$;
2. The difference between the observed mean value of $R'(t)$ and the summation of theoretical mean values of $W'_{b,i}(t)$ components,

$$\left[E[R'(t)] - n \times E(W'_{b,i})\right] = \left[E[R'(t)] - [-\frac{n}{2}Var(W'_{b,i})]\right]$$

   is used as z_0;
3. From Eq. (9), one can get the error term, Z, by obtaining the difference between observed log rainfall, R'(t) and the summation of $W'_{b,i}(t)$'s. Thus the difference between observed variance of $R'(t)$ and theoretical variance, $\left[Var[R'(t)] - n \times Var(W'_{b,i})\right]$, is used as θ, the variance of the noise term, ε.
4. The $1 \times n$ matrix, $[1 \quad 1 \quad \ldots \quad 1]$ is used as $\underline{H}$.
5. The parameters of the marginal distribution, $\underline{m}(t)$ and $\underline{\Sigma}(t)$, of $W'_{b,i}(t)$'s in Eq. (28) and (29) are obtained from given simulation condition.

Thus, with given mean vector, $\underline{m}(t|t-1)$, and covariance matrix, $\underline{\Sigma}(t|t-1)$, obtained from the past data up to time t-1, we can estimate the conditional mean vector, $\underline{m}(t|t)$, and covariance matrix, $\underline{\Sigma}(t|t)$, according to Eq. (23) and (24), respectively. The predicted mean vector and covariance matrix can be obtained based on Eq. (25) through Eq. (29). The correlation distance, t_0, is equal to b^{n-1} such that the lowest correlation for $W'_{b,i}(t)$ with time lag 1 has been set to be exp[-1] no matter what values of n are used in simulation and the highest correlation is always to be $\exp\left(-\frac{1}{2^{n-1}}\right)$.

3.4 Results and Discussion

As shown on Fig. 6 and Fig. 7, each $W'_{b,i}(t)$ component was plotted with the estimated values from Bayesian estimation and predicted values at prediction made for lag 1. We can see the higher frequency of $W'_{b,i}(t)$ for higher i because the correlation decreases as i increases, and thus $W'_{b,i}(t)$ with lower i has higher memory than $W'_{b,i}(t)$ with higher i. For this case of time lag equal to 1, the predicted values are close to estimated one for the case of lower i, but for the case of higher i, the predicted value is more close to the marginal distribution at time t, which is due to the correlation coefficient $\rho_i(\Delta)$ for higher i is less than that of lower i(see Fig. 4). As shown on Eq. (28), the correlation coefficient serves as a weighting factor for marginal distribution and

conditional distribution while prediction are made, and makes the weighting fraction of conditional distribution lower and marginal distribution higher while i increases. In addition, the case of $\Delta = 10$ was plotted on Fig. 8. Comparing the case of Fig. 6 and Fig. 8, correlation between each estimated value and predicted value is decreased with increasing time lag, thus, with less weighting fraction of conditional distribution, the predicted value approaches to mean value of marginal distribution and then almost nothing can be done for prediction while lag larger than 10. The special case on higher i=5 in Fig. 8 also shows that the predicted values are almost equal to the mean values of the marginal distribution while time lag increases or i increases.

The comparison of the predicted value of rainfall, which is the summation of predicted mean value of $W'_{b,i}(t)$ components plus z_0, and the observed synthetic rainfall were plotted on Fig. 9 and 10 for $\Delta = 0$, 1, 2, 3, 5 and 10, in which the predicted values of $\Delta = 0$ is equal to the estimated values from estimation step. From Fig. 9 and 10, one can see a better prediction with smaller error while prediction is made for time lag equal to 1, and much higher prediction error shown in the case of time lag equal to 10. Notice that for the case $\Delta = 0$, the estimation of posterior (conditional) distribution was made by given prior distribution and observation, so there is still little difference from estimated value and observed value, as shown in Fig. 9(a)

The prediction error is used to describe the error between observation and prediction and can be written as

$$\varepsilon_{prd} = R'(t) - R'_{prd}(t) \tag{30}$$

where the predicted value of log-rainfall $R'_{prd}(t)$ is the summation of predicted mean value of each $W'_{b,i}(t)$ at each time plus z_0. The mean and variance of Eq.(30) are plotted on Fig. 11 for the case of total simulation time equal to 2^{12}, n_d=1 and n=5. From Fig. 11, we see the mean value of prediction error is very close to 0 and variance of prediction error increases with increasing time lag, and then approaches a value that is equal to the variance of observed rainfall when prediction is made beyond time lag equal to around10, which means while time lag is larger than 10, the prediction step almost do nothing for prediction. Notice that in Fig. 11, the prediction error at time lag equal to 0 is calculated from the error between $R'(t)$ and the estimated value $R'(t|t)$.

Here, we also introduce a quantity, namely explained variance, written as

$$Var_{\exp} = \frac{Var[R'(t)] - Var[\varepsilon_{prd}]}{Var[R'(t)]} \tag{31}$$

which is the difference between the observed variance of log rainfall and the variance of prediction error normalized by the observed variance. Eq.(31) is plotted on Fig.12. with different prediction time lag in linear and log-log scale for the case of parameters the same as simulation in Fig. 11. As shown on Fig. 12(a), we see the $Var_{\exp}$ almost equal to

1 as prediction is made from zero time lag and decrease to 0 while time lag is higher than 10. If one omits the values of larger time lag that the prediction error approaches observation variance, say the time lag after 10, one can see a straight line with negative slope on the plot of $Var_{\exp}$, versus prediction time lag on log-log scale as shown on Fig. 12(b), .

In order to see how the dressing measure effects the prediction, other simulations are also made with different n_d, which make the variance, θ, of noise term become larger as n_d increases. As shown on Fig. 13, the predictions are made with simulations of n_d=1, 3, and 5, for the case n = 7. One can see that increase of n_d, which result in the increase of variance used for error term in Bayesian estimator, makes prediction error approach to observed variance ($Var_{\exp}$=0) more rapidly. This is because, at estimation step, the term $\left|\underline{H}\underline{\Sigma}(t|t-1)\underline{H}^T+\theta\right|$ in Eq. (23) increases and thus the estimated conditional distribution is more close to the prior guess (the prior distribution) while θ is larger.

Besides looking at the prediction error in the case of n=5, we also made prediction for the case of n=3, 5, 7, and 9. In this case, the simulations are made with different n, and the same parameters used are n_d=1, and simulation time is 2^{12}=4096. The realizations of these four simulations are illustrated on Fig. 14, while the mean and variance of prediction errors are plotted for different prediction time lag on Fig. 15. As shown on Fig.

15, all the mean values of prediction error for each case are almost equal to zero, and the variance of prediction error approaches to the observed variance while predictions are made at larger time lag. The case of n=3 approaches to observed variance at time lag around 3 while the case of n=9 at time lag beyond 50. The predictability is sensitive to the number of n, especially when the case n is large. The explained variance, as shown on Fig. 16, of the case with higher n decreases to zero slower than that of lower n, which is due to the higher correlation of $W'_{b,i}$ with small i in the case of large n, e.g. $W'_{b,i}$ in the case of n=9 was highly correlated (see Fig. 6) compared to all the other $W'_{b,i}$ in all cases here. Hence, the better prediction can be made with smaller error for larger n because of higher correlation.

In order to see the predictability for different $W'_{b,i}$ which is with different correlation, we plot the variance of prediction error for each $W'_{b,i}$ normalized by the observed variance, as shown on Fig. 17, with n = 3, 4, 5, 6, 7, and 9 for prediction time lag = 1. It shows that for each case of different n, the lowest value for the variance of prediction error can be made for smallest i, and the highest variance of prediction error appeared while predicting for the largest i, thus one can see that the better prediction for each $W'_{b,i}$ can be made for lower i due to higher correlation.

Chapter 4. Conclusion

From this study, the CLC process has been used for temporal rainfall modeling for the purpose that the scaling properties can be displayed, like discrete random cascade, and that the rainfall field is continuous in time, instead of discreteness for discrete cascade. With CLC model, Kalman filter can be used as a good approach of short-term rainfall prediction for the case of rainfall field with longer correlation length. This work of prediction for temporal rainfall can be extended to the case of space-time rainfall model.

Over and Gupta [*1996*] have developed a model that is dealing with space-time rainfall modeling. In this model, the spatial variations of the variables $W_i(\underline{x}, t)$ is defined by the cascade structure, while the temporal variation is assumed to have Markov property. However, the property that the time process $W_i(\underline{x}, t)$ are independent copies of the same random process W(t) at any given spatial location $\underline{x}$ makes the time series of precipitation at a point non-multifractal. This is an important difference with the model that is multifractal in both space and time.

Extension of the continuous lognormal cascade (CLC) to space-time can be made by assuming that the component at level k, $W_k(\underline{x}, t)$, is an independent realization of a continuous space/time random field $W_k(b_x^k \underline{x}, b_t^k t)$ where the branching number, b_x and b_t

are constants greater than 1 and $W(\underline{x}, t)$ is a given random field. The fields $W(\underline{x}, t)$ with autoregressive (AR) structure would be most convenient to be handled numerically (see *Bennett, 1979*). The space-time CLC model still has some resemblance with rainfall model proposed by *Over and Gupta [1996]*, but former has multifractal properties in both space and time. Then, in order to make this model more realistic, various issues will have to be addressed: (1) The lacunarity of rainfall, which could be dealt with by adding "beta component" to the lognormal multifractal model, which will be described in the next paragraph; (2) The evolutionary character of precipitation, which requires use of an adaptive version of the Kalman filter [e.g., *Strobach, 1990*]; (3) Storm motion and tracking, which is often the source of large rainfall forecasting errors.

Another issue is that rainfall is lacunar, which means that rainfall is intermittent in space and time. Therefore one must devise a method to deal with zeros for dry periods in the rainfall sequences, which depends on the rainfall model. In simple multifractal models, of the so-called beta-log-normal (Beta-LN) type, the actual rainfall process is viewed as the product of a lognormal multifractal process and an independent beta process. The former has been described in this study, and the latter has only two values, one is zero and the other is positive such that the mean value is 1. In the case of a beta-LN model, one must separately deal with the LN and beta components of the process.

For the LN component one can proceed as described before, and treat zero values as missing observations. Hence, during dry periods, one should suppress the estimation step of the Kalman filter for the LN component, and, in addition, predict the beta component of the process, again recognizing the multifractal structure of that component. There is also dependence between the LN and beta components, as the beta parameters depend on the mean rainfall intensity.

A simple way to deal with lacunarity in rainfall prediction is to use alternating renewal model of the dry and wet periods, with exponentially distributed wet periods and Weibull distributed dry periods. These distributions have been previously used by *Eagleson [1972], and Molini et al. [2001]*, and are sufficient if the objective is to reproduce rainfall extremes and many properties at sub-synoptic scales [*Veneziano and Iacobellis, 2002*]. Once the distributions of wet and dry periods are assumed, one can make prediction for the state of wet/dry conditions in the future by given current wet/dry conditions and the time elapsed since the last change in condition, which can be done by using the associated hazard functions.

Finally, these issues can be demonstrated by dealing with real rainfall data, which will require more extensive parameterization for the realistic rainfall and more detail study.

References

1. Bennett, R. J. (1979), *Spatial Time Series: Analysis, Forecasting*, Control, Pion Limited, London.
2. Benzi, R., G. Palandin, G. Parisi, and A. Vulpiani (1984), "On the Multifractal Nature of Fully Developed Turbulence and Chaotic Systems," *J. Phys. A Math. Gen.*, 17: 3521-3531.
3. Calvet, L. and A. Fisher (2001), "Forecasting Multifractal Volatility," *J. of Econometrics*, 105: 27-58.
4. Deidda, R., R. Benzi, and F. Siccardi (1999), "Multifractal Modeling of Anomalous Scaling Laws in Rainfall," *Water Resour. Res.*, 35(6): 1853-1867.
5. Deidda, R. (2000), "Rainfall Downscaling in a Space-Time Multifractal Framework" *Water Resour. Res.*, 36(7): 1779-1794.
6. Eagleson, P. S. (1972), "Dynamics of Flood Frequency," *Water Resour. Res.*, 8: 878-898.
7. Gupta, V. K. and E. C. Waymire, 1990, "Multiscaling Properties of Spatial Rainfall and River Flow Distributions," *J. Geophys. Res.*, 95(D3): 1999-2009.
8. Gupta, V. K. and E. C. Waymire, 1993, "A Statistical Analysis of Mesoscale Rainfall as a Random Cascade," *J. Appl. Meteorology* , 32(2): 251-267.
9. Ladoy, P., F. Schmitt, D. Schertzer, and S. Lovejoy (1993), "Variabilite' temporelle multifractale des observations pluviometriques a Nimes," *C. R. Acad. Sci. Paris*, 317, II : 775-782.
10. LeCam, L. (1961), "A Stochastic Description of Precipitation," in J. Neyman (ed.), *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 3,

Berkeley (CA), pp. 165-176.

11. Lovejoy, S. and D. Scherzer, (1995), "Multifractals and Rain," in A. W. Kundzewicz (ed), *Uncertainty Concepts in Hydrology and Hydrological Modeling*, Cambridge Press, pp. 62-103.
12. Marsan, D, D. Scherzer, and S. Lovejoy (1996), "Causal Space-Time Multifractal Processes: Predictability and Forecasting of Rain Fields," *J. Geophys. Res.*, 101(D21): 26,333-26,346.
13. Molini, A., P. La Barbera, and L. G.. Lanza, (2000), "A Few Stochastic Properties of Rainfall Intermittency in Space and Time," *2nd Plinius Conference on Mediterranean Storm, Eur. Geophys. Soc.*, Siena, Italy.
14. Over, T. M. and V. K. Gupta, (1994), "Statistical Analysis of Mesoscale Rainfall: Dependence of a Random Cascade Generator on Large-Scale Forcing" *J. Applied Meteorology*, 33: 1526-1542.
15. Over, T. M. and V. K. Gupta, (1996), "A Space-Time Theory of Mesoscale Rainfall Using Random Cascade," *J. Grophys. Res.*, 101(D21): 26319-26331.
16. Strobach, P. (1990), *Linear Prediction Theory: A Mathematical Basis for Adaptive Systems*, Springer-Verlag, Berlin.
17. Tissier, Y., S. Lovejoy, and D. Schertzer (1993), "Universal Multifractal: Theory and Observation for Rain and Clouds," *J. Applied Meteorology*, 32(2): 223-250.
18. Veneziano, D. (1999), "Basic Properties and Characterization of Stochastically Self-similar Processes in R^d," *Fractals*, 7: 59-78.
19. Veneziano, D and V. Iacobellis, (2002), "Multi-scaling Pulse Representation of Temporal Rainfall, " *Water Resour. Res.*, 38(8): 13.
20. Waymire, E. C. and V. K. Gupta (1981), "The Mathematical Structure of Rainfall

Representation. 1. A Review of the Stochastic Rainfall Models, 2. A Review of the Point Process Theory, 3. Some Application of the Point Process Theory to Rainfall Processes" *Water Resour. Res.*, 17(5): 1261-1294.

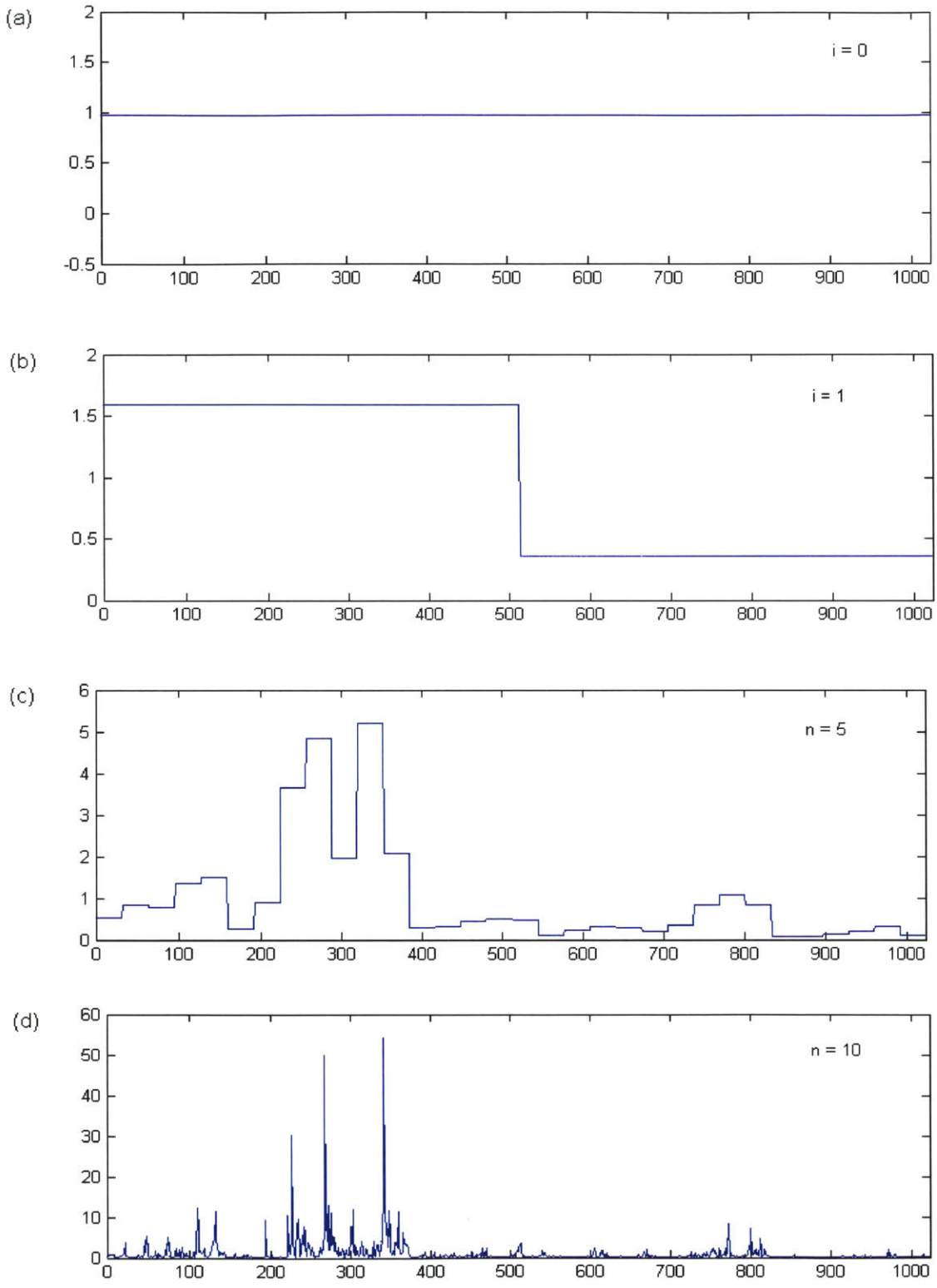

Figure 1. Discrete random cascade construction of (a) n=0, (b) n=1, (c) n=5, and (d) n=10, with b=2

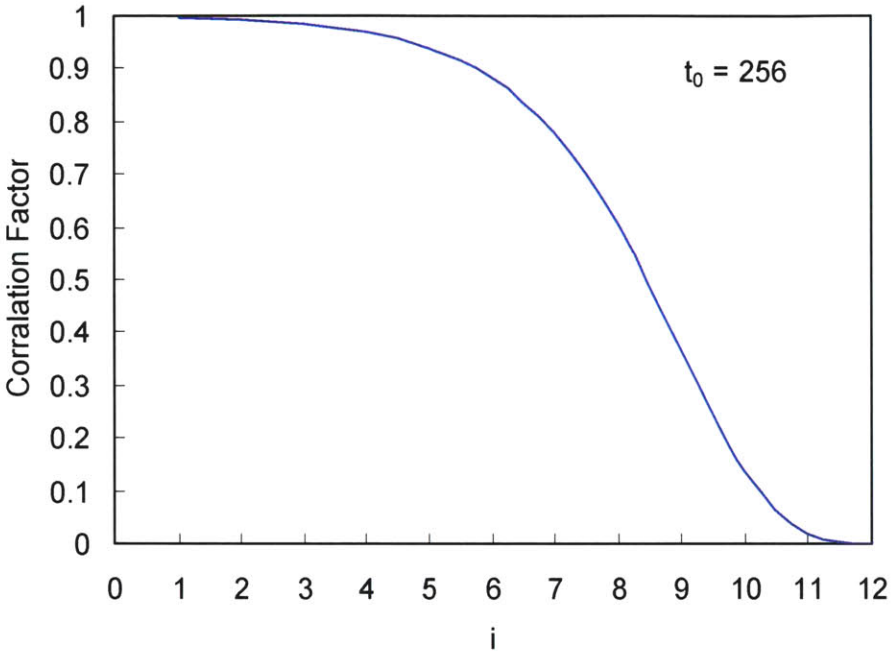

Figure 2. Plot of correlation factor ϕ_i, with different values of *i*

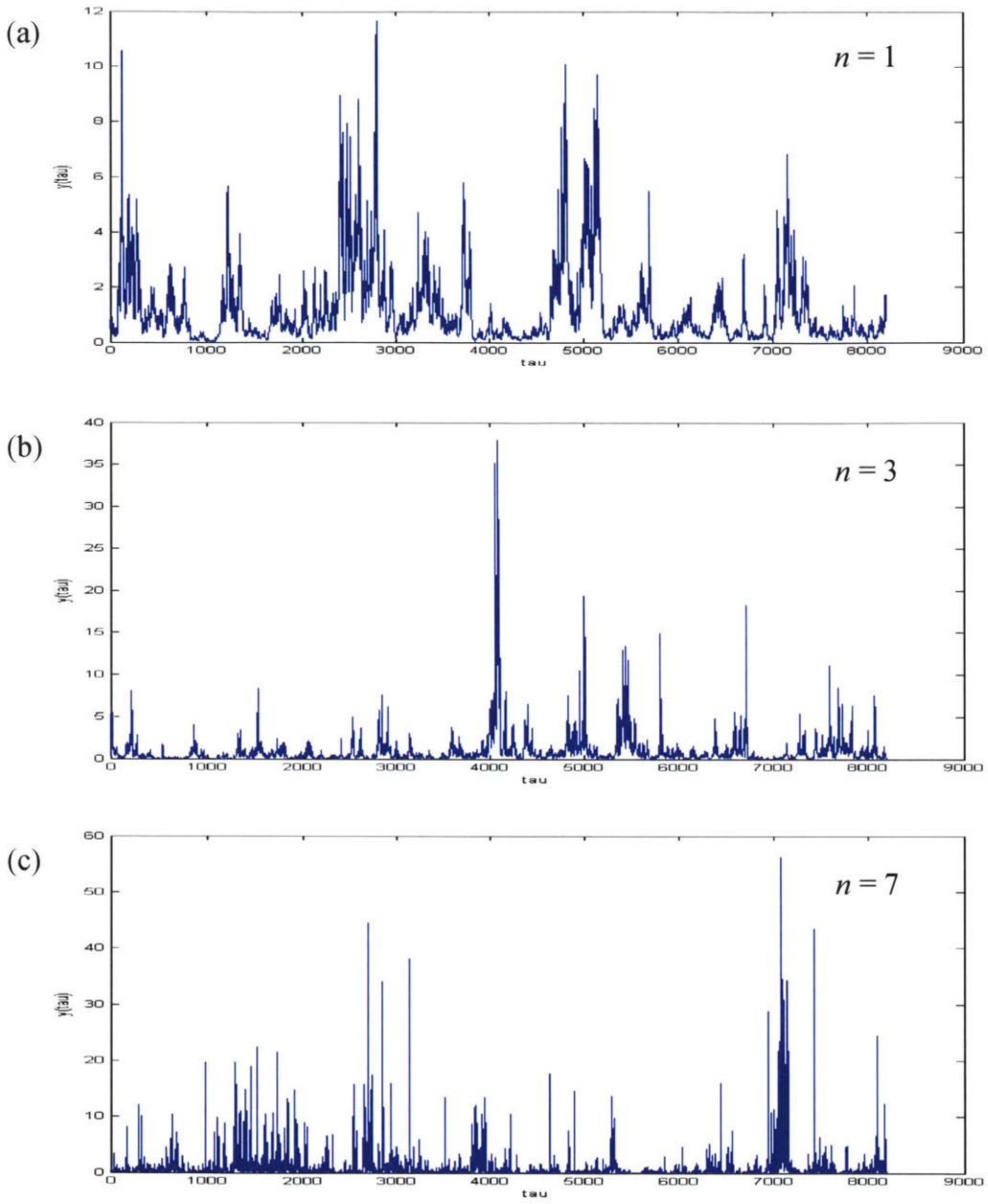

Figure 3. Simulation of $Y(\tau)$ with b=2 and total τ = 8196 for the case of (a) n = 1, (b) n = 3, and (c) n = 7

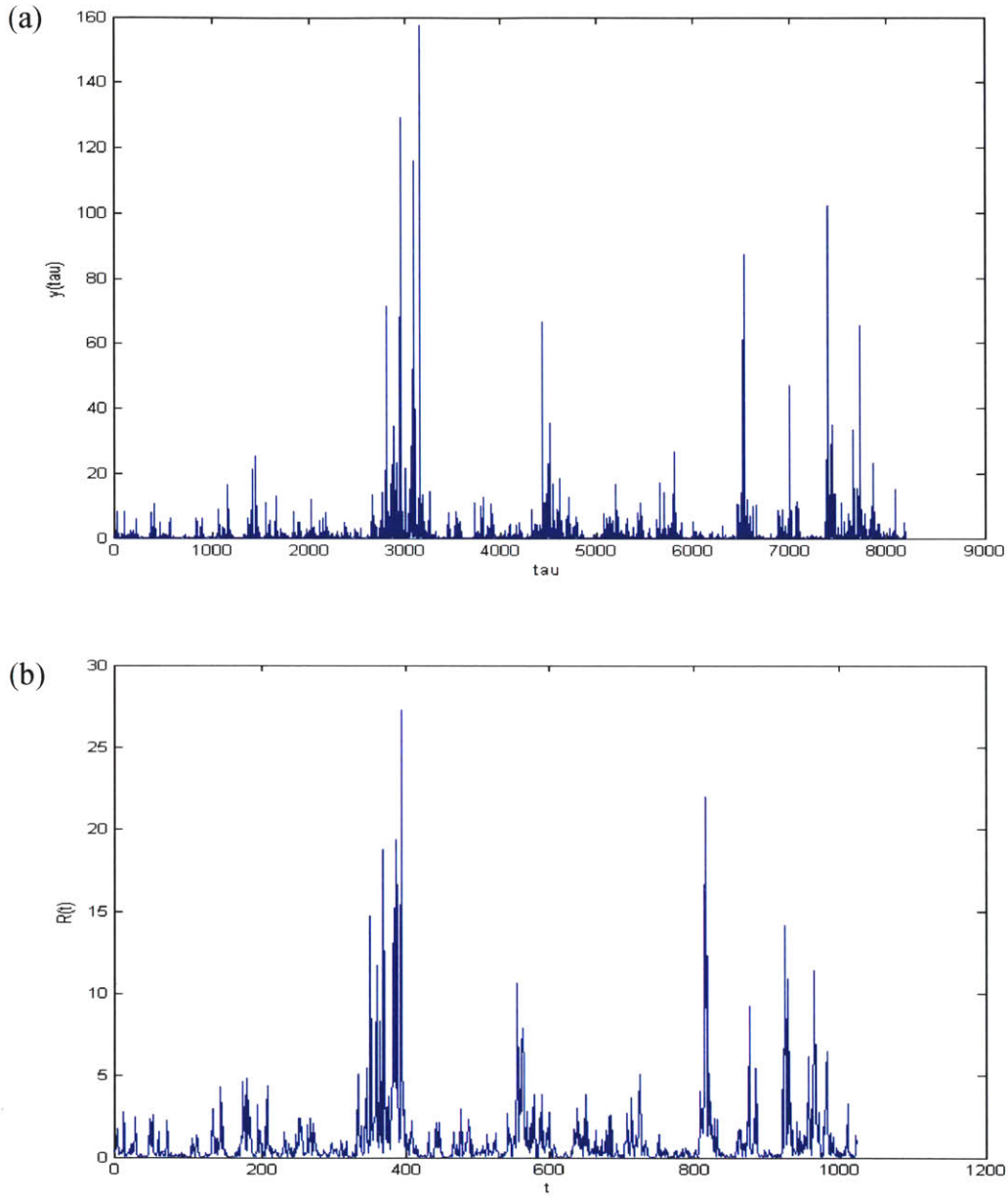

Figure 4. Simulation of (a), $Y(\tau)$ and (b) $R(t)$ with n = 9, n_d = 3 and all the other parameters as in Fig. 3..

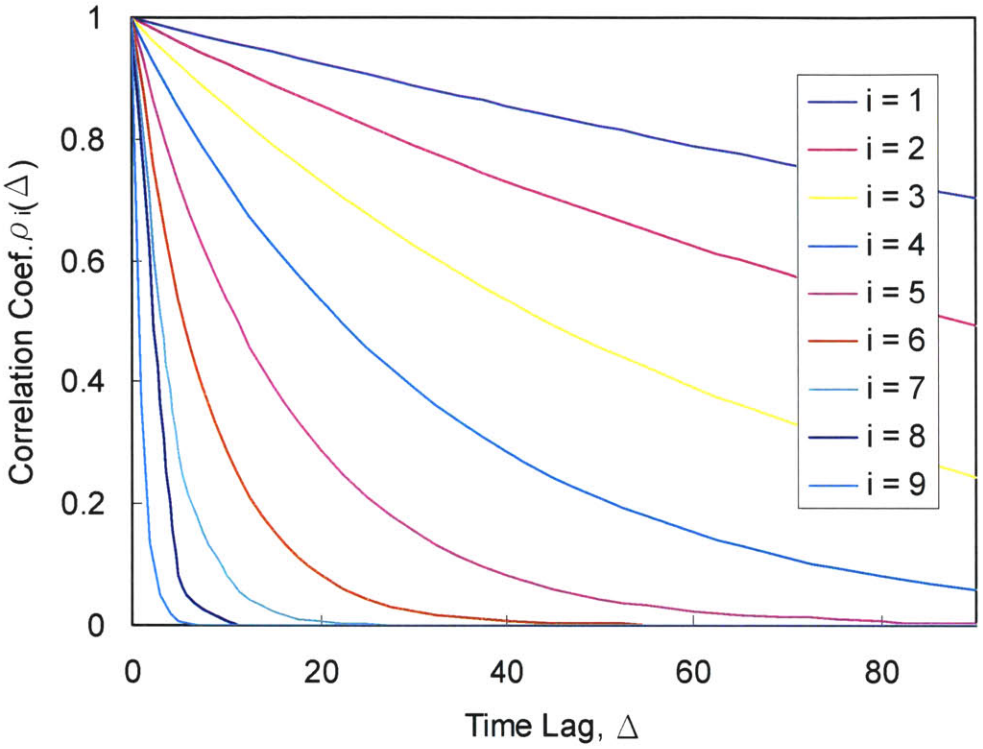

Figure 5. Correlation coefficient $\rho_i(\Delta)$ with different *i* and different time lag Δ

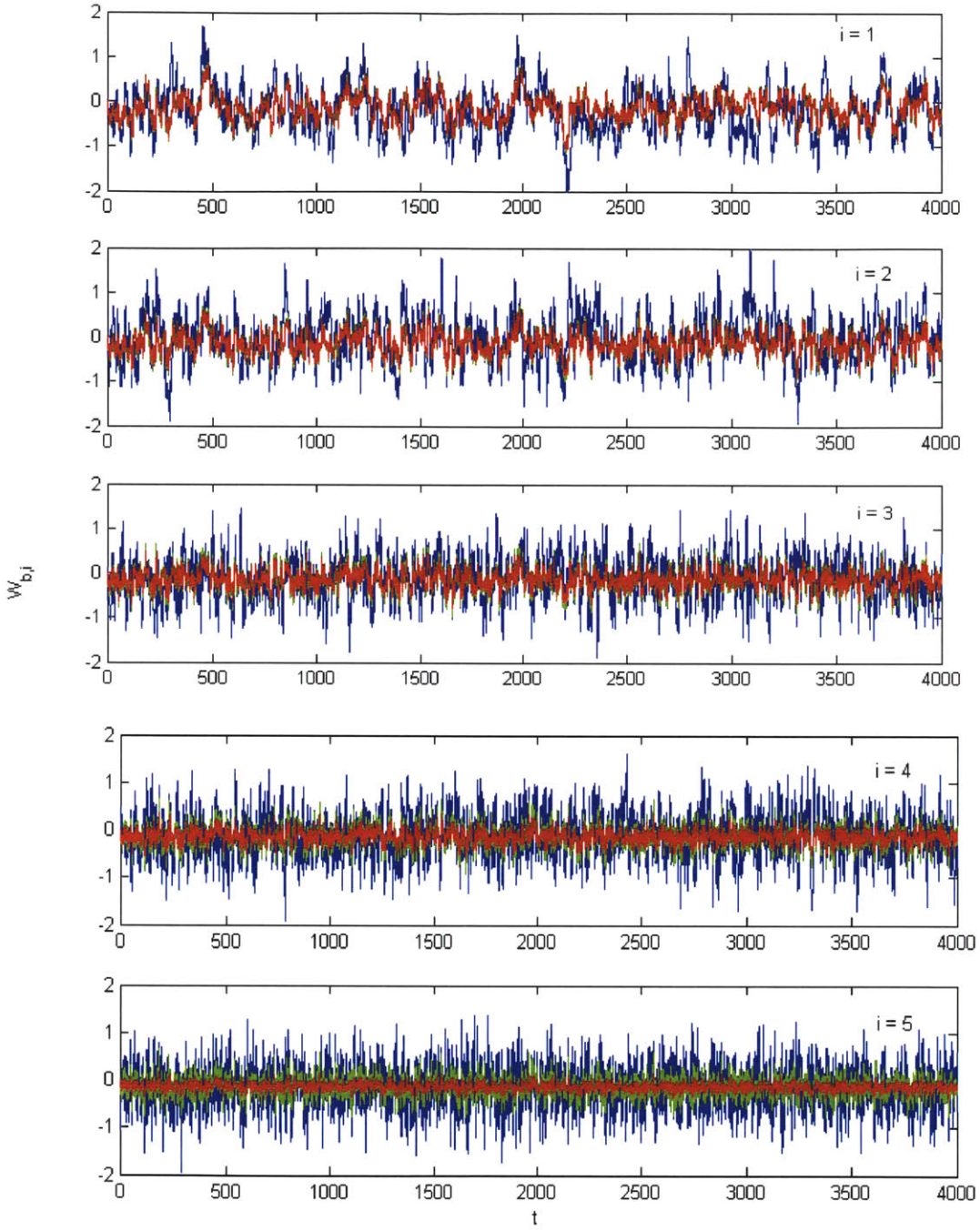

Figure 6. Comparison of $W_{b,i}(t)$ (blue) of observed R'(t), estimated values (green) and predicted value (red) of time lag = 1 for i = 1, 2, 3, 4 and 5 for the case of n_d = 1, simulation time = 4096 and n = 5

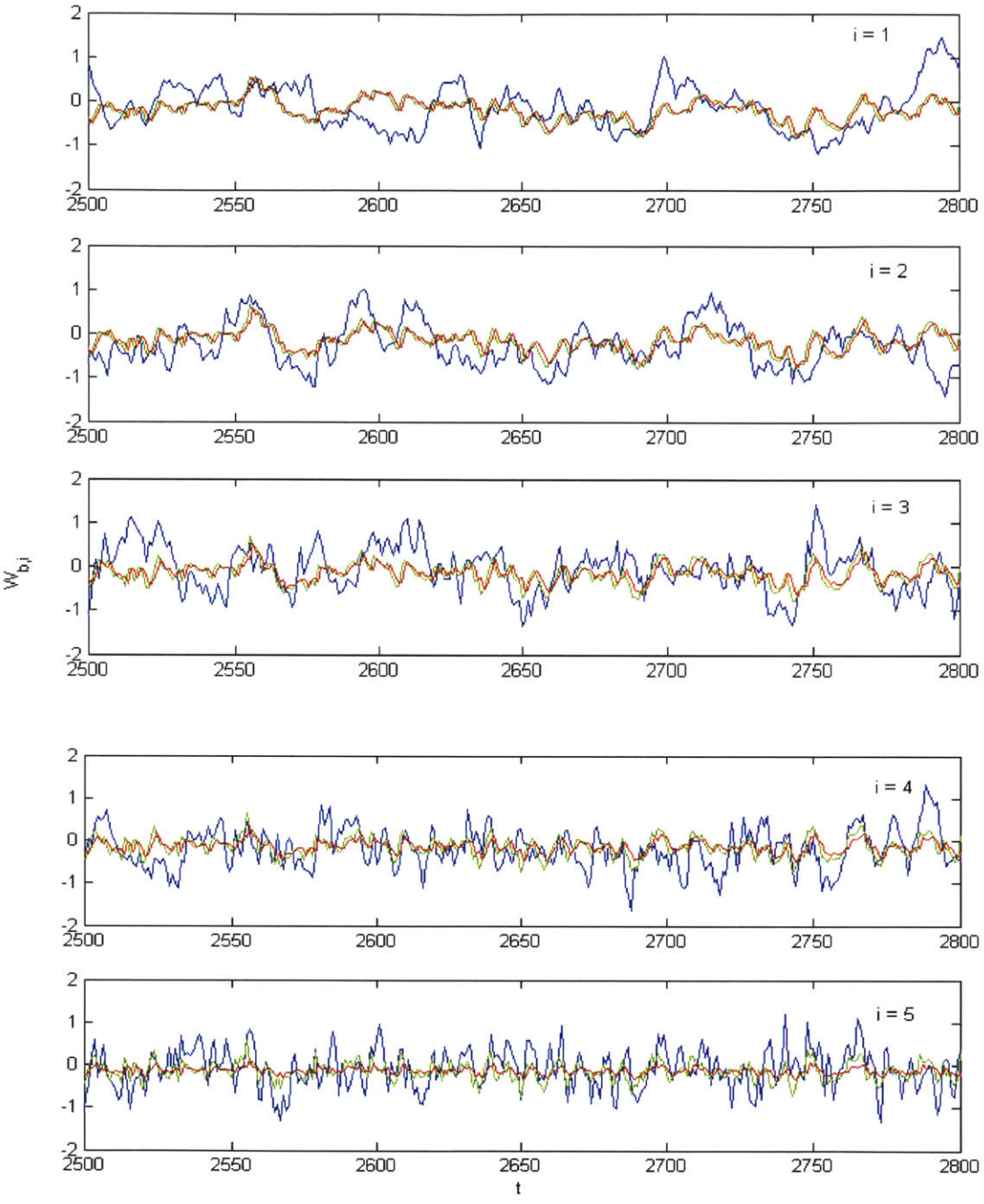

Figure 7. Comparison of $W_{b,i}(t)$ (blue) of observed R'(t) with estimated values (green) and predicted value (red) of time lag = 1 for the case i = 1, 2, 3, 4 and 5, between t = 2500 and 2800

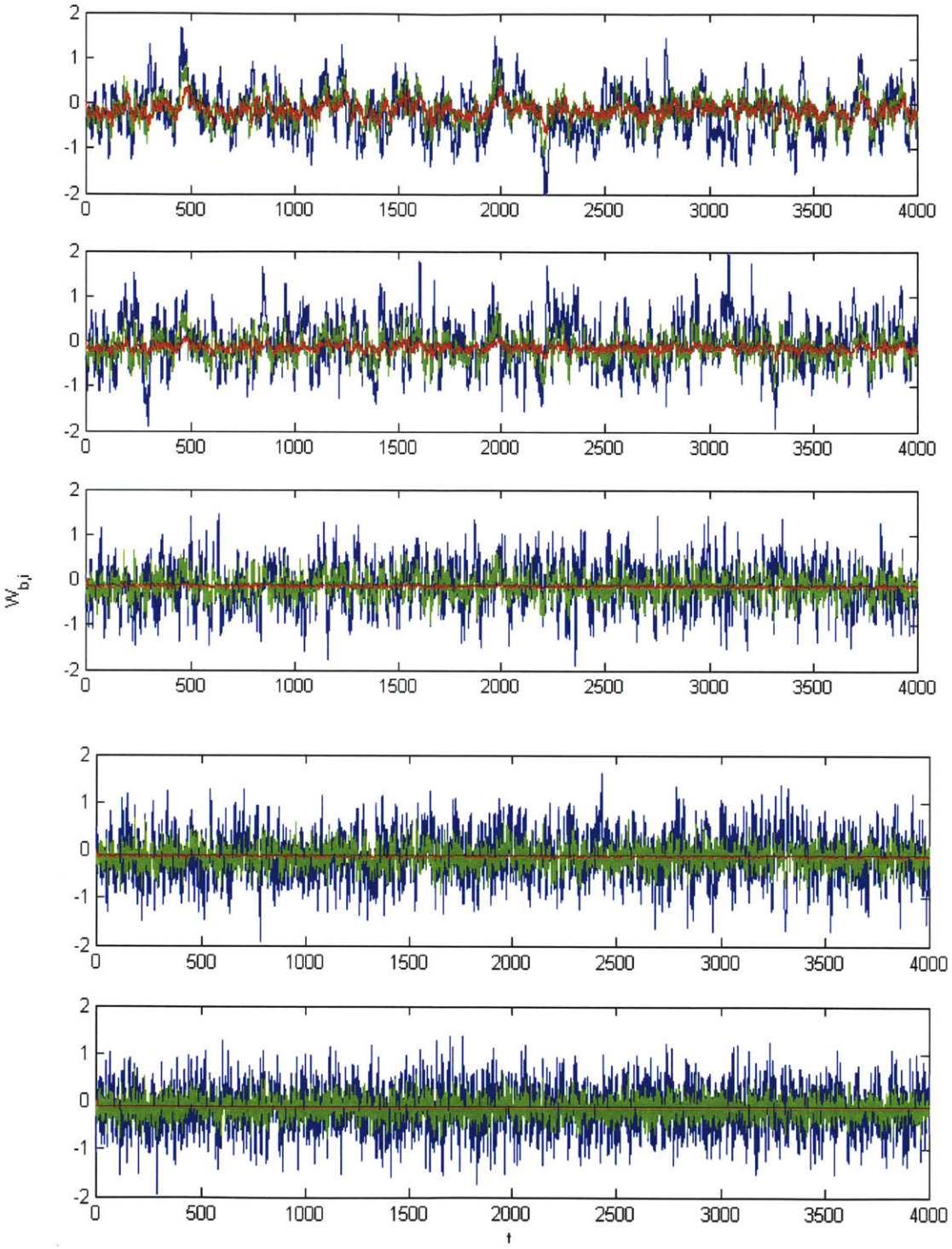

Figure 8. Comparison of $W_{b,i}(t)$ (blue) of observed R'(t) with estimated values (green) and predicted value (red) of time lag = 10 for the case i = 1, 2, 3, 4, and 5, with n_d = 1, simulation time = 4096 and n = 5

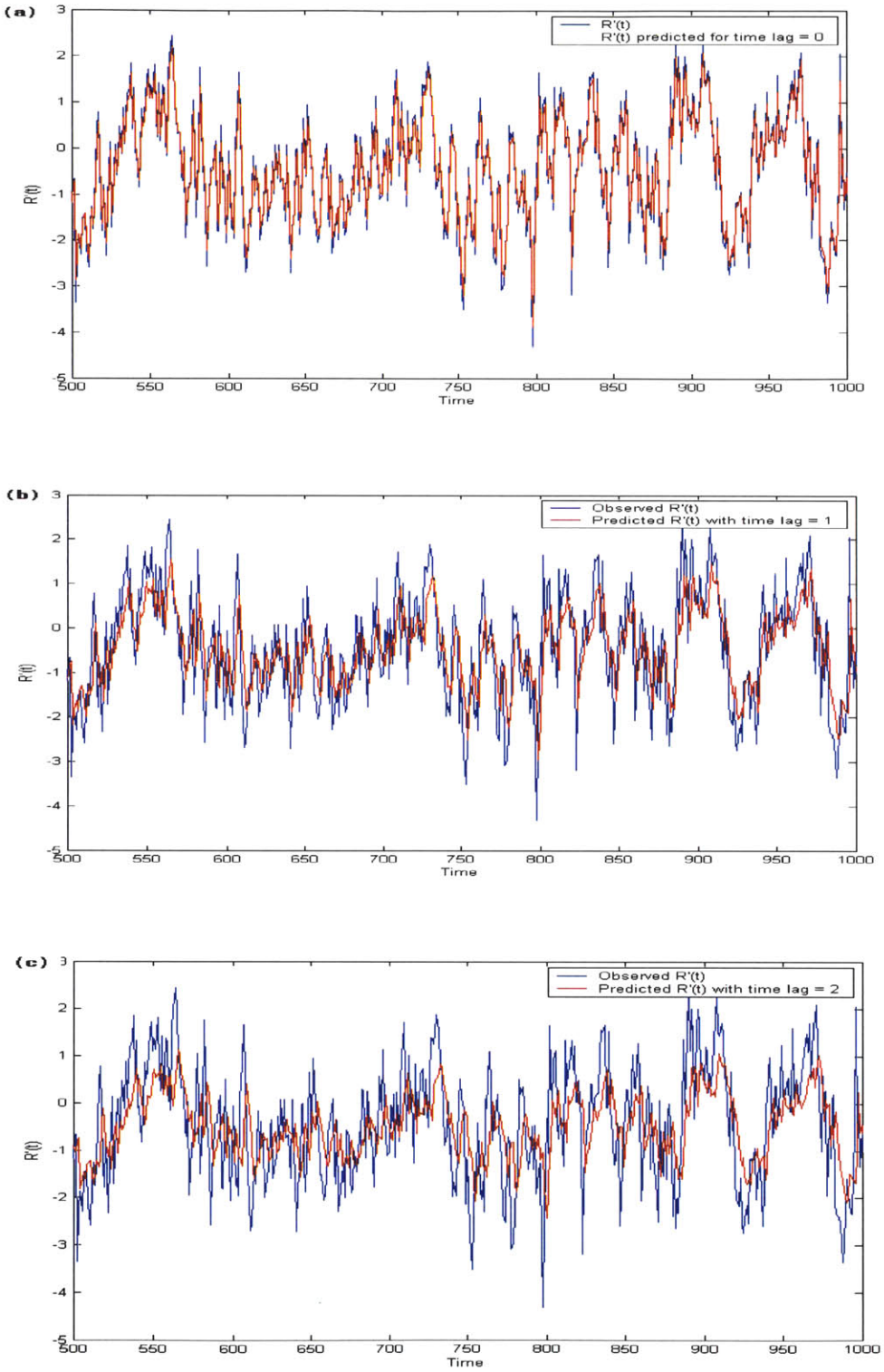

Figure 9. Comparison of observed log rainfall, R'(t) (blue) with predicted value (red) of (a) time lag = 0, (b) time lag = 1, and (c) time lag = 2 for the case of simulation time = 4096, n = 5, and n_d = 1

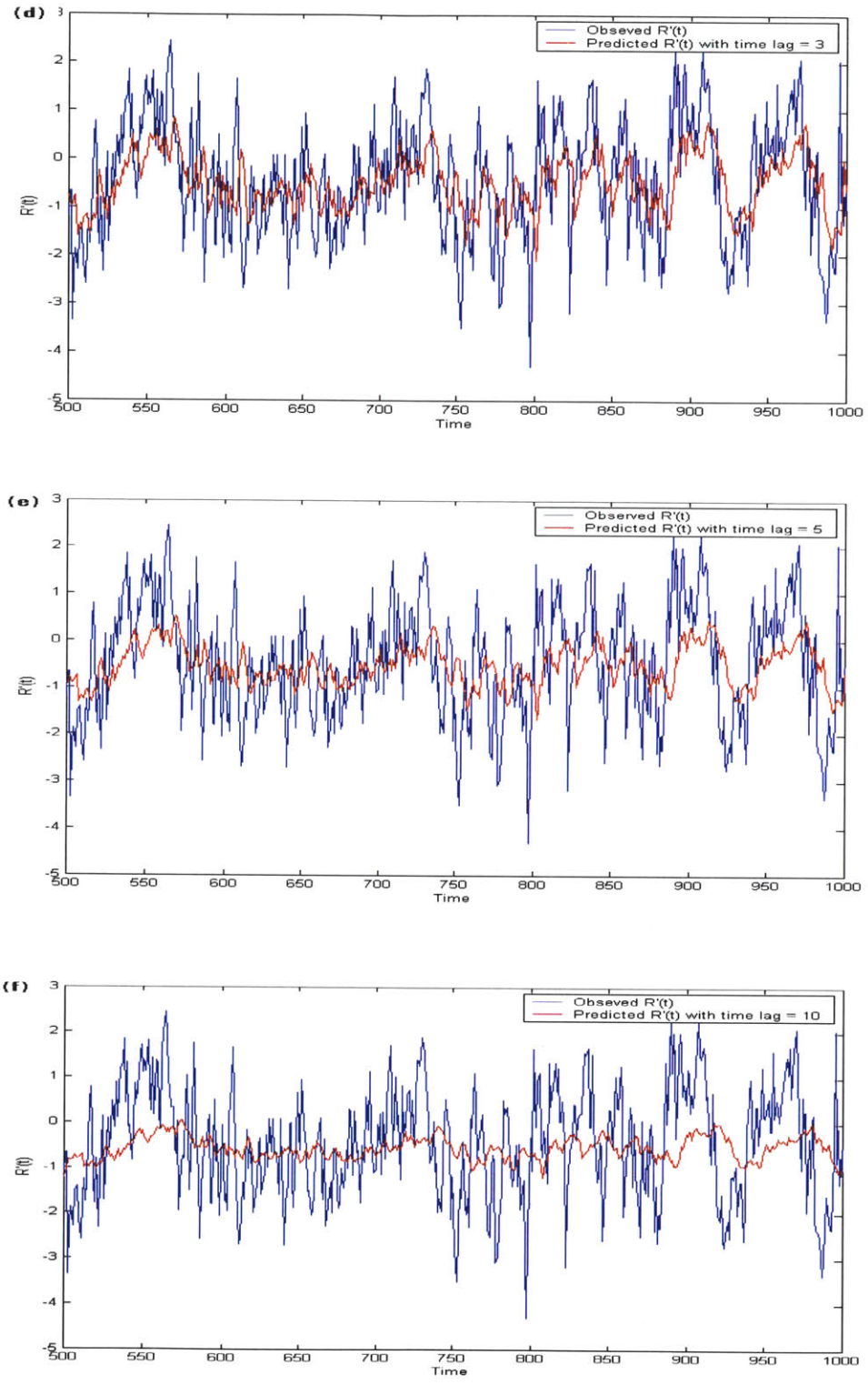

Figure 9 (continued). Comparison of observed log rainfall, R'(t) (blue) with predicted value (red) of (d) time lag = 3, (e) time lag = 5, and (f) time lag = 10 for the case of simulation time = 4096, n = 5, and n_d = 1

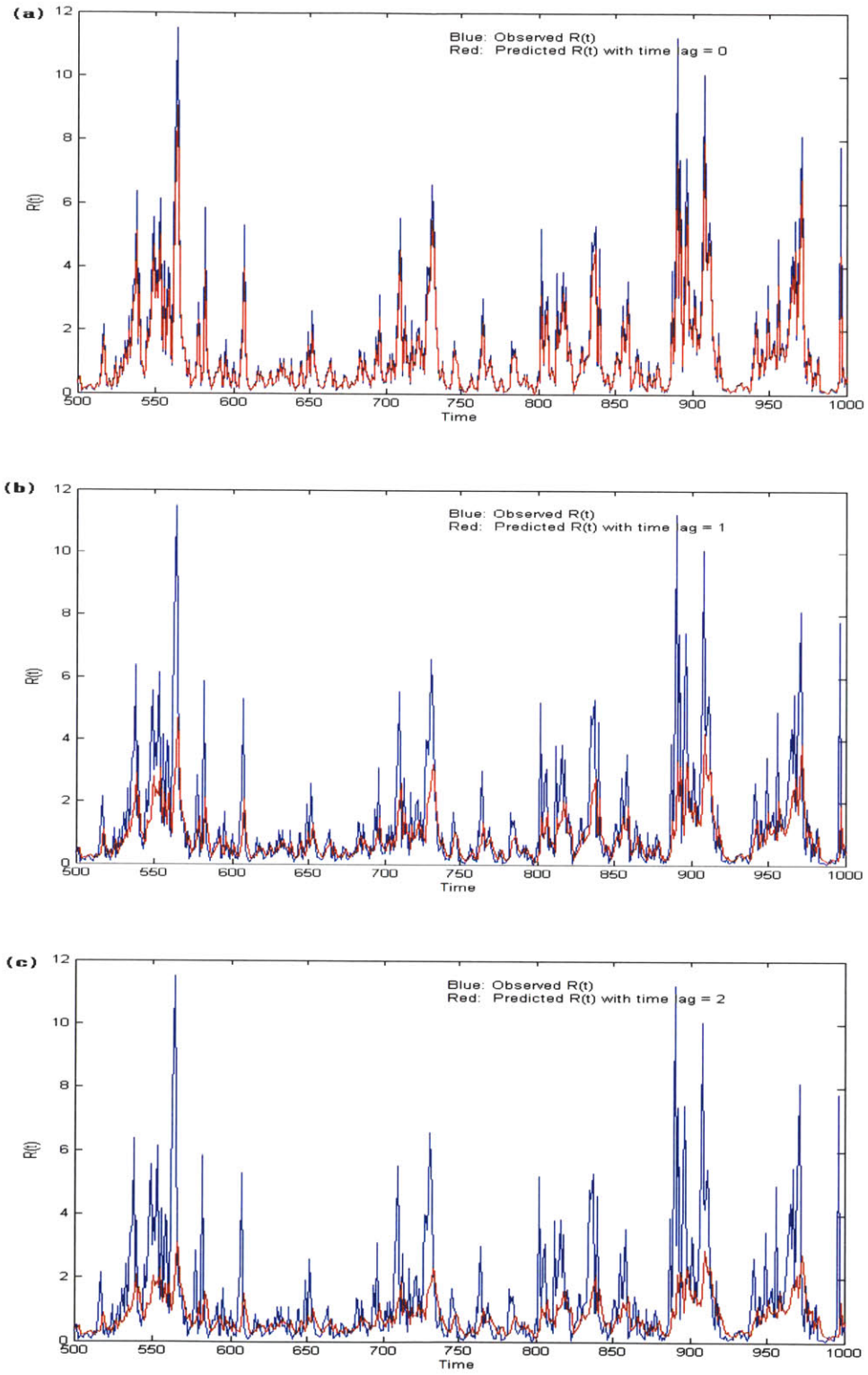

Figure 10. Comparison of observed rainfall, R(t) (blue) with predicted value (red) of (a) time lag = 0, (b) time lag = 1, and (c) time lag = 2 for the case of simulation time = 4096, n = 5, and n_d = 1

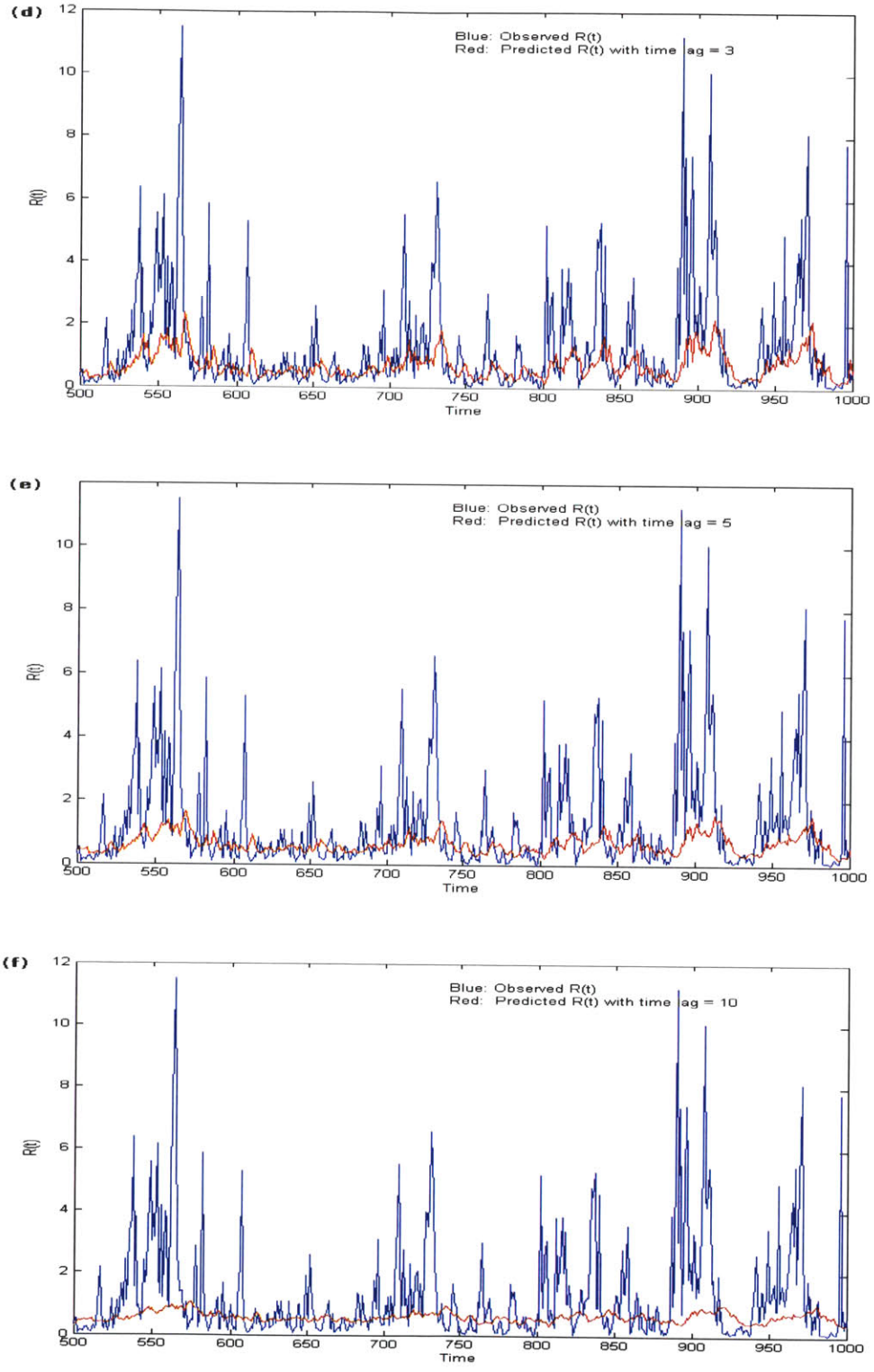

Figure 10 (continued). Comparison of observed rainfall, R(t) (blue) with predicted value (red) of (a) time lag = 3, (b) time lag = 5, and (c) time lag = 10 for the case of simulation time = 4096, n = 5, and n_d = 1

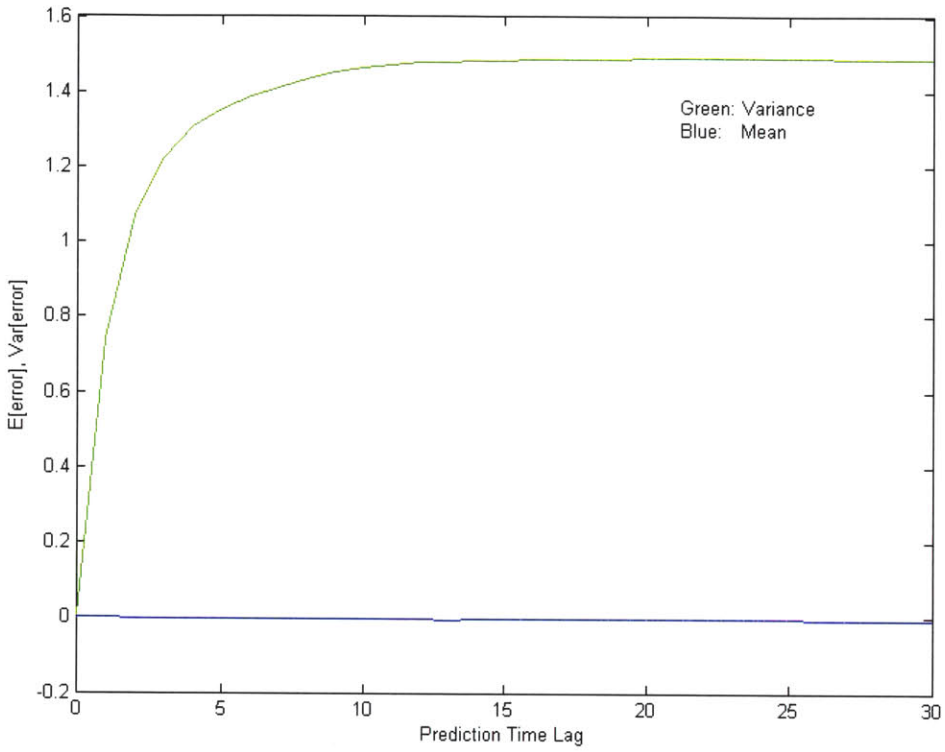

Figure 11. Plot of mean and variance of prediction error versus prediction time lag from 1 to 30, with simulation condition of n = 5, n_d = 1 and total simulation time = 2^{12}

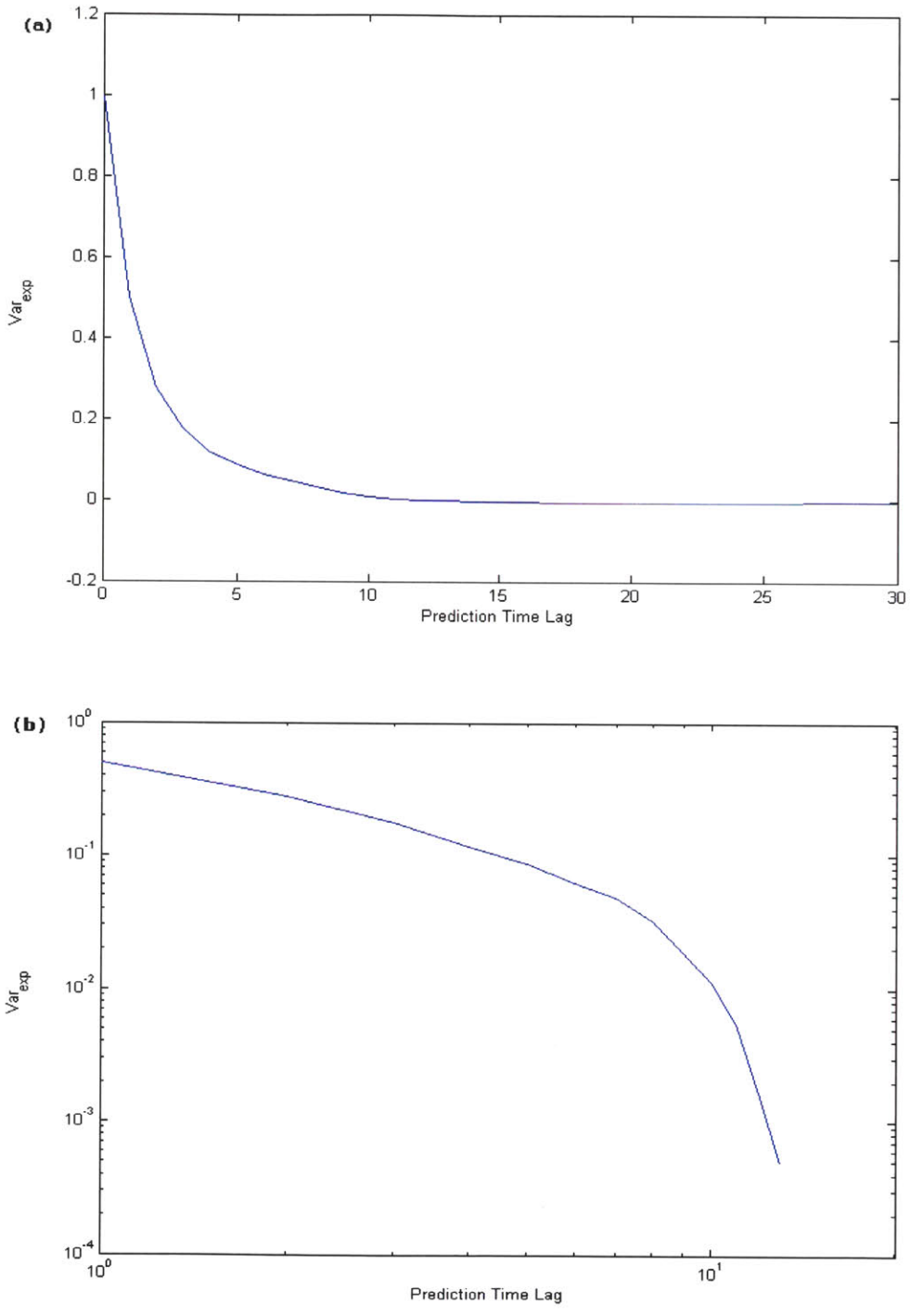

Figure 12. Var_{exp} versus prediction time lag from 1 to 30, with simulation condition of n = 5, n_d = 1 and total simulation time = 2^{12}

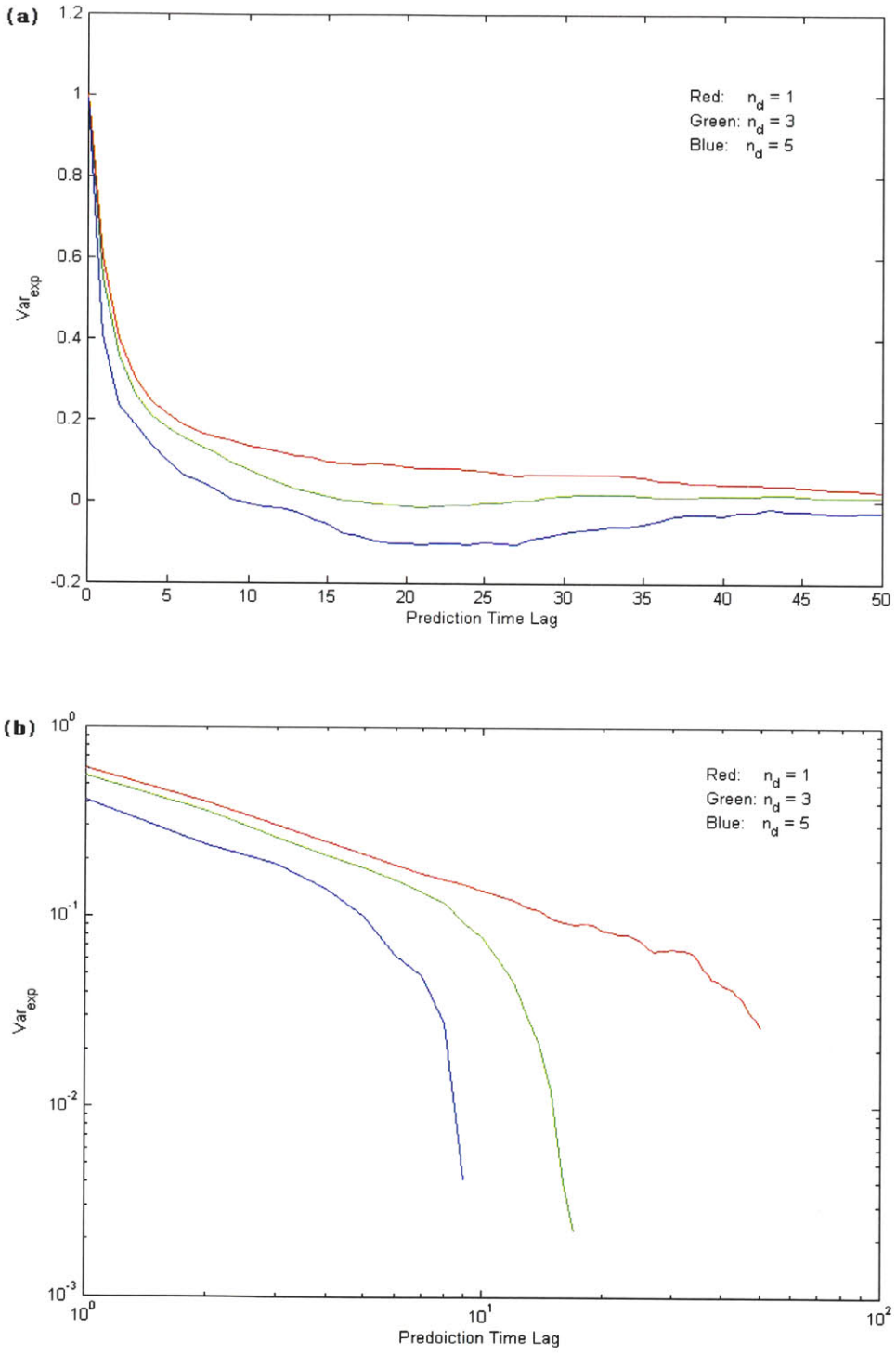

Figure 13. Var_{exp} of time lag from 1 to 50 for the case of n = 7, n_d = 1, 3, and 5, and simulation time = 4096

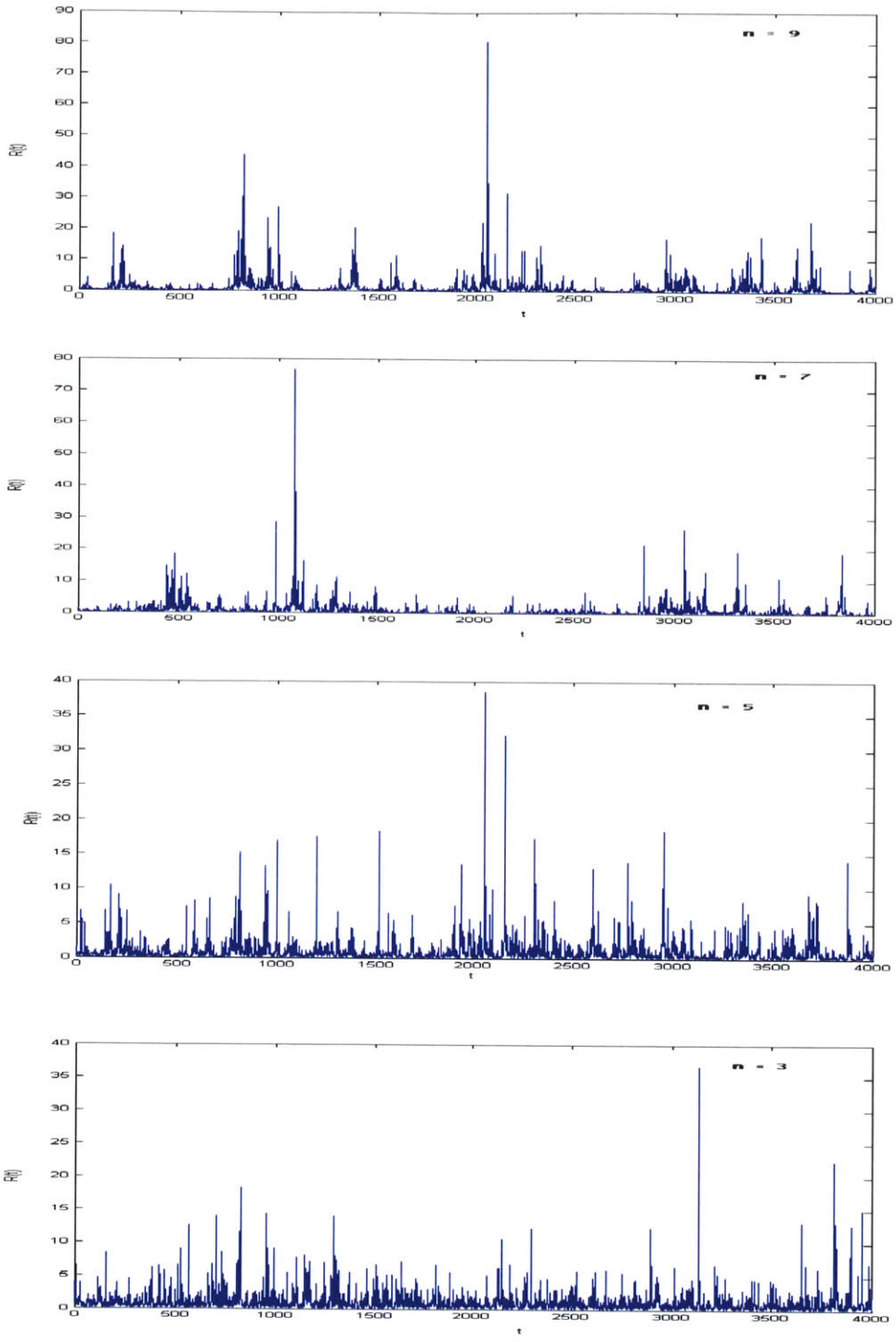

Figure 14. Realization of R(t) of n = 3, 5, 7, and 9, with n_d = 1 and simulation time = 4096

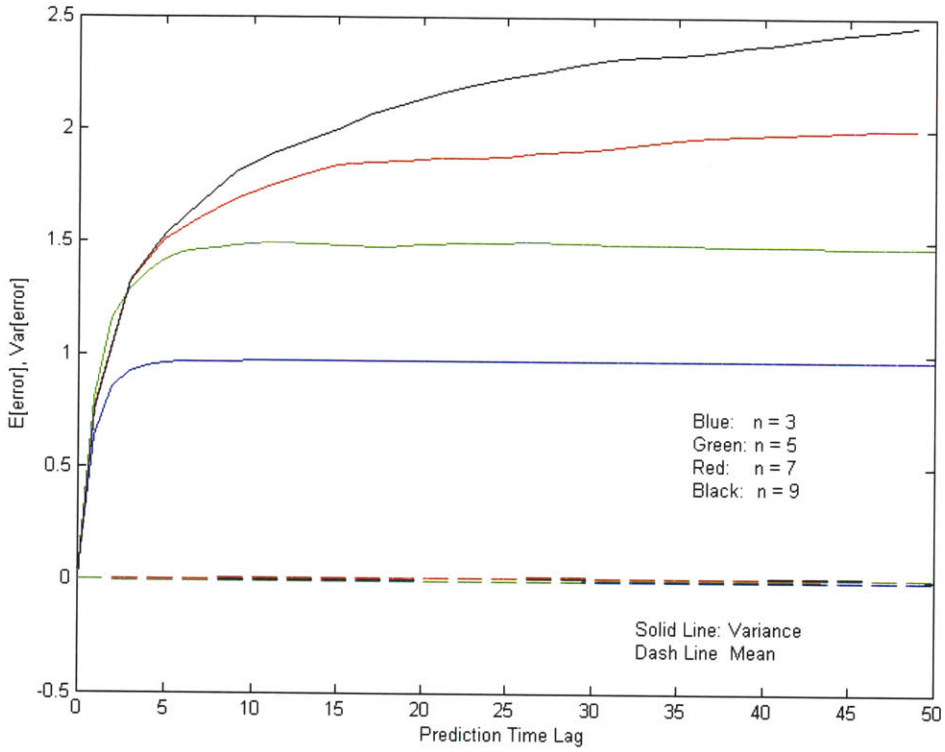

Figure 15. Plot of mean and variance of prediction error versus time lag for n = 3, 5, 7, and 9

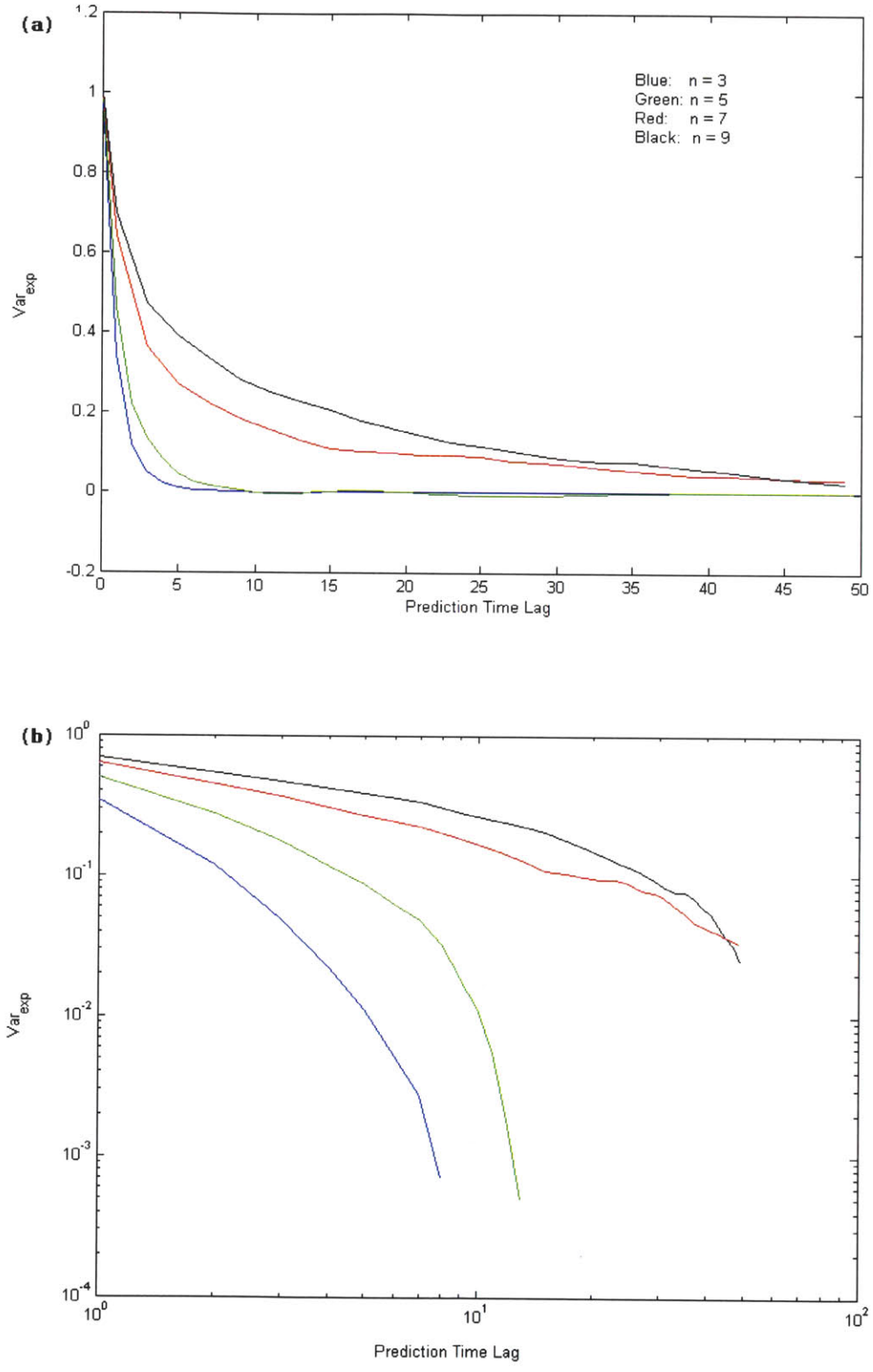

Figure 16. Plot of Var_{exp} of time lag from 1 to 50 for the case of n = 3, 5, 7, and 9 with $n_d = 1$ and simulation time = 4096

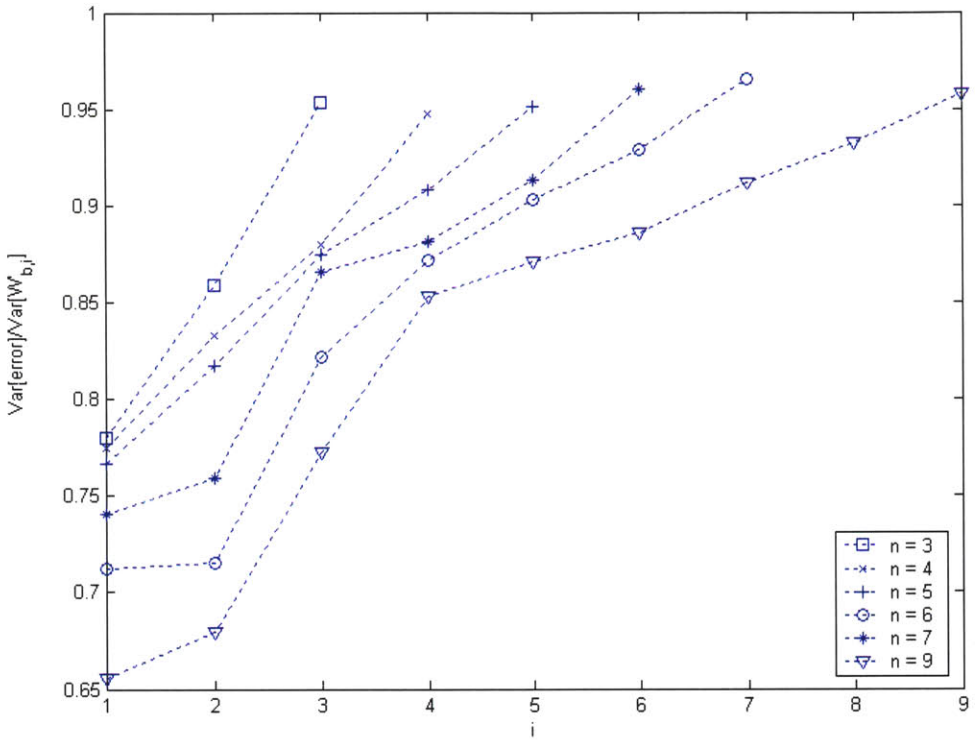

Figure 17. Normalized variance of prediction error for each $W_{b,i}$ component versus i with n = 3, 4, 5, 6, 7, and 9