MINISTRY OF EDUCATION AND TRAINING
DANANG UNIVERSITY

NGUYEN VAN HUNG

RESEARCH ON THE FIRST ORDER GAMMA AUTOREGRESSIVE [GAR(1)] MODEL TO APPLY IN THE FIELD OF HYDROLOGY

SPECIALIZATION: COMPUTER SCIENCE
CODE: 62.48.01.01

SUMMARY OF DOCTORAL DISSERTATION

DA NANG - 2016

The doctoral dissertation has been fulfilled at

Danang University

Science Advisors:

1. Associate Professor, Dr. Sc. Tran Quoc Chien

2. Professor, Dr. Huynh Ngoc Phien

Reviewer 1: Professor, Dr. Nguyen Thanh Thuy,
Hanoi University of Technology.

Reviewer 2: Associate Professor, Dr. Nguyen Mau Han,
Hue University of Science.

Reviewer 3: Dr. Pham Minh Tuan,
DaNang University.of Technology.

The dissertation is defended at the Examination Committee at the level of Danang University on June 24, 2016.

The dissertation can be referred at:

- Vietnam National Library;
- The Center of Information and Documentation of Danang University.

INTRODUCTION

Nowadays, computer science plays a very important role in the development of worldwide, has deeply impact on most of the fields of engineering, socio-economic. There were many works in the field of computer science research on telecom-informatics, biomedical-informatics already bringing tremendous efficiency to human life, meanwhile, works research on hydrological-informatics are still shortcomings. The purpose of this study aims to contribute to the development of hydrological-informatics now and in the future. To reach this purpose, the objectives of this study are as follow:

- Research on GAR(1) model, overview of works related to: GAR(1) model, stochastic simulation method, methods for generating random variates, models for simulating of streamflows and reservoir capacity problem.

- Study of the algorithms for generating GAR(1) variables includes: algorithms generate the random variables with the uniform distribution, exponential distribution, normal distribution, Poisson distribution and the gamma distribution.

- Study of the models for simulating of monthly and annual streamflows and investigation on the mean range of reservoir storage with infinite capacity.

CHATER 1
THE GENERAL PROBLEMS

To reach the objectives of the study: ***Research on The first order gamma autoregressive [GAR(1)] model and to apply in the field of hydrology***, the author studies documents, works have been published in local and abroad related to the following issues:

- Theoretically: The basic research on probability theory, study of the algorithms to generate random variables, methods, models and algorithms used to simulate the monthly and annual streamflows and the reservoir problems.

- Reality: The results related to the experiments, simulating the streamflows at the hydrological gauging stations and reservoir capacity.

1.1. Several Basic Problems of Probability Theory

This section presents the basic theory of probability includes the concept of random variable, distribution, probability density function and the numerical characteristics of random variables such as: the expectation, variance, skewness coefficient and the kurtosis coefficient, and as a basis for further study.

1.2. The Gamma Distribution

1.2.1. The Probability Density Function

A continuous random variable X is said to have a three-parameter gamma distribution if its density can be expressed as:

$$f(x) = \frac{(x-c)^{a-1}e^{-(x-c)/b}}{b^a\Gamma(a)}, \qquad (1.1)$$

where $a > 0,\ b > 0,\ c \geq 0,\ x \geq c,$ and $a,\ b,\ c$ are respectively the shape, scale, and the location parameter. The gamma function $\Gamma(a)$ is defined by:

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt, \quad a > 0,$$

when $c = 0$ we have the two-parameter gamma distribution, and, when $c = 0$ and $b = 1$ we have the one-parameter gamma distribution.

By transformation method, the gamma variables with two parameters or three parameters can be converted into the gamma variables with one parameter. For the three-parameter variables, the transformed variables can be obtained by: $y=(x-c)/b$ or $x=c+by$. For two-parameter variables the transformation used is: $y=x/b$ or $x=by$. Hence, y follows the one-parameter gamma distribution.

1.2.2. The Statistical Descriptors

The statistical descriptors of the three-parameter gamma distribution are given by the following formulas:

Expectation: $E(X) = a,$ (1.2)

Variance: $Var(X) = a,$ (1.3)

Skewness: $g = \frac{2}{\sqrt{a}}.$ (1.4)

1.3. The First-order Autoregressive [GAR(1)] Model with Gamma Variables

1.3.1. GAR(1)Model

The model by Lawrance and Lewis(1981) has the following form:

$$X_i = \Phi X_{i-1} + e_i, \quad (1.5)$$

where X_i *is* the random variable representing the dependent processes at time i, Φ *is* autoregressive coefficient and e_i *is* an independent variable to be specified. X_i has a marginal distribution given by a three-parameter gamma density function defined as Eq.(1.1). The process defined by Eq.(1.5) is denoted as the GAR(1) model. To simulate the process, the parameters of the model must be known and e_i can be generated by certain generators (unit uniform, exponential and Poisson generator).

1.3.2. Estimation of GAR(1) Model Parameters

Fernandez and Salas(1990) have presented a procedure for bias correction based on computer simulation studies, applicable for the parameters of GAR(1) model. The stationary linear GAR(1) process of eq.(1.5) has four parameters, namely a, b, c and Φ. By using the method of moments, these parameters and the population moments of the variable X_i have the following relationships:

$$M = c + ab. \quad (1.6)$$

$$S^2 = ab^2, \quad (1.7)$$

$$G = 2/\sqrt{a}, \quad (1.8)$$

$$R = \Phi, \quad (1.9)$$

where M, S^2, G, R are the mean, variance, skewness coefficient, and the lag-one autocorrelation coefficient, respectively. These population statisticals can be estimated based on a sample $\{X_1, X_2, \ldots, X_N\}$ by:

$$m = \frac{1}{N}\sum_{i=1}^{N} X_i, \quad (1.10)$$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - m)^2, \quad (1.11)$$

$$g = \frac{N}{(N-1)(N-2)s^3}\sum_{i=1}^{N}(X_i - m)^3, \quad (1.12)$$

$$r = \frac{1}{(N-1)s^2}\sum_{i=1}^{N-1}(X_i - m)(X_{i+1} - m), \quad (1.13)$$

where *m*, *s*, *g* and *r* are estimators of *M*, *S*, *G* and *R* respectively and *N* is sample size. As the variables are dependent and nonnormal, some of these estimators are biased. Hence some correction needs to be made and after that we obtain the unbiased estimators of *M, R, S* and *G*. Once all these values are computed, Eqs.(1.6)-(1.9) are used to estimate the set of model parameters *a*, *b*, *c* and Φ, respectively.

1.4. Generating of GAR(1) Variables

To generate GAR(1) variables, the algorithms for generating of random variables having unit uniform distribution, exponential distribution, normal distribution, Poisson distribution and gamma distribution need be used. Various algorithms have been suggested to generate the random variables having gamma distribution and divided into two cases: (1) For shape parameter $a \leq 1$, and, (2) For shape parameter $a > 1$. Several works suggested algorithms for generating gamma variables with any value of shape parameter such as the work of Marsaglia and Tsang (2000), and recently, as remarked by Hong Liangjie (2012), the algorithm proposed by Marsaglia and Tsang (2000) is ease coding and having fastest speed and was installed in the GSL library and Matlab software "gamrnd".

1.5. Streamflow Simulation Problem

The problem of streamflow simulation is based on annual or monthly historical data which were observed at hydrological stations, using the model to generate sequences of data with length of *n* having the same numerical characteristics, namely mean value, standard deviation, skewness coefficient and correlation coefficient of historical data. The parameters of the historical series of monthly flows (i.e. mean value, standard deviation, skewness coefficient) are computed by the following expressions:

$$m_j = \frac{1}{N}\sum_{i=1}^{N} X_{i,j}\,, \qquad j = 1..12$$

$$s_j^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_{i,j} - m_j)^2\,, \qquad j = 1..12$$

$$g_j = \frac{N}{(N-1)(N-2)s_j^3}\sum_{i=1}^{N}(X_{i,j} - m_j)^3, j = 1..12.$$

The models using for streamflow simulation are classified into parametric and nonparametric models. Parametric models are divided into categories: independent and dependent of historical data. Starting with the assumption that history data is independent and having defined probability distribution, several models have been proposed, and in which, the Thomas-Fiering model using for streamflow simulation with any probability distribution type is commonly used. With the diversity of climate, many works determined the streamflows are often follow a dependent and skew distribution, and for this case, Fernandez and Salas(1990) showed that GAR(1) model is very effective in annual streamflow simulation.

1.6. Reservoir Capacity Problem

There are many problems in the study of reservoir such as planning, designing, operating or multi-reservoir operating. For the problems of planning, designing reservoirs, important issue is to determine the capacity of reservoir based on the inflows and the outflows of reservoir. Studies of reservoir capacity depending on the cases, namely finite, semi-finite, and infinite. A finite capacity reservoir allows both spillage and emptiness, while a semi-finite capacity reservoir allows either spillage or emptiness only. An infinite capacity reservoir allows neither spillage nor emptiness in the sense that it will never spill or run dry throughout its life time of n years and as shown in the work of Salas-La Cruz(1972), this assumption is suitable for planning and design studies of large capacity reservoirs. (hundred million m^3 and up). However, with climate change being recognized widely nowadays, extreme conditions of rainfall and runoffs, resulting in long periods of droughts and big floods, will occur. These conditions call for the construction of reservoirs with big storage capacity for flood protection and for adequate water supply during drought periods. As such, range analysis becomes an appropriate method for use again.

CONCLUSION OF CHAPTER 1

From the systematic study of themed works published, the author discovered the following shortcomings:

There is no study, evaluation, selection of the appropriate algorithms to generate GAR(1) variables, no suggested model using for monthly streamflow simulation with GAR(1) process and how to determine the mean range of reservoir storage with GAR(1) inflows.

From the foregoing shortcomings, the research orientations are: considers the effectiveness and selects the appropriate algorithms for generating GAR(1) variables, studies the numerical characteristic of the sum of GAR(1) variables, investigates the monthly and annual streamflow simulations with GAR(1) variables and the mean range of reservoir storage with GAR(1) inflows.

CHAPTER 2
ALGORITHMS FOR GENERATING GAR(1) VARIABLES

This chapter presents the algorithms for generating GAR(1) variables. By means of theoretical and simulation methods, the basic theory and the algorithms for generating GAR(1) variables are studied, implemented and tested.

2.1. Investigation of Several Algorithms for Generating GAR(1) Variables

To apply the GAR(1) model in practice, needs to generate the GAR(1) variables based on the statistical sample. To generate

GAR(1) variables should incorporate random variable generators with the unit uniform distribution, exponential distribution, normal distribution, Poisson distribution and the gamma distribution.

2.2. Proposed Algorithm to Generate The Gamma Variates

The algorithm by Minh(1988) was used to generate variates having a gamma distribution with shape parameter $a>1$ only. Based on the result of Marsaglia and Tsang (2000), the method which is an improvement of Minh's algorithm to generate gamma random variables for all values of shape parameter proposed by Hung, Trang and Chien(2014) denoted IMGAG algorithm as follows:

(1) If $a>1$ using Minh's algorithm with shape a to generate X, go to step (3);

(2) If $1\geq a>0$ using Minh's algorithm with shape $a+1$ to generate X^*, compute $X = X^* U^{1/a}$ with $U\sim$U(0,1);

(3) Deliver X;

(4) End.

2.3. Proposed Additional Criterion for Evaluating The Effectiveness of Random Variable Generators

In practice, the evaluation of the effectiveness of a random variable generator is mainly based on the following criteria: the complexity and ease to implement of the algorithm. In addition to the above criteria; Hung, Trang and Chien (2014) proposed additional criterion to evaluate the effectiveness of different algorithms used to generate random variables with a specific type of distribution as follows: using the algorithm to generate the sequence of random number and evaluating the randomness and the preservation of the numerical characteristics of the distribution based on the mean, variance and the skewness of the series of generated data.

2.4. Computer Simulation

2.4.1. Simulation Methods

To generate the gamma random variables, the algorithms were used: Ahrens (1974) for the case of shape parameter $a\leq 1$, Tadikamalla (1978) for the case of shape parameter *a*>1, IMGAG (2014) and Marsaglia (2000) for all values of shape parameter *a*. The algorithms were implemented in the C language and with the different values of shape parameter (from 0.1 to 500), uses each algorithm to generate series of 10,000 gamma random numbers. Based on the series of generated random numbers, the statistical parameters: mean value, variance and skewness coefficient computed by using the formulas (1:10) - (1:12). The correlation coefficient computed using the formula (1.13).

2.4.2. Experimental Results

From the simulation experiments, the results are given in tables 2.1 - 2.3 and showed in figures 2.1 - 2.3 as follow:

Table 2.1. Mean values of 10,000 generated gamma variables using algorithms: IMGAG, Marsaglia and Ahrens

a	IMGAG		Marsaglia		Ahrens	
	D.gen.	% Err.	D.gen.	% Err.	D.gen.	% Err.
0.1	0.099	*0.78*	0.114	*14.32*	0.098	*2.13*
0.2	0.195	2.39	0.230	*15.02*	0.199	*0.55*
0.3	0.296	*1.27*	0.343	*14.38*	0.297	*1.09*
0.4	0.390	*2.57*	0.450	*12.67*	0.394	*1.54*
0.5	0.498	*0.41*	0.564	*12.79*	0.502	*0.34*
0.6	0.603	*0.58*	0.665	*10.90*	0.592	*1.26*
0.7	0.693	*1.04*	0.778	*11.14*	0.700	*0.00*
0.8	0.798	0.30	0.867	*8.43*	0.794	*0.78*
0.9	0.914	*1.55*	0.980	*8.94*	0.886	*1.54*
1.0	0.984	*1.60*	1.350	*35.03*	0.995	*0.53*

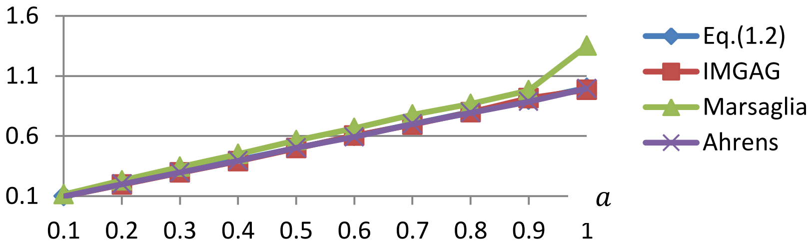

Figure 2.1: Mean values with shape parameters $a \leq 1$

Table 2.2. Variances of 10,000 generated gamma variables using algorithms: IMGAG, Marsaglia and Ahrens

a	IMGAG		Marsaglia		Ahrens	
	D.gen.	% Err.	D.gen.	% Err.	D.gen.	% Err.
0.1	0.098	*1.79*	0.094	*6.44*	0.102	*2.13*
0.2	0.192	*4.18*	0.183	*8.54*	0.196	*2.25*
0.3	0.273	*8.03*	0.270	*10.08*	0.290	*3.34*
0.4	0.373	*6.78*	0.346	*14.89*	0.396	*1.01*
0.5	0.483	*3.42*	0.416	*16.71*	0.502	*0.36*
0.6	0.604	*0.70*	0.506	*15.59*	0.578	*3.67*
0.7	0.668	*4.53*	0.562	*19.74*	0.696	*0.52*
0.8	0.795	0.64	0.609	*23.92*	0.763	*4.60*
0.9	0.937	*4.12*	0.684	*23.99*	0.872	*3.09*
1.0	0.961	*3.86*	1.351	*35.06*	0.991	*0.86*

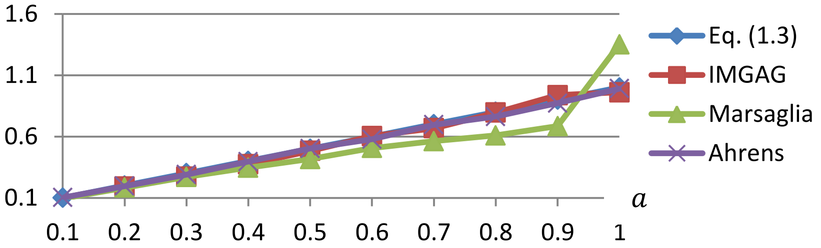

Figure 2.2: Variances with shape parameters $a \leq 1$

Table 2.3. Skewness coefficients of 10,000 generated gamma variables using algorithms: IMGAG, Marsaglia and Ahrens

a	Skew ness	IMGAG		Marsaglia		Ahrens	
		S.gen.	% Err.	S.gen.	% Err.	S.gen.	% Err.
0.1	6.235	6.752	*6.75*	4.524	*28.47*	6.614	*4.57*
0.2	4.472	4.633	*3.36*	2.938	*34.30*	4.363	*2.44*
0.3	3.651	3.530	*3.34*	2.429	*33.47*	3.521	*3.58*
0.4	3.162	3.187	*0.78*	2.235	*29.31*	3.276	*3.59*
0.5	2.828	2.898	*2.45*	1.912	*32.40*	2.840	*0.42*
0.6	2.582	2.480	*3.94*	1.872	*27.51*	2.486	*3.73*
0.7	2.390	2.422	*1.30*	1.653	*30.87*	2.323	*2.82*
0.8	2.236	2.283	*2.10*	1.525	*31.78*	2.074	*7.24*
0.9	2.108	2.048	*2.86*	1.393	*33.93*	2.011	*4.59*
1.0	2.000	2.046	*2.28*	1.698	*15.08*	1.917	*4.13*

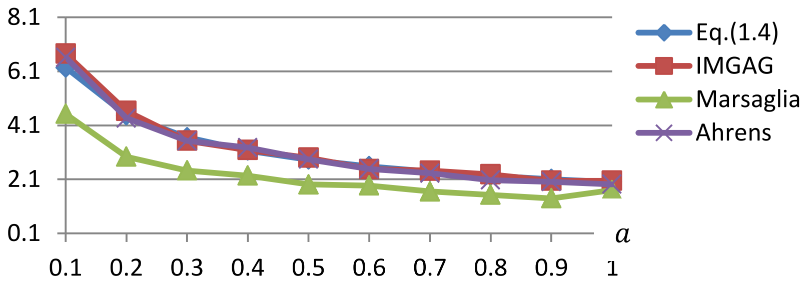

Figure 2.3: Skewness coefficients with shape parameters $a \leq 1$

For shape parameter $a>1$, using algorithms: IMGAG, Marsaglia, Tadikamalla and obtained the tables and figures corresponding.

CONCLUSION OF CHAPTER 2

In chapter 2, the author obtained the following results: study of algorithms used to generate random variables having the unit uniform distribution, normal distribution, exponential distribution, Poisson distribution and the gamma distribution. Proposed IMGAG algorithm to generate the gamma variables with any value of shape parameter $a>0$, and proposed additional criterion to evaluate the effectiveness of the random variable generators by using computer simulation to generate series of random numbers, based on the series of generated data, test the randomness and evaluates the preservation of the numerical characteristics of the distribution based on the mean, variance and the skewness of the series of generated data. The details will be discussed in the conclusions of the dissertation.

CHAPTER 3
COMPUTER SIMULATION OF STREAMFLOWS WITH GAR(1) PROCESS

This chapter presents the research on the models and the algorithms are used to simulate the streamflows. The author uses GAR(1) model, studied Thomas-Fiering model, and, proposed two models: GAR(1)-Monthly and GAR(1)-Fragments used to simulate the monthly streamflows. By means of computer simulation, the models and algorithms were tested and evaluated in terms of the preservation of statistical parameters, including the mean value, standard deviation and the skewness coefficient of historical data.

3.1. Problem of Streamflow Simulation

Based on historical streamflows observed in the gauging stations, the streamflow simulation problem is to evaluate the preservation of the four important descriptors, namely, the mean, standard deviation, skewness coefficient and the correlation coefficient of each streamflow sequence by using the model to generate the sequence of streamflow (monthly or annual) with length of *n* large enough.

3.2. Thomas-Fiering Model (Th.Fiering)

Based on statistical sample of monthly streamflow of *N* years (*N*-called statistical sample size) at a gauge station, The basic model of Thomas-Fiering used to describe the sequence of monthly streamflow is written as:

$$Q_{i,j} = Q_j^* + b_j\left(Q_{i,j-1} - Q_{j-1}^*\right) + s_j(1 - r_j^2)^{\frac{1}{2}}t_{i,j}, \qquad (3.1)$$

where $Q_{i,j}$ is the monthly streamflow in month *j* of year *i*; b_j is the regression coefficient for estimating the flow in month *j* from that in month *j*-1; Q_j^* and s_j are the mean and standard deviation of the historical streamflow in month *j*, respectively; r_j is the correlation coefficient between historical streamflow sequences in months *j* and *j*-1 and $t_{i,j}$ is a random variable with zero mean and unit variance.

3.3. Method of Fragments

Svanidze [12] presented a method in which the monthly flows are standardized year by year so that the sum of the monthly flows in any year equals unity. This is done by dividing the monthly flows in a year by the corresponding annual flow. By doing so, from a record of *N* years, one will have *N* fragments of twelve monthly flows. The annual flows obtained from an annual model can be disaggregated by selecting the fragments at random. Since the monthly parameters were not preserved well, Srikanthan and McMahon[11] suggested a way to improve this preservation by selecting the appropriate fragment for each flow in the annual flow series. This was done as follows: The annual flows from the historical record were ranked according to increasing magnitude, and *N* classes were formed. The first class has the lower limit at zero while class *N* has no upper limit. The intermediate class limits are obtained by averaging two successive annual flows in the ranked series. The corresponding fragments were then assigned to each class. The annual flows were then checked one by one for the class to which they belong and disaggregated using the corresponding fragment.

3.4. Proposed Models Using for Monthly Streamflow Simulation with GAR(1) Process

3.4.1. Gar(1)-Monthly Model (GAR(1)-M)

The GAR(1) model has been found to be very good for the case of annual data: $X_i = \Phi X_{i-1} + e_i$.

According to the results of Hung, Phien and Chien (2014), for the case of historical monthly data of *N* years, each sequence of data of the same month, say j, of *N* years long forms a sequence of data in month j, and the GAR(1) model can be applied to simulate these monthly data. So the GAR(1)-Monthly model is as follows:

$$X_{i,j} = \Phi_j X_{i-1,j} + e_i, j = 1..12 \quad (3.2)$$

where: $X_{i,j}$ is the random variable representing the dependent processes at time *i* of month j, Φ_j *is* autoregressive coefficient of month j and e_i is an independent variable to be specified. Each sequence of dependent gamma variable represents a sequence of data of same month over years. The system of equations in (3.2) constitutes a model for use to simulate monthly streamflows.

As the result of Hung, Phien and Chien (2014), in reality, the correlation coefficient (r_j) between monthly flows into consecutive years may be negative and this may give rise to a negative value of the autoregressive coefficient (Φ_j), therefore a modification of the correlation coefficient of month j is needed to make the GAR(1) model applicable: $r_j = - r_j$ if $r_j < 0$.

- ***Simulation Algorithm:***

(1) Initialize and update the array of historical montly $A[N][12]$, N (the number of years of historical data), n (the number of years of generated data);

(2) Initialize the array of generated monthly data $Q[n][12]$;

(3) Using formulas (1.6) - (1.13) and biased adjusting of the estimators to compute 12 sets of parameters *a, b, c* and Φ of GAR(1)-Monthly model (each the set of parameters *a*, *b*, *c* and Φ corresponding to one series of historical monthly data over the years);

(4) For j = 1 to 12: if $r_j < 0$ compute $r_j = - r_j$;

for i = 1 to n: compute $Q_{i,j} = X_{i,j}$ (using GAR(1) model to generate e_i and compute $X_{i,j}$);

(5) End.

3.4.2. GAR(1)-Fragments Model (GAR(1)-F)

Hung, Phien and Chien (2014) research and applied GAR(1) model for monthly flows, the model is obtained by a combination of the GAR(1) model with the fragments method. From the historical record of monthly data (of *N* years long), the historical record of annual flow with *N* years, the classes and the fragments are formed. The annual flow obtained from the GAR(1) model will be disaggregated to obtain the monthly flow by using the corresponding fragments. Based on historical record of monthly flow, the GAR(1)-fragments model generates monthly flows in the following algorithm:

- ***Simulation Algorithm:***

(1) Initialize and update the array of historical montly $A[N][12]$, N (the number of years of historical data), n (the number of years of generated data);

(2) Initialize the array of generated monthly data $Q[n][12]$;

(3) Seperate the historical series becomes N classes, each class is one year of history;

(4) Sort N classes according to increasing magnitude of historical annual streamflow A_i

($A_i=\sum_{j=1}^{12} A_{i,j}$, $A_{i,j}$: $A_{i,j}$ is the monthly streamflow in month j of year i, after sorting A_1 corresponding to smallest annual flow, A_N corresponding to largest annual flow;

(5) Compute the upper bound U_i of two successive classes:

$U_i = \frac{A_i + A_{i+1}}{2}$, i = 1,2,..N-1. U_N has arbitrary large value;

(6) Compute parameters: shape, scale, location and autoregressive coefficient of GAR(1) model based on the historical annual streamflow;

(7) Generate a random number X_1 has three-parameter gamma distribution (the parameters were computed as in Step 6);

(8) Select the class has the smallest upper bound is greater than or equal to X_1 (so called ith class);

(9) Compute $Q_{1,j} = M_{i,j} * X_1$,$Q_{1,j}$ is the monthly streamflow in month j of year 1; $M_{i,j} = A_{i,j} /A_i$, $M_{i,j}$ is the fragment of historical monthly streamflow in month j of year i;

(10) Compute $Q_{k,j}$: k=2,...,n (n: number of years to genarate), use GAR(1) model to generate e_k and compute X_k, $k=2,..,n$. Select the class having the smallest upper bound greater than or equal to X_k (so called i^{th} class), then $Q_{k,,j} = M_{i,j} * X_k$;

(11) End.

3.5. Computer Simulation

3.5.1. The Data Used and Simulation Method

Based on the results of chapter 1, using the suitable algorithms for generating the ramdom variables follow Thomas-Fiering model, GAR(1)-Monthly model and GAR(1)-Fragments model. The historical flows (m^3/s) at Thanh My gauge station on Vu Gia river, Nong Son gauge station on Thu Bon river in Quang Nam province from 1980 to 2010 and Yen Bai gauge station on Thao river in Yen

Bai province from 1958 to 2011 were used. The algorithms were coded in C language. For each model and at each station, a moderate sample of 1000 years of data was generated on computer using the referred algorithms.

3.5.2. Emperimental Results

Experimental results are given in tables 3.1 - 3.4 and showed as figures 3.1 - 3.3:

Table 3.1. Mean values at Nong Son station

Month	History	GAR(1)-M	GAR(1)-F	Th.Fiering
1	248.96	245.40	220.25	267.63
2	138.21	137.85	136.53	147.64
3	94.05	93.01	94.06	101.39
4	76.45	76.84	66.42	87.16
5	107.30	106.38	97.66	121.01
6	94.54	94.15	93.68	101.73
7	70.33	71.44	74.95	74.84
8	85.02	85.60	91.32	93.60
9	195.59	195.30	174.61	94.19
10	697.19	705.26	778.81	754.37
11	1041.81	1039.30	1074.54	1116.12
12	619.97	622.19	559.19	659.08

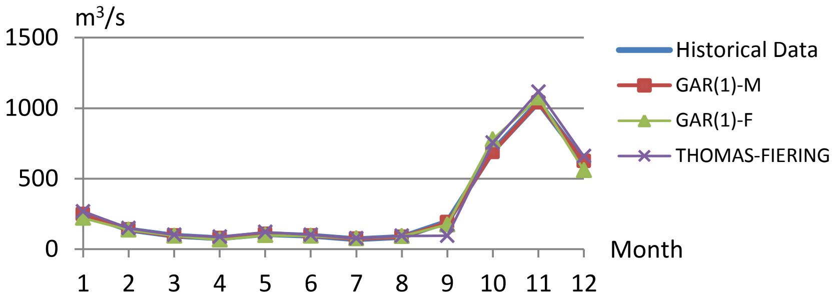

Figure 3.1: Mean values at Nong Son station

Table 3.2. Standard deviations at Nong Son station

Month	History	GAR(1)-M	GAR(1)-F	Th.Fiering
1	110.97	104.54	87.42	79.22
2	46.07	45.50	37.07	34.23

3	33.30	32.67	30.37	24.61
4	39.32	40.82	34.25	29.29
5	60.89	63.72	53.22	45.05
6	39.63	38.2	32.01	29.01
7	25.65	26.07	29.32	19.35
8	48.82	49.52	71.14	36.02
9	174.70	178.68	88.39	18.56
10	354.16	376.42	438.79	244.56
11	549.65	544.42	534.59	401.98
12	329.72	334.52	311.34	235.41

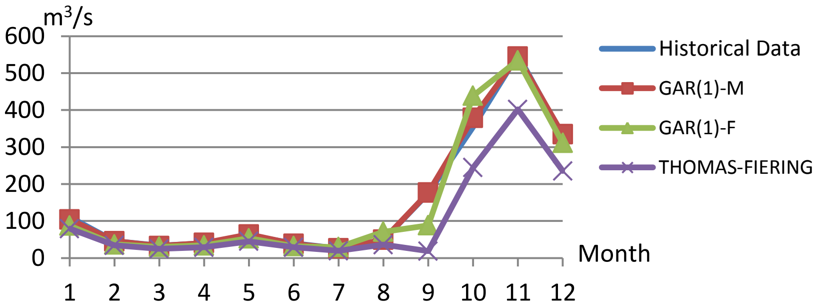

Figure 3.2: Standard deviations at Nong Son station

Table 3.3. Skewness coefficients at Nong Son station

Month	History	GAR(1)-M	GAR(1)-F	Th.Fiering
1	1.54	1.53	1.51	0.67
2	1.09	1.23	0.95	0.57
3	0.87	1.20	0.73	0.43
4	1.70	1.98	2.18	0.48
5	0.79	1.00	0.78	0.35
6	0.77	0.80	0.93	0.34
7	0.47	0.64	1.32	0.22
8	1.55	1.76	3.44	0.62
9	3.08	5.17	2.32	1.73
10	0.23	-0.01	-0.12	0.22
11	0.68	0.66	1.66	0.42
12	0.84	1.12	0.96	0.55

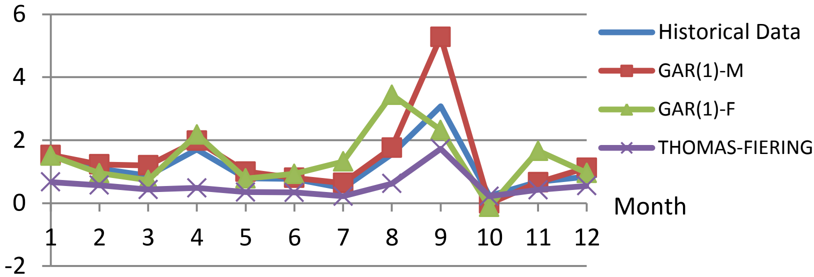

Figure 3.3: Skewness coefficients at Nong Son station

Table 3.4. Statistical parameters of annual data at Nong Son station

Parameters	History	GAR(1)-M	GAR(1)-F	Th.Fiering
Mean	3469.72	3454.17	3467.92	3588.66
Stand. Deviation	1030.77	729.03	1025.29	664.64
Skewness	0.76	0.32	0.78	0.08

Similarly at the Thanh My and Yen Bai gauge stations, the author obtained the tables and figures corresponding also.

CONCLUSION OF CHAPTER 3

In chapter 3, the author obtained the following results: proposed the GAR(1)-Monthly and GAR(1)-Fragments models using for computer simulation of monthly streamflows. By computer simulation, the statistical descriptors such as the mean, standard deviation and the skewness coefficient obtained from generated monthly data by the GAR(1)-Monthly model are closer to their historical values than those obtained by the GAR(1)-Fragments and Thomas-Fiering models.

CHAPTER 4
THE MEAN RANGE OF RESERVOIR STORAGE WITH GAR(1) PROCESS

The contents of this chapter presents the study of reservoir storage problem. By theoretical analysis, the author obtained the closed forms of the expectation and the variance of the sum of GAR(1) variables. Combining the approximate formula of Phien

(1978) and the obtained closed form of the variance of the sum of GAR(1) variables, and from that, the author proposed the approximate expression for the mean range of reservoir storage with GAR(1) inflows. By computer simulation of GAR(1) model to generate the annual inflows, and the mean range of reservoir storage were obtained with the different of parameters and were compared with that results obtained from the approximate expression.

4.1. The Storage of Reservoir

4.1.1. General Storage Equation of Reservoir

Let $\{z_i\}$ be a sequence of random variables with $E(z_i) = 0$ then the cumulative or partial sum, S_i, the maximum partial sum or surplus, M_n, the minimum partial sum or deficit, m_n , and the range, R_n , of the cumulative sums are defined respectively as

$$S_i = z_1 + z_2 + \dots + z_i , i = 1,2,\dots,n \tag{4.1}$$

$$M_n = max(0, S_1, S_2, \dots, S_n), \tag{4.2}$$

$$m_n = \min(0, S_1, S_2, \dots, S_n), \tag{4.3}$$

$$R_n = M_n - m_n, \tag{4.4}$$

it is clear that $M_n \geq 0, m_n \leq 0$ and $E(S_i)$= 0.

4.1.2. The Mean Range with Independent Inflows

The range has been investigated with assuming that the inflows (I_k) discharges to the reservoir are distributed as independent variables. In order to avoid the dependent of range on each distribution type, a new variable is introduced. This is done by standardizing I_k:

$$Z_k = \frac{I_k - \mu}{\sigma} = \frac{z_k}{\sigma},$$

where σ is the standard deviation of I_k. It is clear that the standardized variable has zero mean and unit variance.

With the introduction of this new variable Z, then if $E(R_n^0)$ and $E(R_n)$ are the expected values of range corresponding to z and Z, respectively, then

$$E(R_n^0) = \sigma E(R_n).$$

By using the multivariate normal distribution function, Salas-La Cruz(1972) showed that the mean range of reservoir storage is as follows:

$$E(R_n) = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} \frac{[Var(S_i)]^{\frac{1}{2}}}{i}.$$

For the case of independent gamma variable Z_k, the skewness coefficient should be taken into account as in the work of Phien(1978), then the approximate formula of the mean range is therefore expressed as a function of *n* and skewness g:

$$E(R_n) = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} \frac{[Var(S_i)]^{\frac{1}{2}}}{i} . e^{-\frac{0.0475g^{1.65}}{0.7(n-1)^{0.6}+2}}. \qquad (4.5)$$

4.2. The Basic Numerical Characteristics of The Sum of GAR(1) variables

The random variable X_i of GAR(1) model is expressed as follows: $X_i = \Phi X_{i-1} + e_i$.

Then the sum of *n* GAR(1) variables is a random variable, denoted S_n is computed by the equation in the following:

$$S_n = \sum_{i=1}^{n} X_i,$$

where: X_i, *i* = 1, 2, …, *n* are GAR(1) variables.

By theoretical analysis, Hung and Chien (2013) obtained the closed forms of the basic numerical characteristics, namely the expectation and the variance of the sum of GAR(1) variables with one-parameter are as follow :

The expectation of the sum of GAR(1) variables denoted as $E(S_n)$, and $E(S_n) = na$.

The variance of the sum of GAR(1) variables denoted as *Var(S_n)*, and $Var(S_n) = na + 2a\sum_{k=1}^{n-1}(n-k)\Phi^k$. (4.6)

4.3. Approximate Expression for The Mean Range

The range to be investigated here is that of the cumulative sums:

$$S_i = \sum_{k=1}^{i} z_k = \sum_{k=1}^{i} (I_k - \mu),$$

in which z_k is the fluctuation $I_k - \mu$ of I_k around its long-term mean μ, and I_k is a dependent gamma variable and follows the GAR(1) model:

$$I_k = \Phi I_{k-1} + e_k.$$

Following the work of Phien (1978), the skewness coefficient is taken into account and the closed form of the variance of the sum of GAR(1) variables obtained by Hung and Chien (2013). Substituting Eq. (4.6) for the variance of the sum of GAR(1) variables into Eq. (4.5) the following approximate expression for the mean range is obtained for standardized variables:

$$E(R_n) = \sqrt{\frac{2}{\pi}}\sum_{i=1}^{n} \frac{\left[i + 2\sum_{k=1}^{i-1}(i-k)\emptyset^k\right]^{\frac{1}{2}}}{i} . e^{-\frac{0.0475g^{1.65}}{0.7(n-1)^{0.6}+2}}. \quad (4.7)$$

4.4. Computer Simulation

4.4.1. The Data Used and Simulation Method

For each value of the skewness coefficient of the gamma distribution and each value of the autoregressive coefficient of the GAR(1) model, a sample of 100,000 GAR(1) variables were generated. Each generated sequence of 50 values is used to compute the range of the reservoir with a life time of 50 years long. Similarly, for reservoirs with shorter life time of *N* years (*N*<50), each sequence of *N* values is used to compute the corresponding value of range. These computed values of the range are then treated as generated data for the range in the simulation study. In these experiments, the skewness coefficient has values in the range [0.5, 3.0] which is valid for most actual data. Using the approximate expression (Eq. 27) and the generated data for the range, the mean ranges were then computed at different values of n, Φ and g.

4.4.2. Experimental results

The results are given in table 4.1 and as figure 4.1:

Table 4.1. Values of the mean range with autoregressive coefficient $\Phi = 0.6$ and skewness coefficient $g = 2.0$

l year	Mean range of reservoir storage computed by		
	Eq. (4.7)	Generated data	% Err.
5	3.225	3.197	0.875
10	5.663	5.624	0.693
15	7.688	7.635	0.694
20	9.452	9.380	0.767
25	11.034	10.945	0.813
30	12.482	12.392	0.726
35	13.823	13.727	0.699
40	15.079	14.979	0.667
45	16.264	16.152	0.693
50	17.389	17.265	0.718

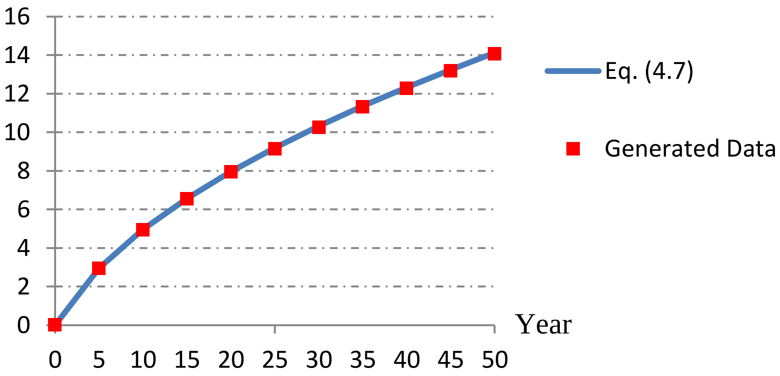

Figure 4.1: Autore. coeff. $\Phi = 0.6$ and skew. coeff. $g = 2.0$

Similarly, the author obtained the tables and figures with different autoregressive coefficients Φ in the range [0.5, 3.0] and skewness coefficient g in the range [0.2, 0.8] corresponding also.

CONCLUSION OF CHAPTER 4

In chapter 4, the obtained results are as follow: By theoretical analysis, a closed form formula for the variance of the sum of GAR(1) variables was derived. This formula is then used along with the empirical formula of Phien (1978) to obtain an approximate expression for the mean value of the reservoir storage. The results computed from the approximate expression can be compared very well with those obtained from generated data.

DISSERTATION CONCLUSIONS

1. The obtained results

From the study of the chapters: The general problems, algorithms to generate GAR(1) variables, computer simulation of streamflows with GAR(1) process and the mean range of reservoir storage with GAR(1)) process presented in the dissertation, the results are:

1.1. Theoretically

- Proposed a algorithm which is the improvement of the algorithm of Minh (1988), namely IMGAG algorithm for generating independent random variables having gamma distribution with all values of shape parameter a>0. Proposed additional criterion to evaluate the effectiveness of a random variable generator by using computer simulation to generate the series of random numbers, and, tests the randomness and considers the preservation of the numerical characteristics of the distribution based on the mean, variance and the skewness of the series of generated data.

- Proposed models: GAR(1)-Monthly and GAR(1)-Fragments using for monthly streamflow simulation.

- Theoretical analysis and derived the closed forms of the expectation and the variance of the sum of GAR(1) variables. Based on the closed form of variance of the sum of GAR(1) variables, combining with the closed form proposed by Salas-La Cruz (1972) and the empirical formula suggested by Phien (1978), the author obtained an approximate expression for the mean range of reservoir storage with GAR(1) inflows.

1.2. Computer Simulation

- For the case of shape parameter $a<1$: The simulation study showed that the numerical characteristics, namely the expectation, variance and the skewness coefficient of the gamma distribution were preserved very well by the IMGAG algorithm and the AHRENS algorithm, much better than those by the MARSAGLIA algorithm. For the case of shape a: $1<a<5$, the TADIKAMALLA algorithm and the IMGAG algorithm preserved the numerical characteristics of the gamma distribution much better than those by the MARSAGLIA algorithm.

- The mean value and the standard deviation of the historical sequences were preserved very well by three models under consideration (GAR(1)-Monthly, GAR(1)-Fragments, Thomas-Fiering), whereas the GAR(1)-Fragments and Thomas-Fiering models do not preserve the skewness coefficient well.

- The mean, standard deviation and the skewness coefficient obtained from generated monthly data by the GAR(1)-Monthly model are closer to their historical values than those obtained by the GAR(1)-Fragments and Thomas-Fiering models.

- In this study, annual data were obtained from generated monthly data by taking the sum of twelve monthly values in a year. Then the mean, standard deviation and the skewness coefficient can be calculated. It was found that these statistical descriptors can be reproduced very well by the GAR(1)-Fragments model, much better than those by the GAR(1)-Monthly and Thomas-Fiering models.

- The mean range computed from the approximate expression is very close to that obtained from the computer. Therefore, this approximate expression (eq. 4.7) may be used in estimating the mean storage of a reservoir with any value of shape parameter of GAR(1) model in practice.

From the obtained results as in above, indicate that the IMGAG algorithm, approximate expression, GAR(1)-Monthly model and the GAR(1)-Fragments model proposed by the author have been tested by computer simulation with the actual data, and proven the effectiveness, and, showing that the IMGAG algorithm for generating of gamma variates, the approximate expression for the mean range of reservoir storage, the GAR(1)-Monthly and GAR(1)-Fragments models using for computer simulation of monthly streamflows with GAR(1) variables can be applied very good in the field of hydrology.

2. Recommendations for Further Study

Besides the obtained results, the following recommendations are suggested for further sudy:

- To generate the GAR(1) variables, the generators are used: The unit uniform, exponential, normal, Poisson and the gamma. This study evaluates the efficiency of the gamma generators only, therefore the evaluation of the effectiveness of generators having the normal distribution and the Poisson distribution will be made for better applying in practice.

- With each model for monthly streamflows and at each gauge station, taking the inspection why the statistical descriptors of several historical monthly data have not been preserved well. Evaluates the preservation of the correlation coefficient of the proposed models.

- Research on the closed form for the mean range of reservoir storage with GAR(1) inflows.

Above are the issues should be further studied and resolved in future ./.

LIST OF THE PUBLISHED PAPERS OF THE AUTHOR

1. Nguyen Van Hung and Tran Quoc Chien, (2013), "*Computer simulation and approximate expression for the mean range of reservoir storage with GAR (1) inflows*" In Proceedings of the Fourth Symposium on Information and Communication Technology, SOICT 2013, DaNang, VietNam, Dec. 5 - 6, 2013 (ACM Digital Library, New York, NY, USA), pp. 11 - 17.
2. Nguyen Van Hung, Huynh Ngoc Phien, Tran Quoc Chien, (2014), *"Computer Simulation of Streamflows with GAR(1)-Monthly and GAR(1)-Fragments Models*", The World of Computer Science and Information Technology Journal (WCSIT, USA), Volume 4, No.11, pp. 150 - 156.
3. Nguyen Van Hung, Ngo Thi Thanh Trang, Tran Quoc Chien, (2014), *"An Improvement of Minh's Algorithm for Generating Gamma Variates with Any Value of Shape Parameter",* Indian Journal of Computer Science and Engineering (IJCSE), Volume 5, No 6, pp. 199 - 205.
4. Nguyễn Văn Hưng, Trần Quốc Chiến, Võ Đình Nam, (2012) *"Nghiên cứu đánh giá các thuật toán sinh biến ngẫu nhiên có phân phối gamma"*, Tạp chí Khoa học và Công nghệ - Đại học Đà Nẵng, Số 10(59), trang 58 - 63.
5. Nguyen Van Hung, Tran Quoc Chien, (2013), *"Computer Simulation of Monthly Streamflows with Thomas-Fiering model and Gar(1)-Fragments model*", Tạp chí Khoa học và Công nghệ - Đại học Đà Nẵng, Số.12(73), (Số tiếng Anh). trang. 46 - 51.
6. Nguyễn Văn Hưng, Ngô Thị Thanh Trang, (2014), *"Mô phỏng lưu lượng dòng chảy hàng tháng với mô hình FGAR(1) và mô hình MGAR(1)",* Tạp chí Khoa học và Công nghệ - Đại học Đà Nẵng, Số 1(74), Quyển 2 (Số đặc biệt dành cho hội nghị RAIT), trang 25 - 29.
7. Nguyễn Văn Hưng, Phan Văn Sơn, Trần Quốc Chiến, (2014), *"Nghiên cứu sự bảo toàn các tham số đặc trưng của mô hình FGAR(1)*", Kỷ yếu Hội thảo Khoa học Hệ thống thông tin, Trường Đại học Sư phạm - Đại học Đà Nẵng, trang 110 - 116.