CORRELOGRAM ANALYSIS AND STOCHASTIC MODELLING OF ANNUAL PEAK FLOWS OF CANADIAN RIVERS

BY

WENBIN LI

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Civil and Geological Engineering

The University of Manitoba

Winnipeg, Manitoba, Canada

ISBN 0-315-92173-0

Name WENBIN LI

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Civil Engineering — 0543 — UMI

SUBJECT TERM — SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Nom ____

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

______________________________ SUJET

☐☐☐☐ CODE DE SUJET

U·M·I

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS
Architecture 0729
Beaux-arts 0357
Bibliothéconomie 0399
Cinéma 0900
Communication verbale 0459
Communications 0708
Danse 0378
Histoire de l'art 0377
Journalisme 0391
Musique 0413
Sciences de l'information 0723
Théâtre 0465

ÉDUCATION
Généralités 515
Administration 0514
Art 0273
Collèges communautaires 0275
Commerce 0688
Économie domestique 0278
Éducation permanente 0516
Éducation préscolaire 0518
Éducation sanitaire 0680
Enseignement agricole 0517
Enseignement bilingue et multiculturel 0282
Enseignement industriel 0521
Enseignement primaire. 0524
Enseignement professionnel 0747
Enseignement religieux 0527
Enseignement secondaire 0533
Enseignement spécial 0529
Enseignement supérieur 0745
Évaluation 0288
Finances 0277
Formation des enseignants 0530
Histoire de l'éducation 0520
Langues et littérature 0279
Lecture 0535
Mathématiques 0280
Musique 0522
Orientation et consultation 0519
Philosophie de l'éducation 0998
Physique 0523
Programmes d'études et enseignement 0727
Psychologie 0525
Sciences 0714
Sciences sociales 0534
Sociologie de l'éducation 0340
Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE
Langues
- Généralités 0679
- Anciennes 0289
- Linguistique 0290
- Modernes 0291

Littérature
- Généralités 0401
- Anciennes 0294
- Comparée 0295
- Médiévale 0297
- Moderne 0298
- Africaine 0316
- Américaine 0591
- Anglaise 0593
- Asiatique 0305
- Canadienne (Anglaise) 0352
- Canadienne (Française) 0355
- Germanique 0311
- Latino-américaine 0312
- Moyen-orientale 0315
- Romane 0313
- Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THÉOLOGIE
Philosophie 0422
Religion
- Généralités 0318
- Clergé 0319
- Études bibliques 0321
- Histoire des religions 0320
- Philosophie de la religion 0322

Théologie 0469

SCIENCES SOCIALES
Anthropologie
- Archéologie 0324
- Culturelle 0326
- Physique 0327

Droit 0398
Économie
- Généralités 0501
- Commerce-Affaires 0505
- Économie agricole 0503
- Économie du travail 0510
- Finances 0508
- Histoire 0509
- Théorie 0511

Études américaines 0323
Études canadiennes 0385
Études féministes 0453
Folklore 0358
Géographie 0366
Gérontologie 0351
Gestion des affaires
- Généralités 0310
- Administration 0454
- Banques 0770
- Comptabilité 0272
- Marketing 0338

Histoire
- Histoire générale 0578
- Ancienne 0579
- Médiévale 0581
- Moderne 0582
- Histoire des noirs 0328
- Africaine 0331
- Canadienne 0334
- États-Unis 0337
- Européenne 0335
- Moyen-orientale 0333
- Latino-américaine 0336
- Asie, Australie et Océanie 0332

Histoire des sciences 0585
Loisirs 0814
Planification urbaine et régionale 0999
Science politique
- Généralités 0615
- Administration publique 0617
- Droit et relations internationales 0616

Sociologie
- Généralités 0626
- Aide et bien-àtre social 0630
- Criminologie et établissements pénitentiaires 0627
- Démographie 0938
- Études de l' individu et de la famille 0628
- Études des relations interethniques et des relations raciales 0631
- Structure et développement social 0700
- Théorie et méthodes 0344
- Travail et relations industrielles 0629

Transports 0709
Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES
Agriculture
- Généralités 0473
- Agronomie. 0285
- Alimentation et technologie alimentaire 0359
- Culture 0479
- Élevage et alimentation 0475
- Exploitation des péturages 0777
- Pathologie animale 0476
- Pathologie végétale 0480
- Physiologie végétale 0817
- Sylviculture et faune 0478
- Technologie du bois 0746

Biologie
- Généralités 0306
- Anatomie 0287
- Biologie (Statistiques) 0308
- Biologie moléculaire 0307
- Botanique 0309
- Cellule 0379
- Écologie 0329
- Entomologie 0353
- Génétique 0369
- Limnologie 0793
- Microbiologie 0410
- Neurologie 0317
- Océanographie 0416
- Physiologie 0433
- Radiation 0821
- Science vétérinaire 0778
- Zoologie 0472

Biophysique
- Généralités 0786
- Medicale 0760

SCIENCES DE LA TERRE
Biogéochimie 0425
Géochimie 0996
Géodésie 0370
Géographie physique 0368
Géologie 0372
Géophysique 0373
Hydrologie 0388
Minéralogie 0411
Océanographie physique 0415
Paléobotanique 0345
Paléoécologie 0426
Paléontologie 0418
Paléozoologie 0985
Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT
Économie domestique 0386
Sciences de l'environnement 0768
Sciences de la santé
- Généralités 0566
- Administration des hipitaux 0769
- Alimentation et nutrition 0570
- Audiologie 0300
- Chimiothérapie 0992
- Dentisterie 0567
- Développement humain 0758
- Enseignement 0350
- Immunologie 0982
- Loisirs 0575
- Médecine du travail et thérapie 0354
- Médecine et chirurgie 0564
- Obstétrique et gynécologie 0380
- Ophtalmologie 0381
- Orthophonie 0460
- Pathologie 0571
- Pharmacie 0572
- Pharmacologie 0419
- Physiothérapie 0382
- Radiologie 0574
- Santé mentale 0347
- Santé publique 0573
- Soins infirmiers 0569
- Toxicologie 0383

SCIENCES PHYSIQUES
Sciences Pures
Chimie
- Genéralités 0485
- Biochimie 487
- Chimie agricole 0749
- Chimie analytique 0486
- Chimie minérale 0488
- Chimie nucléaire 0738
- Chimie organique 0490
- Chimie pharmaceutique 0491
- Physique 0494
- PolymÇres 0495
- Radiation 0754

Mathématiques 0405
Physique
- Généralités 0605
- Acoustique 0986
- Astronomie et astrophysique 0606
- Electronique et électricité 0607
- Fluides et plasma 0759
- Météorologie 0608
- Optique 0752
- Particules (Physique nucléaire) 0798
- Physique atomique 0748
- Physique de l'état solide 0611
- Physique moléculaire 0609
- Physique nucléaire 0610
- Radiation 0756

Statistiques 0463

Sciences Appliqués Et Technologie
Informatique 0984
Ingénierie
- Généralités 0537
- Agricole 0539
- Automobile 0540
- Biomédicale 0541
- Chaleur et ther modynamique 0348
- Conditionnement (Emballage) 0549
- Génie aérospatial 0538
- Génie chimique 0542
- Génie civil 0543
- Génie électronique et électrique 0544
- Génie industriel 0546
- Génie mécanique 0548
- Génie nucléaire 0552
- Ingénierie des systömes 0790
- Mécanique navale 0547
- Métallurgie 0743
- Science des matériaux 0794
- Technique du pétrole 0765
- Technique minière 0551
- Techniques sanitaires et municipales 0554
- Technologie hydraulique 0545

Mécanique appliquée 0346
Géotechnologie 0428
Matières plastiques (Technologie) 0795
Recherche opérationnelle 0796
Textiles et tissus (Technologie) 0794

PSYCHOLOGIE
Généralités 0621
Personnalité 0625
Psychobiologie 0349
Psychologie clinique 0622
Psychologie du comportement 0384
Psychologie du développement 0620
Psychologie expérimentale 0623
Psychologie industrielle 0624
Psychologie physiologique 0989
Psychologie sociale 0451
Psychométrie 0632

CORRELOGRAM ANALYSIS AND STOCHASTIC MODELLING OF ANNUAL PEAK FLOWS OF CANADIAN RIVERS

BY

WENBIN LI

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

THESIS COPYRIGHT LICENSE

I, WENBIN LI, in the interest of facilitating research by others at the University of Manitoba (hereinafter referred to as the "University") and elsewhere, and in satisfaction of a condition to being awarded a master 's degree as a graduate student at the University, hereby grant a license on the terms and conditions following to the University to reproduce my thesis, or any substantial part thereof, in any material form whatever, which thesis is to be submitted in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

Terms of Conditions

1. This license shall continue for the full terms of the copyright or for so long as may be legally permitted.

2. The Universal Copyright Notice shall be contained in reproduction and copies made under the authority of this license.

3. It shall be permissable for the University to submit reproduction and copies in any material form to the National Library of Canada or like institutions.

4. This license does not permit the sale of authorized reproduction and copies at a profit, but it does permit the collection by the University or other institutions concerned of charges covering actual reproduction and copying costs.

5. All production and copies made under the authority of this license shall bear a statement to the effect that the reproduction or copy in question has been made available by authority of the copyright owner solely for the purpose of private study and research and may not be reproduced and copied

except as permitted by the copyright laws and without written authorization from the copyright owner.

6. The effective date of this license will be the date that thesis is submitted to the University. (In exceptional cases where adequate cause can be shown to delay reproduction and copying, a request may be made of the Dean of Graduate Studies to restrict reproduction of copying of the thesis for a period of up to three years after it submission and the Dean, in consultation with the student's advisor, shall determine the period of restriction if any.)

7. I save and indemnify the University for any liability, losses, damages and awards resulting from any actions, suites, demands or claims. respecting an alleged or proven copyright infringement by the contents of my thesis or respecting defamatory statement contained within my thesis, and arising from the exercise by the University of its rights hereunder.

Signed this 17 day of Dec. 19 93.

17 · Dec · 1993 ______________________

(date) (Signature of Student)

Student Number: ______

ABSTRACT

A correlogram analysis was performed on annual peak flows of more than two hundred unregulated rivers in Canada. These rivers showed a very low lag-one autocorrelation but a significant Hurst phenomenon. Though Anderson's test indicates that the null hypothesis of time independence can not be rejected, these series are certainly time dependent because of their significant long term persistence. This demonstrates that in stochastic modelling of annual peak flow series, the Hurst phenomenon should not be disregarded.

To improve the estimation of the parameters for such modelling, the bias in estimating the Hurst parameter by standard methods was discussed. A bias correction formula was derived by means of Monte Carlo simulation. It was also found that conventional methods of correcting for bias in the autocorrelation function are quite inadequate when the series shows a significant long term persistence. This required bias correction was found to be a function of the Hurst parameter.

To reduce the sample variance in the estimation of the Hurst parameter, the information obtained from basic data set was used in the Bayesian updating of the estimate for a single river. This technique is demonstrated for two rivers.

Finally, the bias correction and the updating were tested by developing several stochastic models with the mixed noise model. It was shown that on average better simulation results can be obtained. It was also shown that the bias correction eliminates the incompatibility between the observed lag-one autocorrelation coefficient and the observed Hurst statistic, which is sometimes a problem in mixed noise models.

ACKNOWLEDGEMENTS

First, I would like to express my sincere appreciation to Professor Caspar Booy, my supervisor and mentor, for his guidance and encouragement throughout the course study, research and thesis preparation. Without the financial aid provided by him, my study would be difficult and this thesis would be impossible.

I would also like to thank other faculty members in Water Resources Engineering, Dr. S.P.Simonovic and Dr. B.J. Lence for their assistance and help during the last two years.

I am grateful to Dr. L.M. Lye for his valuable advice and for providing the basic annual peak flow data set.

Gratitude is also extended to my friends, Yujuin Yang, Zheng Lu, Xiao jiang, Cheng Gingang, Dong Qi and Ansari Khan for their encouragement, help and friendship.

Lastly, my sincere thanks to my parents and my wife Ning Zhong for their considerable understanding and love through the most difficult times.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 PROBLEMS AND RESEARCH OBJECTIVES

This study deals with flood risk or more precisely, with its mathematical expression in the form of flood frequency distributions or flood frequency curves. The determination of these flood frequency curves is of course a standard engineering practice. But there are several engineering problems associated with frequency curves for which a satisfactory solution has not been found. Two of them are addressed in this thesis.

The first concerns the confidence one can have in the flood frequency distribution with its estimated parameters. This is important if design decisions are to be based on these estimates.

The other is how to make better use of all available statistical information for parameter estimation, rather than basing the estimation solely on a single record of annual floods. One approach is regional flood frequency studies, the other is making use of additional information through Bayesian statistics. Here the latter approach is taken.

When the flood frequency curve is well defined by the available data then the probable error involved in the risk analysis itself is small compared to the stochastic uncertainty and has little or no influence on design decisions. This is the case when there is a record of reasonable length and no serial correlation in the data series, which is the standard engineering assumption.

It has been shown (Booy and Lye; 1989) that long term serial correlation exists in several annual flood series of Canadian rivers and that this may lead to a substantial probable error in the risk assessment which ought to be taken into account in the design. They also showed that many annual flood series show significant long term persistence, though the first order serial correlation is not significantly different from zero.

Mathematical confidence limits can be established for relatively simple distributions if one assumes no serial correlation or a very simple correlation structure. If the distribution is more complicated or the correlation structure is complex, then one needs simulation to establish confidence limits on risk assessments and on the total risk that a flood protection structure will fail within the planning horizon.

Simulation is also necessary for the parameter estimation. Formulas for the estimation of the correlation parameters have a large bias which depends on the serial correlation structure. Adequate bias correction factors have been established for first order correlation assuming a first order Markov model but not for higher order coefficients when a more complicated correlation structures exists.

Finally, simulation is necessary to establish appropriate likelihood functions for a Bayesian approach to the estimation of the parameter that describes the long term correlation structure, for which the Hurst coefficient K was chosen in this study.

It can thus be said that the development of adequate simulation techniques is essential for the solution of the two problems that were identified above; this will therefore be a main topic in this thesis.

In simulation, both the choice of the model and the parameter estimation are crucial. The approach taken in this study will be :

A. to examine the correlation structure of a large number of annual flood series of Canadian rivers;

B. to review available time series models and to select an appropriate model type;

C. to estimate the model parameters that govern short and long term correlation;

D. to apply the simulation to flood series of selected rivers;

E. to compare the observed autocorrelation function (ACF) of the rivers with the ACF obtained by simulation.

In connection with the parameter estimation, it should be noted that there is substantial bias in the formulas for the calculation of the Hurst statistic and the ACF. Correction for this bias is therefore the subject of two separate chapters.

1.2 OUTLINE OF THESIS

The study begins in the second chapter with the analysis of the correlation structure of more than 200 data series obtained from unregulated rivers in Canada. In this chapter, the autocorrelation function (ACF) and the Hurst statistic are calculated and the serial correlation structure is examined.

Chapter three deals with the selection of a suitable stochastic model for simulation. Different stochastic models including the autoregressive model (AR), the autoregressive moving average model (ARMA), the fractional gaussian noise model (FGN), the fast fractional gaussian noise generator (FFGN), the broken line process (BL), the ARMA-Markov Model (AM), and the mixed noise model (MN) are reviewed. It is argued that the fast fractional noise model and the mixed noise model are the most suitable for use in this study.

Chapter four deals with the estimation of the Hurst statistic for relatively short time series. The Hurst statistic K as well as the Hurst coefficient H are briefly discussed and bias correction formulas for K are derived from simulation. Finally, this formula is verified.

Chapter five deals with bias correction formulas for the autocorrelation coefficients based on so called long memory models. Monte Carlo simulation results are used for the derivation of valid bias correction formulas that can be used for Canadian rivers.

In chapter six, Bayes' theorem is used in the development of a methodology of updating the Hurst statistic for a single Canadian river while incorporating in the analysis the information of all rivers. This Bayesian method improves the estimation of the Hurst statistic.

Chapter seven shows applications of the procedure for specific stations. In the procedure, the Bayesian updated Hurst statistic and the bias corrected lag-one autocorrelation coefficient are utilized.

Chapter eight presents conclusions and recommendations for the future studies.

CHAPTER 2.

CORRELATION STRUCTURE OF ANNUAL PEAK FLOW SERIES FOR CANADIAN RIVERS

2.1. INTRODUCTION

In conventional flood frequency analysis, the correlation structure of the data has no effect on risk analysis since it has no effect on the frequency curve derived from the data. It is true that textbooks caution against the use of frequency curves when there is serial correlation in the data, but they seldom say what to do instead. In flood risk analysis, this did not appear to be a major problem. Conventional tests confirmed that for annual peak flows, the first order correlation is negligible. Any occasional departure was therefore (probably correctly) attributed to chance except for rivers with unusual amounts of basin storage. Engineers have thus traditionally paid little attention to correlation structure of the data. Nevertheless, the correlation structure can be quite important for engineering design because correlation increases the variance of the sample statistics. This means that there is a much greater uncertainty in parameter estimates. One can also put it this way that the correlation decreases the information content of the record. Long term serial correlation is especially important for parameter variability. Mandelbrot and Wallis (1969 c) showed that the variance of the sample mean is a function of the Hurst statistic:

$$Var(\overline{X}) = \sigma_x^2 \; n^{2h-2} \qquad \textbf{(2-1)}$$

where σ_x^2 is the true variance of X and h is Hurst coefficient; n is the sample length. With a sample of 50 independent data, the variance of the sample mean is $\sigma_x^2/50$ since h=0.5 for independent data. With a Hurst coefficient of h = 0.7, which is about average for many Canadian rivers, the variance of the sample mean is increased by a factor 4.78. Parameter uncertainty is rather important in design since it increases the total flood risk the designer must provide for to attain adequate safety against flooding. Put simply one can write:

Variance in predicted peak = variance calculated from the accepted model

+ variance due to model uncertainty

In this chapter, the correlation structure of annual peak flow data of 209 rivers across Canada is examined. The method for estimating the autocorrelation function given by Jenkins and Watts (1968) and Box and Jenkins (1970) is used to obtain the autocorrelation function of each peak flow series. The results are listed and are also shown graphically. Anderson's test was used to determine significant departures from zero first order correlation. After this, the long term serial correlation as measured by the Hurst statistic K is used and described. It will be shown that for most Canadian rivers, the peak flow series show the Hurst phenomenon to a very significant degree.

2.2. GENERAL DESCRIPTION OF DATA

A set of annual peak flow data from more than 200 rivers in Canada were obtained from Dr. L.M. Lye from Memorial University at St. John's, Newfoundland. The data were selected on the basis of sample length (at least 30 years) and the absence of river flow regulation. The selection covers almost all provinces of Canada and includes big rivers as well as small creeks. The shortest annual peak flow series is 30 years and the longest one is 100 years. The average length of these series is 43 years. Details of these data are listed Table 2-1. Figure 2-1 shows the distribution of the sample length.

Table 2-1. Canadian Rivers and Their Short and Long Term Persistence

Name of River	Years	H(g10)	K	K'	R(1)	R(1)'
1.Adams River Near Squilax	42	0.5639	0.6604		0.2282	0.2785
2.Ashnola River at Athabasca	42	0.5725	0.5919		-.3399	-.3494
3.Athabasca River at Athabasca	47	0.5800	0.5595		-.2022	-.1964
4.Ausable River near Springbank	43	0.4971	0.5640		-.101	-.0857
5.Badger Creek Near Cartwright	30	0.8892	0.7183		-.1298	-.1113
6.Battle River Near Unwin	36	0.5335	0.6053		.09263	0.1355
7.Bear River East Branch at Bear River	35	0.5118	0.6674		0.3418	0.4182
8.Beaverbank River Near Kinsac	67	0.7675	0.7189		-.0868	-.0765
9.Berens River at Outlet of Long Lake	31	0.6391	0.6436		-.0332	-.0010
10.Birdtail Creek Near Birtle	35	0.6645	0.6315		-.0051	0.0265
11.Boundary Creek Near Porthill	61	0.7989	0.7563		0.1820	0.2123
12.Bowron River Near Wells	33	0.7482	0.6600		0.0251	0.0630
13.Bulkley River at Quick	58	0.6710	0.6385		0.1993	0.2326
14.Cariboo River Below Kangaroo Creek	31	0.7924	0.7150		0.1965	0.2626
15.Carrot River Near Armley	34	0.4573	0.5274		.01835	0.0541
16.Cascade River Near Banff	30	1.1118	0.7550		0.2484	0.3251
17.Castor River at Russell	41	0.7905	0.7229		0.2921	0.3507
18.Chilko River at Outlet of Chilko Lake	60	0.8277	0.7384		-.0347	-.0193
19.Arrow River Near Arrow River	30	0.6927	0.6150		-.1332	-.1152
20.Athabasca River Below Mcmurray	31	0.7267	0.6779		-.0886	-.0647
21.Atlin River Near Atlin	39	0.5389	0.6393		.01063	0.0404
22.Babine River at Babine	41	0.6375	0.7002		.09777	0.1354
23.Barnes Creek Near Needles	38	0.8807	0.7446		.1896	0.2413
24.Beaver River at Cold Lake Reserve	33	0.6087	0.6926		0.2173	0.2818
25.Beaurivage(Riviere) A Saint Etienne	37	0.8466	0.7905		0.2070	0.2624
26.Bell (Riviere) A Senneterre-2	36	0.3507	0.5124		-.1848	-0.1767
27.Big Sheep Creek Near Rossland	40	0.5993	0.6320		.06092	0.0955
28.Black River Near Washago	73	0.7231	0.7284		.07516	0.0940
29.Bow River at Banff	80	0.6825	0.6390		-.1306	-0.1469
30.Broken Head River Near Beausejour	46	0.7931	0.6963		0.1124	+.1243
31.Campbell River,Outlet of Cambell Lake	38	0.6175	0.6955		-.0580	-.0354
32.Carrick Creek Near Carlsrume	35	0.9613	0.7361		0.0572	0.0968
33.Carrot River Near Smoky Burn	34	0.4143	0.5066		-.0540	-.0279
34.Castle River Near Beaver Mines	44	0.8109	0.7140		.04381	0.0732
35.Chilliwalk River at Vedder Crossing	32	0.7975	0.6581		-.0401	-.0101

Name of River	Years	H(g10)	K	K'	R(1)	R(1)'
36.Chilliwalk River at Outlet Chilliwalk Lake	32	0.5270	0.6256		-.0317	-.0005
37.Clam Harbour River Near Birchtown	31	0.3934	0.6276		-.0317	0.0007
38.Clearwater River Above Limestone Creek	30	0.9644	0.7985		.2685	.3483
39.Clearwater River Near Rocky Mountain House	32	0.7240	0.6983		0.1169	.1693
40.Clearwater River at Outlet of Clearlake	38	0.8685	0.8022		.2999	.3646
41.Clearwater River Near Clearwater Station	39	0.5839	0.6330		0.0748	0.1119
42.Columbia River at Nicholson	77	0.9682	0.7498		-.0484	-.0373
43.Columbia River Near Faimont Hot Springs	43	0.8190	0.6828		-.2790	-.2820
44.Columbia River at Donald	44	0.7797	0.7012		-.2858	-.2894
45.Conjuring Creek Near Russell	30	0.4206	0.5600		-.2106	-.2045
46.Cooks Creek Near East Selkirk	32	0.7324	0.6186		-.1070	-.0866
47.Cottonwood River Near Cinema	34	0.6425	0.6313		.06415	0.1060
48.Crow River at Frank	39	0.9721	0.7845		0.1463	0.1916
49.Cypress Creek Near Clearwater	30	0.5044	0.5481		-.2938	-.3005
50.Dease River at Mcdame	30	0.5163	0.5362		-.3205	-.3313
51.Deer Creek at Deer Park	30	0.9536	0.7456		-.1115	-.0902
52.Drywood Creek Near Twin Butte	52	0.7835	0.7268		0.0780	0.1053
53.Duncan River Near Howser	33	1.0938	0.8194		0.1944	0.2557
54.East River at St.Margarets Bay	63	0.7261	0.6757		-.1112	-.1018
55.East Prairie River Near Enilda	30	0.6806	0.7427		0.1644	.2282
56.East Humber River Near Pine Grove	35	1.1215	0.7795		0.0047	0.0376
57.Elbow River at Bragg Creek	54	0.7214	0.6738		0.0293	0.0516
58.Elbow River above Glenmore dam	44	0.5852	0.6609		0.0329	0.0612
59.English River at Umfreville	67	0.7618	0.7258		-.0176	-.0028
60.English River Near Sioux Lookout	60	0.7426	0.7733		0.1364	0.1640
61.Etomami River Near Bertwell	34	0.6875	0.6240		0.0590	0.1002
62.Fish Creek Near Priddis	33	1.0790	0.8035		0.1819	0.2415
63.Fish Creek Near Propect Hill	37	0.6988	0.7178		0.0685	0.1071
64.Flathead River at Flathead	60	0.8359	0.7789		0.1973	0.2293
65.St.Francis River at Outlet Glasier Lake	37	0.8835	0.7677		0.1623	0.2123
66.Fraser River at Hansard	36	0.7025	0.6685		-.1257	-.1102
67.Fraser River at Shelley	39	0.8054	0.6762		-.1259	-.1117
68.Fraser River at Mcbride	36	0.8774	0.6989		-.0702	-.0477
70.Grander River at Big Chute	39	0.7137	0.6331		-.0052	0.0228

Name Of River	Years	H(g10)	K	K'	R(1)	R(1)'
71.Garnish River Near Garnish	30	0.7239	0.6185		0.2657	0.3450
72.Ghost River Near Black Rocky Mountain	40	0.6572	0.6821		0.2540	0.3100
73.Gods River Below Allen Rapids	39	0.6055	0.6931		0.2699	0.3260
74.Grand River at Loch Lomond	68	0.8109	0.7048		-.0489	-.0363
75.Grass River at Wekusko Falls	31	0.6185	0.6272		0.1645	0.2259
76.Green River Near Pemberto	38	0.5879	0.6727		.06643	0.1037
77.Hall(Riviere)Press D'East Hereford	40	1.0108	0.7268		-.0989	-.0821
78.Harricana(Riviere) A Amos	56	0.3985	0.5258		-.1505	-.1428
79.Harrison River Near Harrison Hot Springs	38	0.3879	0.5947		.00617	0.0363
80.Highwood River at Diebel's Ranch	38	0.8852	0.7180		-.0215	0.0054
81.Homathko River at the Mouth	32	0.8555	0.7839		.0702	0.1160
82.Horse Creek at International Boundary	43	0.7181	0.6571		-.0036	0.0216
83.Icelandic River Near Riverton	30	0.5131	0.5293		-.2476	-.2473
84.Incomappleux River Near Beaton	37	0.6914	0.6567		.04873	.0849
85.Indian Brook at Indian Falls	34	0.3807	0.6571		0.2186	.2811
86.Iskut River Below Johnson River	30	0.5348	0.5785		-.0367	-.0039
87.Island Lake River Near Island Lake	32	0.7501	0.7021		-.0520	-.0237
88.Kabinakagami River at Highway No.11	36	0.5071	0.5519		-.1717	-.1619
89.Kettle River Near Ferry	60	0.7633	0.7527		0.1983	.2303
90.Kettle River Near Laurier	59	0.7313	0.7220		0.1098	.136
91.Kinojevis En Aval Du Lac Preissac	33	0.5395	0.5843		-.0936	-.0720
92.Kluane River at Outlet Kluane Lake	36	0.7005	0.6800		-.1872	-.1794
93.Kootenay River at Kootenay Crossing	41	0.4748	0.5534		-.0995	-.0832
94.Kootenay River at Newgate	42	0.9720	0.7584		.04978	.0813
95.Kootenay River Near Skookumchuck	39	0.5147	0.5817		-.2436	-.2429
96.Lahave River at West Northfield	73	0.7942	0.7140		-.0210	-.0077
97.Lardeau River at Marblehead	43	0.6527	0.6466		-.2206	-.2146
98.Mcleod River Above Embarras River	34	0.3267	0.5153		-.0828	-.0605
99.Lepreau River at Lepreau	72	0.5265	0.5692		.00535	0.0204
100.Liard River at Lower Crossing	42	0.6184	0.6690		.1743	0.2190
101.Lillooet River Near Pemberton	63	0.4526	0.6158		.1032	0.1271
102.Lindeman Creek Near Bennett	34	0.7379	0.6518		-.1570	-.1446
103.Little Saskatchewan River,Minnedosa	30	0.7543	0.6715		0.1449	0.2057
104.Little Southwest Miramachi River at Lyttleton	37	0.5654	0.6749		0.1385	0.1856
105.Lobstick River Near Styal	32	0.8577	0.7344		0.0576	0.1015

Name of River	Years	H(g10)	K	K'	R(1)	R(1)'
106.Lodge Creek at International Boundary	35	0.5349	.6562		0.2210	0.2818
107.Lodge Creek,Alberta Boundary	38	0.8884	.7225		.05767	0.0939
108.Manyberries Creek,Brodin's Farm	45	0.6785	.6794		.0044	0.0292
109.Northeast Margaree River,Margaree Vally	72	0.7531	.7516		0.1291	0.1514
110.Southwest Margaree River ,Upper Margaree	70	0.8125	.7495		0.1360	0.1594
111.St.Mary River at Wycliffe	43	0.6276	.6772		-.0206	0.0030
112.Mceachern Creek at International Boundary	53	0.5952	.6448		-.0789	-.0649
113.Mckinnon Creek Near Mccreary	30	0.7579	.6983		0.1738	0.2390
114.Middle Brook Near Gambo	30	0.6463	.6914		0.0642	0.1125
115.Mille Iles En Aval Du Lac Des Duex Montagnes	35	0.4971	.6210		-.2037	-.1977
116.Mink Creek Near Ethelbert	34	0.5709	.6937		.0663	0.1085
117.Missinaibi River at Mattice	69	0.8351	.7712		0.1887	0.2157
118.Mistaya River Near Saskatchewan Crossing	38	0.4971	.6598		.0133	0.0443
119.Moose River Near Red Pass	34	0.8151	.7127		-.0075	0.0249
120.Moyie River at Eastport	59	1.0197	.8409		0.1353	0.1633
121.Nagayami River at Highway No.11	38	0.4075	.5377		-.0407	-.0161
122.Namakan River at Outlet of lac Lacroix	66	0.7046	.7171		0.2492	0.2814
123.Nass River Above Shumal Creek	32	0.6238	.6018		-.2679	-.2705
124.Petite Nation ,Pres De Cote-saint-pierre	43	0.6133	.6372		.03753	0.0670
125.Neebing River Near Thunderbay	35	0.9802	.8005		0.2478	0.3120
126.Nith River at New Hamburg	38	0.8503	.7199		.00853	0.0389
127.Nith River Near Canning	42	0.7410	.6548		-.0589	-.0388
128.Northeast Pond River at Northeast Pond	35	0.8106	.7017		.08063	0.1233
129.North Magnetawan River Near Burk's Fall	73	0.5391	.5837		-.0597	-.0486
130.Nottawasaga River Near Baxter	40	0.6800	.6472		0.1030	0.1422
131.North Pine River Near Pine River	35	0.7269	.7036		0.2463	0.3103
132.East Oakville Creek Near Omagh	32	0.6930	.6829		-.0805	-.0563
133.Oldman River Near Waldron's Corner	39	0.9004	.7482		-.0775	-.0578
134.Overflowing River at Overflowing River	33	0.7787	.7003		0.1562	0.2122
135.Pembina River at Janie	31	0.7175	.6467		-.1347	-.1176
136.Pembina River Near Entwistle	34	0.3956	.5792		-.1152	-.0972
137.Pembina River Below Paddy Creek	33	0.3942	.5678		-.0476	-.0197
138.Petite Nation a Portage-de-la-nation	46	0.7562	.6849		.09022	0.1226
139.Pigeon River at Outlet of Round Lake	31	0.7551	.6684		-.0659	-.0385
140.Pigeon River at Middle Falls	65	0.6838	.6812		.0170	0.0345

Name of River	Years	H(g10)	K	K'	R(1)	R(1)'
141.Pine Creek Near Pine Creek Station	30	0.7998	.6801		-.0455	-.0140
142.Pipers Hole River at Mother's Brook	36	0.9610	.8146		0.3354	0.4086
143.Poplar River at International Boundary	56	0.5626	.6685		-.1574	-.1503
144.Prairie Creek Near Rocky Moutain House	37	0.5804	.6310		0.1459	0.1939
145.Quesnel River at Likely	64	0.8453	.7243		0.1848	0.2138
146.Quesnel River Near Quesnel	50	0.8127	.7378		0.1724	0.2091
147.Red Deer River Near The Mouth	33	0.6652	.6659		0.1673	0.2249
148.Chilko River Near Red Stone	62	0.5764	.6054		-.1595	-.1533
149.Richelieu(Riviere) A Saint-Jean	36	0.6868	.6351		-.1066	-.0887
150.Richelieu(Riviere) Aux Rapides Fryers	51	0.6900	.6479		0.1974	0.2355
151.Roaring River Near Minitonas	30	0.6838	.7052		0.1043	0.1588
152.Rock Creek below Horse Creek Near Int.Bound.	32	0.8416	.7562		-.0258	0.0063
153.Rocky River Near Colinet	39	0.5450	.5980		-.2251	-.2213
154.Rolph Creek Near Kimball	53	0.8103	.7965		0.2027	0.2397
155.Roseau River Near Dominion City	49	0.4725	.5546		.02324	0.0475
156.Rodeau River Near Earibou	67	0.6820	.6694		0.1826	0.2101
157.Roseway River at Lower Ohio	71	0.8918	.7327		.07736	0.0969
158.Saint John River at Fort Kent	62	0.8569	.7374		0.2014	0.2325
159.Salmon River Near Prince George	36	0.7581	.7027		-.0022	0.0288
160.Salmo River Near Salmon	40	0.4230	.6076		0.1234	0.1649
161.Saugeen River Near Walkerton	74	0.6198	.6364		0.1237	0.1451
162.Saugeen River Near Port Elgin	74	0.5806	.6187		0.0062	0.0208
163.Seal River Below Great Island	31	0.3510	.5743		.0156	0.0549
164.Shekak River at Highway No.11	37	0.5812	.6366		-.0786	-.0579
165.Shell River Near Inglis	38	0.9097	.7856		0.2539	0.3132
166.Shogomoc Stream Near Trans Canada Highway	45	0.5511	.6526		.07212	0.1035
167.Sikanni Chief River Near Fort Nelson	44	0.3531	.4962		-.0570	-.0377
168.Similkameen River at Princeton	44	0.7095	.7003		.07097	0.1031
169.Skeena River at U S K	41	0.7935	.6456		-.2528	-.2531
170.Skootamatta River Near Actinolite	30	1.0142	.7801		.04956	.0956
171.Slocan River Near Crescent Vally	64	0.8164	.7274		0.1289	0.1542
172.South Thompson River at Chase	48	0.6202	.6957		0.1690	0.2071
173.Southwest Margaree River Near upper Margaree	70	0.8089	.7495		0.1346	0.1579
174.South Nation River at Spencerville	41	0.6441	.6416		.06859	0.1030
175.Sprague Creek Near Sprague	43	0.7109	.6883		0.2268	0.2757

Name of River	Years	H(g10)	K	K'	R(1)	R(1)'
176.St.Mary River Near Marysville	41	0.5471	.6109		-.0858	-.0681
177.Stellako River at Glenannan	39	0.9911	.8293		0.2139	0.2669
178.Stikine River at Telegrach Creek	34	0.7809	.7007		-.0530	-.0268
179.St.Mary River at Stillwater	73	0.5462	.6637		-.0165	-.0030
180.Stony Creek Near Neepawa	30	0.8888	.6813		-.0862	-.0610
181.Stuart River Near Fort St.James	56	0.8111	.7505		0.2457	0.2838
182.Sturgeon River Near Barwick	35	0.6754	.6930		0.3589	0.4375
183.Sturgeon River Near Fort Saskatchewan	54	0.5847	.6102		-.1451	-.1367
184.Swiftcurrent Creek at Many Glacier	54	0.6259	.6139		.03356	0.0562
185.Swiftcurrent Creek Below Rock Creek	34	0.8621	.6915		-.0936	-.0727
186.Sydenham River Near Alvinston	40	0.8438	.6897		-.0190	0.0067
187.Sydenham River Near Owen Sound	43	0.7576	.5943		-.0319	-0.0095
188.Teslin River Near Telsin	41	0.6869	.6453		-.1369	-.1247
189.Tetagouche River Near West Bathurst	37	0.7163	.6598		-.0493	-.0250
190.North Thompson River at Mclure	30	0.5275	.6482		0.1830	0.2496
191.North Thompson River Near Barriere	44	0.8778	.7409		.04448	0.0739
192.Thompson River Near Spences Bridge	37	0.7392	.7412		0.2600	0.3218
193.Torrent River at Bristol's Pool	30	0.5500	.6047		-.0465	-.0152
194.Turtle River Near Mine Centre	58	0.6817	.7051		.0783	0.1026
195.Turtle River Near Laurier	40	0.6544	.6383		.0650	0.1000
196.Twenty Miles Creek at Balls Falls	32	0.4056	.5278		-.1438	-.1286
197.Upper Humber River Near Reidville	60	0.5651	.6462		0.1725	0.2027
198.Upsalquith River at Upsalquitch	45	0.6417	.6569		.07145	0.1028
199.Waterhen River Near Waterhen	34	0.8230	.7846		0.5991	0.7123
200.Waterhen River Near Waterton Park	41	0.7510	.6867		.0943	0.1315
201.Whitemouth River Near Whitemouth	42	0.8578	.6741		-.0074	0.0181
202.Whitewater Creek Near International Boundary	53	0.4874	.5840		.0492	0.0736
203.Wilson River Near Dauphin	31	0.8419	.7103		.0188	0.0586
204.Wolf Creek at Highway No.16A	34	0.8340	.7236		-.1358	-.1206
205.Woody River Near Bowsman	35	0.8092	.7316		.2062	0.2651
206.York(Riviere) A Sunny Bank	36	0.7543	.7581		.2382	0.2992
207.Yukon River Above Frank Creek	36	0.6211	.6450		-.2139	-.2094
208.Yukon River at Carmacks	37	0.6174	.6530		.0393	0.0744
209.Red River at Emerson	98	0.8321	.7630		0.1225	0.1384

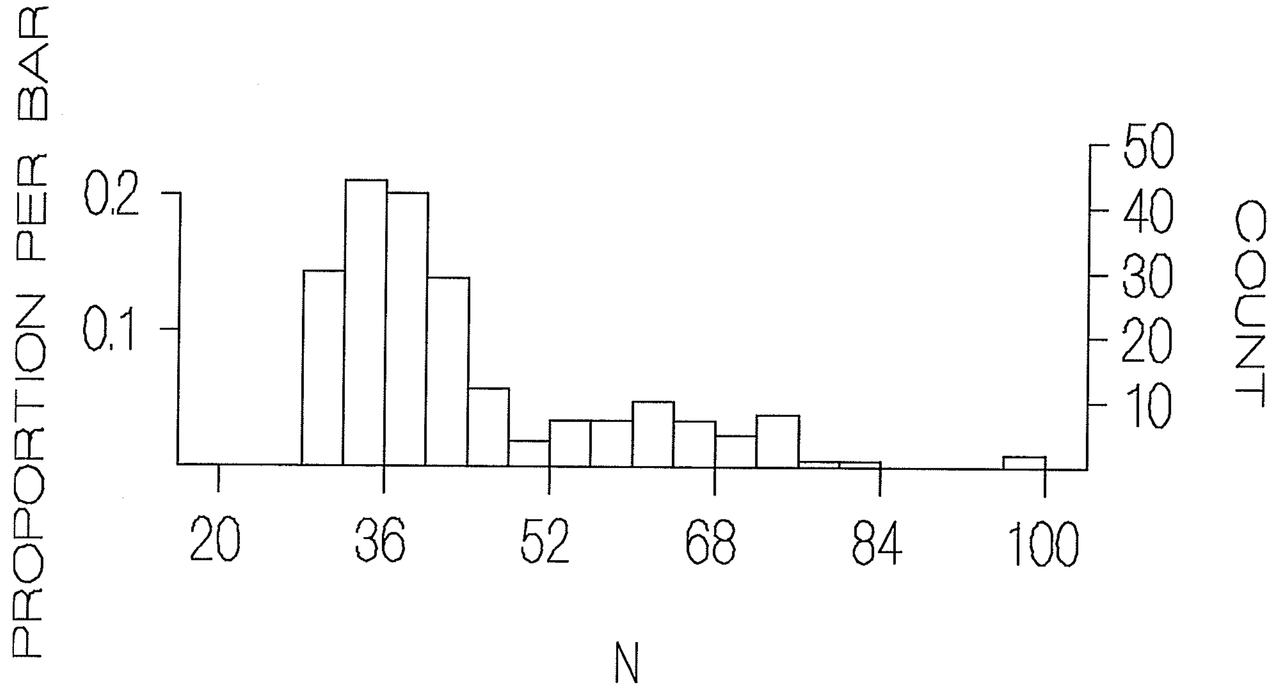


Figure 2-1. Distribution of Sample Size of the data set

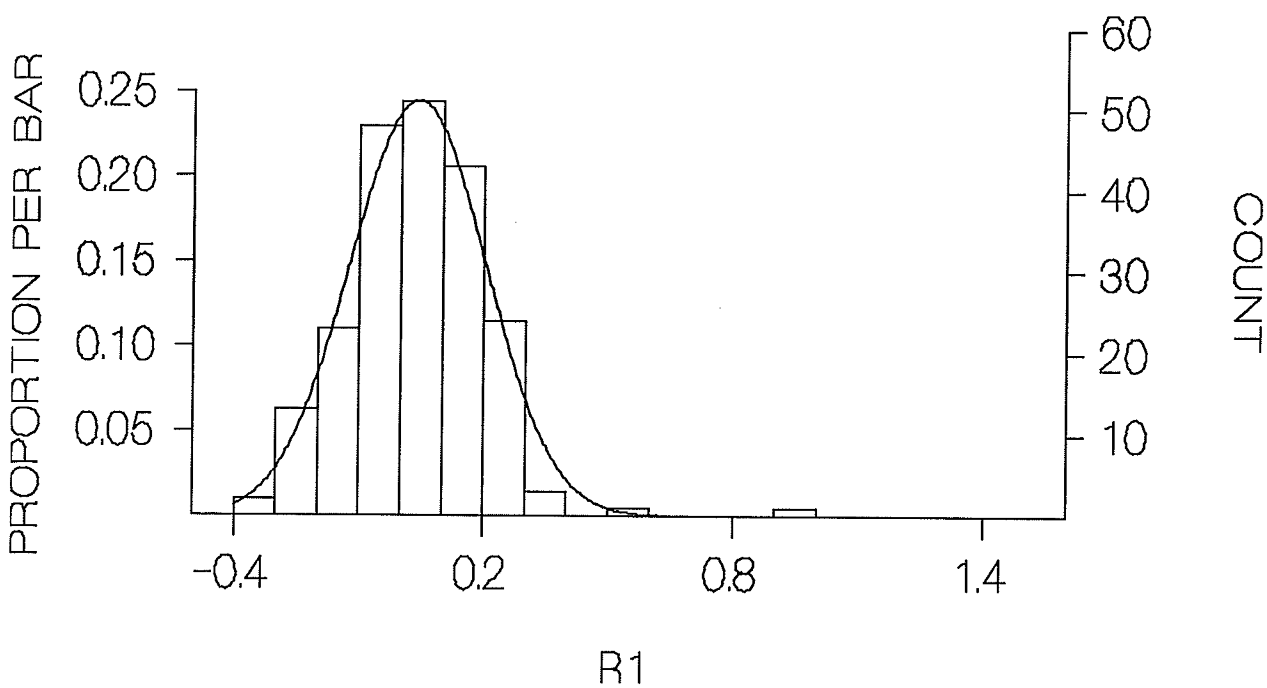


Figure 2-2. Distribution of Lag-one Autocorrelation of the data set (no bias correction)

2.3 ESTIMATION OF THE LAG-ONE AUTOCORRELATION FOR THE DATA SET

The autocorrelation function measures the degree of linear self-dependence of a time series. The autocorrelation function is very important in stochastic hydrology since it shows both short term and long term correlation. The estimation of the correlation coefficient for different values of the lag presents a problem the estimators commonly used show considerable bias.

The most commonly recommended formula for estimating the lag k autocorrelation r_k for a record of length n is given by:

$$r_k = \frac{\sum_{i=1}^{n-k} x_i x_{i+k} - [1/(n-k)] \sum_{i=1}^{n-k} x_i \sum_{i=1}^{n-k} x_{i+k}}{\left[\sum_{i=1}^{n-k} x_i^2 - 1/(n-k) (\sum_{i=1}^{n-k} x_i)^2 \right]^{0.5} \left[\sum_{i=1}^{n-k} x_{i+k}^2 - 1/(n-k) (\sum_{i=1}^{n-k} x_{i+k})^2 \right]^{0.5}} \tag{2-2}$$

Jenkins and Watts (1968) and Box and Jenkins (1970) suggested that a more efficient estimator is given by:

$$r_k = \frac{\sum_{i=1}^{n-k} [x_i - (1/n) \sum_{i=1}^{n} x_i] \quad [x_{i+k} - (1/n) \sum_{i=1}^{n} x_i]}{\sum_{i=1}^{n} [x_i - (1/n) \sum_{i=1}^{n} x_i]^2} \tag{2-3}$$

In this study, equation 2-3 was used to estimate the lag-one autocorrelation coefficient for each peak flow series in the data set. The results r^1 are listed in Table 2-1. Apparently the correlation coefficients range from -0.3399 for the Ashnola River at Athabasca to the 0.5991 for the Waterhen River near Waterhen. The distribution of the lag-one autocorrelation coefficient is presented in Figure 2-2. The mean value is 0.039 and the standard deviation is about 0.16. It can be seen that most rivers have lag-one autocorrelation between -0.3 and 0.4.

2.4. BIAS CORRECTION OF THE CORRELATION ESTIMATE FOR SHORT MEMORY MODELS

It can be shown that both equation 2-2 and 2-3 underestimate the true correlation coefficients because of bias especially when the true correlation is relatively large. A bias correction formula should therefore be used to get better estimates of the parameter.

According to Wallis(1972), Kendall derived a bias correction to order n^{-1} based on the definition of ρ_k given in equation (2-3). For the first order Markov process, this correction was expressed as

$$E(r_k) = \rho^k - \frac{1}{(n-k)}\left\{\frac{1+\rho}{1-\rho}(1-\rho^k)+2k\rho^k\right\} \qquad \textbf{(2-4)}$$

When k=1,

$$E(r_1) = \rho - \left(\frac{1}{n-1}\right)(1+3\rho) \qquad \textbf{(2-5)}$$

Kendall also derived a correction of ρ_k based on circular definition of a time series for the same lag-one Markov process

$$E(r_k) = \rho^k - \frac{1}{n}\left[\frac{1+\rho}{1-\rho}(1-\rho^k)+3k\rho^k - k\rho^{n-k}\right] \qquad \textbf{(2-6)}$$

when k=1

$$E(r_1) = \rho - \left(\frac{1}{n}\right)(1+4\rho) \qquad \textbf{(2-7)}$$

The adequacy of bias correction given by (2-5) and (2-7) was demonstrated by Wallis (1972) for series generated by the AR(1) Markov model. The adequacy of the bias correction was verified in this study from samples generated with AR(1) and AR(2) process. The results are shown in Table (2-2),(2-3),(2-4) and (2-5). It can be seen that the bias correction gives very good results for the synthetic series. It was therefore used for the calculation of the bias corrected lag-one correlation for the Canadian rivers. Corrected values R(1)′ for all rivers are shown in Table (2-1). It was found that the mean lag-one autocorrelation of these rivers is 0.070 and the standard deviation is 0.1792. The distribution of the bias-corrected correlation coefficient for the data set can be seen in Figure (2-3).

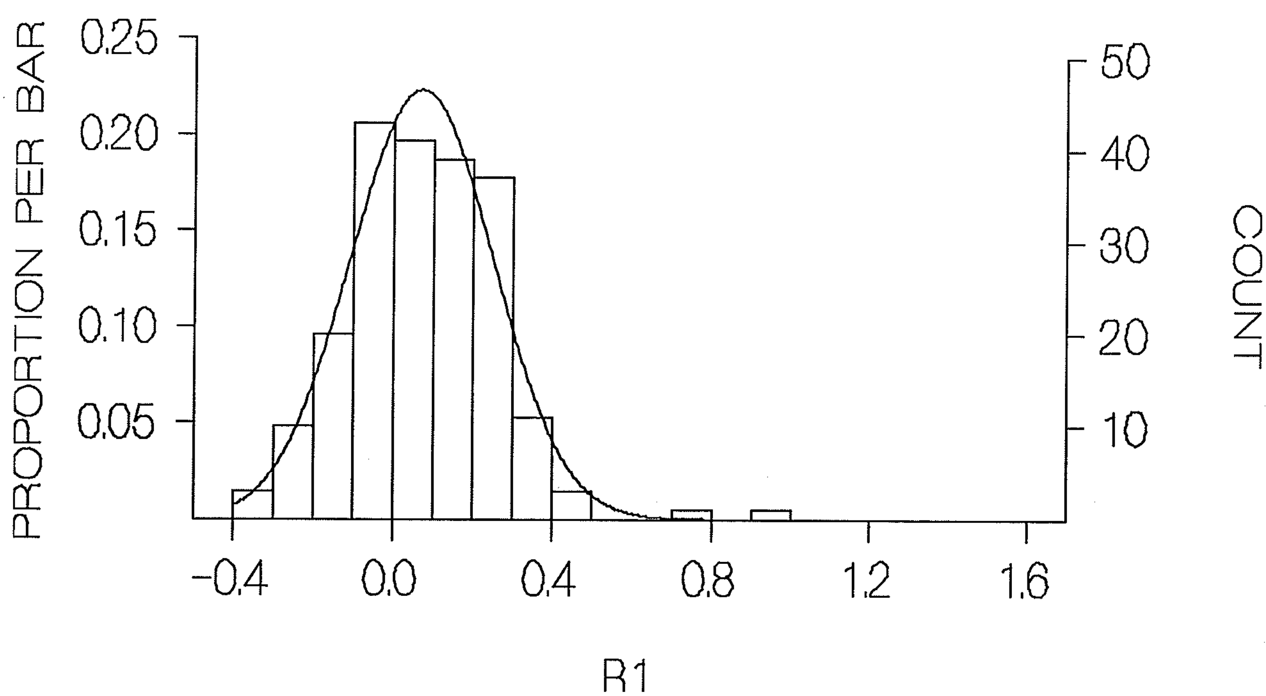


Figure 2-3. Bias Corrected Lag-one Autocorrelation of the Data Set

Table 2-2. Value of $(\rho_1 - r_1)$ obtained before correction data generated by AR(1) model

	n=25	n=50	n=75	n=100	n=125	n=150	n=175	n=200	n=225	n=250
-0.9	-0.100	-0.050	-0.036	-0.029	-0.024	-0.020	-0.016	-0.014	-0.013	-0.012
-0.5	-0.040	-0.027	-0.018	-0.017	-0.018	-0.015	-0.010	-0.008	-0.008	-0.008
-0.1	0.015	0.004	0.002	0.000	-0.006	-0.006	-0.001	-0.002	-0.002	-0.002
0.1	0.043	0.020	0.012	0.009	0.001	0.001	0.001	0.003	0.002	0.002
0.5	0.104	0.054	0.033	0.029	0.018	0.015	0.014	0.011	0.009	0.007
0.9	0.206	0.106	0.066	0.048	0.038	0.033	0.026	0.024	0.019	0.017

each entry is based on 200 sequences

Table 2-3. Value of $(\rho_1 - r_1)$ obtained after correction data generated by AR(1) model

	n=25	n=50	n=75	n=100	n=125	n=150	n=175	n=200	n=225	n=250
-0.9	0.005	0.002	-0.001	-0.003	-0.003	-0.003	-0.001	-0.001	-0.002	-0.002
-0.5	0.000	-0.007	-0.005	-0.007	-0.010	-0.009	-0.004	-0.003	-0.003	-0.004
-0.1	-0.010	-0.008	-0.007	-0.007	-0.011	-0.010	-0.005	-0.005	-0.005	-0.005
0.1	-0.015	-0.008	-0.007	-0.005	-0.010	-0.009	-0.005	-0.005	-0.005	-0.005
0.5	-0.019	-0.007	-0.007	-0.001	-0.006	-0.005	-0.003	-0.004	-0.004	-0.005
0.9	0.027	0.015	0.005	0.002	0.001	0.002	0.000	0.002	-0.001	-0.002

each entry is based on 200 sequences

Table 2-4. Value of $(\rho_1 - r_1)$ obtained before correction
data generated by AR(2) model

	n=25	n=50	n=75	n=100	n=125	n=150	n=175	n=200	n=225	n=250
-0.9	-0.087	-0.049	-0.034	-0.025	-0.018	-0.015	-0.014	-0.012	0.012	0.012
-0.5	-0.029	-0.018	-0.017	-0.013	-0.006	-0.005	-0.005	-0.005	-0.005	-0.005
0.2	0.056	0.032	0.011	0.008	0.009	0.007	0.005	0.004	0.003	0.002
0.5	0.089	0.052	0.027	0.020	0.017	0.013	0.010	0.008	0.007	0.005
0.9	0.190	0.104	0.065	0.049	0.037	0.033	0.025	0.021	0.020	0.017

each entry is based on 200 sequences .

Table 2-5. Value of $(\rho_1 - r_1)$ obtained after correction
data generated by AR(2) model

	n=25	n=50	n=75	n=100	n=125	n=150	n=175	n=200	n=225	n=250
-0.9	0.020	0.003	0.001	0.001	0.003	0.002	0.001	0.001	0.000	0.001
-0.5	0.014	0.002	-0.004	-0.003	0.002	0.001	0.001	0.000	-0.001	-0.001
0.2	-0.019	-0.005	-0.013	-0.010	-0.006	-0.006	-0.005	-0.004	-0.005	-0.006
0.5	-0.036	-0.009	-0.014	-0.011	-0.007	-0.007	-0.008	-0.007	-0.006	-0.007
0.9	0.007	0.013	0.004	0.003	0.000	0.002	-0.001	-0.002	-0.001	-0.001

each entry is based on 200 sequences

2.5. ANDERSON'S TEST

Anderson's Test (J.D. Salas, 1980) is one of the many available tests used to determine the significance of observed correlation coefficients. It is easy to use and preferred by many researchers like Salas, Delleur, Yevjevich and Lane (1980). The confidence limits at the 95% level for the null hypothesis of zero correlation are

$$r_{k(.95)} = \frac{-1 \pm 1.96\sqrt{N-k-1}}{N-k} \tag{2-8}$$

where N is the sample size and k is the lag.

The correlation coefficients of the data set have been calculated and the lag-one autocorrelations coefficient are listed in Table 2-1. Applying Anderson test, we can generally not reject the null hypothesis of no correlation for the single river, so that these rivers would be regarded as time independent series on the basis of this test. Figure 2-5-1 to 2-5-6 present some results of Anderson's test for typical Canadian rivers.

It should be noted, however, that the first order correlations may be small so that they will not be distinguished from zero by Anderson's test. To check this, a t-test was performed in the mean of the first order correlation of the observed set, which is 0.039. The standard deviation of the lag-one correlation is 0.16. At the significance level of α, the hypothesis of no autocorrelation can be rejected if

$$abs(\overline{X}) > t_{\alpha/2}(n-1)\ \frac{s}{\sqrt{n}}$$

At the significance level of 1% and 5%, the hypothesis of zero mean is rejected. It may be concluded that at least some of the data series have significant first order correlation. This point is important because the assumption of a long term serial correlation structure is not compatible with the assumption of zero first order correlation.

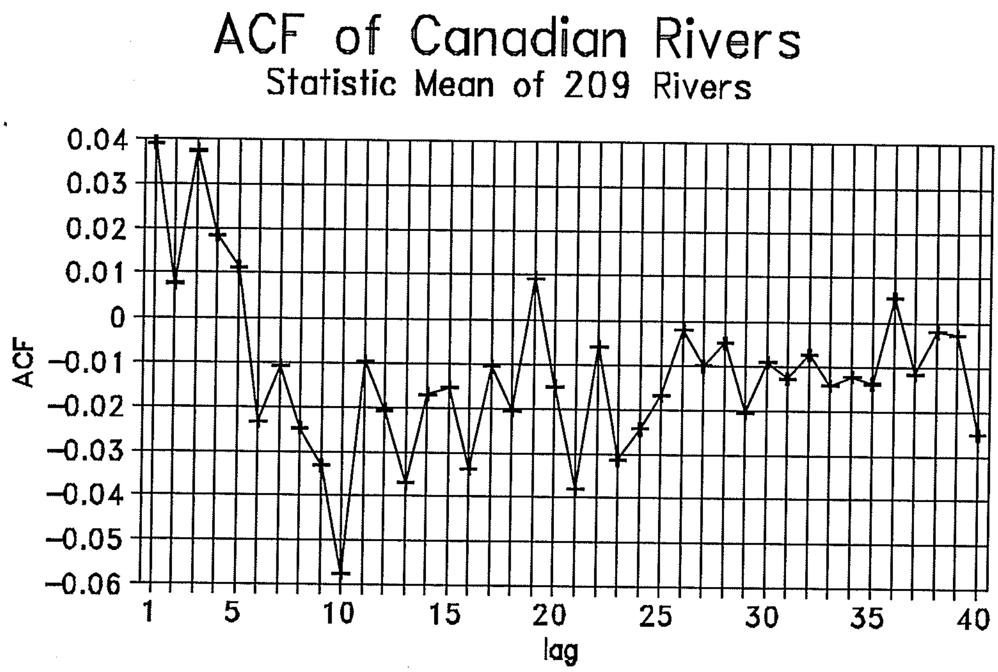


Figure 2-4. Mean ACF of the Basic Data Set

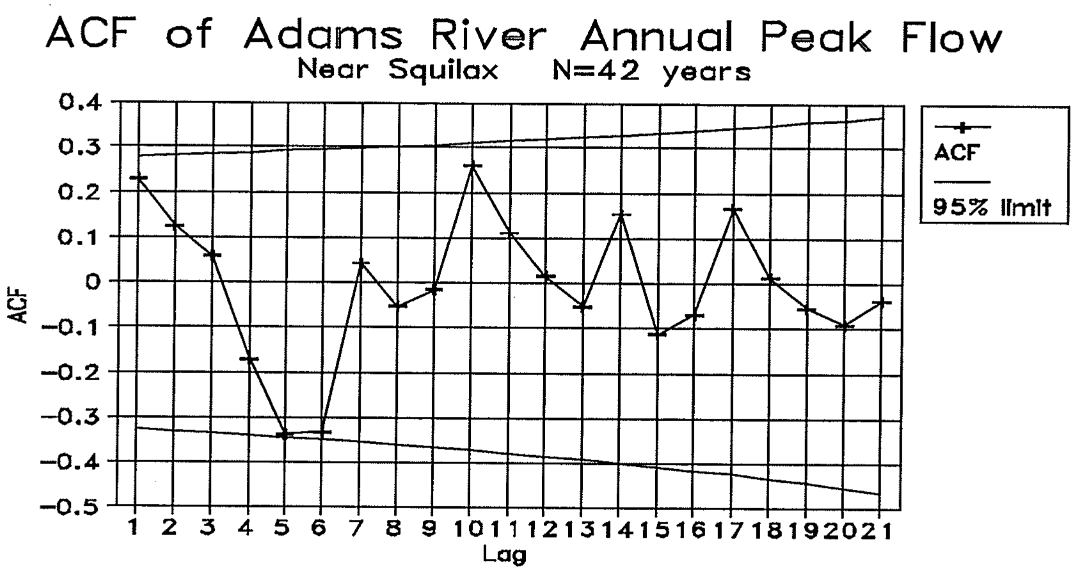


Figure 2-5-1. ACF of Adams River Annual Peak Flow Near Squilax

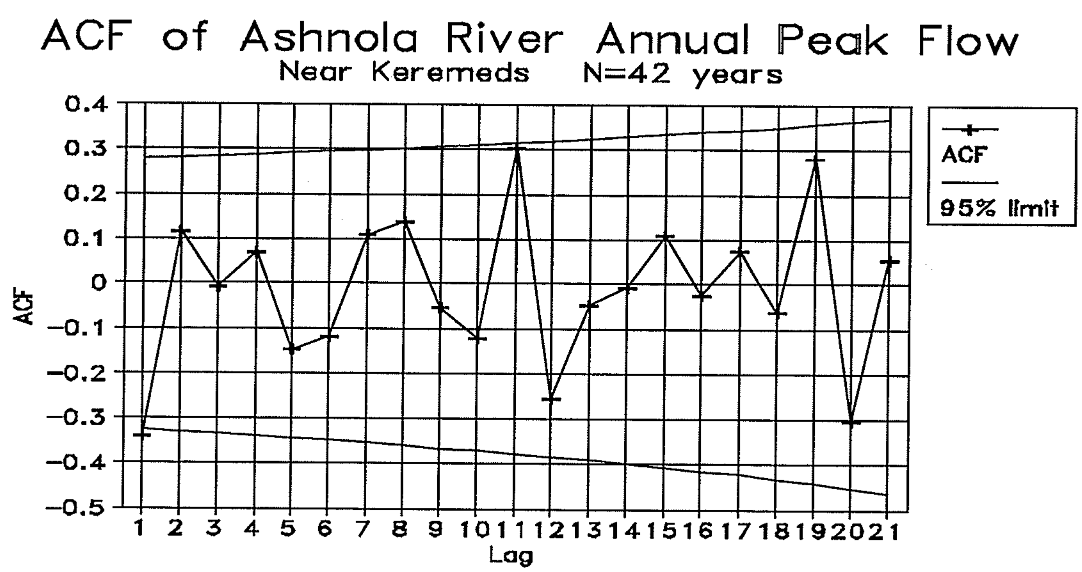


Figure 2-5-2. ACF of Ashnola River Annual Peak Flow at Athabasca

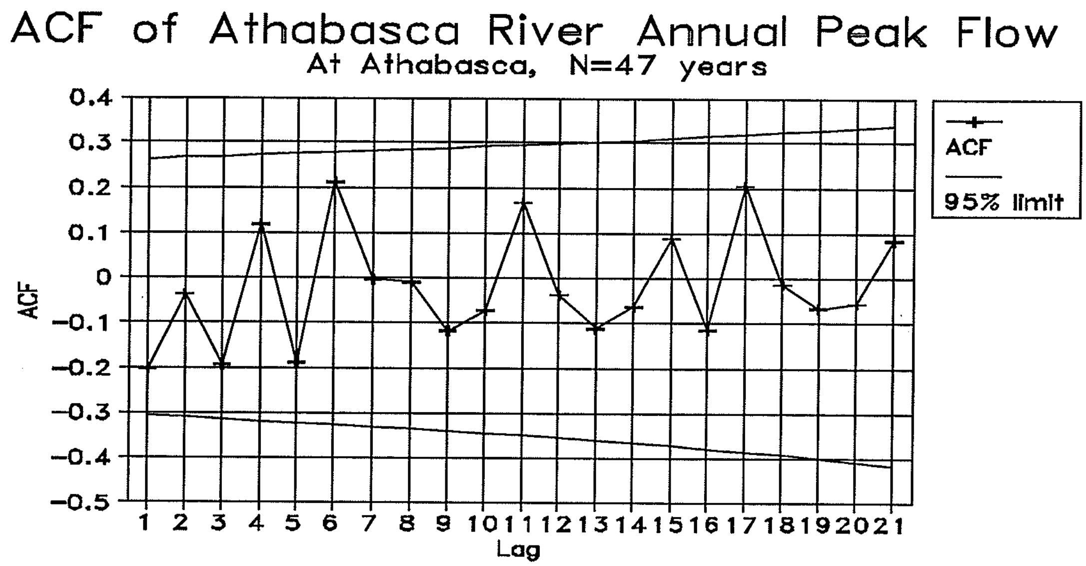


Figure 2-5-3. ACF of Athabasca River Annual Peak Flow at Athabasca

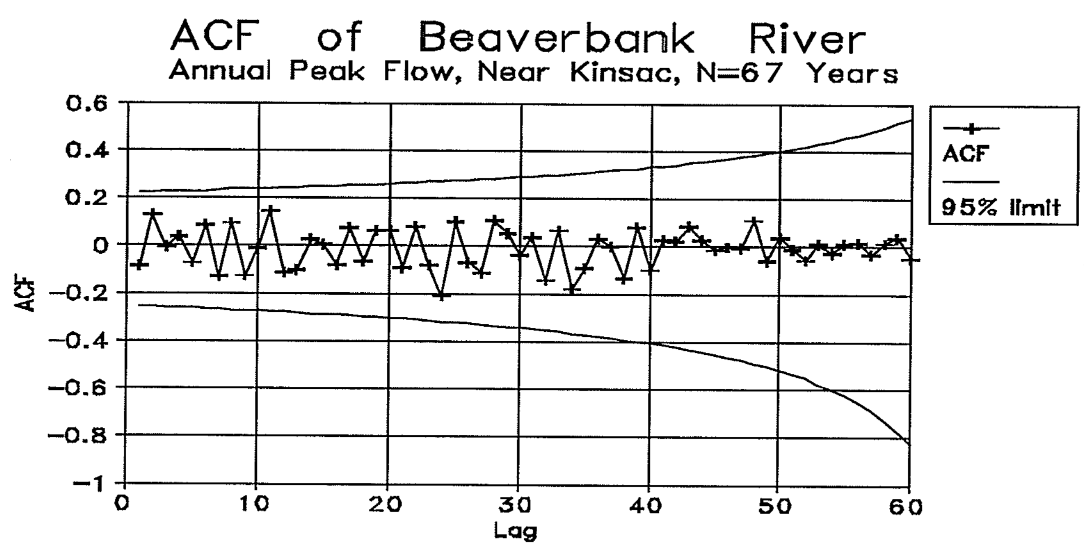


Figure 2-5-4. ACF of Beaverbank River Annual Peak Flow Near Kinsac

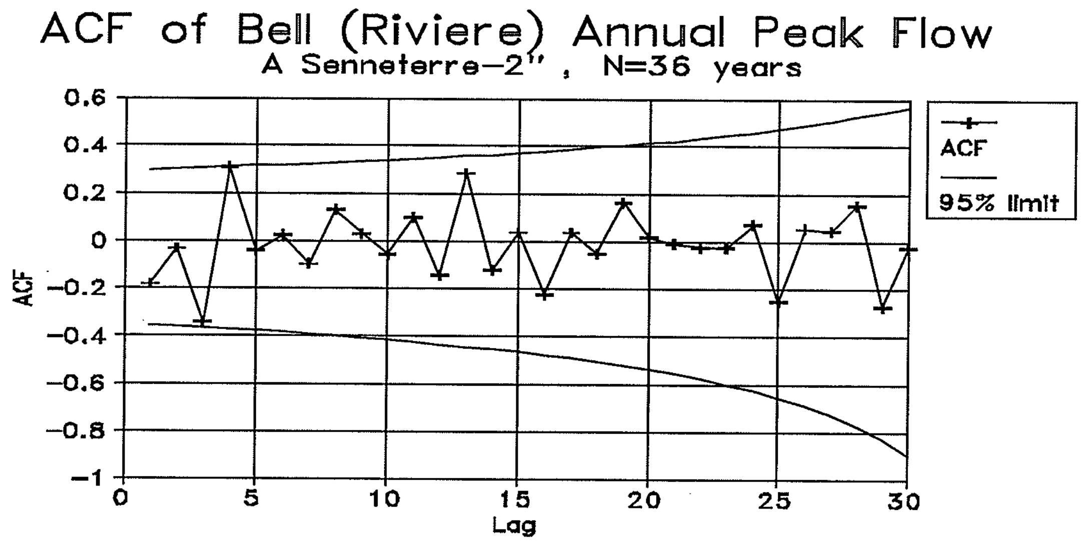


Figure 2-5-5. ACF of Bell Riviere A Senneterre-2" Annual Peak Flow

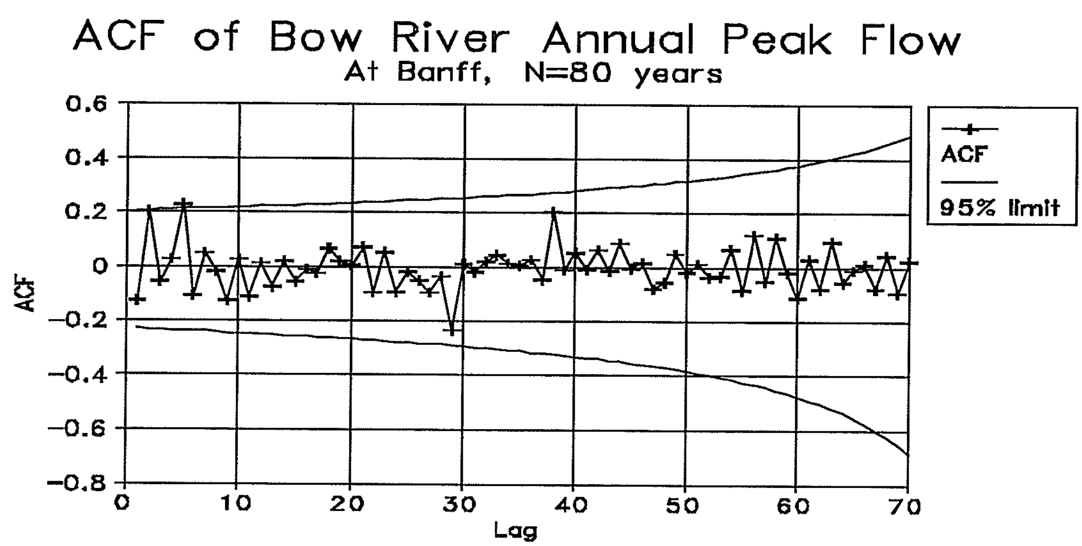


Figure 2-5-6. ACF of Bow River Annual Peak Flow at Banff

2.6. ESTIMATION OF THE HURST PARAMETER USING K AND H

In the study of long term reservoir storage for the river Nile in its large equatorial lakes, H. E. Hurst (1951) defined a parameter h which is a measure of long term serial correlation. He also introduced an estimator of h, the so called Hurst statistic K. Hurst found that for many hydrological series, K is about 0.72, which is significantly larger than the value of 0.5 which one would expect from long data series with no correlation or only short term correlation. Later studies showed that K is a biased estimator so that even for independent data one would obtain a K value of about 0.6. Moreover K was shown to be dependent on the length of the series and the first order serial correlation. Nevertheless, the observed values of K for many natural series indicated a grouping of high and low values of flows that cannot be attributed to chance. This grouping , which is not cyclic but irregular, is called the Hurst phenomenon.

The coefficient K was originally defined by Hurst as follows:

Let X be a time series. The range R of the accumulated departures from the sample mean of X during N years is a variable which increases with the sample length N. Dividing this range by the sample standard deviation s, one obtains the rescaled range R/s. Hurst showed mathematically that the mean rescaled range is proportional to the square root of the length of the series if the data are independent. He suggested equation 2-9 for the estimation of this exponent for actual time series.

$$E\ (\ R/s\ ) = \left(\frac{N}{2}\right)^{K} \qquad \textbf{(2-9)}$$

He found, that for many natural phenomena, K is significantly larger than 0.5. While this was at first attributed to correlation and short records, later studies confirmed that the so-called Hurst phenomenon is indicative of a long term correlation structure that is fundamentally different from short term correlation structure such as the low order autoregressive or moving average series.

The formula Hurst used was criticized since he used only a one parameter fit, which means that the lines relating R and N are straight lines on logarithmic paper and are forced to pass all through the point R/s=1 and N=2.

In the discussion of Hurst's paper, V.T. Chow (1951) proposed a two parameter approximation of the relation between the rescaled range and the series length, which can be written as

$$E\ (\ R/s\ )\ =\ C\ N^{h} \qquad \textbf{(2-10)}$$

where C is a constant. The parameter h became known as the Hurst h. To estimate the parameter h from a sample, one divides the series of length N into subseries of the length n in order to determine the mean rescaled range as a function of sample length n. The estimator H, called Hurst's statistic H, is determined by linear regression after taking the logarithm. Following Chow's study, Wallis and Matalas (1970) explored several ways of subdividing time series. They adopted two procedures for calculating H called F-Hurst and G-Hurst.

To use the F-Hurst procedure, a time series of length N may be divided into (N-n+1) subseries for a given n. Each subseries produces a point for the regression analysis of the relationship. This operation allows (N-n+1) of the N time points to serve as the starting time of a subseries of length n. With the length taking integer from 5 to N, the total number of subseries is (N-3)(N-4)/2. Thus to determine the Hurst statistic by means of this procedure, an enormous amount of computation is needed.

To use the G-Hurst procedure, a much smaller number of series are used. In this procedure, uniformly spaced starting points are chosen with n taking minimum value of 10. The relatively larger sample variance makes this sample fraction quite acceptable. In this study, the G-Hurst procedure was used with less than 15 regression points.

The Hurst coefficient K and the statistic H of the basic data set were estimated and their distributions are presented in Figure 2-6 and 2-7. The mean of H for the set is 0.70 and the standard deviation is 0.1649. The mean of the Hurst K is 0.6756 and the standard deviation is 0.07256. Both K and H would thus seem to indicate that Canadian rivers have significant Hurst phenomenon ($K>0.5$, $H>0.5$). The conclusion that the rivers show a significant Hurst phenomenon may be premature, however, because there is considerable bias in the estimation of the parameters in that even for independent series values of H and K in excess of 0.5 are obtained. It was also found that the Hurst H has about 5 times as much variance as the Hurst K. Since K and H are both biased estimators of Hurst statistic and K has less variance and is easier to calculate, K is preferred to measure the Hurst phenomenon.

To demonstrate that the set of Canadian rivers has a significant Hurst phenomenon, the values of K obtained for them were compared with those obtained for 2090 independent series with the same sample size as the data set. This procedure was followed because the bias of the estimates is dependent on the length of the series so that the mean of K for a set of independent data with varying length is not known in advance. The sample distribution of the calculated Hurst K is graphically shown in Figure 2-8. The mean of K is equal to 0.6365 and the standard deviation is equal to 0.07647. As we can see from Figure 2-9, evidently, the distribution of K from the independent series is different from that basic data set. A t-test was performed, assuming as the null hypothesis that these samples have the same mean of K. This hypothesis was rejected at the 0.1% significant level. It may be concluded that the peak flows in the data set are generally not time independent series even though their lag-one serial correlations are very low.

Since the observed distribution of K for the data set is incompatible with the assumption that the series are independent, it was decided to investigate whether the distribution is compatible with another assumed value of the Hurst parameter. To do this, a mixed noise model was employed to generate 2090 series with h = 0.6 and h = 0.8. The results are shown in Figure 2-11 which suggests that the observations are quite consistent with the assumption that for all rivers, h is equal to 0.6. In fact, the assumption that h varies for different rivers would seem to be contradicted by the fact that the variance for the 2090 series with h=0.6 is somewhat larger than the variance of the observed series. This conclusion, however, cannot be substantiated because of the inherent variability of the sample variance. The sample variance of K is probably much more important than the difference of K between different series. For this reason, it is important to use additional information to reduce the uncertainty in the estimation of K.

In order to investigate the variance of K, a mixed noise model was also used to generate 2090 synthetic series. Some results are shown in Table 2-6. The F-test was used to test the null hypothesis that there is no difference in the variance between the independent series and the Canadian river peak flows. The null hypothesis cannot be rejected at the significant level of 5%. So the variance of the Hurst K for different cases may be seen as from the same process. But due to its variability, we did not get exactly the same value.

Comparing the distribution of the Hurst K obtained from the data set and the one generated by the mixed noise model with h equal to 0.6 suggests that the mean of the model Hurst of the data set is approximate 0.6.

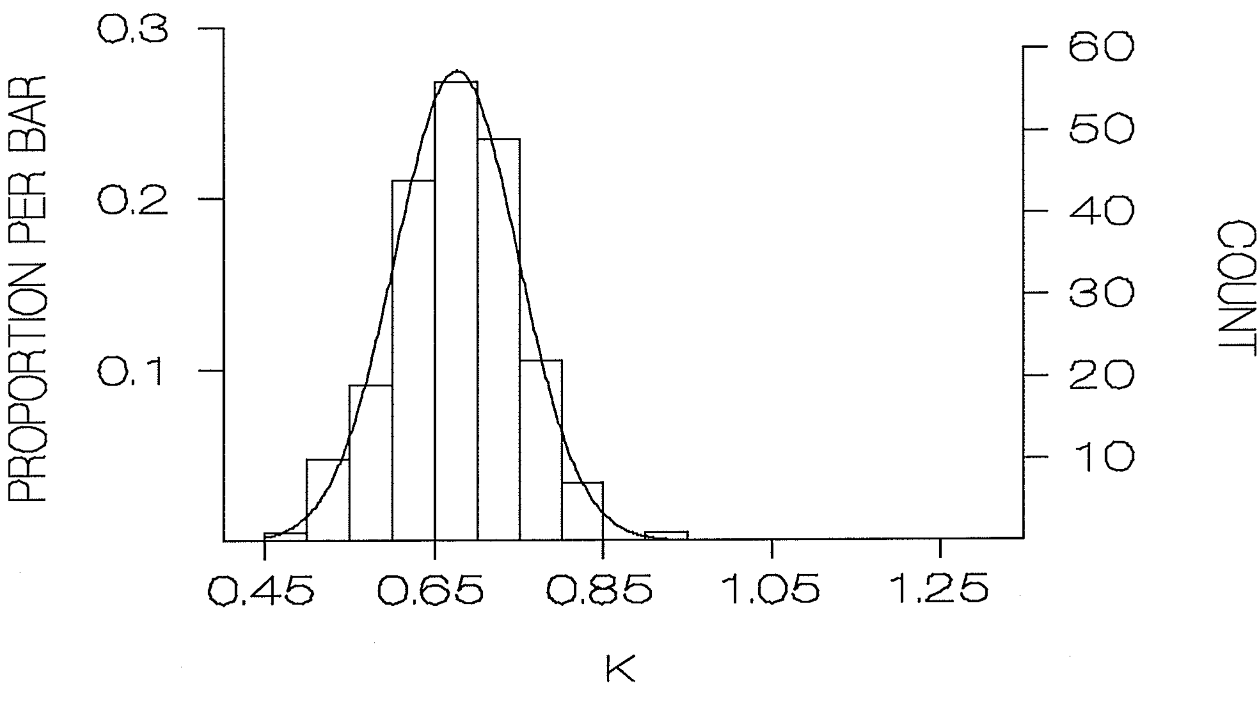


Figure 2-6. Hurst K Distribution of Annual Peak Flows of the Data Set

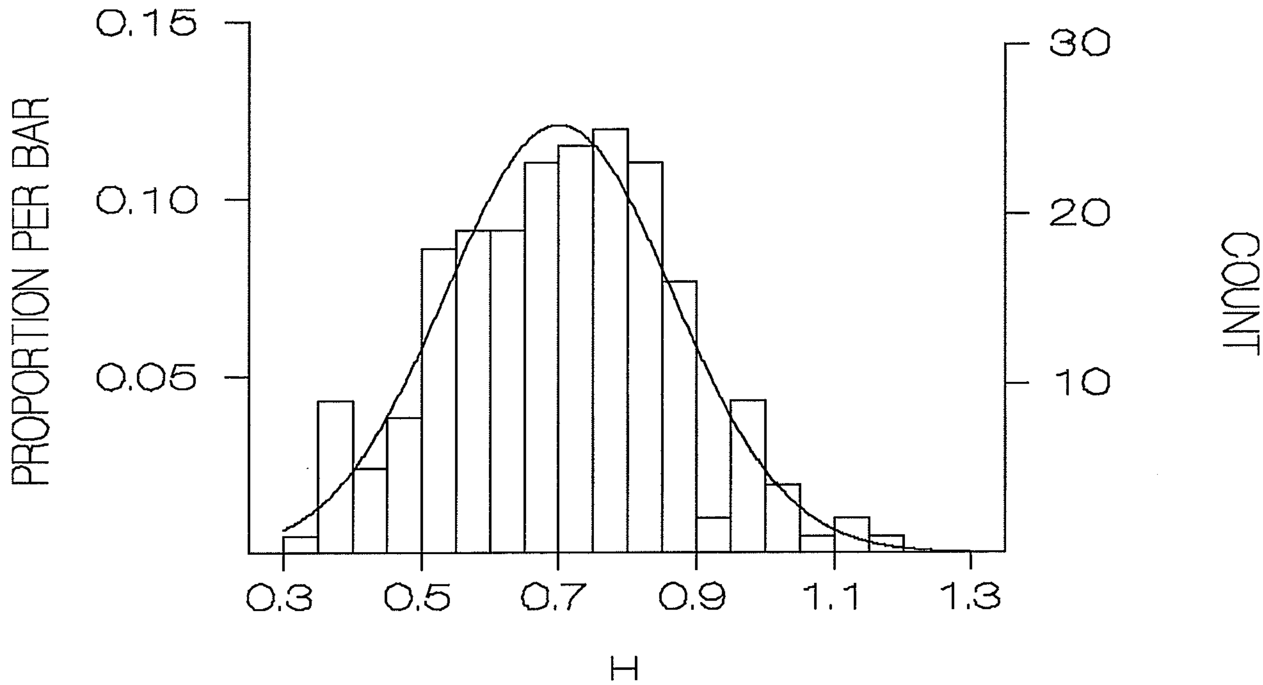


Figure 2-7. Hurst H(g10) Distribution of Annual Peak Flows of the Data Set

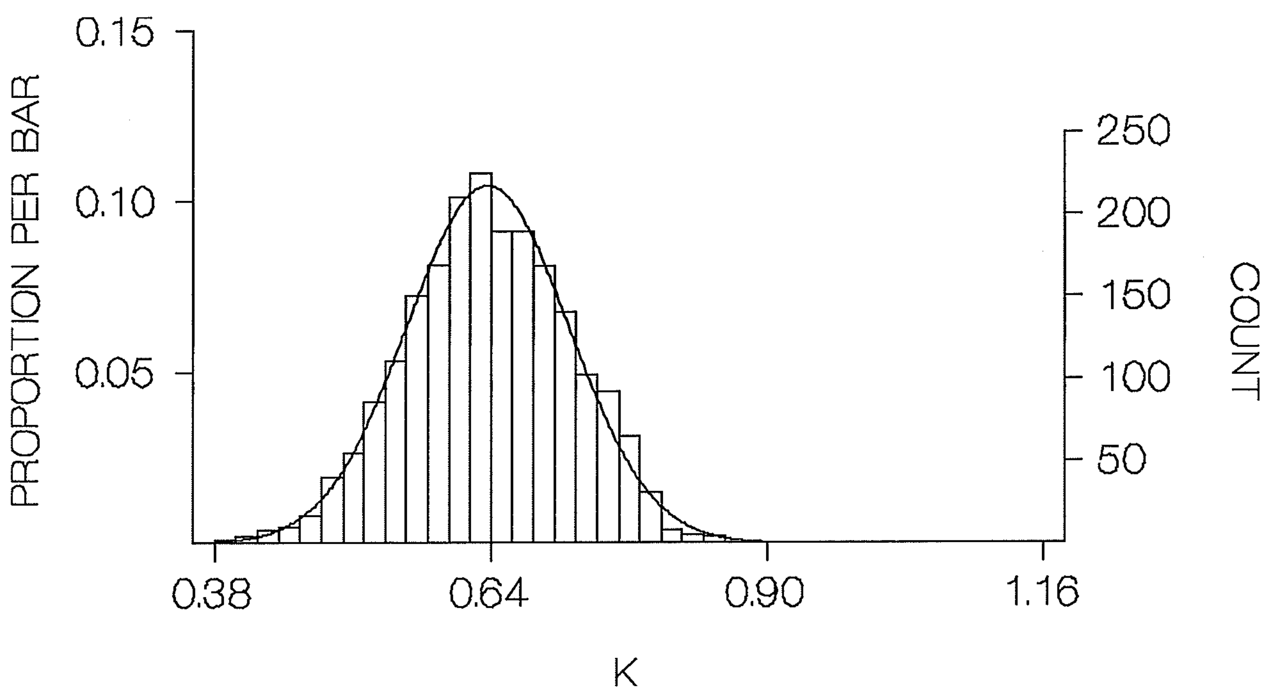


Figure 2-8. Hurst K Distribution of Time Independent Series

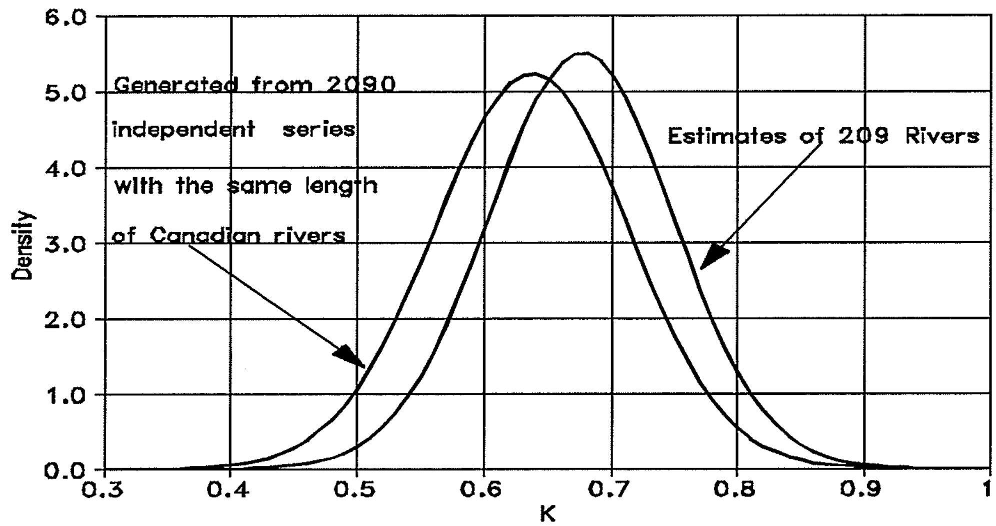


Figure 2-10. Comparison of K Distributions Between the Data Set and Independent Series

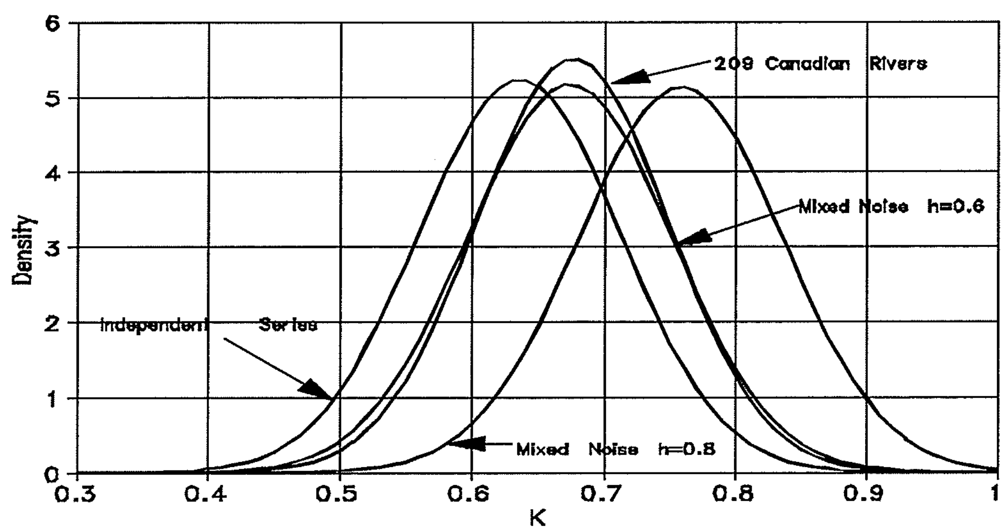


Figure 2-11. Comparison of Various K Distributions

Table 2-6. Comparison of K between Generated series and Canadian Flows

Sample Series	Mean of K	Standard Deviation
Peak Flows of the Data Set	0.6756	0.07256
2090 Independent Series	0.6365	0.07647
2090 Series Generated by Mixed Noise Model with h = 0.60	0.6735	0.07740
2090 Series Generated by Mixed Noise Model with h = 0.80	0.7579	0.07793

2.7. CONCLUSION

On the basis of Anderson's test, the null hypothesis that annual peak flow series are time independent can not be rejected for most Canadian rivers. However, the values of the Hurst K showed significant long term persistence for many rivers. This contradiction arises because Anderson's test is not sensitive enough to show small correlation in relatively short series. The null hypothesis was rejected for the data series "on average" by means of a t-test. The Hurst statistic, on the other hand, measures the long term correlation only. When the short and the long term correlations are considered simultaneously, the conclusion is that one may expect a significant correlation structure.

Comparing the sample distributions of the Hurst coefficient K from historical data, independent series and series generated by the mixed noise model with different values of the Hurst parameter h shows the need for a better estimation procedure for h since K for individual rivers is a biased estimator and its sample variance is large. The first order correlation is certainly much smaller than the value assumed in the fractional noise models or their approximations.

The information contained in the data set can be used as a prior distribution and Bayesian estimation procedures can be applied to update this prior with the sample information for a given river. This requires a suitable simulation procedure (a) to remove the bias from the estimation of the serial correlation and the Hurst statistic , (b) to develop suitable prior distribution of K, and (c) to develop the likelihood functions for the Bayesian estimation. Developing such a procedure is the topic of the next chapter.

CHAPTER 3. SIMULATION MODELS FOR ANNUAL PEAK FLOWS

3.1 INTRODUCTION

A simulation model for annual peak flows must preserve the observed short term serial correlation as well as to the main features of the long term serial correlation structure. In this review, we will begin with a discussion of the autoregressive (AR and ARMA) models although these are basically short memory models. The AR and ARMA models are based on intuitive types of time dependence and they are the basis of more sophisticated models that can more adequately represent long term serial correlation. Next, the fractional gaussian noise model (FGN) is discussed, which was the first model that could adequately reproduce the Hurst phenomenon. Attempts to reduce the excessive amount of computer time needed and to obtain more realistic short term correlation for FGN led to the development of the fast fractional gaussian noise generator, broken line process, ARMA-Markov process and the mixed noise model. These models will also be discussed.

3.2. AUTOREGRESSIVE MODELS (AR)

Autoregressive (AR) models have been extensively used in hydrology and water resources since the early 1960's for modelling annual and periodic hydrologic time series. The application of these models has been attractive in hydrology mainly because

(1) the autoregressive form has an intuitive type of time dependence (the value of a variable at present time depends on the values at previous times);

(2) they are the simplest models to use.

The historical development of AR models in hydrology may be divided into two periods, which are that of the 1960 decade initiated mainly by the work of Thomas and Fiering (1962) and Yevjevich (1963), and the 1970 decade motivated by the publication of the book of Box and Jenkins (1970). During the first period, the estimation of parameters was based on the method of moments and the test of goodness of fit of the models was based on the correlogram analysis. Since 1970, researchers in hydrology have used more refined methodologies suggested by Box and Jenkins and others, especially for improving the estimates of parameters of the model, for verifying the conditions to be met by those parameters, for verifying or checking the assumptions of the model and for selecting among competing models.

AR models may have constant parameters or parameters varying with time. Constant parameters apply to annual values and others may be applied for time series of intervals that are a fraction of a year.

For a stationary time series y_t, which is normally distributed with mean μ and variance σ^2 and has an autoregressive (Markovian) correlation (or time dependence structure) with constant parameters, the autoregressive model of order p, denoted by AR(p), can be written as:

$$Y_t = \mu + \phi_1 (Y_{t-1} - \mu) + \dots\dots + \phi_p (Y_{t-p} - \mu) + \varepsilon_t \qquad (3\text{-}1)$$

where ε_t is the time independent (not correlated) series which is independent of Y_t, and it is also normally distributed with mean 0 and variance σ_ε^2. The coefficients $\phi_1, \dots\dots \phi_p$ are called the autoregression coefficients. Parameters to be estimated are ($\mu, \sigma^2, \phi_1, \dots\dots \phi_p, \sigma_\varepsilon^2$).

Although various forms of AR models have been used in the field of stochastic hydrology (J. D. Salas et al, 1980), They are actually the same model. The use of one form or another depends mainly on personal performance and convenience. When p is set as one, we get

$$Y_t=\mu+\phi_1(Y_{t-1}-\mu)+\varepsilon_t \tag{3-2}$$

This model is called the first-order autoregressive or first order Markov model. The variable Y at time t is a function of the variable Y at time t−1 plus a random part. Similarly, the AR(2) model can be expressed as

$$Y_t=\mu+\phi_1(Y_{t-1}-\mu)+\phi_2(Y_{t-2}-\mu)+\varepsilon_t \tag{3-3}$$

Properties of AR models with constant parameters can be presented as follows:

The expected value and variance are

$$\mathbf{E}(\ Y_t\ ) = \mu \tag{3-4}$$

$$\mathbf{E}(\ \varepsilon_t\ ) = 0 \tag{3-5}$$

$$\mathbf{Var}(\ Y_t\ ) = \sigma^2 \tag{3-6}$$

$$\mathbf{Var}(\ \varepsilon_t\ ) = \sigma_\varepsilon^{\ 2} \tag{3-7}$$

When they are estimated from historical data, the variance of σ^2 and $\sigma_\varepsilon^{\ 2}$ are related by

$$\sigma_\varepsilon^{\ 2}=\sigma^2[1-\sum_{j=1}^{p}\phi_j\, p_j] \tag{3-8}$$

where ϕ_j is the j^{th} autoregression coefficient and p_j is the lag-j autocorrelation coefficient of variable Y_t.

The autocorrelation function of the variable Y_t is obtained by multiplying both sides of the model form by Y_{t-k} and taking the expectation term by term.

$$\rho_k=\phi_1\rho_{k-1}+\phi_2\rho_{k-2}+......+\phi_p\rho_{k-p} \tag{3-9}$$

Where $k>0$

For AR(1) model

$$\rho_k=\phi_1^{\,k} \tag{3-10}$$

For AR(2) model

$$\rho_k=\phi_1\rho_{k-1}+\phi_2\rho_{k-2} \tag{3-11}$$

3.3. AUTOREGRESSIVE-MOVING AVERAGE MODELS (ARMA)

Some hydrological series can be modeled by autoregressive-moving average models, which are mixed autoregressive and moving average processes. By applying this model, in some cases, we use the least number of parameters. As the parameters are estimated from historical data, the idea of parsimony in the number of parameters is particularly attractive.

It is assumed that the hydrologic time series, to be modeled by an ARMA process, is stationary and approximately normal. Otherwise, the appropriate transformation of the original variable is firstly performed. Let's consider the values of a hydrologic time series Y_t, Y_{t+1}, Y_{t+2}, ... at equally spaced time t, t+1, t+2, ..., and let the deviation from the mean be

$$Z_t = Y_t - \mu \tag{3.12}$$

The series Z_t may be represented as an infinite weighted sum of independent random variables ϵ_t, ϵ_{t-1}, ϵ_{t-2}, ...

$$Z_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \ldots\ldots \tag{3.13}$$

If we make Z_t as a finite weighted sum of ϵ_t, ϵ_{t-1}, ... ϵ_{t-q}, then

$$Z_t = \epsilon_t - \Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \ldots\ldots - \Theta_q \epsilon_{t-q} \tag{3.14}$$

and this is usually called the MA(q) model.

An autoregressive model of order p and a moving average model of order q may be combined to obtain the "mixed autoregressive-moving average" (ARMA) model of order (p,q), it is defined by

$$Z_t = \varepsilon_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots\ldots + \phi_p Z_{t-p} - \Theta_1 \varepsilon_{t-1} - \Theta_2 \varepsilon_{t-2} - \ldots\ldots - \Theta_q \varepsilon_{t-q} \tag{3.15}$$

The parameters of the model are μ, σ_ε^2, ϕ_1, ...ϕ_p, Θ_1, ...Θ_q. A total of p+q+2 parameters must be evaluated from data.

O'Connell (1971,1974) proposed ARMA(1,1) model as an approximation to fractional gaussian noise (FGN). The autocorrelation function of the process is given by

$$\rho_1 = \frac{(\phi-\theta)(1-\phi\theta)}{1-2\phi\theta+\theta^2} \tag{3-16}$$

For $k \geq 2$

$$\rho_k = \phi \rho_{k-1} \tag{3-17}$$

This autocorrelation coefficient decays exponentially from lag-one autocorrelation at a rate controlled by ϕ.

O'Connell (1974) showed that with $0.8 < \phi < 0.99$ and $0.5 < \Theta < 0.95$, the ARMA(1,1) was in certain instances able to model long term serial correlation. When the Hurst coefficient h is asymptotically equal to 0.5 for the ARMA(1,1) process, careful choice of ϕ and Θ will produce values of h substantially greater than 0.5.

This model generally has two principal advantages. First, it contains 2 parameters, secondly, it uses less computing time than FFGN and BL process. The disadvantage of this model is that the Hurst statistic h is not explicitly used in generating mechanism and no equivalence has been found between this model's parameters and the Hurst statistic h

(Lettenmaier and Burges, 1977a; Lye, 1988).

3.4. FRACTIONAL GAUSSIAN NOISE MODEL (FGN)

To explain the Hurst phenomenon, Mandelbrot (1965) first used a self-similar stochastic process which was called fractional gaussian noise to successfully reproduce the so called long term persistence. The properties of this model were published by Mandelbrot and Van Ness (1968) where the terminology of fractional brownian motion and fractional gaussian noise (FGN) was introduced to the hydrological community. (Mandelbrot et al, 1968, Mandelbrot and Wallis, 1969 a, b, c, d, e and Lawrence and Kottegoda, 1977).

FGN employs an autocorrelation function which is independent of any observed correlogram but will automatically reproduce the desired Hurst coefficient h. Fractional gaussian noises with zero mean and unit variance are defined as Gaussian random processes $X_d(t, H)$ which have the following covariance

$$C(s,H)=0.5[|s+1|^{2H}-2|s|^{2H}+|s-1|^{2H}] \tag{3-18}$$

where H is a single parameter in the model, normalized to vary from 0 to 1, called the model Hurst H.

By definition, FGN are continuous parameter processes with "infinite memories", which means the correlation between successive values may be small, large lag autocorrelations are even smaller but are such that their accumulative effect is non-negligible. Notably the rescaled range

R/s is proportional to s^H with H not equal to 0.5. For the definition and a discussion of R/s, see Hurst (1951), Mandelbrot and Wallis (1969 a, b, c). If long run dependence is above H=0.5 (Feller, 1951), construction of a sample function of FGN would unfortunately involve an infinite number of operations, and approximations are needed. The discrete-time fractional noise can be deduced from a Brownian motion B(u) by the formula

$$\Delta B_H(t) = B_H(t) - B_H(t-1) = \int_{-\infty}^{t} K_H(t-u)\, dB(u) \tag{3-19}$$

with the 'kernel function' $K_H(u)$ given by

for $0 < u < 1$, $\mathbf{K_H(u) = u^{(H-0.5)}}$ **(3-20)**

for $u > 1$, $\mathbf{K_H(u) = [\ u^{H-0.5} - (u-1)^{H-0.5}\ ]}$ **(3-21)**

The integral in equation (3-19), which defines $B_H(t)$-$B_H(t-1)$, cannot be evaluated exactly on a computer. Its numerical evaluation necessarily involves three approximations:

1. The span $-\infty < u < t$ must be replaced by finite span $-M+t < u < t$. This introduces a low frequency error term. Its most striking consequence is to make the asymptotic behaviour of R(t,s)/s(t,s) follow an $\sqrt{s}$ law rather than the s^H law;

2. A discrete grid ε must be selected for the variable of integration u; in particular, the infinite small dB(u) must be replaced by a finite difference $B(u+\varepsilon)$-B(u). This introduces a high frequency error term;

3. Once the grid $\varepsilon > 0$ is selected, there is little point in computing the kernel K_H exactly. Simplification of the analytic form of K_H need not introduce major additional errors;

4. Finally, another discrete grid must be selected for $K(t-u)[B(u+\varepsilon)-B(u)]$, the values of this integrand being computed to a finite number of decimals.

Two typical approximations to $B_H(t)$ were mentioned in Mandelbrot and Wallis' paper (Mandelbrot and Wallis 1969), which were called type 1 and type 2 approximations. They are finite weighted moving average in which the number of Gaussian variables to average is roughly proportional to the size T of the desired sample.

Type 1 Approximation (FN1)

In type 1 approximation, grid ε is taken as 0.10 and continually varying Kernel $K_H(u)$ is replaced by a stepwise varying Kernel $K_1(u/H,M)$ as follows:

when $u > M$, $K_1(u/H,M) = 0$

when $u < M$ and u is an exact multiple of 1/10, $K_1(u/H,M) = K_H(u)$

when $u < M$ and u is not an exact multiple of 1/10, and if w is the smallest multiple of 1/10 whose value exceeds u, then

$$K_1(u/H,M) = K_H(w)$$

So, type 1 approximation of fractional noise can be written as

$$F_1(t/H,M) = \sum_{n=10(t-M)}^{10t-1} K_1(t-\frac{n}{10}/H,M)G(\frac{n}{10}) \qquad \text{(2-22)}$$

where M is a very large number and G(n/10) represents independent Gaussian process with zero mean and variance 1/10.

Type 2 Approximation (FN2)

Quite extensive computation is needed when we choose ε=0.1 in type 1 approximation. To simplify them, one could select a rougher grid, say ε equal to 1. For $0.5 < H < 1$, type 2 approximating Kernel $K_2(u/H,M)$ is selected as follows:

when $u > M$, $K_2(u/H,M) = 0$;

when $u < M$ and u is an integer, $K_2(u/H,M) = (H-0.5)\ u^{H-0.5}$;

when $u < M$ and u is not an integer, and if w is the smaller integer whose value exceeds u, $K_2(u/H,M)$ is the function of $(H-0.5)\ w^{H-0.5}$.

The discrete time average corresponding to this Kernel takes the form:

$$F_2(t/H,M)=(H-0.5)\sum_{u=t-m}^{t-1}(t-u)^{H-1.5}G(u) \qquad (2\text{-}23)$$

Type 1 and type 2 functions share essentially identical very low frequency properties but have different high frequency properties. So, type 1 approximation may be expected to be best when high frequency effects are slight, that is when H is near 1.0. From the discussion in Mandelbrot's paper, we know that the two approximations indeed differ little for H=0.9. When however, H reaches 0.5, the kernel K_1 vanishes except for $u<1.0$, whereas K_2 remains non-zero throughout $u<M$. The two approximation therefore differ greatly for H=0.5, and the approximation K_2 is of limited scope. Its usefulness is due to its formal simplicity combined with the fact that empirical values of H are often near 1. For $0< H <0.5$, high-frequency effects are very strong, type 2 approximation is a bit more complicated to implement. In writing the approximate Kernel K_2, one must ensure that summation of $K_2(u/H,M)$ from u=0 to infinite either vanishes or is very close to zero.

3.5. FAST FRACTIONAL GAUSSIAN NOISE GENERATOR (FFGN)

Since type 1 and type 2 approximations to fractional gaussian noise presented practical and psychological drawbacks, Mandelbrot developed a new approximation to be designated $X_f(t,H)$ and called it fast fractional gaussian noise (FFGN). To construct a sample of FFGN $X_f(t,H)$, one needs H and two additional convenience parameters, the base B>1 and the quality factor Q. A sample of T values of $X_f(t,H)$, normalized to have zero mean and unit variance, is then constructed as the sum of a high frequency Markov-Gauss term and a low frequency that is the weighted sum of a number of N(t) of independent Markov-Gauss processes.

The low frequency term $X_L(t,H)$ can be defined as

$$X_L(t,H)=\sum_{n=1}^{N(T)} W_n X(t,r_n/MG) \qquad (3\text{-}24)$$

where $X(t,r_n/MG)$ is the Markov-Gauss process of variance 1 and covariance r_n^s with $r=\exp(-B^{-n})$. The weight given to $X_t(t,r_n/MG)$ is W_n with

$$W_n^2=\frac{H(2H-1)(B^{1-H}-B^{-1+H})}{\Gamma(3-2H)}B^{-2(1-H)n} \qquad (3\text{-}25)$$

Γ being the gamma function. Thus each term depends on B and H. The number of terms satisfies

$$N(T) = \| \operatorname{Log}(QT)/\operatorname{Log}B \| \qquad (3\text{-}26)$$

which also depends on the desired sample size T; $\|X\|$ designates the smallest integer larger than X. The base B and the quality factor (or relative memory) Q together determine the quality of this approximation.

To generate $X(t,r_n/MG)$, one needs to generate a sequence of Gaussian variables of zero mean and unit variance $G_n(t)$

$$X(1,r_n/MG) = G_n(1)$$

$$X(2,r_n/MG) = r_n\, X(1,r_n/MG) + (1-r_n^2)^{0.5}\, G_n(2)$$

....

$$X(t,r_n/MG) = r_n\, X(t-1,r_n/MG) + (1-r_n^2)^{0.5}\, G_n(t)$$

We can see that each operation needs the latest value of $X(t-1,r_n/MG)$ and a gaussian random variable which is independent of any values used before. When $B \rightarrow 1$ and $Q \rightarrow \infty$, the function N(T) increases and the quality of the low frequency approximation improves. It was recommended in Mandelbrot's paper to take B equal to 2, 3, or 4, and let the value of Q depends on H and two further convenience parameter th2 and th3 as shown in Table 3-1. The role TQ plays here is analogous to that of M in type 1 and type 2 approximations of fractional gaussian noise (Mandelbrot and Wallis, 1969 a).

Table 3-1. Values of Quality Factor Q for Various Thresholds th and values of H

	H=.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
th2=0.1 th3=0.1	0.5	1.0	1.7	2.4	3.2	4.0	4.7	5.4	6.0
th2=0.05 th3=0.1	0.5	1.1	1.9	3.0	4.5	6.1	7.9	9.6	11.1
th2=0.1 th3=0.05	1.0	2.0	3.1	4.2	5.1	5.9	6.6	7.1	7.5
th2=0.05 th3=0.05	1.0	2.3	4.0	6.1	8.2	10.2	12.0	13.5	11.5

The generation of short run, high frequency terms $X_h(t,H)$ is independent of T and can be selected to fit either the data or the discrete fractional gaussian noise model (DFGN) optimally. The term $X_h(t,h)$ will usually be a Markov-Gauss process, but for H near 1 it is sufficient to settle for an independent gaussian term that fits DFGN best; Its variance depends on H and B as follows:

$$Variance = \frac{1-B^{-(1-H)}H(2H-1)}{\Gamma(3-2H)} \qquad (3\text{-}27)$$

If one takes for $X_h(t,H)$ the recommended Markov-Gauss process, its variance should be as described above and its correlation of lag-one should be

$$2^{2H-1}-1+\sum_{n=1}^{N(T)} W_n(1-r_n)-\frac{B^{-(1-H)}H(2H-1)}{\Gamma(3-2H)} \qquad (3\text{-}28)$$

3.6. BROKEN LINE PROCESS (BL)

The broken line process (BL) was firstly introduced by Ditlevsen (1969) and was developed by Mejia (1971), Rodriguez-Iturbe et al (1972), Mejia et al (1972 a b), and Gareia (1972). Their purpose had been to show that hydrology and turbulence theory could take account of long-run effects without using the fractional noise. Also, since fractional gaussian noise needs extensive computing resources, BL is one of the alternative processes to reproduce the Hurst phenomenon (Meljia et al, 1972, 1974).

The BL model consists of a sum of series of simple BL processes, each of which results from linear interpolation between equally-spaced normally and independently distributed (NID) variables with random displacement of the starting point of the series in order to make it stationary. The simple BL can be stated as

$$\xi(t)=\xi(t'-ca)=\sum_{j=0}^{\infty}[\eta_j+\frac{\eta_{j+1}-\eta_j}{a}(t'-ja)]I(t') \qquad (3\text{-}29)$$

where η = independent and identically distributed random variables with zero mean and variance σ^2;

c = a random variable uniformly distributed over the interval (0,1);

a = time distance between the η_j, also referred to as the memory parameter and

$I(t') = 1 \qquad J_a \leq t' \leq (j+1)a$

$I(t') = 0 \qquad$ otherwise

The mean of this process is 0 and its variance is 2/3 σ^2. The autocorrelation function of the process is

$$\rho_s = 1\text{-}0.75\,(s/a)^2\,[\,2\text{-}\,2/a] \qquad 0 \leq s \leq a$$

$$\rho_s = 0.25\,[\,2 - s/a\,]^3 \qquad a \leq s \leq 2a$$

$$\rho_s = 0 \qquad 2a \leq s \qquad (2\text{-}30)$$

The BL process is obtained from a weighted summation of NT simple BL process. This procedure is similar to that in FFGN. It is found that only a quarter of time required for FFGN is needed for BL model.

3.7. ARMA-MARKOV MODEL (AM)

While the ARMA model does give a low-cost alternative to FFGN model, the unsuccessfulness to preserve a given population Hurst coefficient as an explicit model parameter may prove to be a drawback in operational applications. The FFGN model, which uses the Hurst coefficient as an explicit input parameter, is relatively expensive to run, especially for

$H \geq 0.8$, where we need a large number of Markov terms which must be summed to give an adequate approximation of infinite memory.

In the ARMA(1,1) model, the decay rate of the autocorrelation function is not dependent on the lag-one correlation coefficient, therefore, if an ARMA(1,1) process is used as one of several independent additive processes, large lag persistence may be achieved by either the lag-one correlation coefficient or the ARMA variance contribution. After some experimentation, Lettenmaier et al found a lag-one Markov process could give an acceptable high frequency fit. The combination of the high frequency lag-one Markov and low frequency ARMA(1,1) process constitutes the ARMA-Markov model. This model has 5 parameters which are the Markov and ARMA variance fractions, the ARMA and Markov lag-one correlation coefficients, and ϕ. The second parameter Θ, Θ of the ARMA process, is uniquely defined by ϕ and the lag-one correlation coefficient of the ARMA process, η_{AM}, (when η_{AM} is substituted for $\eta(1)$ and $\eta(2)$) if the invertibility condition (Box and Jenkins, 1970) $|\Theta| < 1$ is imposed. The generating equation for a standard ARMA-Markov process which has zero mean and unit variance is

$$X_t = \rho_M X_{t-1}{}^{(M)} + \varepsilon_t{}^{(M)} + \phi X_{t-1}{}^{(AM)} - \theta \varepsilon_{t-1}{}^{(AM)} + \varepsilon_t{}^{(AM)} \qquad \textbf{(2-31)}$$

where $\varepsilon_t^{(M)}$ and $\varepsilon_t^{(AM)}$ are independent processes having variance $C_1(1-\rho_m^2)$ and $C_2[(1-\phi^2)/(1+\Theta^2-2\Theta\phi)]$ respectively.

For the standard process the autocorrelation function is fitted to the theoretical FGN autocorrelation function at three lags. Lag-one correlation coefficient may be arbitrarily specified. The parameters of this model can be obtained by solving the following equations:

$$C_1+C_2=1 \tag{3-32}$$

$$\rho_1=C_1\rho_M+C_2\rho_{AM} \tag{3-33}$$

$$\rho_f(K_1,H)=C_1\rho_M^{|K_1|}+C_2\rho_{AM}\Phi^{|K_1|-1} \tag{3-34}$$

$$\rho_f(K_2,H)=C_1\rho_M^{|K_2|}+C_2\rho_{AM}\phi^{|K_2|-1} \tag{3-35}$$

$$\rho_f(K_3,H)=C_1\rho_M^{|K_3|}+C_2\rho_{AM}\phi^{|K_3|-1} \tag{3-36}$$

where C_1, C_2, ρ_M, ρ_{AM} and ϕ are all constrained to lie between zero and one, ρ_1 is the desired lag-one correlation coefficient, ρ_f is the correlation function of FGN defined by

$$\rho_f(K,H)=0.5[(|K|+1)^{2H}-2|K|^{2H}+(|K|-1)^{2H}]\approx H(2H-1)|K|^{2H-2} \tag{3-37}$$

ρ_M and ρ_{AM} are the autocorrelation coefficients of the Markov and ARMA processes respectively, and K_1, K_2 and K_3 are the arbitrary lags at which the theoretical FGN autocorrelation function matches that of the ARMA-Markov process. Lettenmaier and Burges (1977a) suggest to take K_1, K_2 and K_3 to be approximately L/8, L/2 and L, where L is the length of the time series being generated. The above equations can be solved by means of Newton's method for a given $\rho(1)$ and H. The second ARMA parameter Θ is obtained from

$$\rho_{AM}=\frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\theta\phi} \tag{3-38}$$

where only the value of Θ which makes $|\Theta| < 1$ is taken. According to Srikanthan (1979) and Lye (1988), though ARMA-Markov model is able to reproduce Hurst statistic as FFGN and BL, it has the following disadvantages:

(a) It does not preserve the mean like FFGN or BL;

(b) It considerably underestimates the first serial correlation coefficient. Both FFGN and BL give better results;

(c) It is not possible to obtain parameter estimates for series with negative first serial correlation;

(d) Parameters are difficult to estimate and the small sample biases of parameters are unknown.

However, the ARMA-Markov model can save a lot of computer time comparing to FFGN and BL. It explicitly uses H and $\rho(1)$ in their parameter derivation.

3.8. THE MIXED NOISE MODEL (MN)

The mixed noise model (MN) (Booy and Lye 1987,1989) is the sum of three or four AR(1) processes with weights and correlation coefficients that are calculated to produce the correlation structure corresponding to a given value of h and a given first order correlation. The model can accurately reproduce the autocorrelation function of FGN within any specified lag interval by a proper choice of the weighted processes. In MN, both the Hurst coefficient h and $\rho(1)$ are explicitly used to derive model parameters. At lag one, the model correlation coefficient can be made to correspond with the observed first order correlation coefficient $\rho(1)$ and at high lags, the correlogram on log-log graph paper follows the straight line typical of fast fractional gaussian noise with appropriate Hurst coefficient (Booy and Lye, 1989). The first AR(1) process models the high frequency effects, the second and third process model intermediate or medium frequency effects and the fourth AR(1) process models the very low frequency effects in the time series.

For the zero mean, unit variance process, the generating function can be written as:

$$X_t = W_a(\rho_a X_{t-1}^{(a)} + \epsilon_t^{(a)}) + W_b(\rho_b X_{t-1}^{(b)} + \epsilon_t^{(b)}) + W_c(\rho_c X_{t-1}^{(c)} + \epsilon_t^{(c)}) + W_d(\rho_d X_{t-1}^{(d)} + \epsilon_t^{(d)}) \tag{3-39}$$

where $\varepsilon_t^{(a)}$, $\varepsilon_t^{(b)}$, $\varepsilon_t^{(c)}$, and $\varepsilon_t^{(d)}$ are normal independent processes having variance $(1-\rho_a^2)$, $(1-\rho_b^2)$, $(1-\rho_c^2)$, and $(1-\rho_d^2)$ respectively. ρ_a, ρ_b, ρ_c, and ρ_d are first order partial autocorrelations of four AR(1) models. W_a, W_b, W_c, and W_d are weights of these four AR(1) models.

The autocorrelation function of a mixed noise model is

$$\rho_{MN}(k) = W_a^2 \rho_a^k + W_b^2 \rho_b^k + W_c^2 \rho_c^k + W_d^2 \rho_d^k \tag{3-40}$$

where W_a^2, W_b^2, W_c^2, and W_d^2 are variance fractions which sum to unity.

The eight parameters are obtained by solving the following equations:

(a) $W_a^2 + W_b^2 + W_c^2 + W_d^2 = 1$

(b) $W_a^2 \rho_a + W_b^2 \rho_b + W_c^2 \rho_c + W_d^2 \rho_d = \rho(1)$

(c) $W_a^2 \rho_a^3 + W_b^2 \rho_b^3 + W_c^2 \rho_c^3 + W_d^2 \rho_d^3 = \rho(3,h)$

(d) $W_a^2 \rho_a^8 + W_b^2 \rho_b^8 + W_c^2 \rho_c^8 + W_d^2 \rho_d^8 = \rho(8,h)$

(e) $$W_a^2\rho_a^{25}+W_b^2\rho_b^{25}+W_c^2\rho_c^{25}+W_d^2\rho_d^{25}=\rho(25,h) \quad \text{......} \quad \textbf{(3-41)}$$

(f) $$W_a^2\rho_a^{70}+W_b^2\rho_b^{70}+W_c^2\rho_c^{70}+W_d^2\rho_d^{70}=\rho(70,h)$$

(g) $$W_a^2\rho_a^{200}+W_b^2\rho_b^{200}+W_c^2\rho_c^{200}+W_d^2\rho_d^{200}=\rho(200,h)$$

(h) $$W_a^2\rho_a^{201}+W_b^2\rho_b^{201}+W_c^2\rho_c^{201}+W_d^2\rho_d^{201}=\rho(201,h)$$

where ρ(n,h) is the theoretical correlation coefficient of the fractional noise given by equation (3-37).

Since the autocorrelation function of an AR(1) process diminishes very rapidly with increasing lag, the above equations can be solved sequentially starting from the low frequency (high lag) end. From (g) and (h), and assuming the terms with ρ_a, ρ_b, and ρ_c to be negligible beyond lag 200, then one gets

$$\frac{W_d^2\rho_d^{201}}{W_d^2\rho_d^{200}} = \frac{\rho\ (201,h)}{\rho(200,h)} \quad \textbf{(3-42)}$$

Solving this equation, we get

$$\rho_d=\frac{\rho(201,h)}{\rho(200,h)} \quad \textbf{(3-43)}$$

For a given h, the right hand side of equation can be easily calculated, so ρ_d can be calculated. Substituting ρ_d into equation (g) or (h) and neglecting the term with ρ_a, ρ_b, and ρ_c, W_d can be obtained.

From (e) and (f), assuming the terms with ρ_a, ρ_b to be negligible beyond lag 25, then we get

$$W_c^2 \rho_c^{25} = \rho(25,h) - W_d^2 \rho_d^{25} \qquad (3\text{-}44)$$

$$W_c^2 \rho_c^{70} = \rho(70,h) - W_d^2 \rho_d^{70} \qquad (3\text{-}45)$$

Solving equation (3-44) and (3-45), we get

$$(3\text{-}46) \qquad \rho_c = \left[\frac{(\rho(70,h) - W_d^2 \rho_d^{70})}{(\rho(25,h) - W_d^2 \rho_d^{25})}\right]^{\frac{1}{45}}$$

where the right hand side is known. Substituting ρ_c into equation (3-44) or (3-45), W_c^2 is obtained.

Similarly, from (c) and (d) and assuming the terms containing ρ_a to be negligible beyond lag 3, one gets

$$W_b^2 \rho_b^3 = \rho(3,h) - W_c^2 \rho_c^3 - W_d^2 \rho_d^3 \qquad (3\text{-}47)$$

$$W_b^2 \rho_b^8 = \rho(8,h) - W_c^2 \rho_c^8 - W_d^2 \rho_d^8 \qquad (3\text{-}48)$$

Then we obtain

$$\rho_b = \left[\frac{(\rho(8,h) - W_c^2 \rho_c^8 - W_d^2 \rho_d^8)}{(\rho(3,h) - W_c^2 \rho_c^3 - W_d^2 \rho_d^3}\right]^{\frac{1}{5}} \qquad (3\text{-}49)$$

where all terms on the right hand side of equation are defined. Then substituting ρ_b into equation (3-47) or (3-48), W_b^2 is obtained.

Finally, from (a), W_a^2 can be calculated

$$W_a^2 = 1 - W_b^2 - W_c^2 - W_d^2 \tag{3-50}$$

From (b), ρ_a can be defined

$$\rho_a = \frac{\rho(1) - W_b^2\rho_b - W_c^2\rho_c - W_d^2\rho_d}{W_a^2} \tag{3-51}$$

It is mentioned in Booy and Lye's paper (1989) that " It's possible that for some combinations of $\rho(1)$ and h, ρ_a may be negative. For these cases, ρ_a may be set equal to zero". According to their opinions, alternatively, we could adjust the correlation coefficient of the first AR(1) component to match any observed $\rho(1)$ in the correlogram. In actual time series, the sample values of $\rho(1)$ and h are both quite variable, so that low $\rho(1)$ values with high value of h may be encountered. To match the $\rho(1)$ may then result in strangely shaped correlogram. The compatibility of combination of $\rho(1)$ and h requires further study.

3.9. CONCLUSIONS

In this chapter, several mathematical models which are capable of reproducing hydrological short or long term persistence were reviewed. AR and ARMA models are basically intuitive types of the time dependent models and they are the bases of the more sophisticated models that can adequately reproduce the long term serial correlation. However, an AR model itself can not reproduce the long term correlation of the historical series very well, even using large number of parameters. But the proper combination of a few AR(1) processes can successfully reproduce the observed long term correlation structure. ARMA(1,1) model is in certain instance able to model the long term serial correlation. This model only contains two parameters and uses less computing time than the fractional noise model and the broken line process. The disadvantage of this model is that the Hurst h is not explicitly used in generating mechanism and no equivalence is found between the model parameters and the Hurst h. The ARMA-Markov model is a successful approximation of the fractional noise with less computation time. But it has several disadvantages: this model does not preserve the mean and it considerably underestimates the first order serial correlation. The parameters of this model are difficult to estimate and the small sample biases of parameters are unknown. The fractional gaussian noise can adequately reproduce the long term correlation. However, it needs an enormous amount of computation and generally it generates a much larger first order autocorrelation than the observed value. As an approximation of fractional gaussian noise, the broken line process needs less computing time, but the model parameters are not easy to derive. The fast fractional gaussian noise model as an approximation of fractional gaussian noise saves a lot of computer resources, and it is very easy to run. Unfortunately, this model inherits the disadvantage of the fractional gaussian noise model, which is the serious overestimation of the first order autocorrelation. Some of those models as mentioned above can reproduce hydrological long term correlation structure, however, don't use Hurst coefficient explicitly in the generating function.

The mixed noise model seems to be the most appropriate for modelling peak flows. Firstly, it is easy to calculate the model parameters; secondly, it uses least computer time; thirdly, the lag-one autocorrelation and the Hurst coefficient of sample series are used explicitly as the basic inputs of the generating functions. For these reasons, this model was selected to model annual peak flows in the following study.

CHAPTER 4. BIAS CORRECTION OF THE HURST K

4.1. GENERAL

In the previous chapter, simulation models for annual peak flows were discussed. The mixed noise model was selected as a suitable model for the further simulation of peak flows for Canadian rivers. To use this model, one needs the Hurst statistic h as well as the lag-one autocorrelation coefficient. However, the available methods for the estimation of these parameters have serious bias. It can be shown that the bias of the ACF is dependent on the Hurst statistic, but the bias of the Hurst statistic is relatively independent of the ACF. Since for the derivation of the unbiased ACF one needs the unbiased Hurst statistic, the bias correction of the Hurst K will first be discussed in this chapter.

Wallis and Matalas (1971) showed that the formulas for the computation of both K and H produce biased estimates of the Hurst statistic. The bias is significant even for very long sample lengths. The set of annual peak flows of Canadian rivers used in this study have a mean length of only 43 years. It is therefore necessary to derive a bias correction procedure. The Monte Carlo method was used to derive it. Long memory models including FFGN and MN were used for generating synthetic series. Since FFGN and MN are not valid for a model Hurst less than 0.5, type 1 approximation of Fractional Noise was used for these values.

4.2. MONTE CARLO SIMULATION

The Monte Carlo simulation was performed for twenty Hurst h values equally spaced over the range from 0 to 1.0. The sample size used for each type of time series was 43 years, which is the average length of the annual peak flows of Canadian rivers. The number of replications for each time series was 500. FFGN and MN were used to generate series for the model Hurst h larger than 0.5. When the model Hurst h is less than 0.5, FFGN and MN are no longer valid; for these cases FN1 was used. The Hurst K as well as H(g10) were estimated for each series, after which the mean and the standard deviation of H and K were calculated. Histograms showing the results of the simulation are shown in Figures 4-1 A to 4-1 O. These results are also listed in Table 4-1. Figures 4-2 and 4-3 show the mean response of K and H as a function of the model Hurst h. It is evident that both K and H have serious bias. In fact, K even has more bias than H, but K has less variance than Hurst H. When the model Hurst h is larger than 0.725, the Hurst K underestimates h, otherwise it overestimates it. The cross-over value for H is 0.765. Most simulation results show a symmetric distribution for K and H unless h is close to 0 or 1.0.

K has the advantage that we never get its estimate that is greater than 1.0. This does not hold for the Hurst H. When the model Hurst is near 1.0, we may get H estimates that are greater than 1.0. This is one reason why we prefer to use the Hurst K instead of H for the remainder of this study. The other reasons are the smaller variance of K, the much simpler and faster computation and the fact that the computation procedure is not to a degree arbitrary as it is for H.

Table 4-1. Mean Responses of H and K for Different Model Hurst h

Model Hurst	Model	Mean Response of H	STD of H	Mean Response of K	STD of K
0.975	FFGN	0.9545	0.1529	0.8677	0.06967
0.925	FFGN	0.8712	0.1704	0.8217	0.07162
0.900	MN			0.7892	
0.875	FFGN	0.8186	0.1717	0.7875	0.07326
0.825	FFGN	0.7905	0.1760	0.7642	0.07801
0.800	MN			0.75222	
0.775	FFGN	0.7674	0.1901	0.7412	0.08292
0.725	FFGN	0.7415	0.1781	0.7242	0.07761
0.700	MN			0.7095	
0.675	FFGN	0.7217	0.1834	0.7084	0.08144
0.625	FFGN	0.6817	0.1718	0.6823	0.07875
0.600	MN			0.6689	
0.575	FFGN	0.6534	0.1697	0.6639	0.07684
0.525	FFGN	0.6203	0.1616	0.6413	0.07338
0.475	FN1	0.5844	0.1726	0.6154	0.07463
0.425	FN1	0.5748	0.1423	0.6030	0.07720
0.325	FN1			0.5618	0.08030
0.225	FN1			0.5300	0.06955
0.100	FN1			0.4943	0.06490
0.025	FN1			0.4677	0.05850

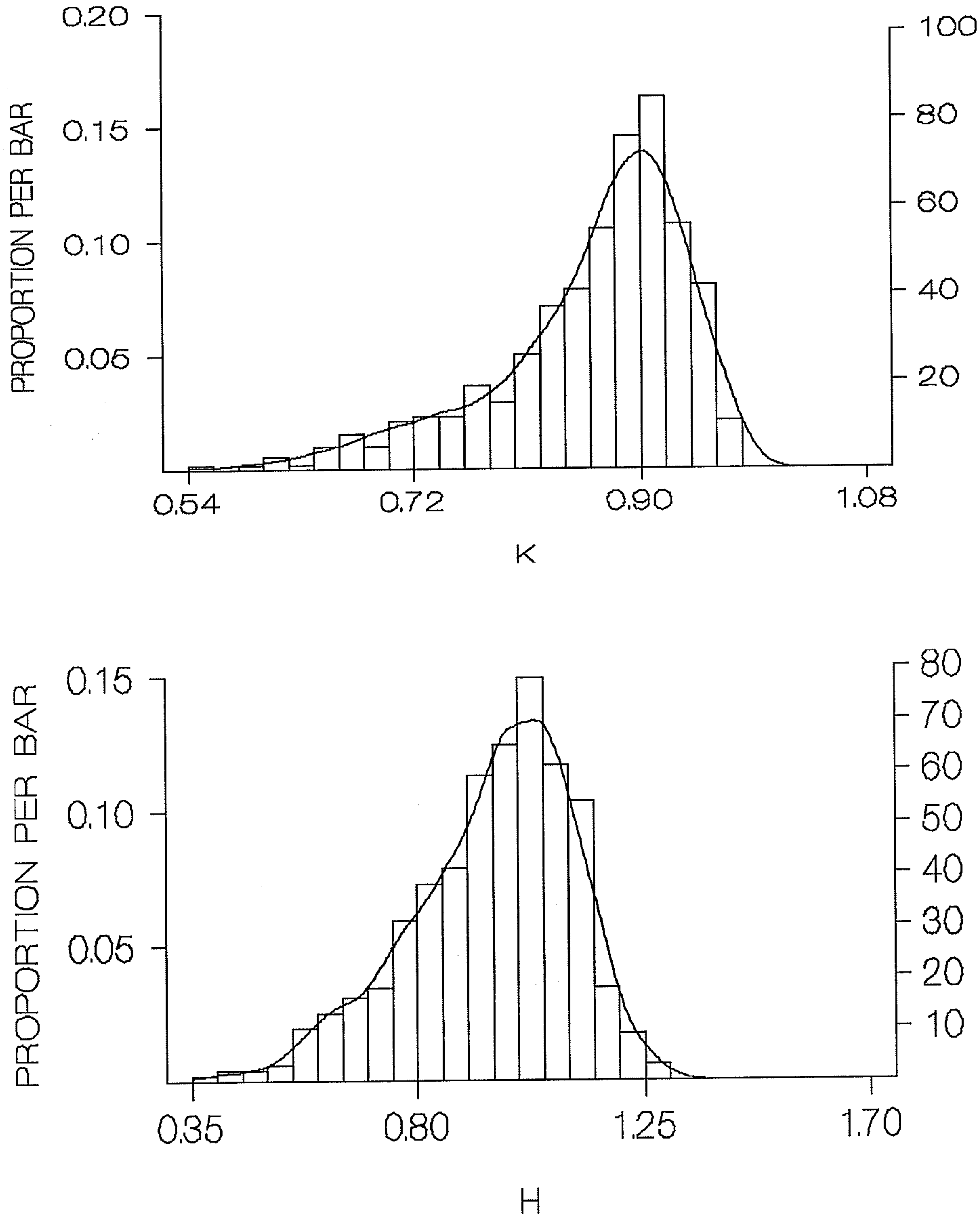
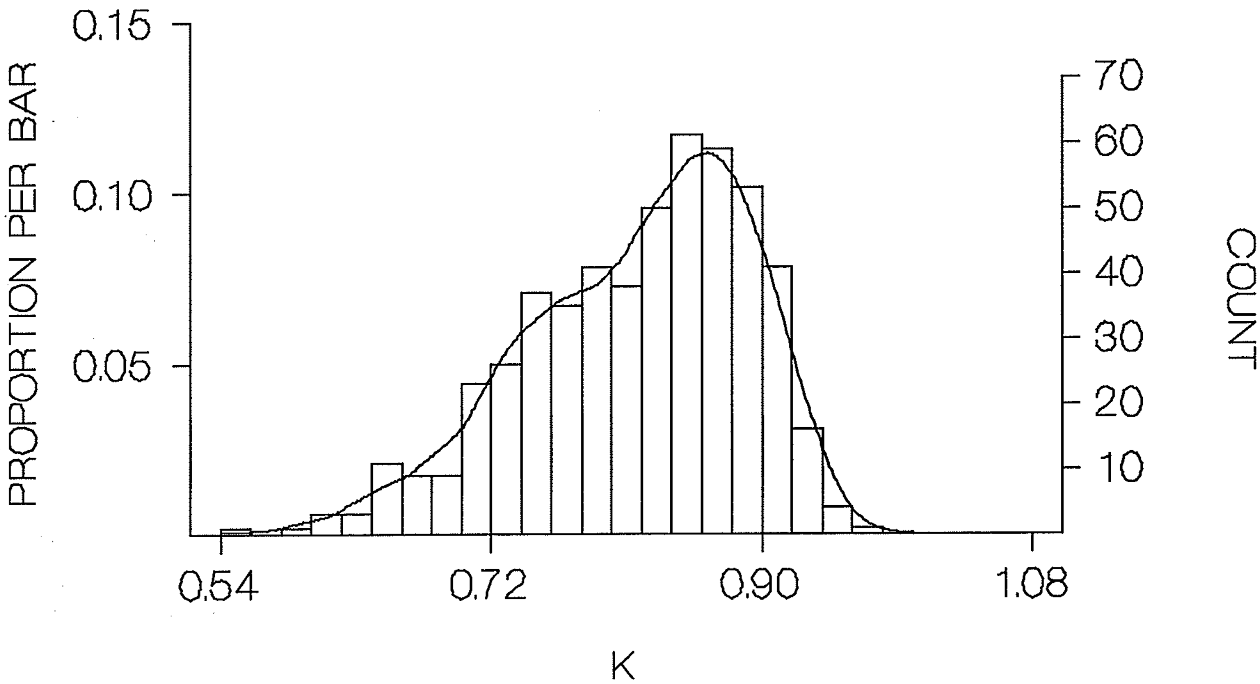


Figure 4-1A. Sample Distributions of K and H for Model Hurst 0.975



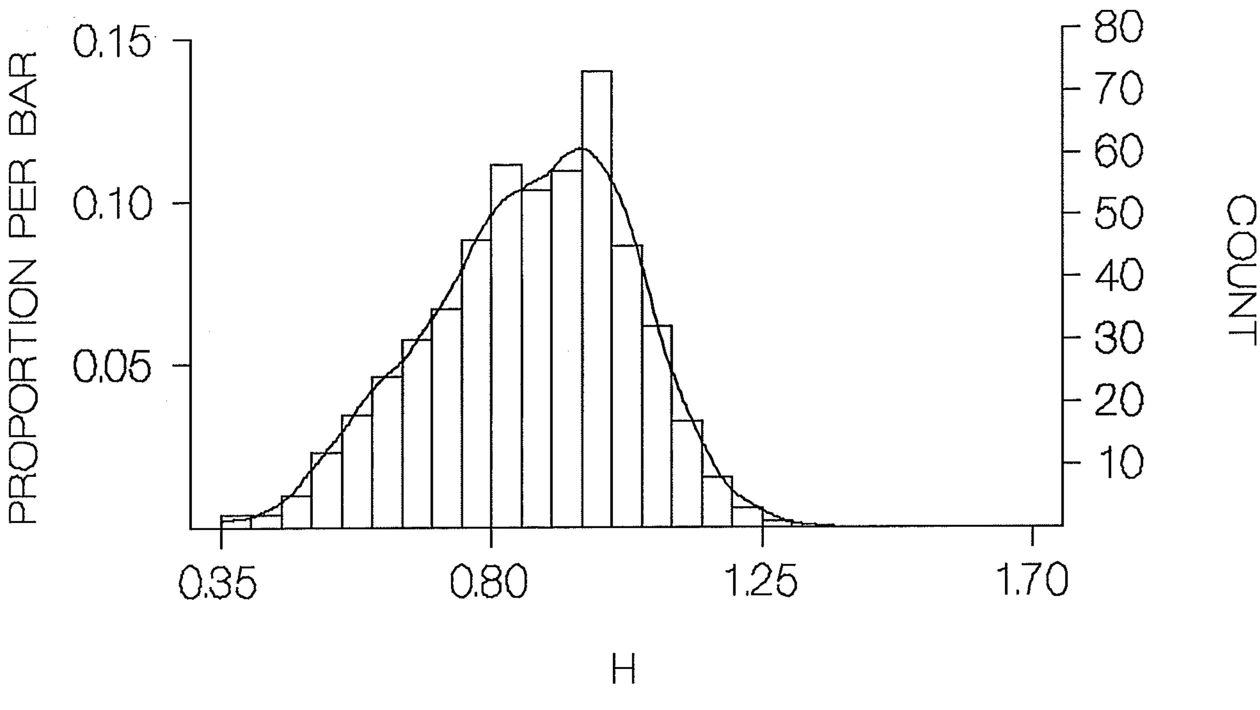


Figure 4-1B. Sample Distributions of K and H for Model Hurst 0.925

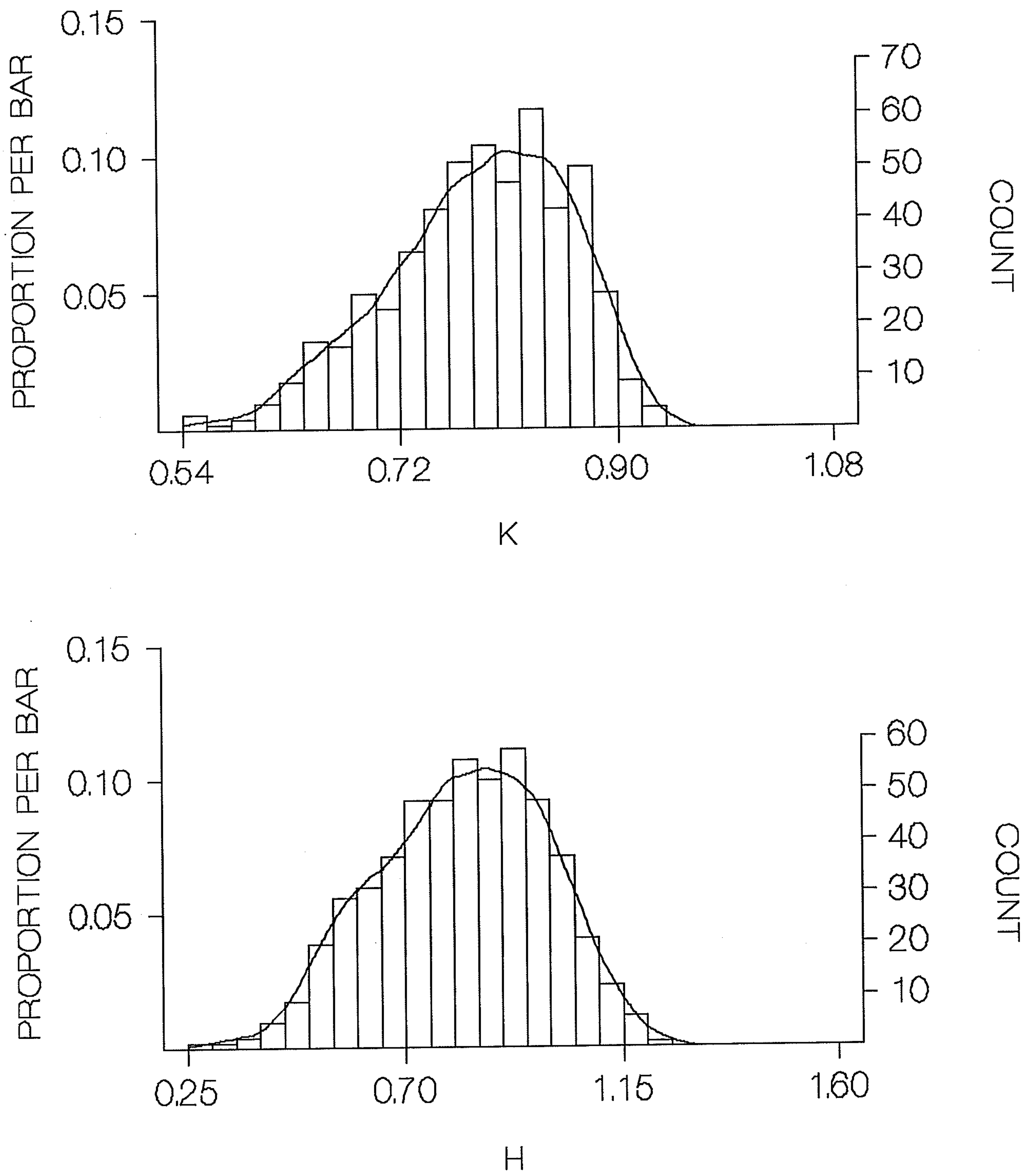
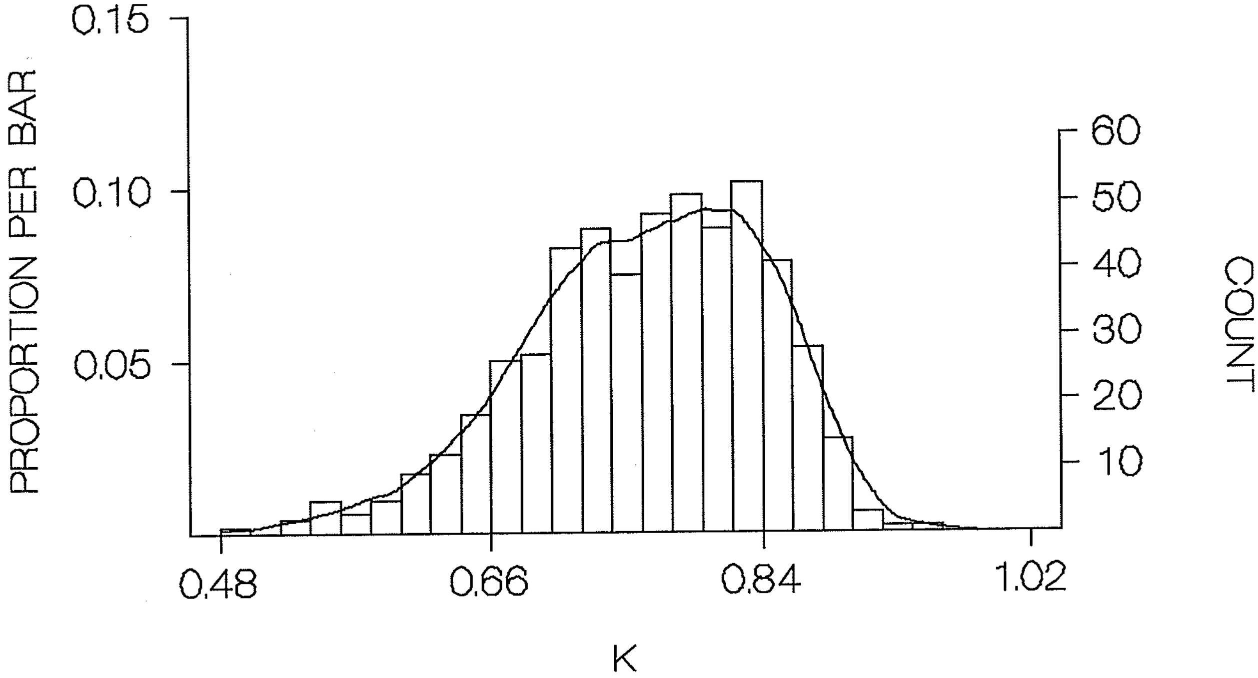


Figure 4-1C. Sample Distributions of K and H for Model Hurst 0.875



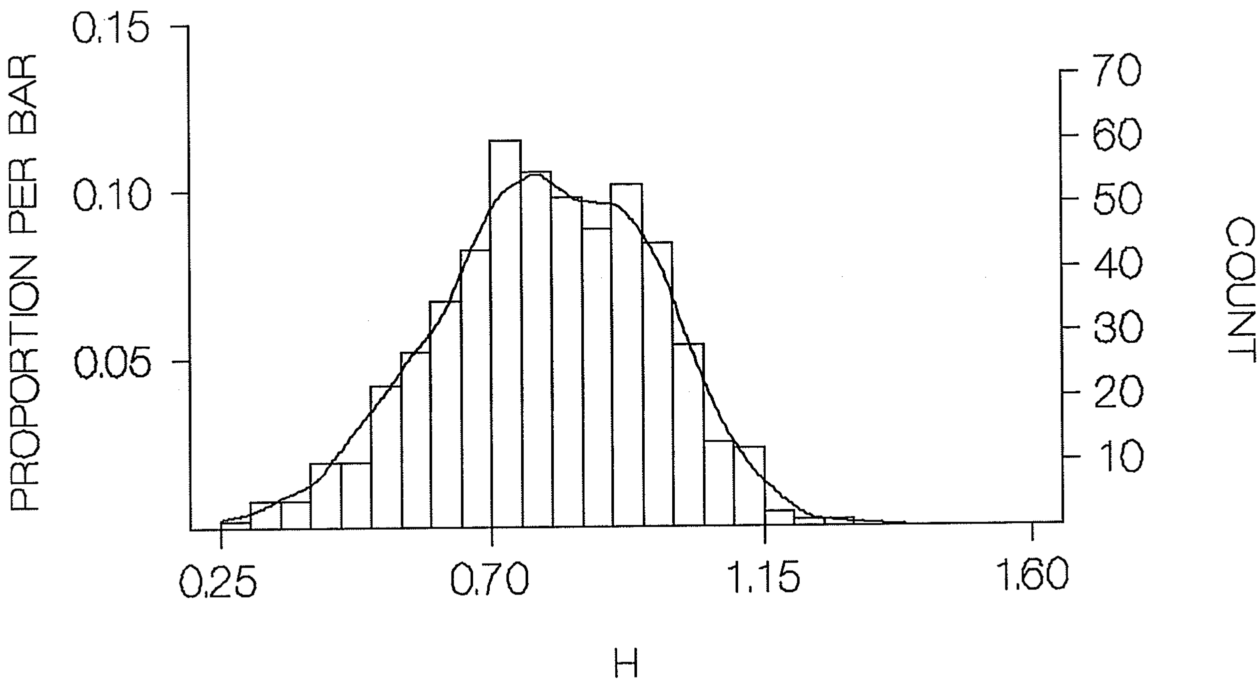


Figure 4-1D. Sample Distributions of K and H for Model Hurst 0.825

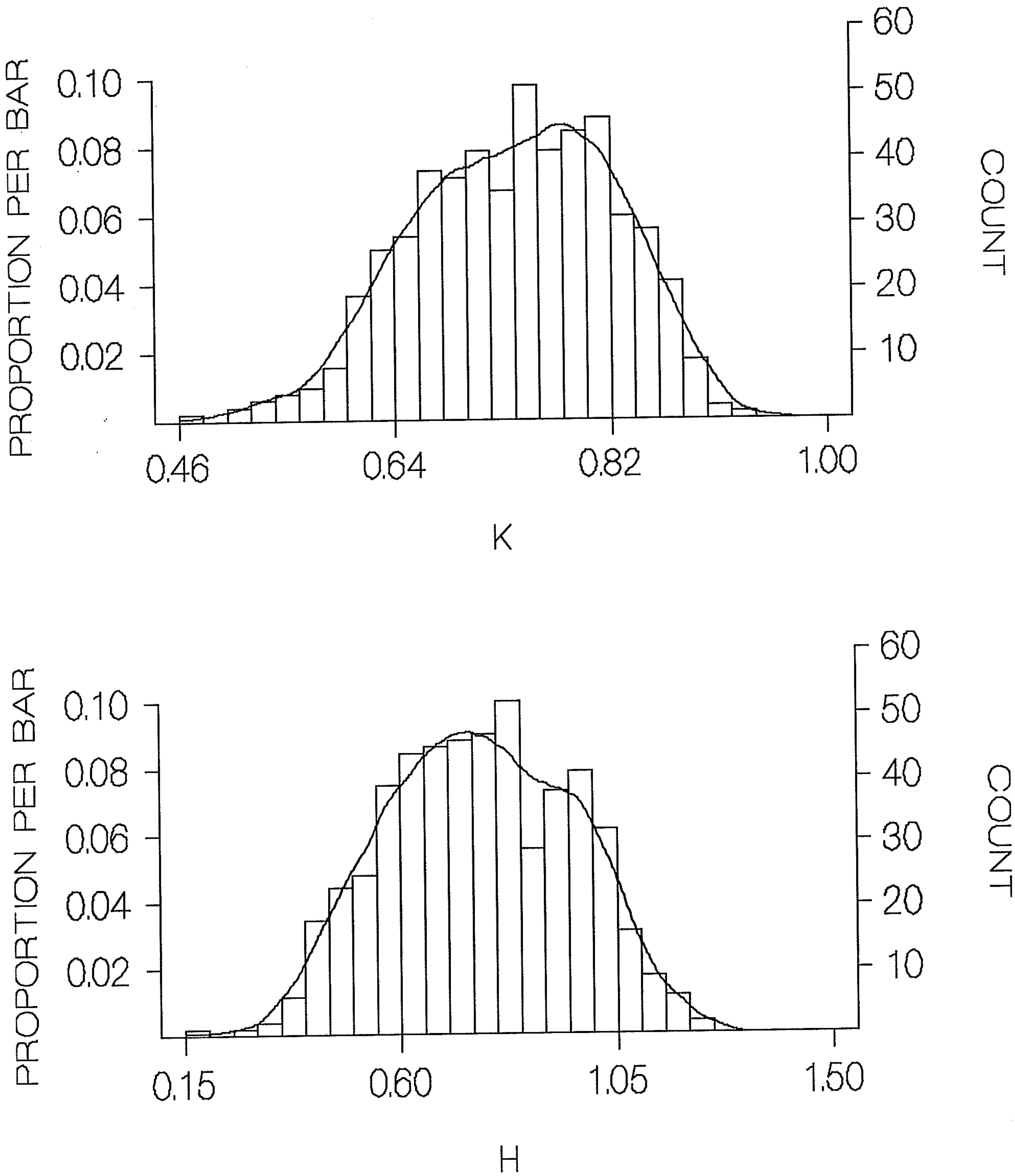
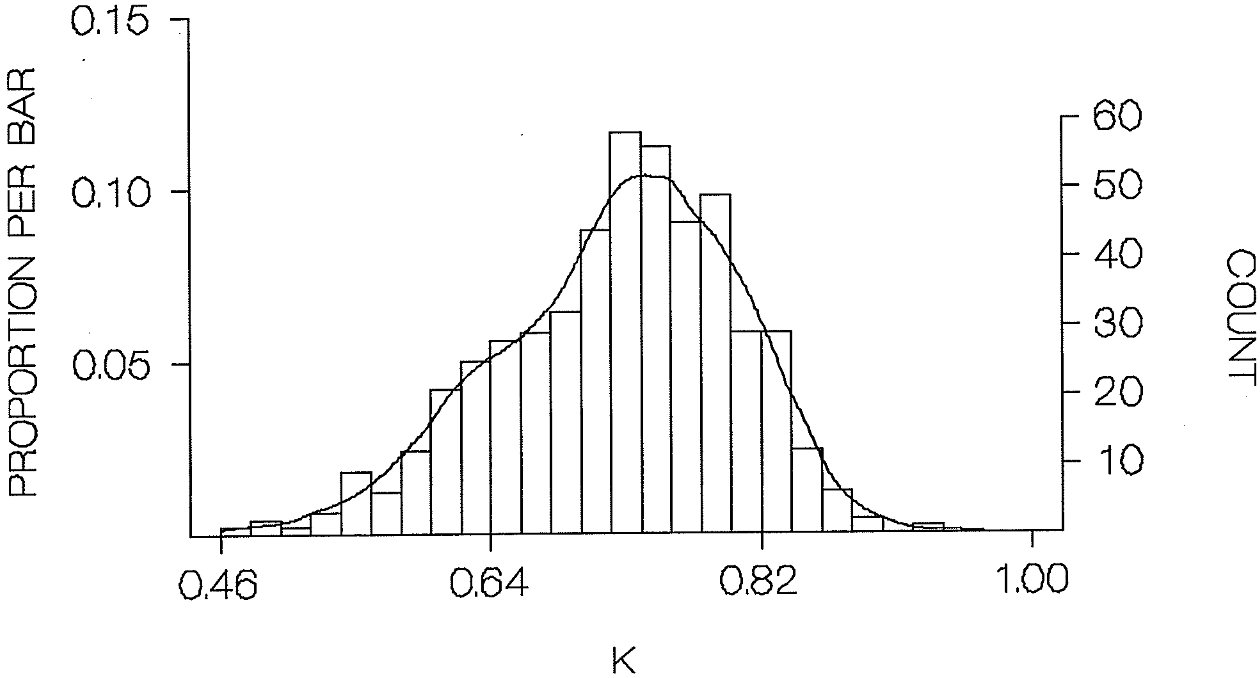


Figure 4-1E. Sample Distributions of K and H for Model Hurst 0.775



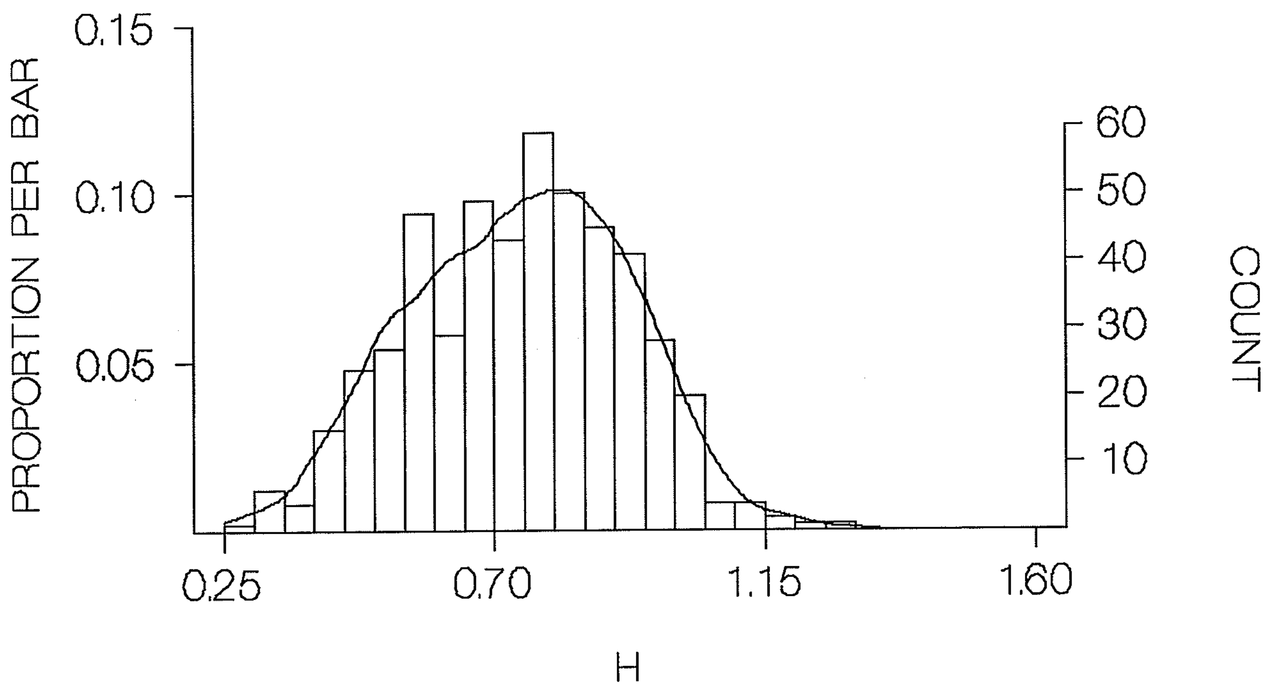
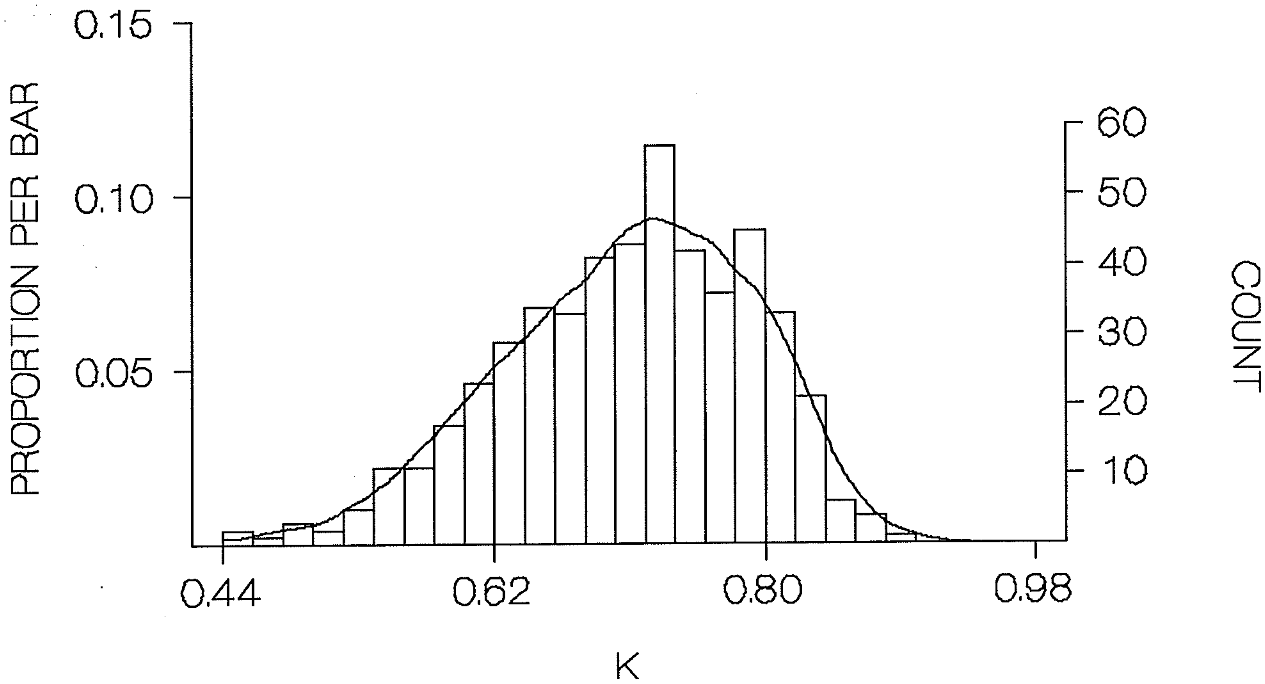


Figure 4-1F. Sample Distributions of K and H for Model Hurst 0.725



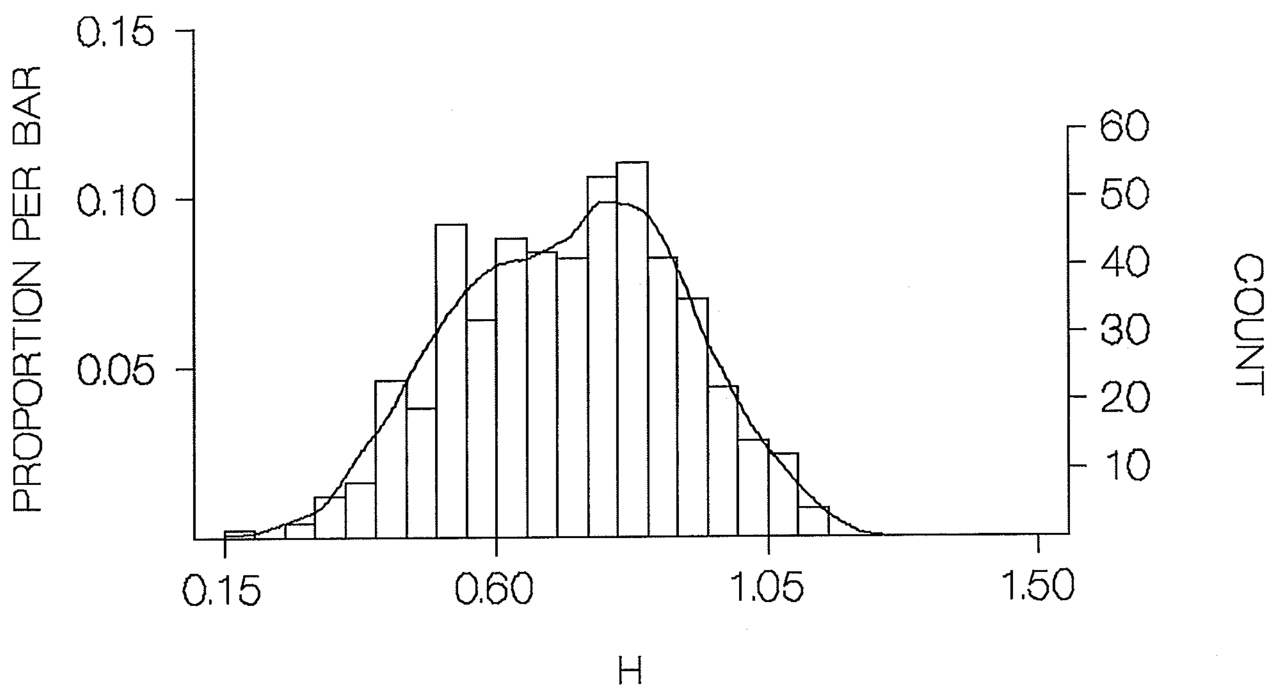


Figure 4-1G. Sample Distributions of K and H for Model Hurst 0.675

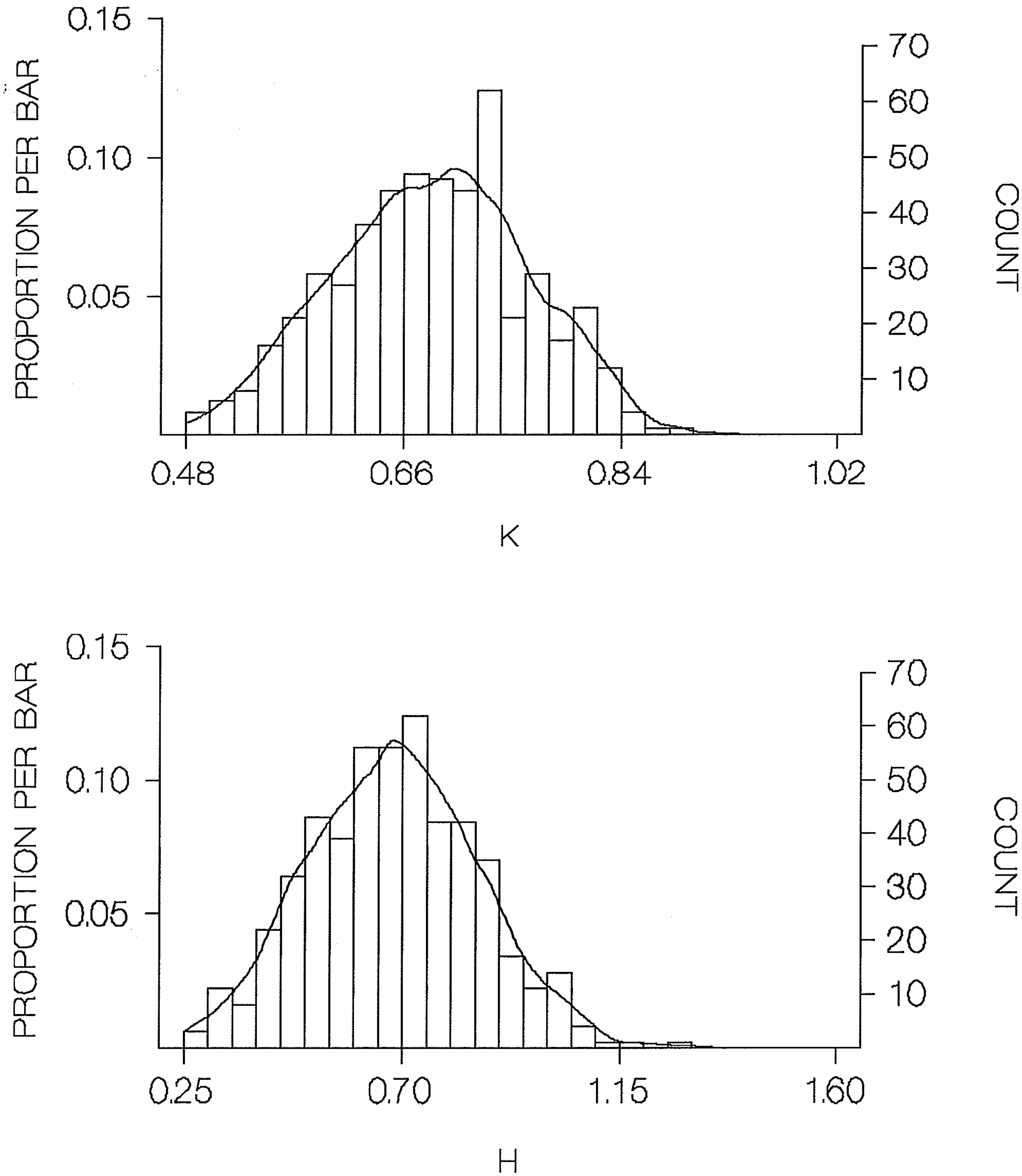


Figure 4-1H. Sample Distributions of K and H for Model Hurst 0.625

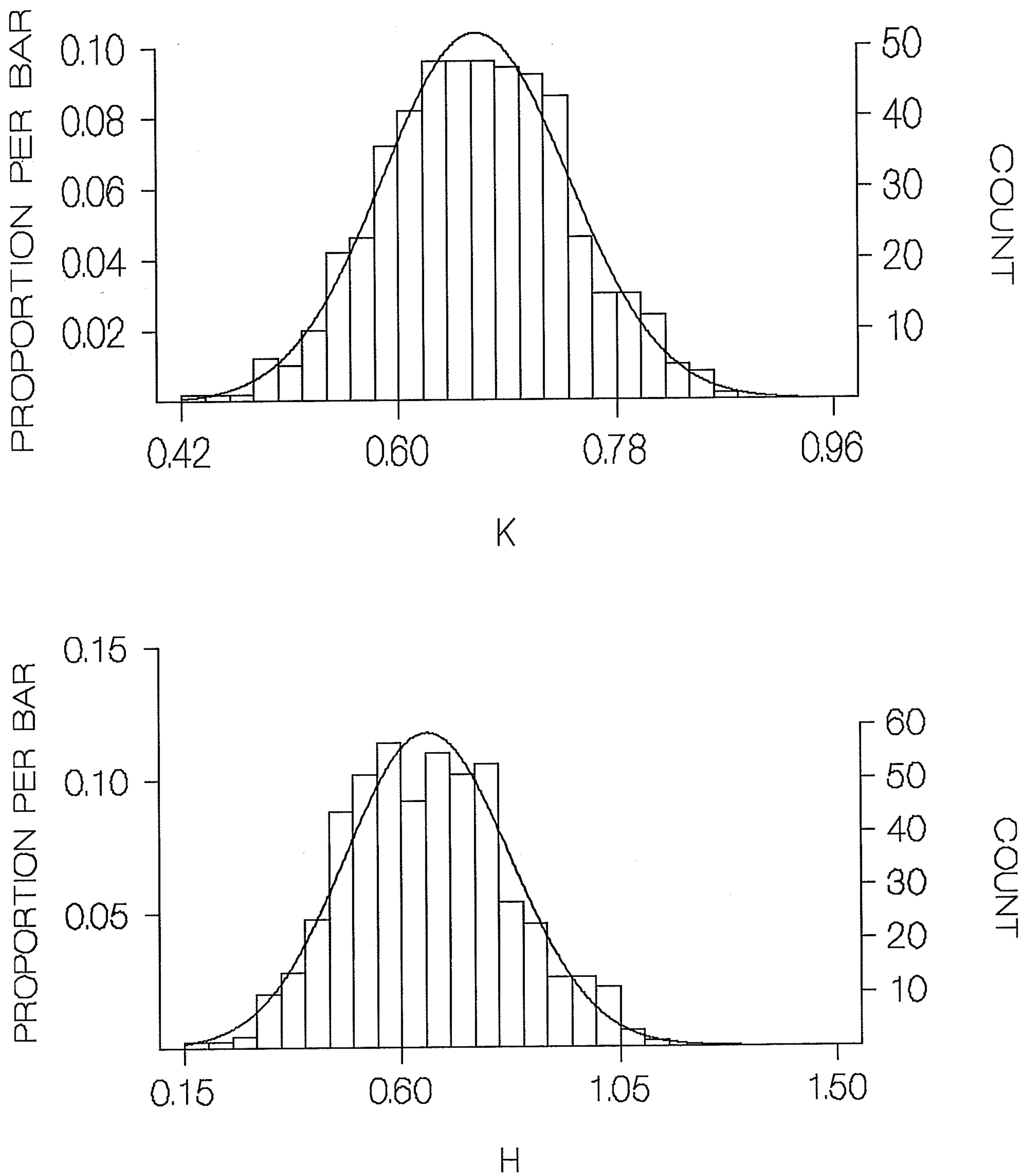
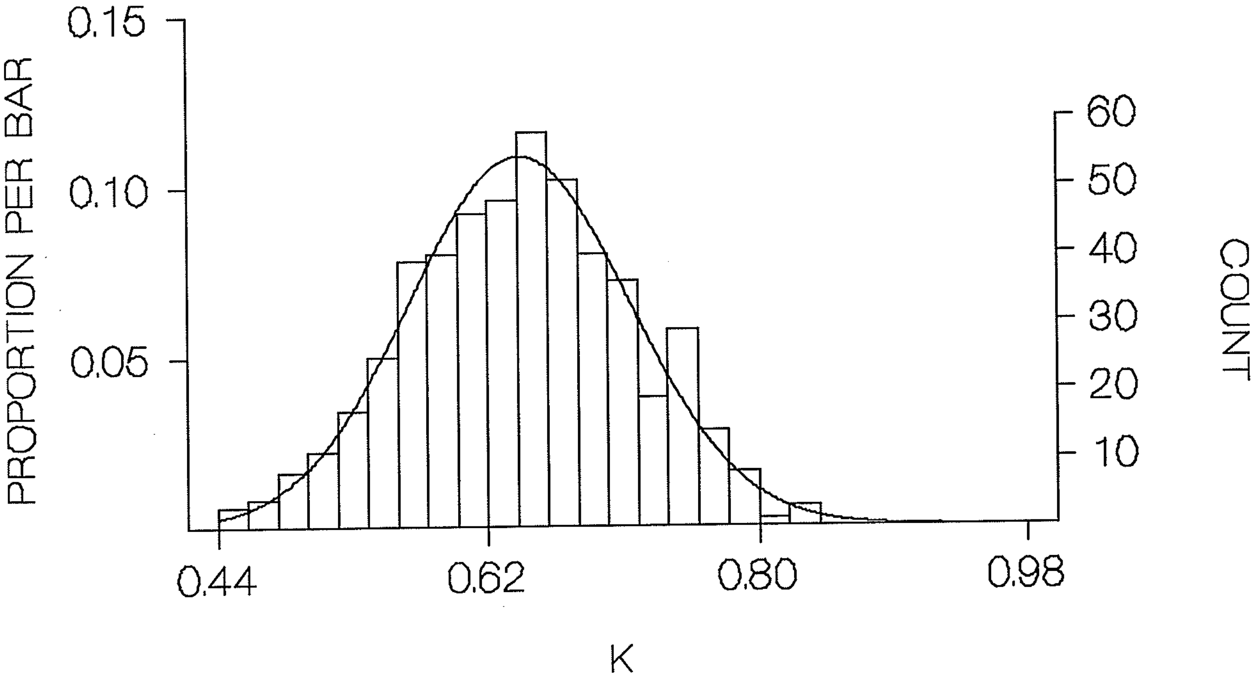


Figure 4-1I. Sample Distributions of K and H for Model Hurst 0.575



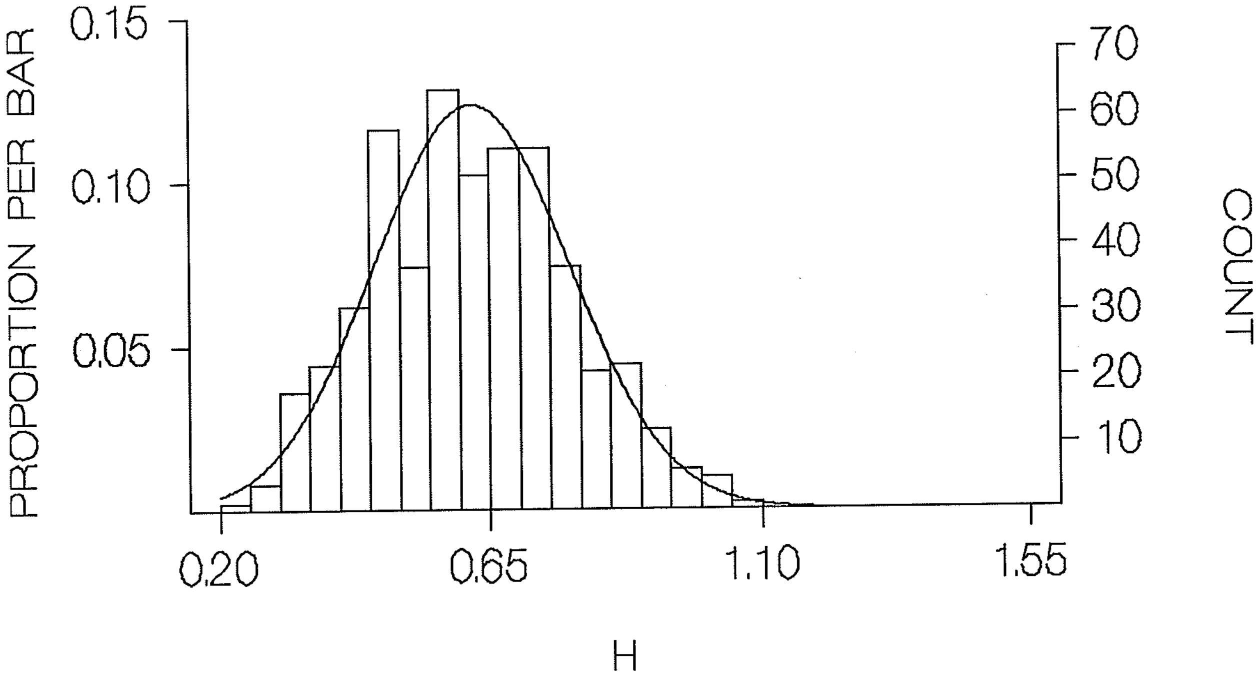


Figure 4-1J. Sample Distributions of K and H for Model Hurst 0.525

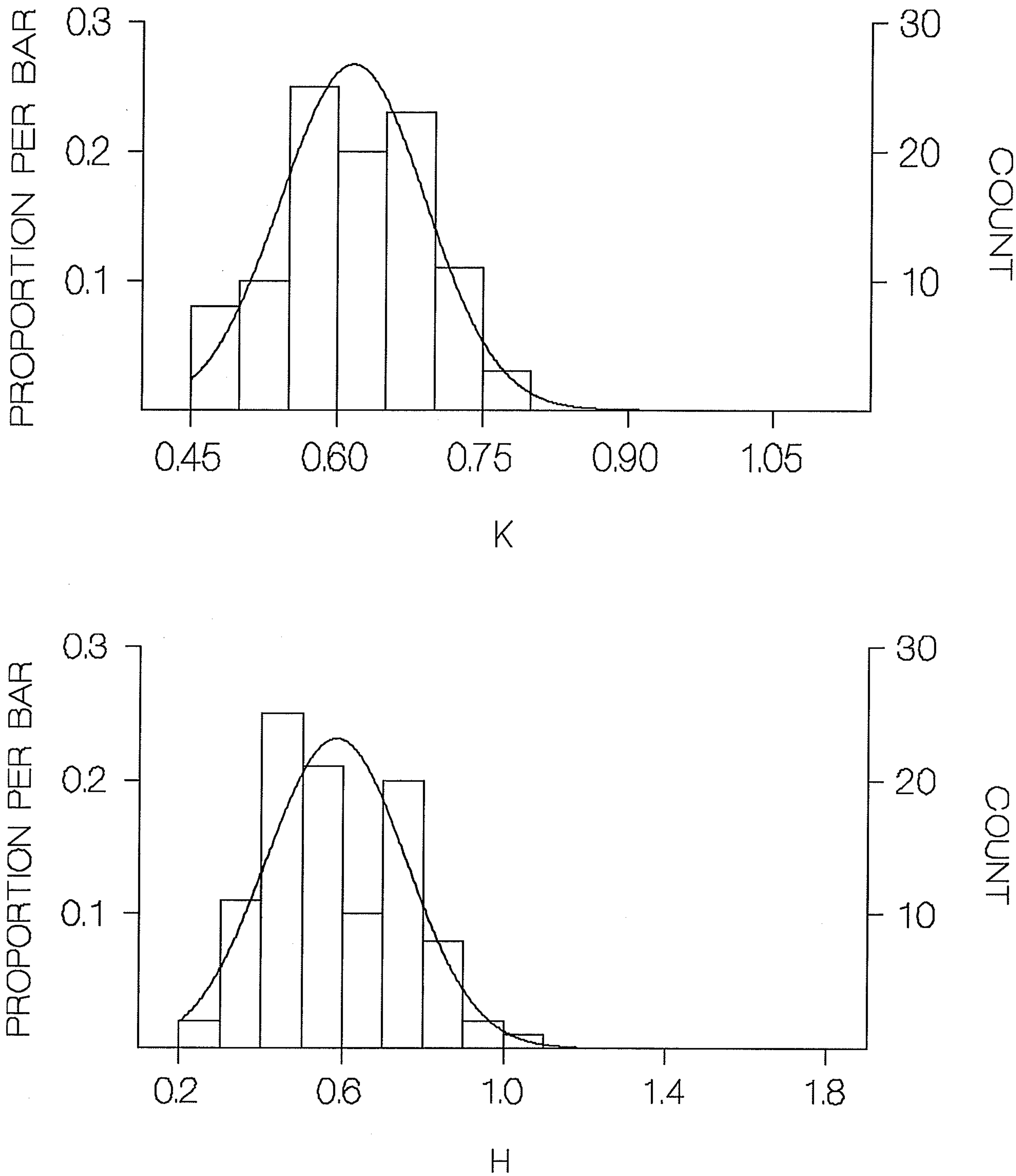
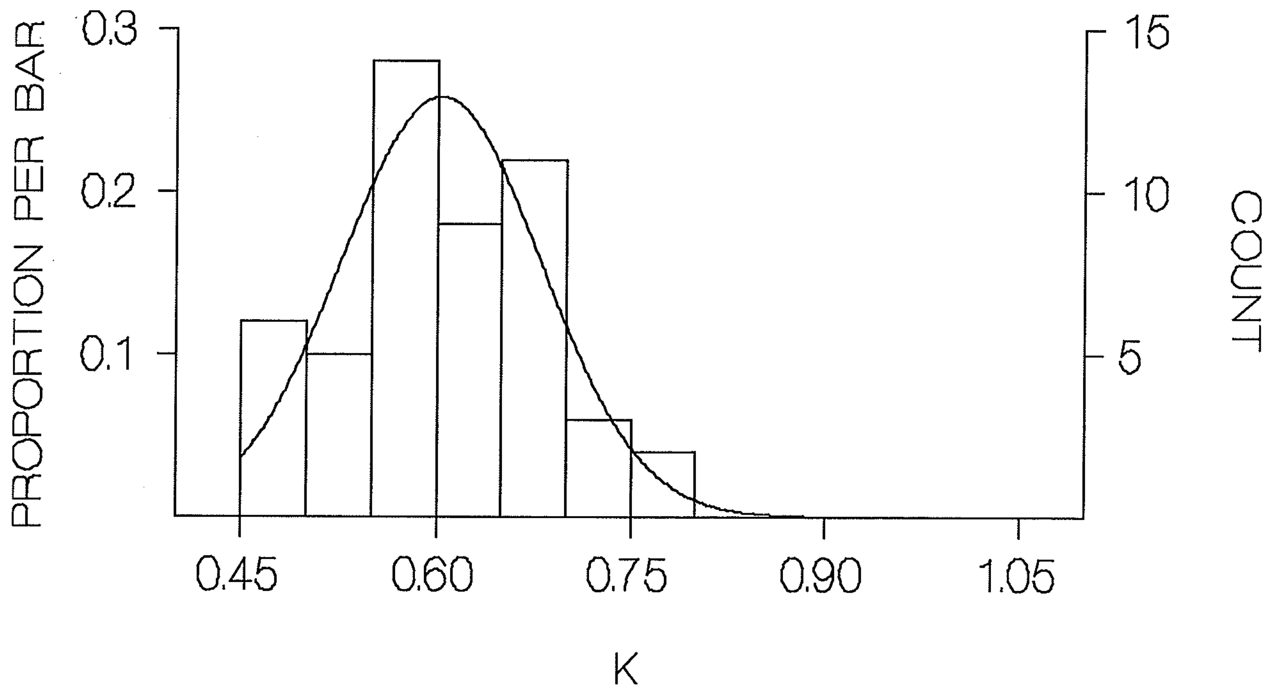


Figure 4-1K. Sample Distributions of K and H for Model Hurst 0.475



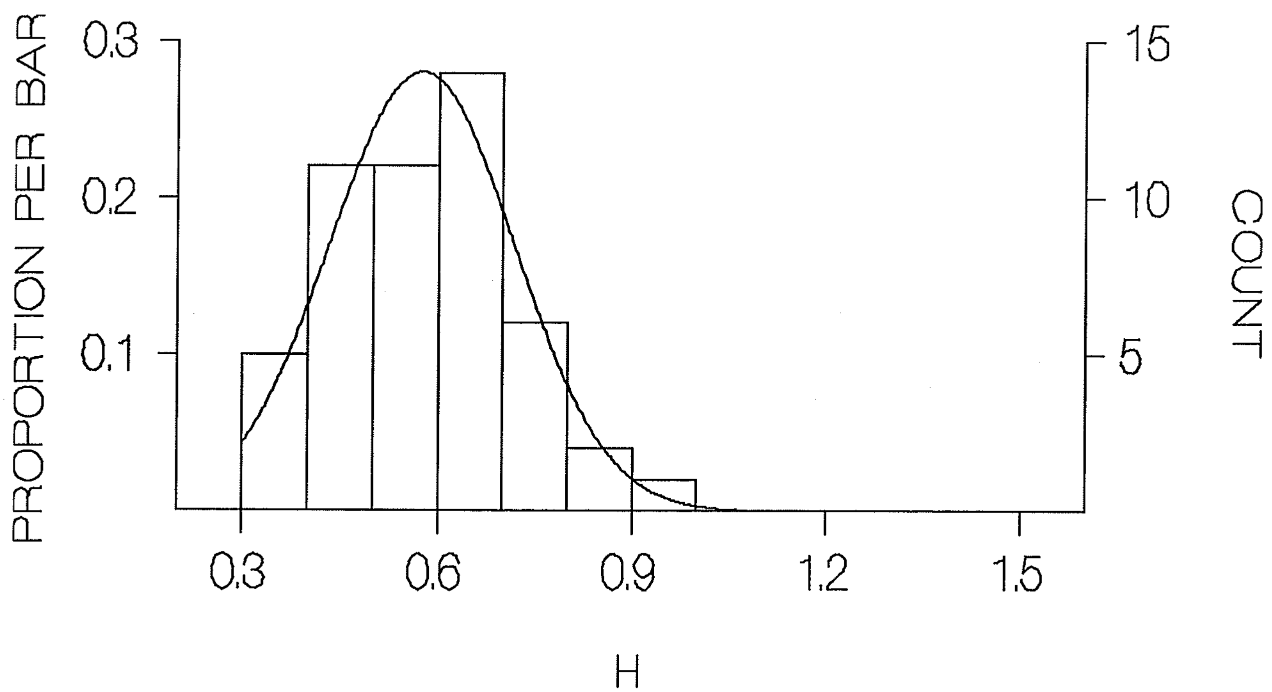


Figure 4-1L. Sample Distributions of K and H for Model Hurst 0.425

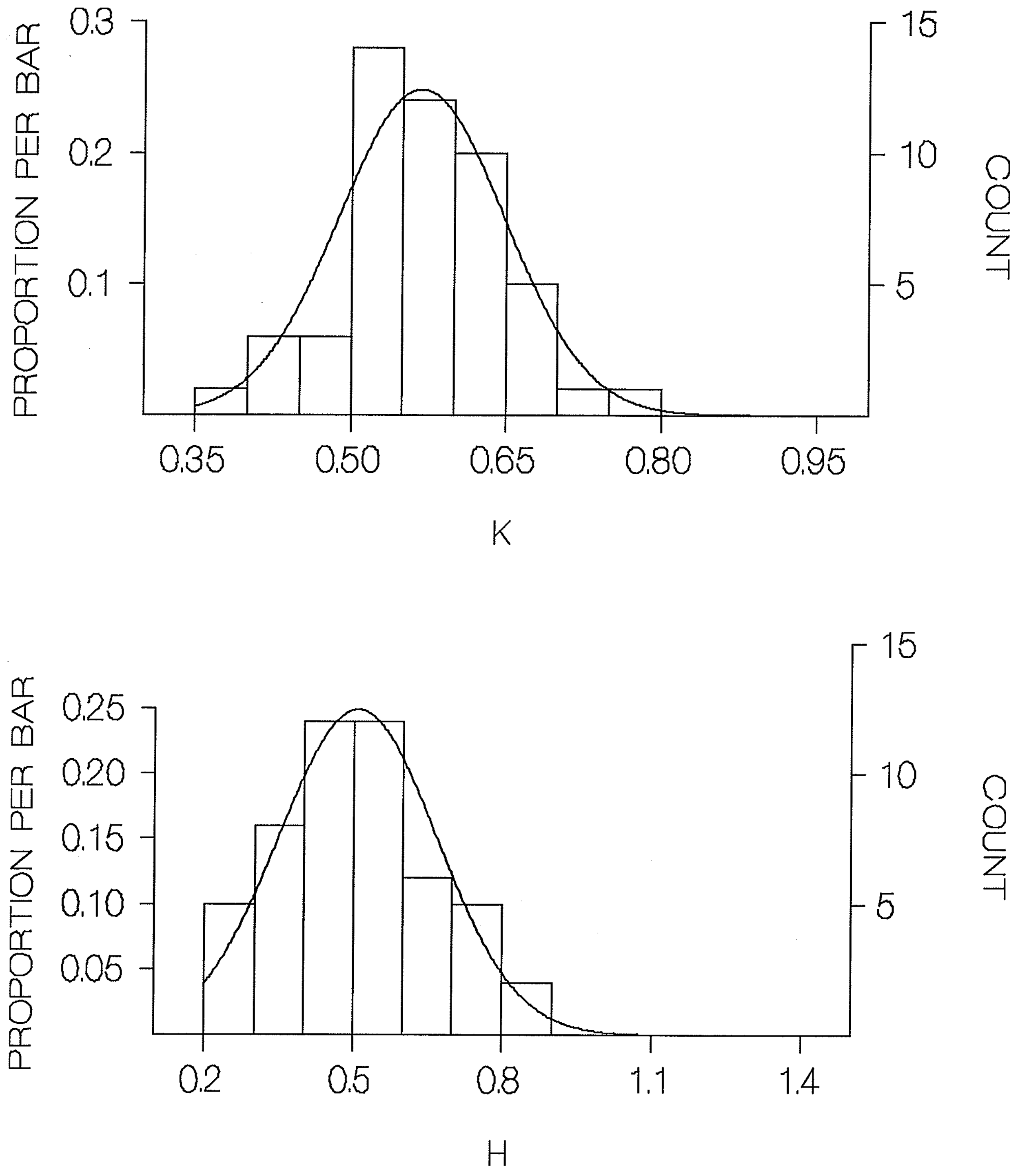
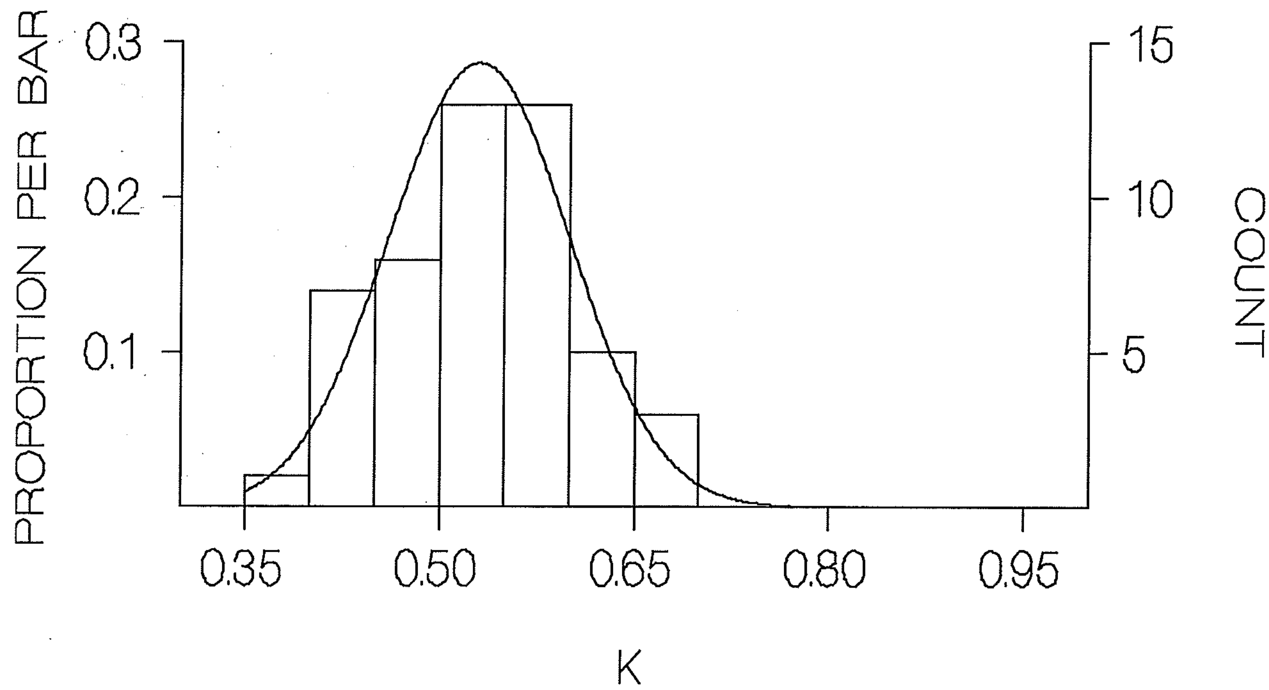


Figure 4-1M. Sample Distributions of K and H for Model Hurst 0.325



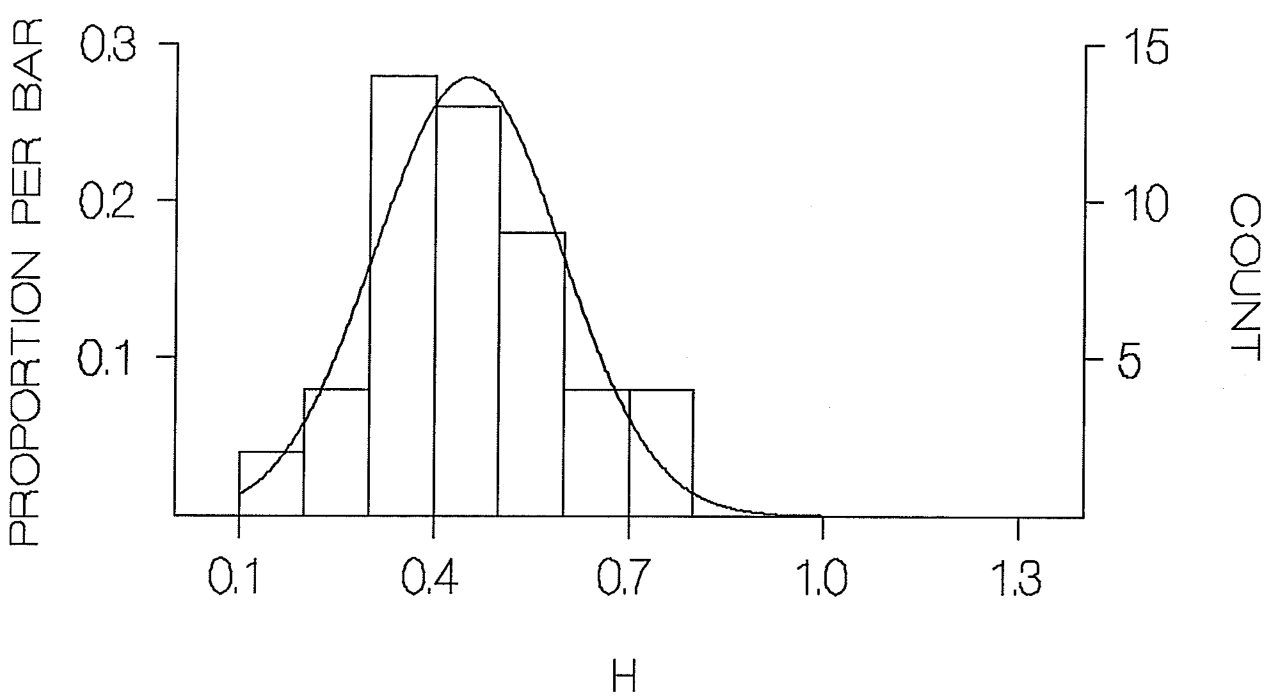


Figure 4-1N. Sample Distributions of K and H for Model Hurst 0.225

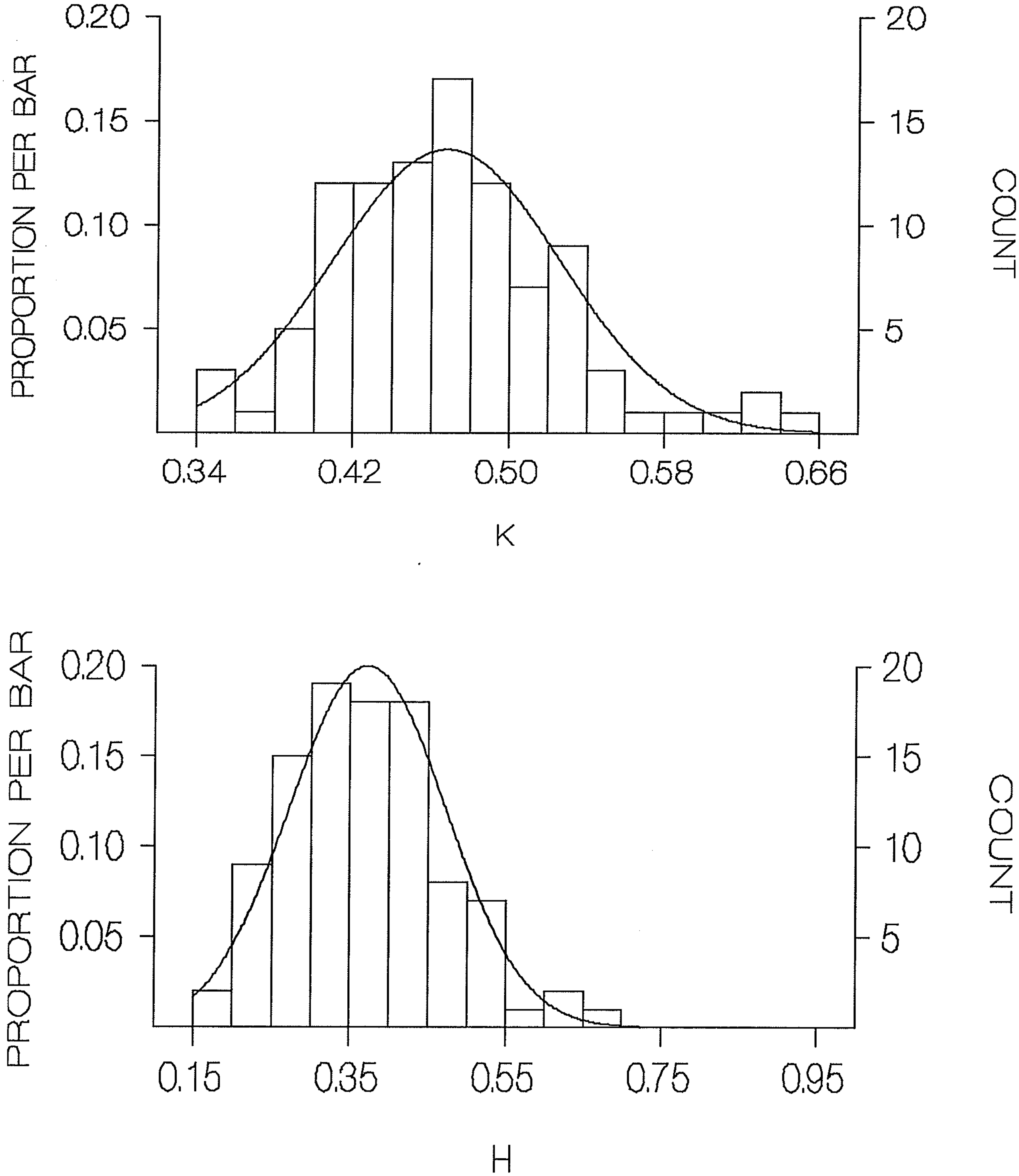


Figure 4-1O. Sample Distributions of K and H for Model Hurst 0.025

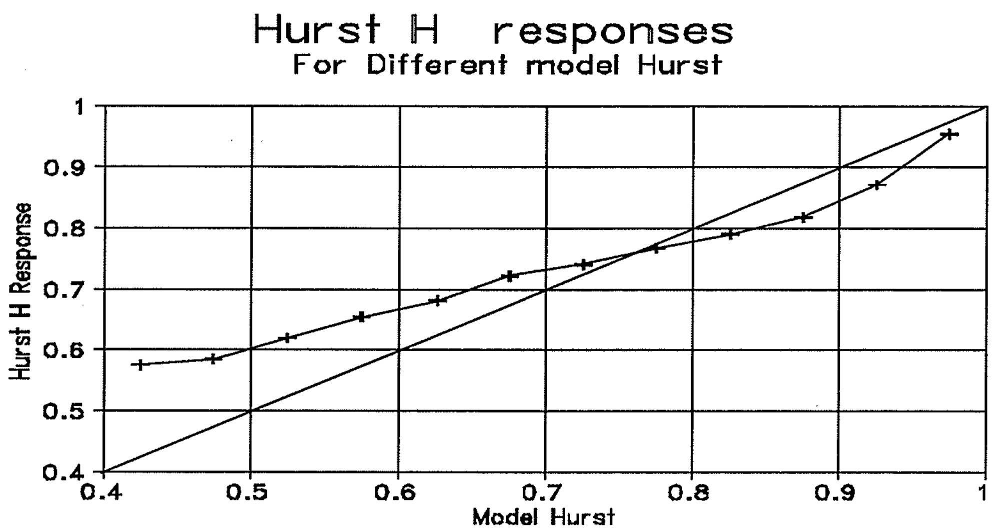


Figure 4-2. Mean Responses of Hurst H for Different Model Hursts.

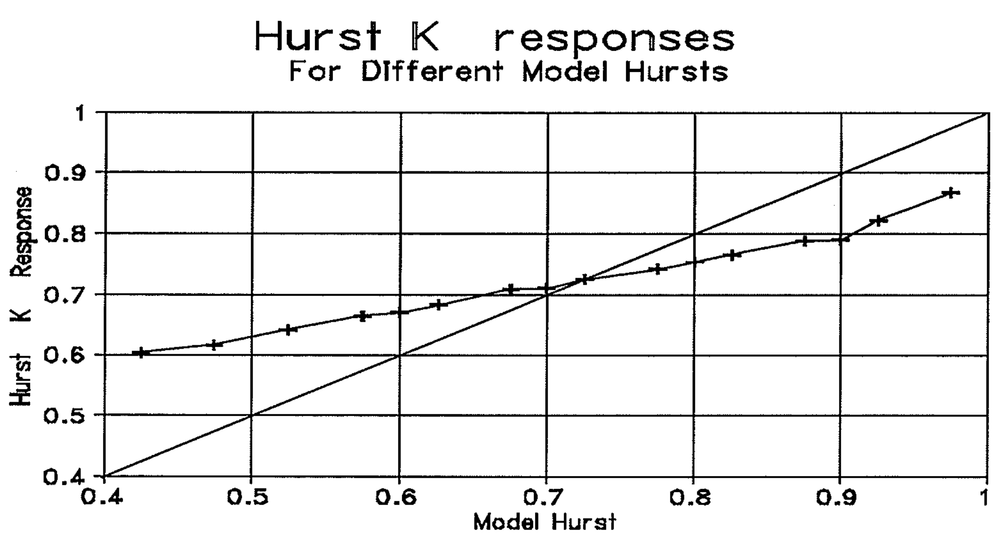


Figure 4-3. Mean Responses of Hurst K for Different Model Hursts.

4.3. BIAS CORRECTION OF THE HURST K ESTIMATOR OF h

Wallis and Matalas (1970) discussed the bias of K of small sample size N. It was found that with the increase of N, the bias decreased albeit rather slowly. Since a bias correction of K is needed for the data set of Canadian peak flows, the average sample size 43 was selected for the simulation process. After Monte Carlo simulation, the mean values of K and H for various model Hurst h versus h were plotted in Figure 4-4. To derive a bias correction formula, regression lines were chosen to approach mean response curve of K in this figure. Having tried various functions, two simple regression lines, one linear and one quadratic were chosen.

Linear Function: $K = 0.4135 * h + 0.4287$ ***(4-1)***

Quadratic Function: $K = 0.4699 + 0.2435 * h + 0.1455 * h^2$ ***(4-2)***

Since this linear regression line did not give a good fit at very low and extremely high K, the quadratic line was preferred to derive the final bias correction formula, which is

$$h = 2.6218 * \sqrt{\frac{\log(\frac{R}{s})}{\log(\frac{N}{2})} - 0.368094} - 0.8370 \qquad \text{(4-3)}$$

After the bias correction, the estimated Hurst parameter h turns out to have a variance of 0.0256 which is larger than the variance of K which was found to be 0.0049. We should mention here that a bias correction was also derived for the estimator H. But after bias correction, the estimated Hurst parameter h had a variance of 0.0675, which is about three times the variance of the parameter as estimated from the K. This is a good reason for using K instead of H for estimation of h.

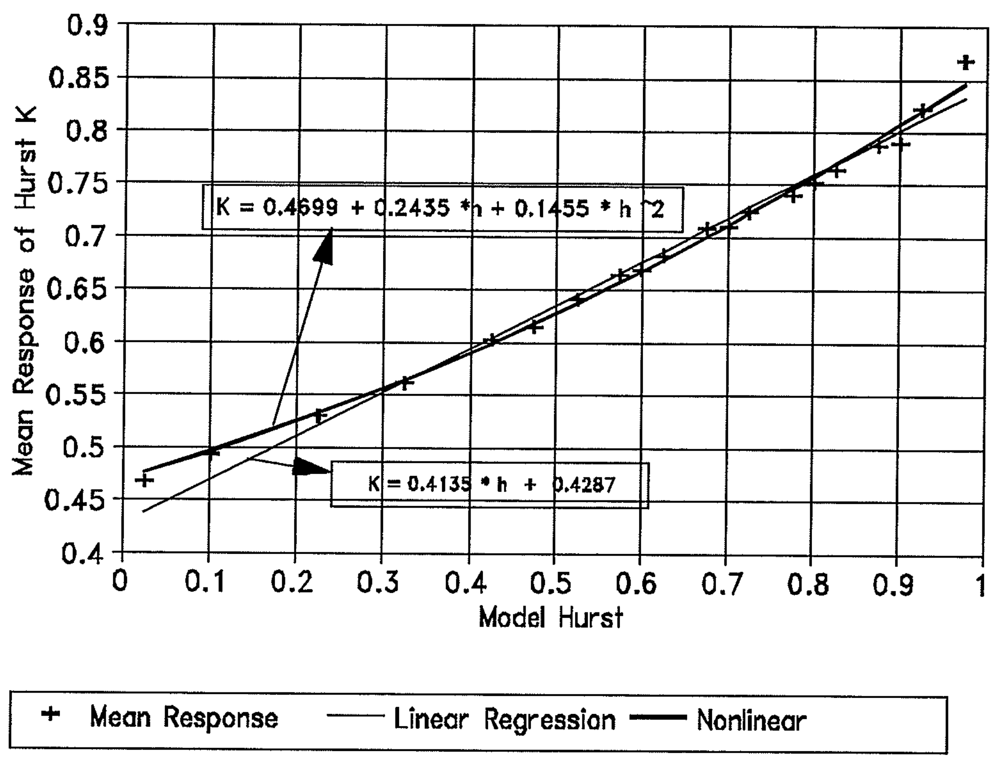


Figure 4-4. Bias Correction of the Hurst K

4.4. VERIFICATION OF THE PROPOSED BIAS CORRECTION

The above mentioned quadratic bias correction formula for K was checked by using both FFGN and MN to generate synthetic series. The results turned out to be satisfactory especially for values of h close to or at 0.60 which is an average values for the observed flow peak series. The estimates are listed in Table 4-2. They show that using K without bias correction can lead to serious misunderstanding of long term serial correlation of these rivers. Although this bias correction formula was based on simulation series of 43 year data, it gives quite good results for series from 30 to 100 years. This was verified by changing the sample size from 30 to 100, and letting model h equal to 0.6. The results are shown in Table 4-3.

Table 4-2. Verification of the Proposed Bias Correction

Model Hurst	Model Type	Length of Series	Observed h
0.85	FFGN	43	0.8271
0.825	FFGN	43	0.8042
0.775	FFGN	43	0.7610
0.725	FFGN	43	0.7170
0.70	FFGN	43	0.6961
0.675	FFGN	43	0.6813
0.625	FFGN	43	0.6241
0.6	FFGN	43	0.6017
0.575	FFGN	43	0.5815
0.525	FFGN	43	0.5279
0.85	MN	43	0.8211
0.70	MN	43	0.6860
0.60	MN	43	0.5937

Table 4-3. Verification of the Bias Correction for Different Sample Size

Model Hurst	Model Type	Length of Series	Observed h
0.60	FFGN	30	0.6045
0.60	FFGN	43	0.6017
0.60	FFGN	50	0.6018
0.60	FFGN	60	0.5901
0.60	FFGN	70	0.5949
0.60	FFGN	80	0.5828
0.60	FFGN	90	0.5771
0.60	FFGN	100	0.5930
0.60	MN	43	0.5937
0.60	MN	100	0.5850

CHAPTER 5.

BIAS CORRECTION OF THE ACF FOR LONG MEMORY MODELS

5.1. GENERAL

Chapter 2 presented the bias correction formula (equation 2-7) for the autocorrelation function (ACF) proposed by Kendall and showed that the correction is quite adequate if the series is generated by short memory models like the AR(1), the AR(2) or the Lag-one Markov model. It was shown in Chapter 2 that the Canadian rivers that were analyzed show the Hurst phenomenon to a significant degree, so that the bias correction formula mentioned in Chapter 2 may not be valid for these rivers. This was shown to be true when FFGN or MN was used to generate synthetic series, after which equation (2-7) was used to get a bias corrected ACF. Comparing the corrected ACF with theoretical ACF, it was found that this formula is indeed quite inadequate for $h > 0.5$. This shows that the Hurst statistic will affect the estimation of ACF from the sample. A bias correction procedure is needed that takes h into account.

5.2. MONTE CARLO SIMULATION USING FFGN AND MN

In this section, the fast fractional gaussian noise model (FFGN) and the mixed noise model (MN) were applied to generate synthetic series for different model Hurst values. After that, equation (2-3) was used to get estimates of autocorrelation coefficients of these series. For each model Hurst value 500 series were generated. The mean ACF of 500 series was calculated for both models and plotted. The curves were compared with the theoretical ACF of the fractional gaussian noise. The latter can be written as

$$\rho_k = 0.5[(k+1)^{2h}-2*k^{2h}+(k-1)^{2h}] \tag{5-1}$$

where h is the model Hurst and k is the number of lags. To make the MN model comparable with the fractional noise model, which does not choose ρ_1 separately, the ρ_1 of the MN was made the same as the ρ_1 of the fractional noise or chosen as an arbitrary value r_1. At high lags, the ACF of MN is practically identical with the theoretical ACF of FFGN.

Figure 5-1 to 5-5 showed some simulation results using these two models. It was found that there was no big difference of observed sample ACF between FFGN and MN except lag-one autocorrelation coefficient, which might be due to the different approximation of fractional gaussian noise by FFGN and MN. It is evident however that bias correction is needed for both. The amount of bias was calculated by subtracting the observed sample ACF from the theoretical curve. Some results from obtained different h values are shown in Figure 5-6. For the relatively low Hurst values 0.6, 0.7, the bias curves are evidently close to a straight line with a maximum value at the lag one. For relatively high values of Hurst, the bias has a maximum value near lag-five; when the lag is larger than five, with the increase of lag, the bias decreases at a constant rate; when the lag is smaller than five, with the increase of lag, the bias increases.

The influence of lag-one autocorrelation on the bias of ρ_1 was also investigated by changing the input value of R(1) in MN keeping the Hurst h the same. Some simulation results are presented in Table 5-1 which shows the bias in r_1 for different values of the model R(1) but the same Hurst h. Fortunately, this change only had a slight influence on the bias of lag-one autocorrelation; with the increase of the lag, this influence gradually became negligible. So, in the derivation of the bias correction formula, the assumption that the ACF bias only depends on long memory persistence and sample size was made.

Table 5-1 Influence of Model Lag-one Autocorrelation on Bias of ACF

Model R(1)	Model Hurst	Observed r_1	Bias of r_1
0.6545	0.9	0.3260	0.3285
0.7000	0.9	0.3824	0.3176
0.7411	0.9	0.4484	0.2927
0.4000	0.8	0.2212	0.1788
0.4500	0.8	0.2659	0.1841
0.5152	0.8	0.3383	0.1769

Since most of Canadian rivers analyzed have relatively short historical records, in the simulation, the sample size was chosen as 43 years which is identical with the average length of 209 rivers.

5.3. BIAS CORRECTION OF THE ACF

Having tried various functions to approach the bias curves, the following were chosen to correct the ACF bias :

For high lag ($k \geq 6$):

$$R_k = r_k - 0.01286k(h-0.5) + 3.35438[h-0.5]^2 + 0.002224 \tag{5-2}$$

where k is the lag;

R is the bias corrected autocorrelation coefficient;

r is the sample estimate of autocorrelation coefficient;

h is the model Hurst.

In addition to the above function, it was recommended that for h<0.65, the following function be used:

$$R_k = r_k + (h-0.5)(0.67436 - 0.01436 * k) \tag{5-3}$$

For low lags ($k \leq 5$) :

$$R_1 = r_1 + 0.001512275e^{5.878188h} \tag{5-4}$$

$$R_2 = r_2 + 0.02 + 0.2615[3^{2h} - 2 * 2^{2h} + 1] \tag{5-5}$$

$$R_3 = r_3 + 0.02 + 0.3370[4^{2h} - 2 * 3^{2h} + 2^{2h}] \tag{5-6}$$

$$R_4 = r_4 + 0.02 + 0.3916[5^{2h} - 2 * 4^{2h} + 3^{2h}] \tag{5-7}$$

$$R_5 = r_5 + 0.02 + 0.4348[6^{2h} - 2 * 5^{2h} + 4^{2h}] \tag{5-8}$$

The bias correction formulas mentioned above were verified. Some results are presented in Figures 5-7, 5-8, 5-9, 5-10, 5-11. It is evident that after bias correction the estimated ACF is very close to the theoretical curve. For the lag-one autocorrelation coefficient, the maximum error is 0.05. After the bias correction, with few exceptions all autocorrelation coefficients tended to have positive values. This makes sense since one would believe that all autocorrelation coefficients of the historical records should be positive. Negative values are likely due to a systematic estimating bias or to parameter variability. It will be shown in Chapter 7 that this bias correction also gives very good results for historical records.

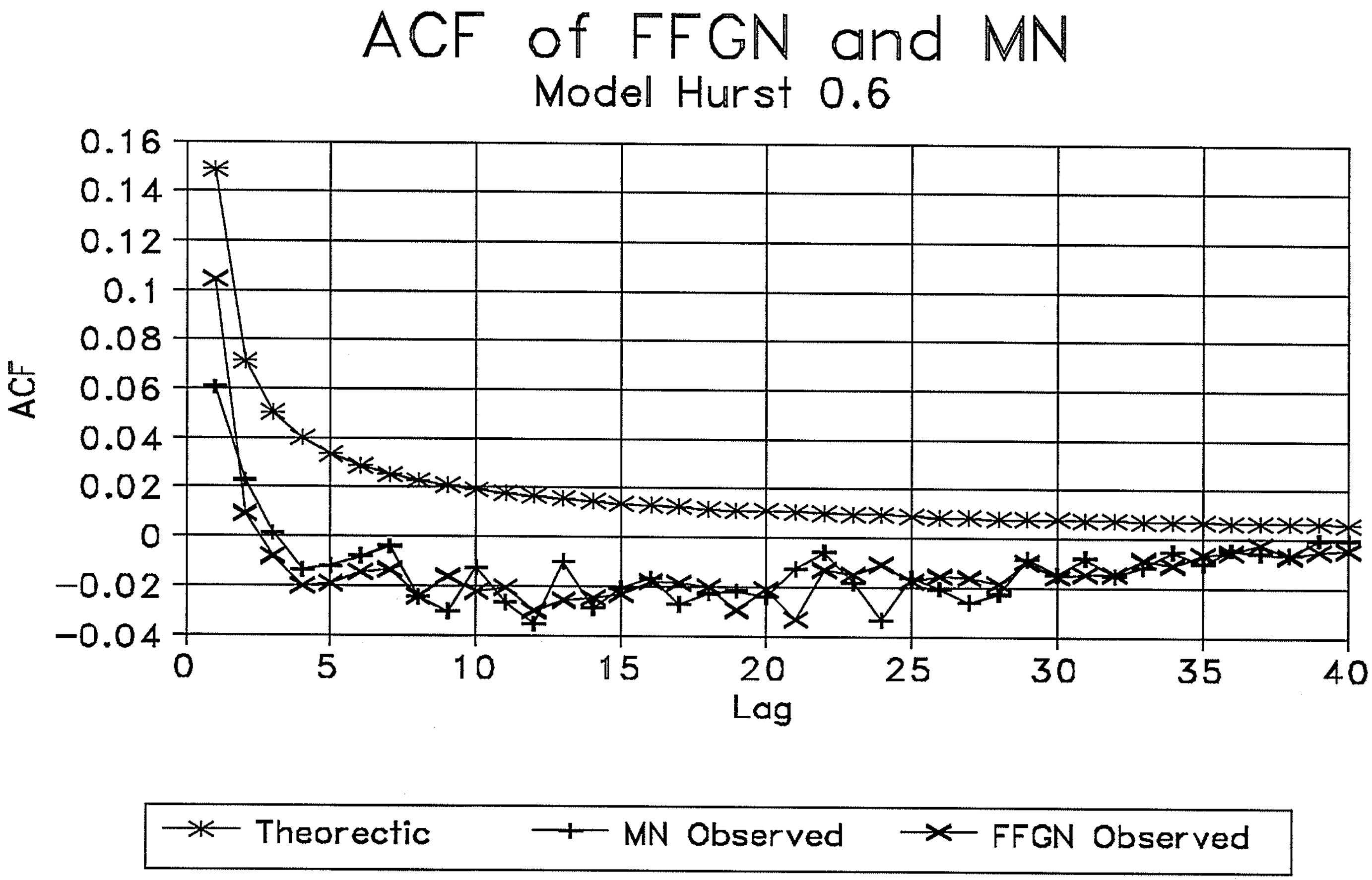


Figure 5-1. ACF of FFGN and MN h=0.6

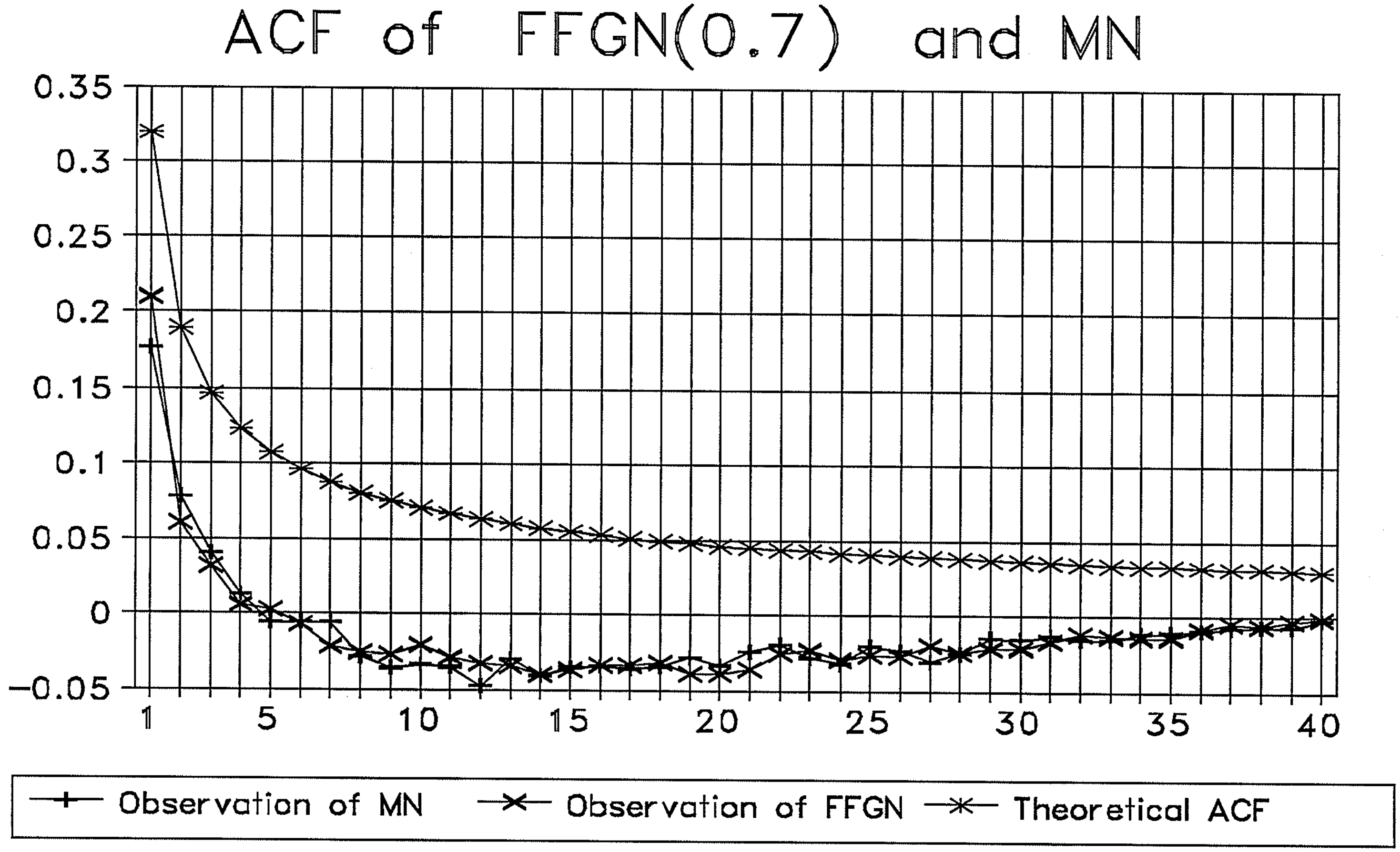


Figure 5-2. ACF of FFGN and MN h=0.7

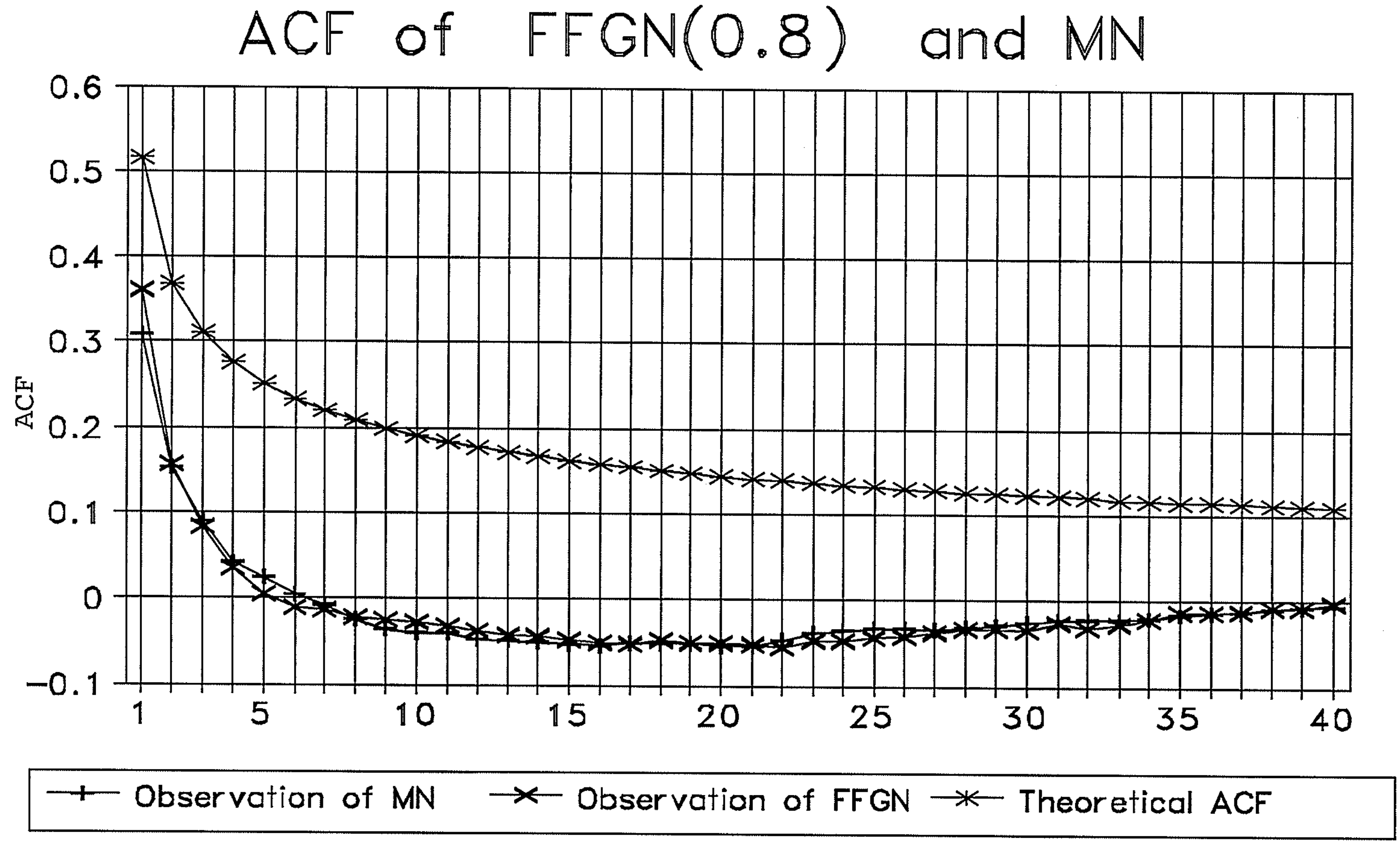


Figure 5.3. ACF of FFGN and MN h=0.8

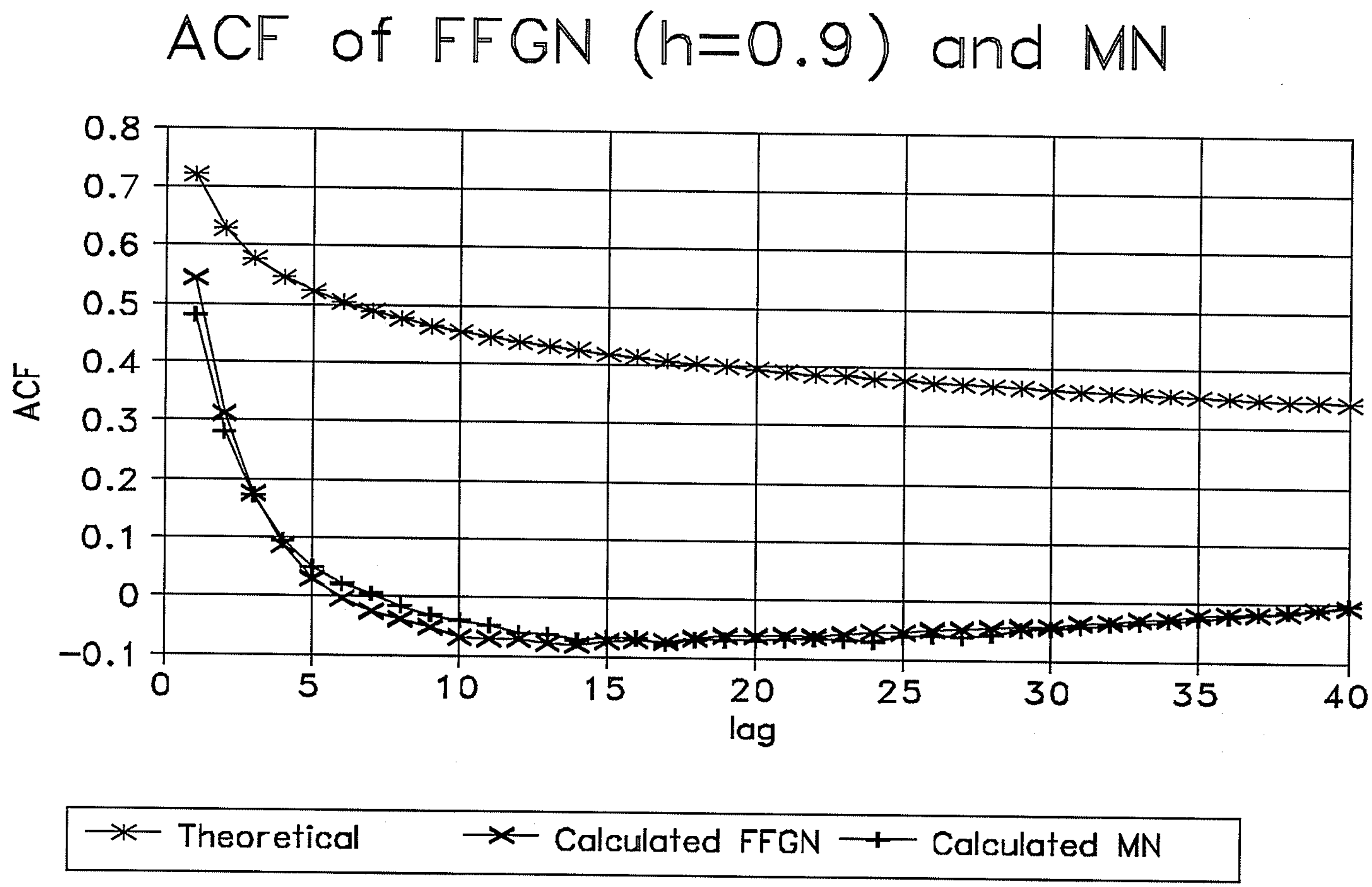


Figure 5.4. ACF of FFGN and MN h=0.9

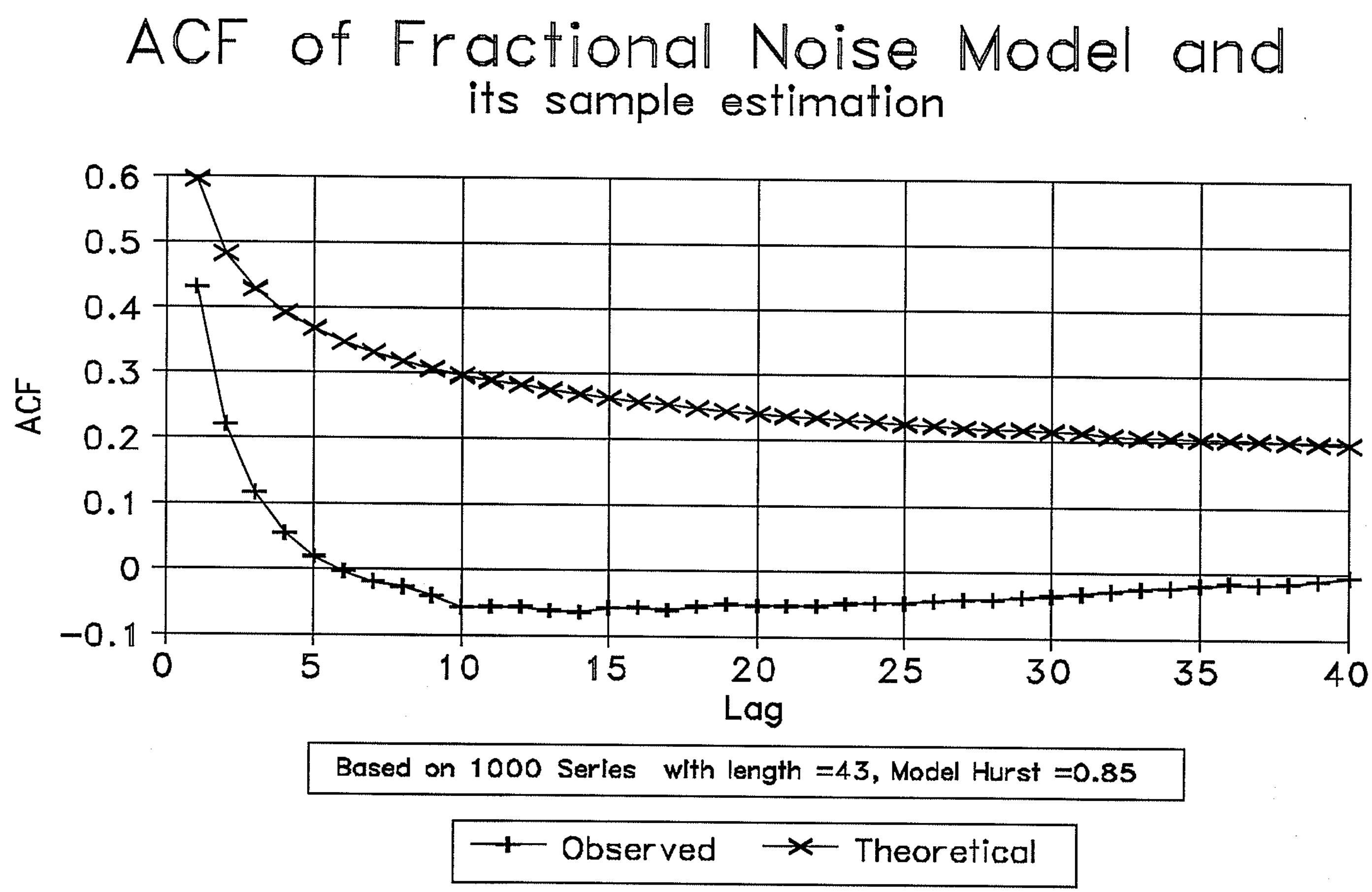


Figure 5.5. ACF of FFGN h=0.85

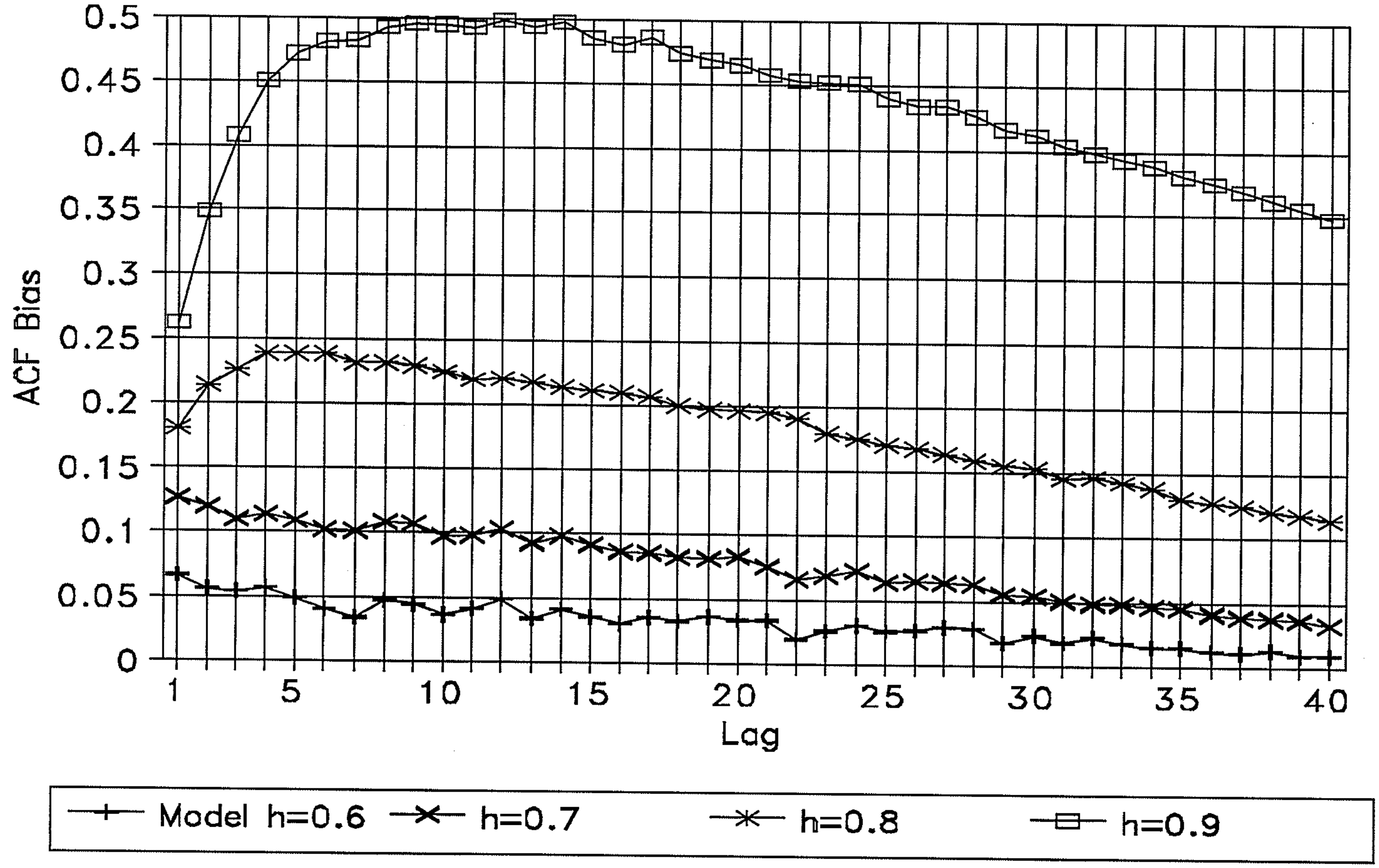


Figure 5.6. ACF Bias of FFGN and MN for Various h

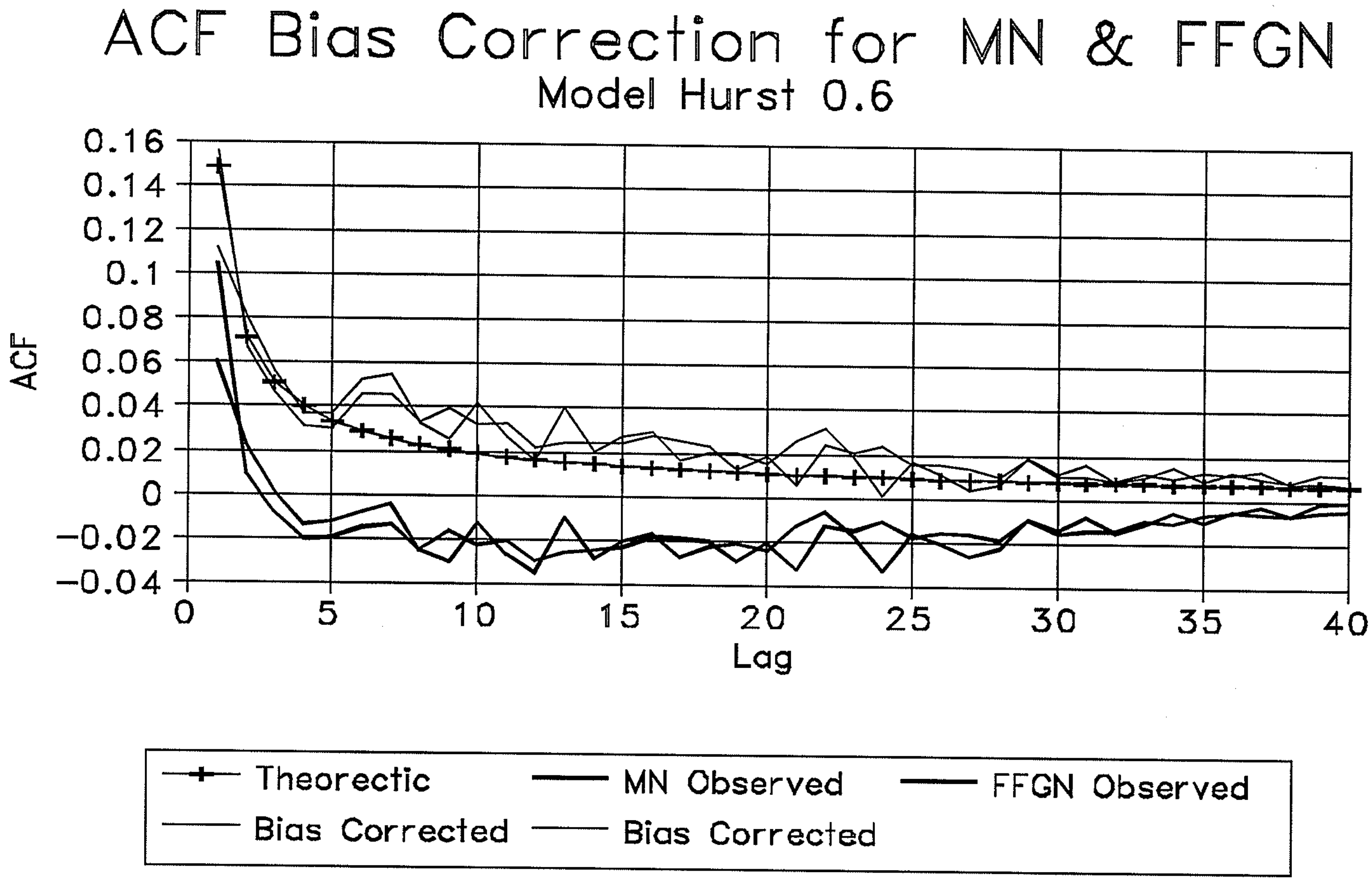


Figure 5-7. ACF Bias Correction for h=0.6.

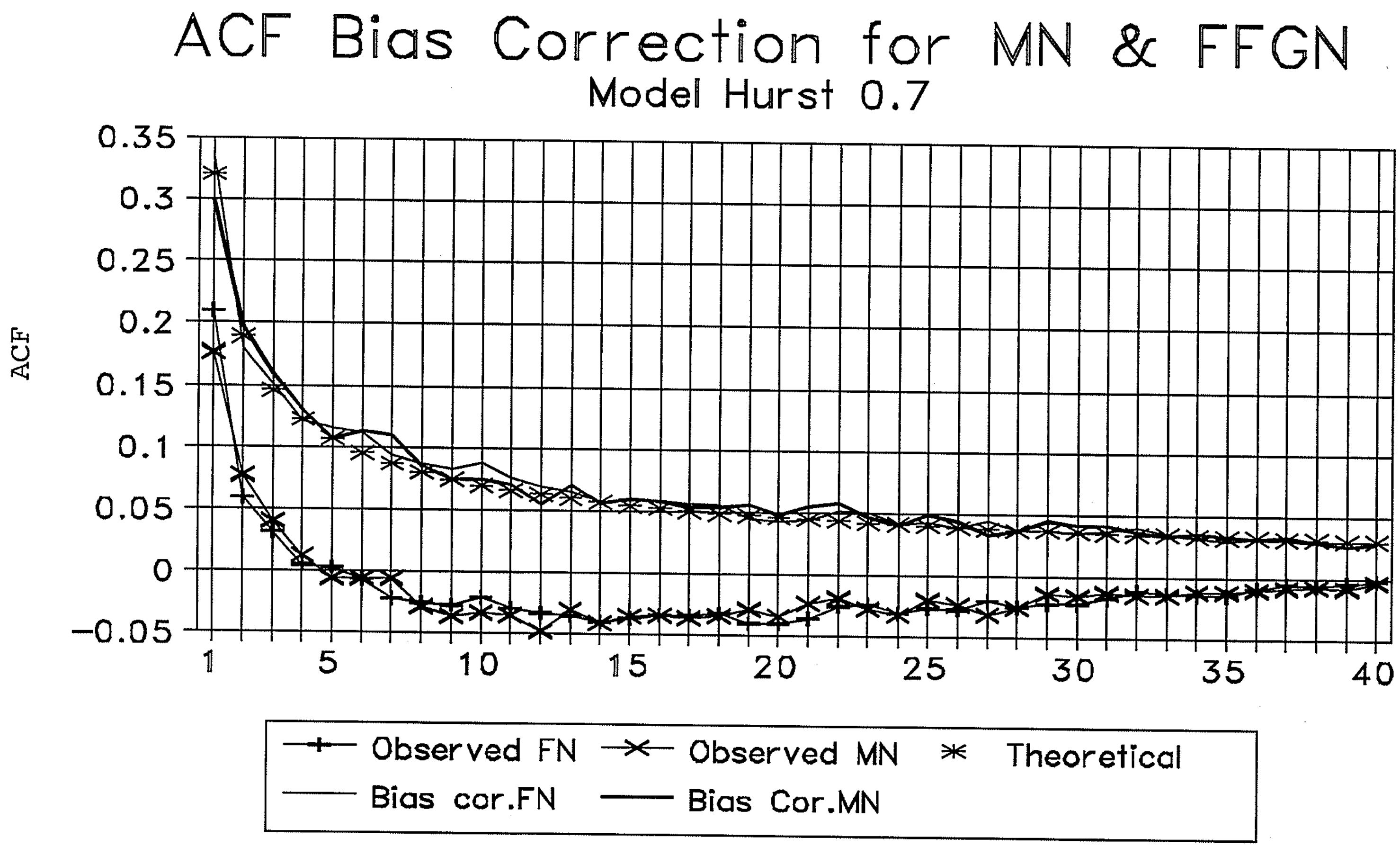


Figure 5-8. ACF Bias Correction for h=0.7.

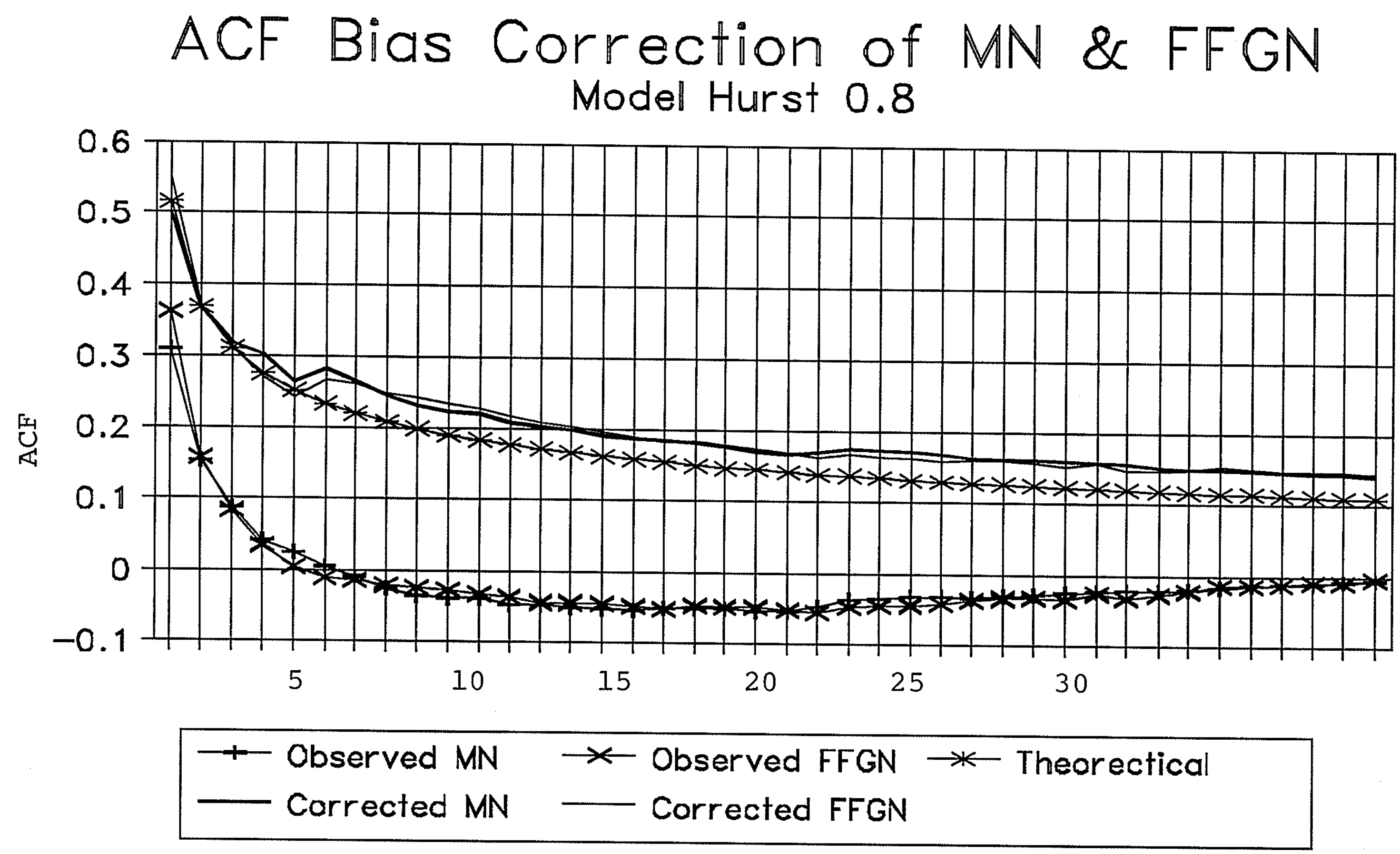


Figure 5-9. ACF Bias Correction for h=0.8.

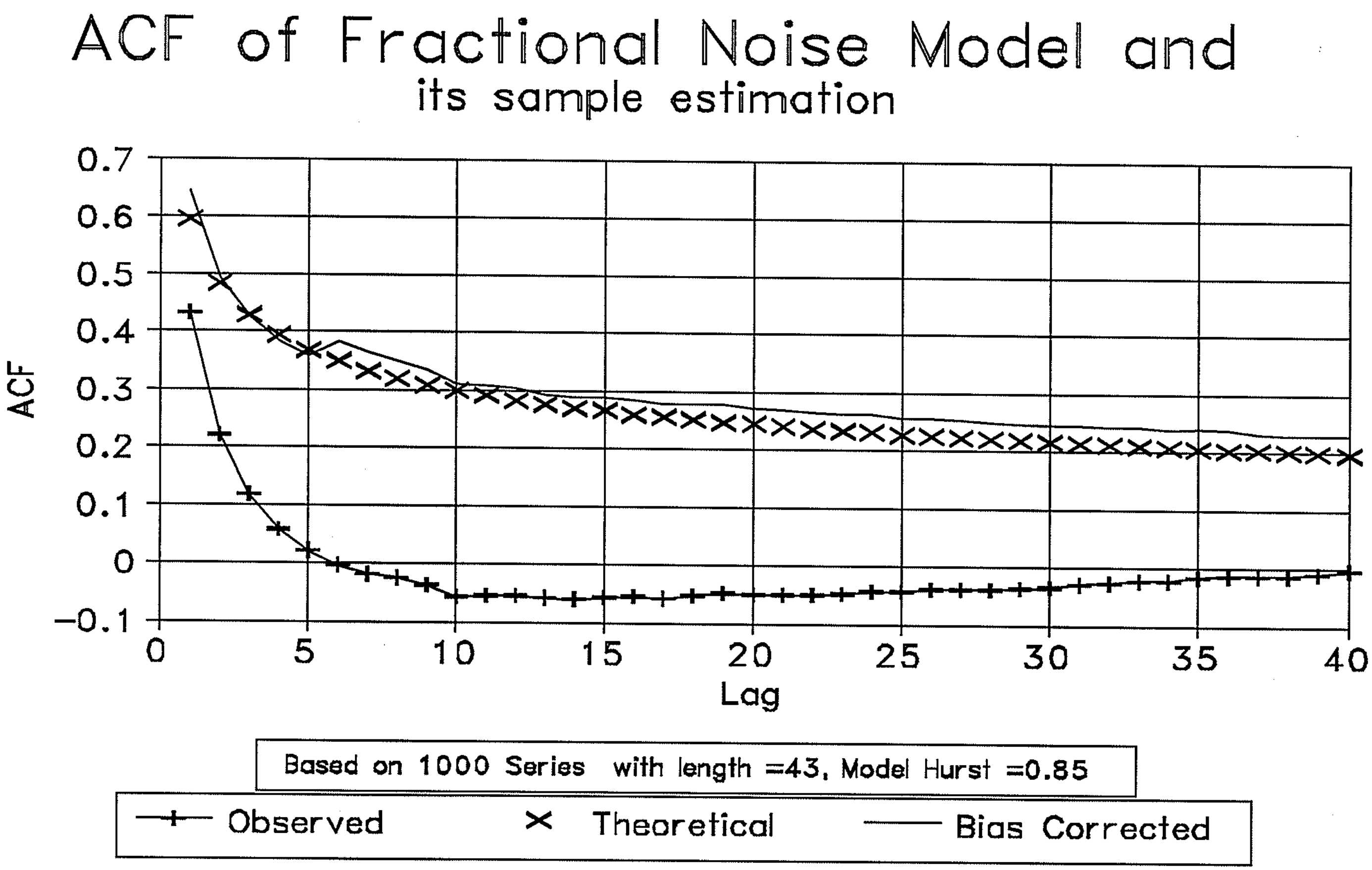


Figure 5-10. ACF Bias Correction for h=0.85

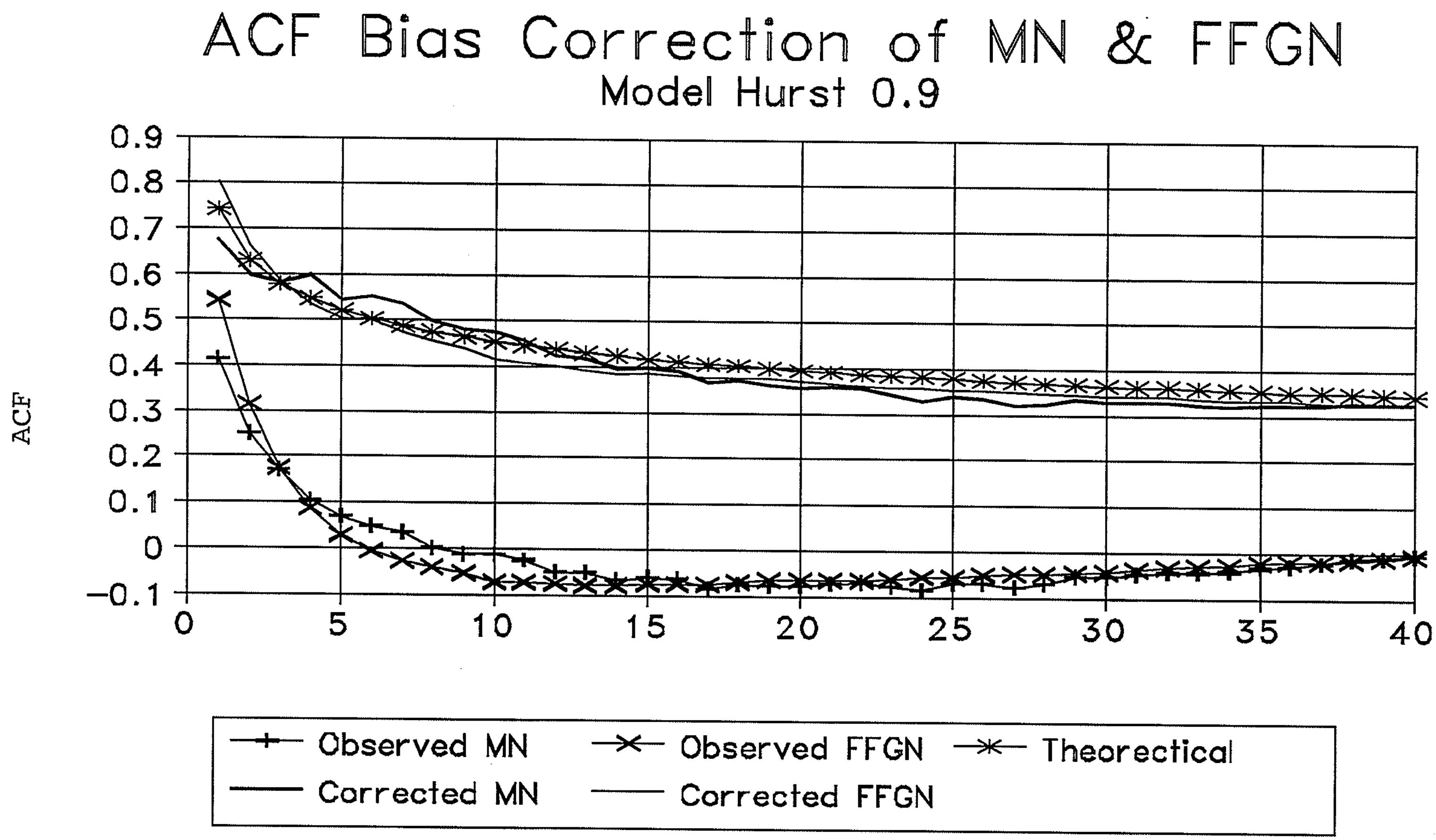


Figure 5-11. ACF Bias Correction for h=0.9

CHAPTER 6.

BAYESIAN UPDATING OF THE HURST STATISTIC

6.1. GENERAL

In the previous chapter, we have discussed simulation models that can preserve the appropriate long term correlation structure and we also completed the work necessary to obtain unbiased estimates of the parameters of these simulation models from the corresponding sample statistics. The elimination of the bias, however, does not reduce the uncertainty in the parameter estimation. This is especially a problem in the estimation of the Hurst parameter, which is quite variable for the typical sample length of the peak flow series under investigation. The difference in the observed values of K for the different drainage basins of the basic data set reflects the sample errors as well as the varying precipitation-runoff characteristics of the basins. One would expect, therefore, that the set of true values of K has less variance than the set of observed values of K. Similarly, the distribution of the corresponding Hurst parameters h can be expected to be more concentrated for the true value than for the "observed" values obtained from the previously discussed bias corrections. This, in turn, implies that an h value that corresponds to a relatively large observed K, is likely over-estimated, while a small observed K is likely to produce an under-estimated value of h.

To obtain estimates of h that reflect this tendency, we note there are two sources of information about h for a particular basin. Firstly, we have the general information of the values of K one can expect for rivers similar to those included in the data set. Secondly, we have a sample of a particular river peak flow series from which K can be estimated. Each of the two sources of information gives probabilistic information about h. This means that Bayes' theorem can be used to combine the information from the two sources so as to reduce the uncertainty in h.

In this chapter, the principle of Bayesian updating from a continuous prior distribution will be first discussed. Next, the procedure when the information is discretized, which is usually the case for practical computer applications, will be described. Finally, two examples of a Bayesian updating of the Hurst h for rivers belonging to the data set will be showed.

6.2. BAYES' THEOREM

This theorem is named after Reverend Thomas Bayes (1702-1761). It can be written in the form of

$$p(\alpha/y)=\frac{p(y/\alpha)\ \ p(\alpha)}{p(y)} \qquad \textbf{(6-1)}$$

where $p(\alpha/y)$ represents the conditional probability of α given the occurrence of event y;

$p(y/\alpha)$ represents the conditional probability of y given the occurrence of event α;

$p(\alpha)$ represents the probability that α occurs;

$p(y)$ represents the probability that y occurs.

Alternatively, this can be written as

$$p(\alpha/y) = N\, p(y/\alpha)\ p(\alpha) \qquad \textbf{(6-2)}$$

where N is called normalizing factor. The theorem can be used to update the estimate of the probability that α occurs when we know that y has occurred ($P(\alpha/y)$), provided that we know the prior probability of α ($P(\alpha)$) and the likelihood function of α ($P(y/\alpha)$). It can also be expressed as

$$p(\alpha/y) \propto l(\alpha)\ p(\alpha) \qquad \textbf{(6-3)}$$

where $l(\alpha)$ represents the likelihood of α. If we call $p(\alpha/y)$ the posterior probability, then

$$\textit{Posterior} \propto \textit{likelihood} * \textit{Prior} \qquad \textbf{(6-4)}$$

This theorem can be used to revise probability estimates in the light of additional evidence.

6.3. BAYESIAN PARAMETER ESTIMATION

When Bayes formula is used to update a parameter estimate then the magnitude of the parameter is subject to a probability assessment and additional information is used to obtain revised probability statements about the parameter which reflects a gain in knowledge or a reduction in the uncertainty about the magnitude. This makes it possible to sequentially use newly available information to update the "prior" probability distribution of the parameter. For a normal prior distribution and likelihood function, this procedure can be stated as follows:

If h is assumed to be normally distributed with mean μ and variance σ^2:

$$h \sim N(\mu,\ \sigma^2) \qquad \textbf{(6-5)}$$

This means

$$f(h) = (2\pi\sigma^2)^{-0.5} e^{(-0.5(h-\mu)^2/\sigma^2)} \qquad \textbf{(6-6)}$$

Suppose that we have the unknown parameter h for which the prior probability distribution can be expressed as:

$$h \sim N(\mu_0, \sigma_0^2) \qquad \textbf{(6-7)}$$

Suppose also that we have an observation x that is normally distributed with mean μ equal to the parameter of interest, that is

$$X \sim N(h,\sigma^2) \qquad \textbf{(6-8)}$$

where μ_0, σ_0 and σ are known. If these assumptions are valid,

$$f(h)=(2\pi\sigma_0{}^2)^{(-0.5)}e^{(-0.5(\mu-\mu_0)^2/\sigma_0{}^2)} \qquad \textbf{(6-9)}$$

$$f(X/h)=(2\pi\sigma^2)^{-0.5}e^{(-0.5(X-\mu)^2/\sigma^2)} \qquad \textbf{(6-10)}$$

f(h/X) can be written as the function of P(h) P(X/h)

$$f(h/X)=(2\pi\sigma_0^2)^{-0.5}e^{(-0.5(h-\mu_0)^2/\sigma_0^2)}*(2\pi\sigma^2)^{-0.5}e^{(-0.5(X-h)^2/\sigma^2)} \qquad \textbf{(6-11)}$$

That means f(h/X) is proportional to

$$e^{(-0.5h^2((\sigma_0^2)^{-1}+(\sigma^2)^{-1})+h(\mu_0/\sigma_0^2+X/\sigma^2))} \qquad \textbf{(6-12)}$$

Regarding f(h/x) as a function of h , it is now convenient to write:

$$f(h/X)=(2\pi\sigma_1^2)^{-0.5}e^{(-0.5(h-\mu_1)^2/\sigma_1^2)} \qquad \textbf{(6-13)}$$

$$\sigma_1^2=\frac{1}{\sigma_0^{-2}+\sigma^{-2}} \qquad \textbf{(6-14)}$$

$$\mu_1=\sigma_1^2\left(\frac{\mu_0}{\sigma_0^2}+\frac{x}{\sigma^2}\right)=\frac{\mu_0\sigma_0^{-2}+x\sigma^{-2}}{\sigma_0^{-2}+\sigma^{-2}} \qquad \textbf{(6-15)}$$

the posterior probability density is

$$f(h/x) \sim N(\mu_1,\sigma_1^2) \qquad \textbf{(6-16)}$$

Often there is insufficient information to ensure that the parameter concerned has a normal or other standard continuous distributions. In that case, it is convenient to use a discrete probability distribution as an approximation. This discretization of the prior and likelihood also makes the procedure simpler and easier on a computer. For a single parameter case, the procedure is as follows:

Let it be assumed that the probabilities of the condition h(i) have been determined in the form of an array P(i) for all value of i. This array is called the prior probability distribution of the parameter h since it has been established prior to or independent of the new information. Let it be further assumed that an event A has been observed and the conditional probabilities of each condition P(A/h(i)) have been determined. This array is called the likelihood of A given the corresponding condition. The observation of event A will change the probability assessment for all conditions in a way that makes the occurrence of A more likely. Bayes' theorem for a one parameter problem can be written in the following form:

$$P(h(i)/A)=N * P(A/h(i)) * P(i) \qquad i=1,2,3.........m \qquad \textbf{(6-17)}$$

This formula expresses that the posterior or updated probability distribution of the parameter h is by definition, the conditional distribution of the parameter h given the observation of A; it is designated P(h(i)/A) for all values of i. Each value in this probability array is calculated as the product of the likelihood of A given the corresponding condition h(i), designated P(A/h(i)), the prior probability of that condition, namely P(i) and a normalizing factor N, which must ensure that the sum of the posterior probability for all conditions adds up to unity.

6.4. BAYESIAN UPDATING OF THE HURST h

In principle, the estimation of Hurst h for a specific drainage basin proceeds in three steps.

First, the prior probability distribution of h is adopted apart from any experimental information about the magnitude of h.

Secondly, this distribution is updated with the information about Hurst h that can be deduced from the set of values of K observed for the basic data set. This would give the probability distribution of h for a specific drainage basin if we know nothing more about it than that the basin is of a Canadian river similar to the rivers in the basic data set.

The distribution obtained in this manner is then updated with the sample information for K obtained from the peak flows for the particular basin under consideration. From the final or posterior distribution of h one can then select a typical value, for example, the mode, or the most likely value as the "best" estimate. Alternatively, depending on the use one wants to make of the estimated parameter, one could choose the mean as the most appropriate estimate.

It can be shown that the estimate of h obtained in this way has a smaller variance than the estimate obtained from the sample alone or from the basic data set alone.

In detail, the estimation procedure is as follows:

First, the possible range of h was divided into m intervals and each interval was represented by its middle value. Supposing the probabilities of condition h_i had been determined in the form of an array $P(h=h_i)$ for all value of i. This array is called the prior (a priori) probability distribution of parameters since it had been established prior to or independent of the new available information. To establish the prior distribution is the most difficult part of this problem. Also, it was assumed that event A had been observed and the probability of A occurrence given each condition h_i could be determined, which was written as $P(A/h_i)$. This array is called the likelihood of A given the corresponding condition, which was sometimes expressed as l(x). So the observation of event A would change the probability assessment for all conditions in such a way that makes A occurring more likely. The Bayes's theorem for parameter h can be written in the following:

$$P(h_i/A) = N * P(A/h_i) * P(h=h_i) \qquad \textbf{(6-18)}$$

$i=1,2,3......m$

where

$P(h_i/A)$------- the posterior or updated probability of parameter h;

N------- a normalizing factor which must ensure that the sum of the posterior probability for all conditions adds up to unity.

To initiate the updating, the assumption was made that the model Hurst parameter h has a uniform distribution between 0.5 and 1, which really means we did not have any knowledge about the distribution of model Hurst except the range. Since there is no physical meaning for h less than 0.5, the range from 0.5 to 1 instead of 0 to 1 was selected. Based on this assumption, we determined the conditional distribution of K, denoted $P(k_j/h_i)$, by means of the Monte Carlo method. We use the data set of the annual peak flows of Canadian rivers as an observed realization or a sample distribution of K, denoted $P(k_j)$. This allows the determination of the posterior distribution of the Hurst h given this observation and the conditional distribution of K.

$$P(h=h_i)=N\sum_{j=1}^{13} P(k_j)*P(k_j/h_i) \tag{6-19}$$

The results are shown in Figure 6-1 and 6-2. This posterior distribution of the model h for a Canadian river in general can be used as the prior distribution of h for a specific Canadian river for which K was observed as B.

$$P(h=h_i/B)=N'P(h=h_i)P(B/h_i) \tag{6-20}$$

Two examples of the updating procedures will be discussed next. One has a relatively low Hurst K and another has a relatively high Hurst K.

Table 6-1A. Probability Distribution of K Response for Various Model Hursts

Based on Monte Carlo Simulation

Response of K	Model Hurst 0.975	Model Hurst 0.925	Model Hurst 0.875	Model Hurst 0.825	Model Hurst 0.775	Model Hurst 0.725
<0.40	0.000	0.000	0.000	0.000	0.000	0.000
0.4-0.45	0.000	0.000	0.000	0.000	0.000	0.000
0.45-0.50	0.000	0.000	0.000	0.002	0.002	0.006
0.50-0.55	0.000	0.000	0.000	0.006	0.006	0.012
0.55-0.60	0.002	0.002	0.010	0.022	0.034	0.050
0.60-0.65	0.008	0.012	0.046	0.052	0.092	0.116
0.65-0.70	0.020	0.042	0.084	0.116	0.162	0.154
0.70-0.75	0.050	0.130	0.152	0.208	0.176	0.260
0.75-0.80	0.072	0.178	0.240	0.230	0.244	0.244
0.80-0.85	0.158	0.228	0.248	0.224	0.184	0.132
0.85-0.90	0.302	0.284	0.194	0.130	0.092	0.024
0.90-0.95	0.324	0.120	0.026	0.010	0.008	0.002
0.95-1.00	0.064	0.004	0.000	0.000	0.000	0.000

Table 6-1B. Probability Distribution of K Response for Various Model Hursts

Based on Monte Carlo Simulation

Response of K	Model Hurst 0.675	Model Hurst 0.625	Model Hurst 0.575	Model Hurst 0.525	Model Hurst 0.475	Model Hurst 0.425
<0.40	0.000	0.000	0.000	0.000	0.000	0.000
0.40-0.45	0.000	0.000	0.000	0.000	0.000	0.000
0.45-0.50	0.012	0.008	0.018	0.030	0.080	0.120
0.50-0.55	0.022	0.042	0.044	0.080	0.100	0.100
0.55-0.60	0.070	0.118	0.146	0.184	0.250	0.280
0.60-0.65	0.134	0.180	0.236	0.242	0.200	0.180
0.65-0.70	0.186	0.224	0.228	0.244	0.230	0.220
0.70-0.75	0.234	0.232	0.200	0.146	0.110	0.060
0.75-0.80	0.212	0.114	0.084	0.066	0.030	0.040
0.80-0.85	0.114	0.070	0.038	0.008	0.000	0.000
0.85-0.90	0.016	0.012	0.006	0.000	0.000	0.000
0.90-0.95	0.000	0.000	0.000	0.000	0.000	0.000
0.95-1.0	0.000	0.000	0.000	0.000	0.000	0.000

Table 6-1C. Probability Distribution of K Response for Various Model Hursts

Based on Monte Carlo Simulation

Response of K	Model Hurst 0.325	Model Hurst 0.225	Model Hurst 0.100	Model Hurst 0.025		
<0.40	0.020	0.020	0.060	0.090		
0.40-0.45	0.060	0.140	0.200	0.330		
0.45-0.50	0.060	0.160	0.200	0.330		
0.50-0.55	0.280	0.260	0.380	0.180		
0.55-0.60	0.240	0.260	0.140	0.030		
0.60-0.65	0.200	0.100	0.020	0.030		
0.65-0.70	0.100	0.060	0.000	0.010		
0.70-0.75	0.020	0.000	0.000	0.000		
0.75-0.80	0.020	0.000	0.000	0.000		
0.80-0.85	0.000	0.000	0.000	0.000		
0.85-0.90	0.000	0.000	0.000	0.000		
0.90-0.95	0.000	0.000	0.000	0.000		
0.95-1.00	0.000	0.000	0.000	0.000		

Table 6-2. Probability Distribution of Hurst K for the Basic Data Set

Hurst K Values	Frequency of 209 Rivers	Probability
<0.4	0	0.000000
0.40-0.45	0	0.000000
0.45-0.50	1	0.004785
0.50-0.55	10	0.047847
0.55-0.60	19	0.090909
0.60-0.65	44	0.210526
0.65-0.70	56	0.267943
0.70-0.75	49	0.234450
0.75-0.80	22	0.105263
0.80-0.85	07	0.033493
0.85-0.90	0	0.000000
0.90-0.95	1	0.004785
0.95-1.00	0	0.000000

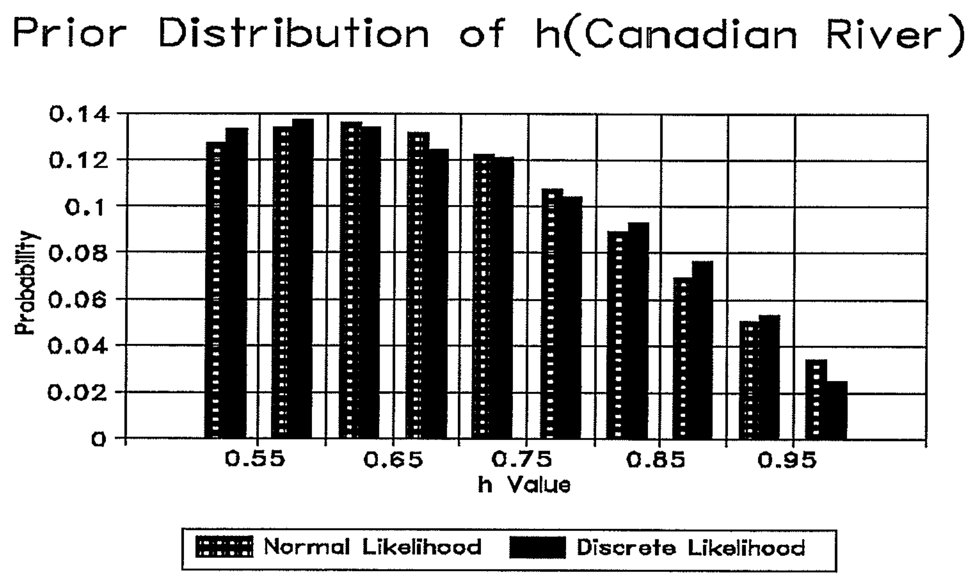


Figure 6-1. Updated Distribution of h for any Canadian river

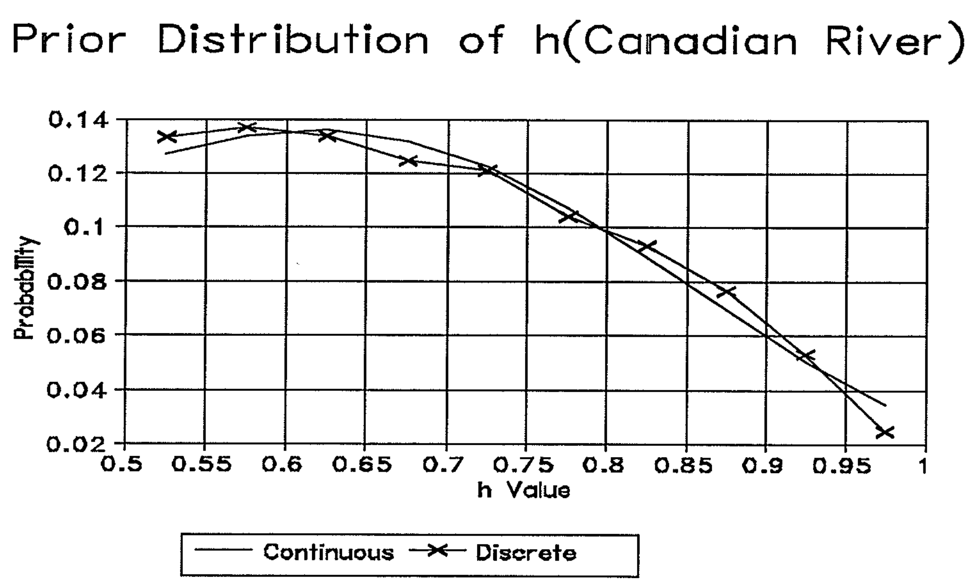


Figure 6-2. Updated Distribution of h for any Canadian river

Example A: The Pipers Hole River at Mother's Brook has 36 year annual peak flow data. Using unbiased estimator we estimated its lag-one autocorrelation coefficient at 0.3354, and the Hurst K to be 0.8146. As prior we used distribution of Hurst h based on the information from the data set, which is shown in Figure 6-1 and 6-2. The likelihood of obtaining the observation K=0.8146 was determined for the range of values of h by simulation. Finally, the posterior distribution of Hurst h was calculated by Bayes' theorem, which is shown in Figure 6-3 and Table 6-3. The most likely value of h for this river turned out to be 0.77. But the distribution is rather flat in the range between 0.75 and 0.85. Based on the sample alone, the Hurst h would have been estimated as 0.91.

Table 6-3. Bayes Updating of h for Pipers Hole River at Mother's Brook

h_i	Prior $P(h=h_i)$	Likelihood $P(B/h_i)$	Normalizing Factor N	Posterior $P(h=h_i/B)$
0.50-0.55	0.133095	0.020064	7.6643	0.020467
0.55-0.60	0.136657	0.047568		0.049822
0.60-0.65	0.133809	0.079152		0.081175
0.65-0.70	0.124182	0.134384		0.127902
0.70-0.75	0.120742	0.155296		0.143712
0.75-0.80	0.103893	0.196480		0.156450
0.80-0.85	0.093110	0.225248		0.160742
0.85-0.90	0.076389	0.246336		0.144222
0.90-0.95	0.053238	0.217600		0.088788
0.95-1.00	0.024885	0.140112		0.026723

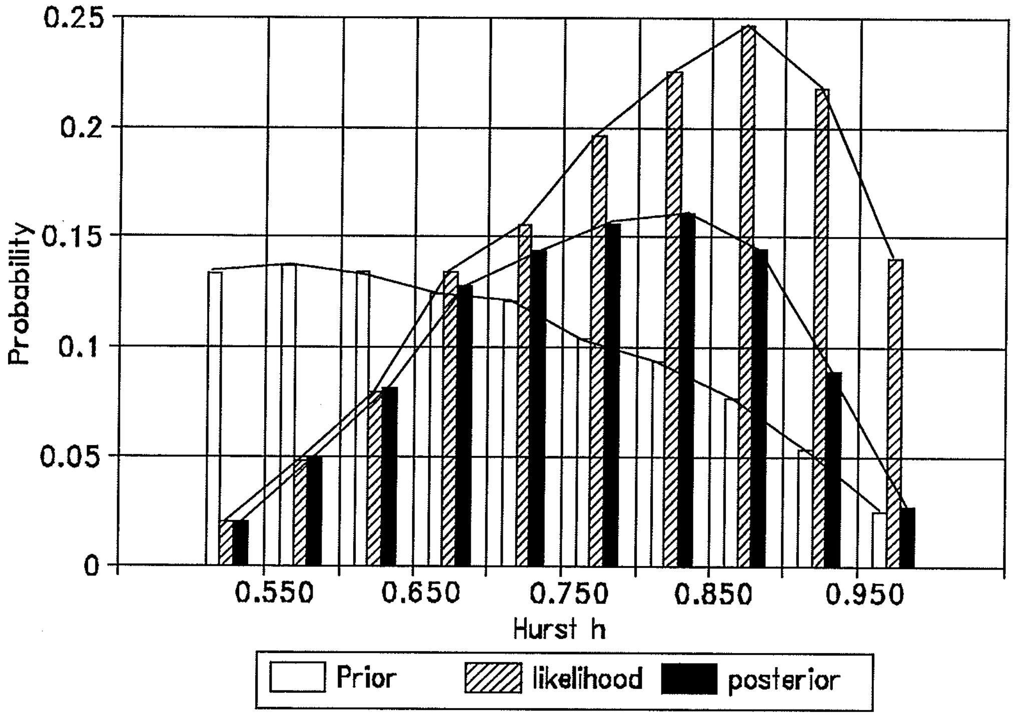


Figure 6-3. Bayesian Updating of the Hurst h for the Pipers Hole River (Observed K = 0.8146)

Example B: the Southwest Margaree River Near Upper Margaree has 70 year annual peak flow data. The lag-one autocorrelation coefficient was estimated at 0.1346, and the Hurst K was estimated as 0.7495. Using the formula mentioned in this section, the likelihood of h for K equals 0.7495 was calculated which is shown in Table 6-4. By means of Bayes' theorem, the posterior distribution of h for this river was obtained which is also listed in Table 6-4. The most likely value of h for this river is obtained as 0.71. The results are shown in a bar graph in Figure 6-4. Based on the sample alone, Hurst h would have been estimated as 0.78.

Table 6-4. Bayesian Updating of h for Southwest Margaree River Near Upper Margaree

h_i	Prior $P(h=h_i)$	Likelihood $P(B/h_i)$	Normalizing Factor N	Posterior $P(h=h_i/B)$
0.50-0.55	0.133095	0.106	5.51	0.0778
0.55-0.60	0.136657	0.142		0.1069
0.60-0.65	0.133809	0.173		0.1276
0.65-0.70	0.124182	0.223		0.1526
0.70-0.75	0.120742	0.252		0.1677
0.75-0.80	0.103893	0.210		0.1202
0.80-0.85	0.093110	0.219		0.1123
0.85-0.90	0.076389	0.196		0.0825
0.90-0.95	0.053238	0.154		0.0452
0.95-1.00	0.024885	0.061		0.0069

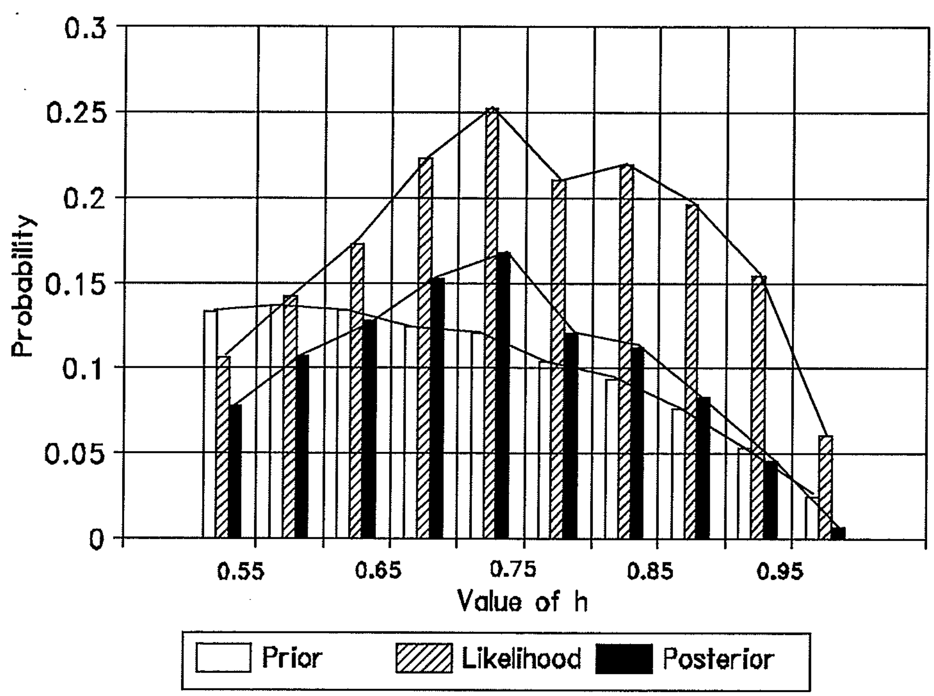


Figure 6-4. Bayesian Updating of the Hurst h for the Southwest Margaree River Near Upper Margaree (observed Hurst K=0.75)

CHAPTER 7.

MODELLING OF ANNUAL PEAK FLOW SERIES

7.1. GENERAL

In the previous chapters, the bias correction formulas for the Hurst statistic and the ACF were derived. A Bayesian updating procedure of the Hurst h was also proposed. Now we have reached the point where we can use all the available information for the estimation of the parameters that are required for the simulation of annual peak flow series.

In hydrological modelling, the model is should be designed to preserve as much as possible the historical mean, the variance, the skewness, the autocorrelation function and the Hurst statistic. Reproducing the mean and the variance is no problem but almost all short memory models fail to reproduce historical long term characteristics, while most long memory models do not preserve the lag-one correlation or do not use the Hurst statistic explicitly. The exception is the mixed noise model (MN), which not only preserves long term persistence but also preserves the short term persistence characteristics (Booy and Lye, 1989). It attempts to reproduce the ACF fractional noise at high lags while correcting it for the first order autocorrelation at low lags. A problem that is occasionally encountered is an incompatibility of the lag-one autocorrelation and the model Hurst h, which results in an unlikely ACF. There are two types of parameter incompatibility. Sometimes one observes a high Hurst statistic with a relatively low lag-one autocorrelation. At other times, one may observe a negative lag-one autocorrelation which seems unrealistic for peak flow series. Both will cause negative partial autocorrelations of one of four AR(1) components. It was suggested by Booy and Lye

(1989) that zero value be assigned to this partial autocorrelation when calculated value was negative. This means that the observed short term correlation is not preserved here. One should keep in mind, however, that parameter incompatibility observed in practice is often caused by using a biased estimator for ρ_1. In addition, both ρ_1 and h are random variables, so that due to chance, one may get a rather unlikely combination. This occurrence is somewhat reduced by using Bayesian updated values of ρ_1 and h.

In this chapter, the annual peak flow series of Canadian rivers will be simulated. Autocorrelation functions will be derived and the results will be compared with those from observed series and with the average autocorrelation function of all 209 available series.

7.2. PARAMETERS USED IN THE MIXED NOISE MODEL

In order to show the entire simulation process, the mixed noise model was applied to the peak flow series of Southwest Margaree River. The observed flow data series is listed in Table 7-1 and shown graphically in Figure 7-1. The uncorrected autocorrelation function of this river was estimated by means of equation (2-3), which is shown in Figure 7-2. As it can be seen from Table 2-1, the uncorrected value for Hurst K is 0.7495 and the lag-one autocorrelation coefficient is 0.1346.

The mixed noise model was also used to reproduce the average ACF of rivers in the data set. From Chapter two, we knew that the average of the lag-one autocorrelation coefficient is 0.039 and the average of the Hurst K is 0.6756. Values of the lag-one to the lag-forty autocorrelation coefficient are listed in Table 7-2.

Table 7-1. Annual Peak Flow Data of Southwest Margaree River Near Upper Margaree

Year	1	2	3	4	5	6	7	8	9	10	11	12
Peak Flow	25.5	42.5	37.5	42.5	48.4	32.0	40.8	52.7	61.2	29.4	55.5	49
Year	13	14	15	16	17	18	19	20	21	22	23	24
Peak Flow	34.5	39.9	34.8	41.6	46.2	42.5	39.1	42.2	34	39.9	32.8	27.6
Year	25	26	27	28	29	30	31	32	33	34	35	36
Peak Flow	32.3	25.6	35.4	32.8	45.3	38.5	42.8	33.7	28.2	40.5	40.8	32.8
Year	37	38	39	40	41	42	43	44	45	46	47	48
Peak Flow	31.7	42.5	39.9	32	28.6	27.5	39.4	36.5	37.7	32.8	31.7	23.2
Year	49	50	51	52	53	54	55	56	57	58	59	60
Peak Flow	44.5	39.4	40.5	32.8	43	53.2	33.1	28.6	40.5	34.3	32.8	41.3
Year	61	62	63	64	65	66	67	68	69	70	71	72
Peak Flow	39.6	53.7	41.7	60.7	42.1	29.5	36.4	52.3	36.8	47.2		

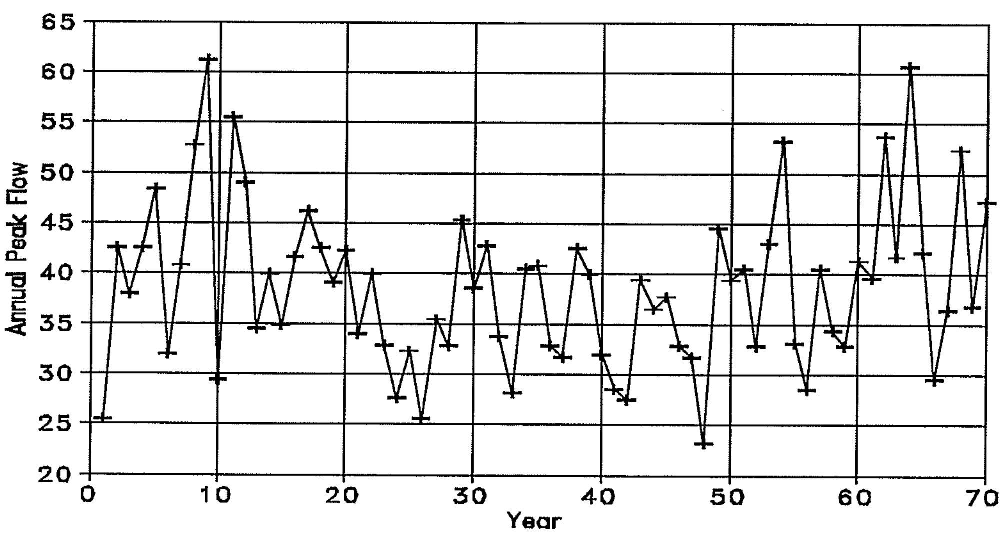


Figure 7-1. Annual Peak Flow of Southwest Margaree River near Upper Margaree

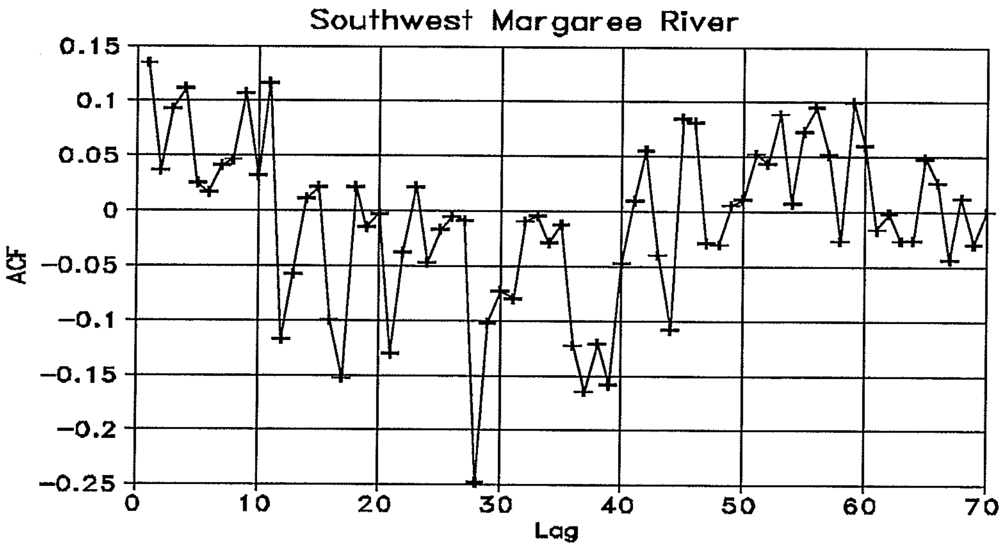


Figure 7-2. Autocorrelation Function of Southwest Margaree River near Upper Margaree

Table 7-2. Mean ACF of Annual Peak Flows of the Data Set

Lag	1	2	3	4	5	6	7	8	9	10
ACF	0.039158	0.007733	0.037505	0.018429	0.01139	-0.02344	-0.01085	-0.02491	-0.03321	-0.05752
Lag	11	12	13	14	15	16	17	18	19	20
ACF	-0.00952	-0.0206	-0.03687	-0.0169	-0.0155	-0.03383	-0.01072	-0.02052	-0.00926	-0.01495
Lag	21	22	23	24	25	26	27	28	29	30
ACF	-0.0379	-0.00598	-0.03123	-0.02439	-0.01671	-0.00196	-0.01008	-0.00464	-0.02057	-0.00919
Lag	31	32	33	34	35	36	37	38	39	40
ACF	-0.01275	-0.0074	-0.01436	-0.01228	-0.0139	0.005354	-0.01164	-0.00193	-0.00283	-0.02511

The difference between this study and others (Lye,1988) is that we use the Bayesian updated Hurst statistic h, which could be obtained by means of the procedure discussed in Chapter Six, as well as a bias corrected lag-one autocorrelation coefficient calculated by the process introduced in Chapter 5, which used the updated h and observed ACF. Parameters used are listed in Table 7-3. These two parameters then were used to calculate eight basic parameters of mixed noise model for each simulation.

Table 7-3. Updated and Bias Corrected Parameters Used in MN

Parameters	Southwest Margaree River	All Canadian Rivers
Hurst h	0.7131	0.6011
Lag-one Autocorrelation	0.2864	0.1053

It may be noted that after bias correction, most of autocorrelation coefficients tend to have positive values. This makes sense as it is believed that all historical data have positive autocorrelations and negative autocorrelations are caused by systematic estimating bias or variability of autocorrelation. These are shown in Figure 7-4 and 7-5.

7.3. DEMONSTRATION OF THE TECHNIQUES

First, Hurst h and lag-one autocorrelation coefficients listed in Table 7-3 were used to calculate eight model parameters of the mixed noise model, which are w_a, w_b, w_c, w_d, ρ_a,ρ_b, ρ_c, ρ_d. Then, five hundred series with identical length equal to the sample size of the Southwest Margaree River were generated. After this, equation (2-3) was used to estimate the sample

ACF of each of these 500 series. Finally, the average ACF of 500 was calculated and the bias correction formulas were applied to get the bias corrected ACF. This ACF was compared with the theoretical ACF of the mixed noise model. This technique was also demonstrated for the average autocorrelation of Canadian rivers. Some results are shown in Figure 7-6 and 7-7. Evidently after bias correction, the generated ACF is very close to the theoretical ACF.

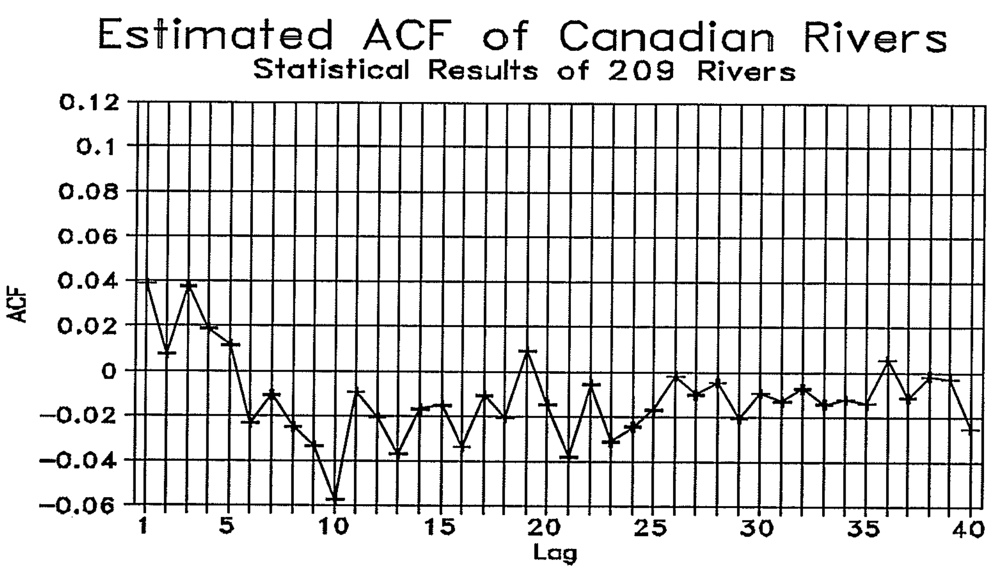


Figure 7-3. Mean ACF of Annual Peak Flows of the Data Set

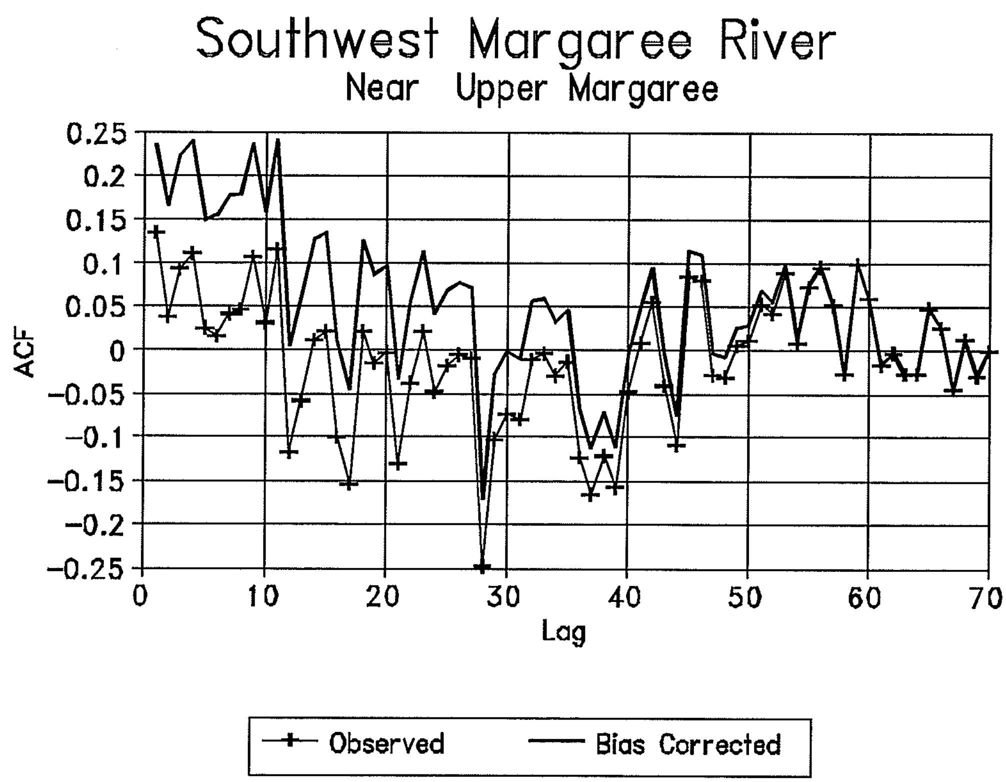


Figure 7-4. Bias Corrected ACF of Southwest Margaree River Near Upper Margaree

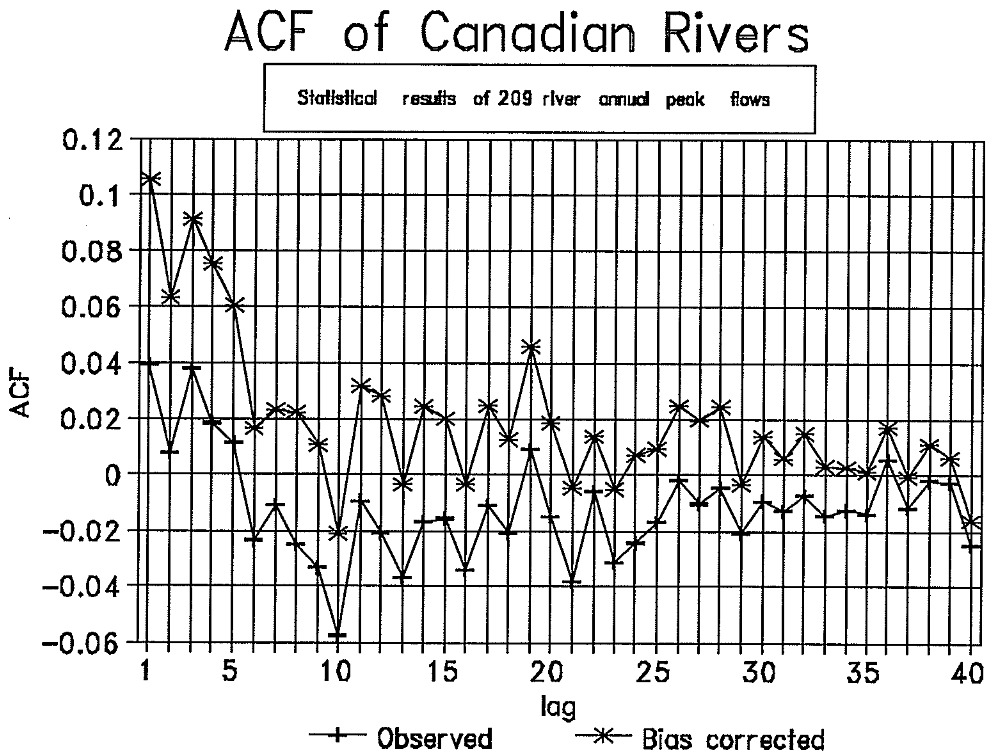


Figure 7-5. Bias Corrected Mean ACF of the Data Set

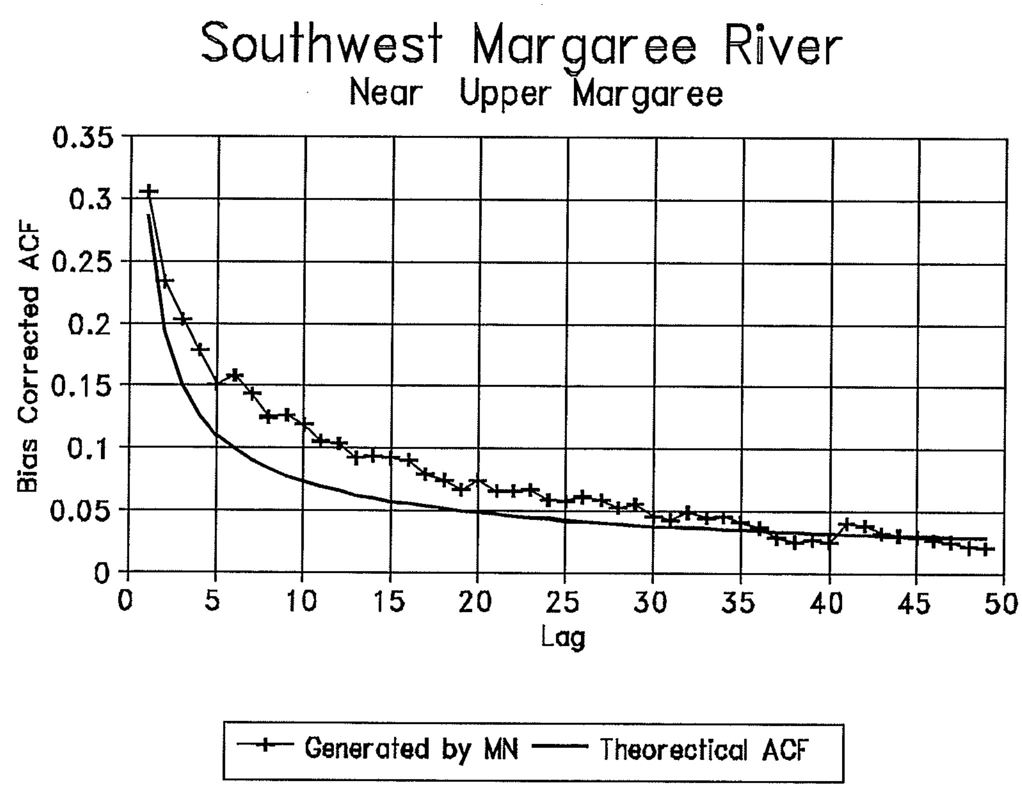


Figure 7-6. Comparison Between MN Generated ACF and Theoretical ACF

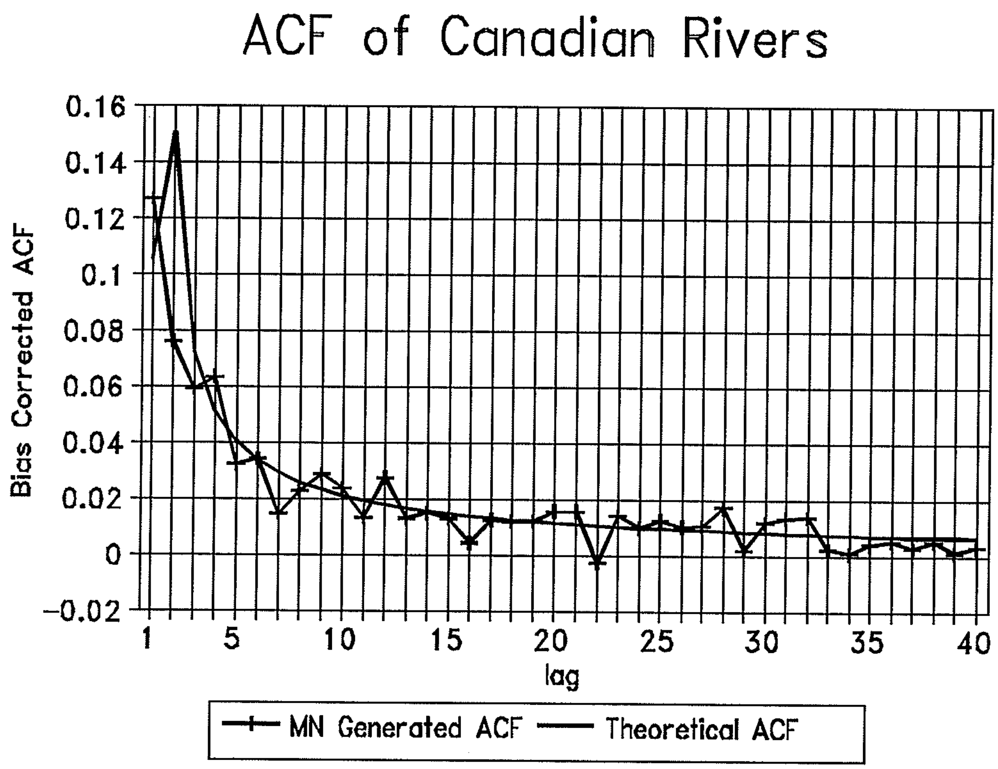


Figure 7-7. Comparison Between MN Generated ACF and Theoretical ACF

7.4. COMPARISON BETWEEN THE GENERATED AND THE SAMPLE ACF

Since it was very difficult to demonstrate the advantage of this procedure based on a single river, the average ACF of 209 Canadian rivers was used. Five hundred series were generated and their ACF were estimated. This generated ACF was compared with the observed ACF from more than 200 rivers. These are shown in Figure 7-8 and 7-9. It is evident that on average, with the use of updated and bias corrected parameters, the mixed noise model can reproduce the correlogram of historical annual peak flows.

7.5. COMPARISON BETWEEN THIS TECHNIQUE AND LYE'S METHOD

The difference between this technique and Dr. Lye's method (L.M.Lye, 1989) is the updated h and bias corrected lag-one autocorrelation were used in this study. One would expect some improvement. To show the improvement using the bias corrected and updated parameters, two procedures were used simultaneously, one with the use of updated and bias corrected parameters and the other with observed K and lag-one autocorrelation coefficient. These procedures were applied to the average correlogram of Canadian rivers. Simulation results are shown in Figure 7-10. Generally, it is hard to find a big difference when one looks at the ACF curve. But when one looks at the lag-one autocorrelation, it is easy to find that results from this study are closer to the sample observations. In addition, the generated average value of Hurst K of 500 series by this technique is about 0.675, it turns out to be 0.71 by Lye's method. Comparing these two procedures, we can find that better simulation results were obtained with the use of updated h and lag-one autocorrelation.

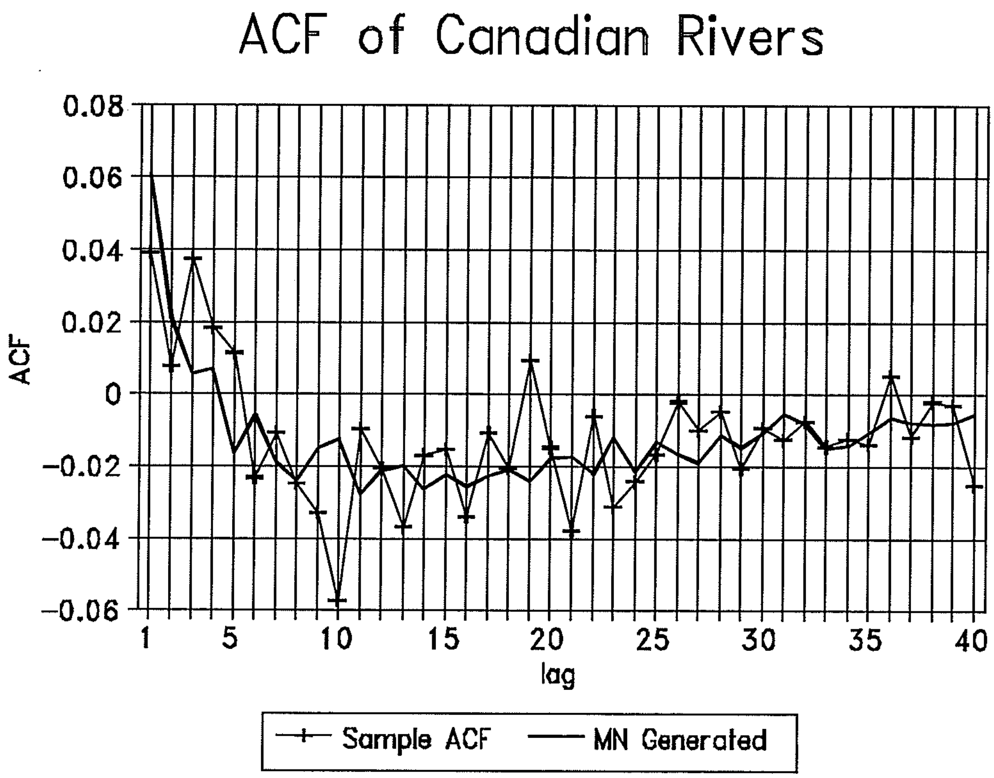


Figure 7-8. Comparison Between the Generated and Sample ACF of the Data Set

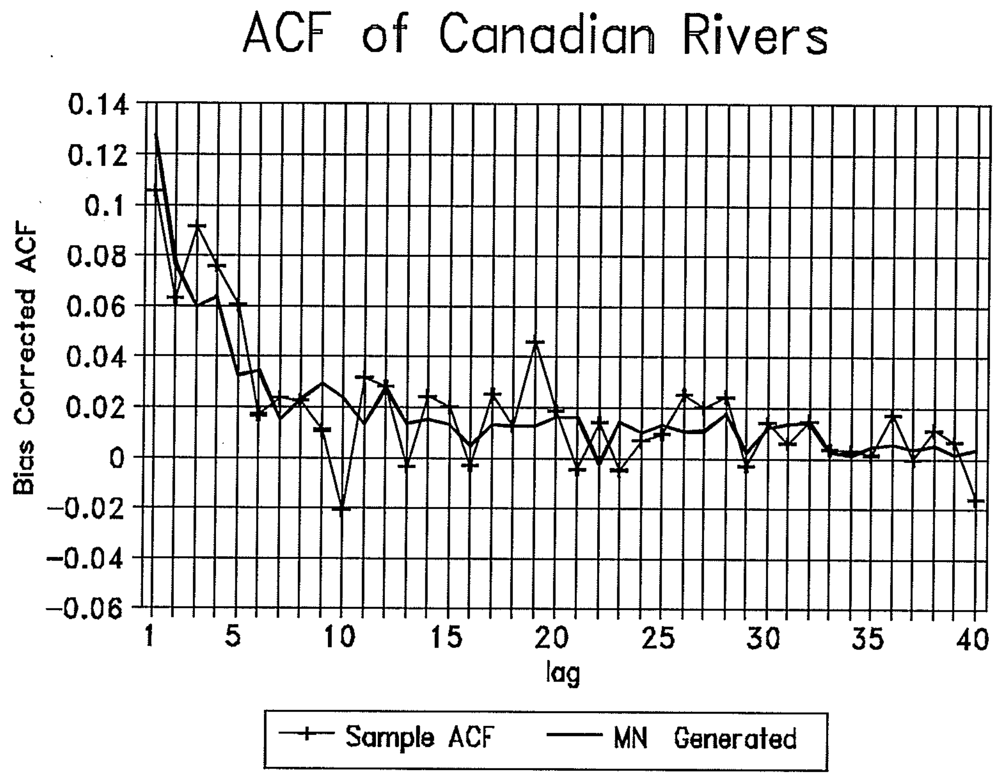


Figure 7-9. Comparison Between the Generated and Sample ACF, Bias Corrected.

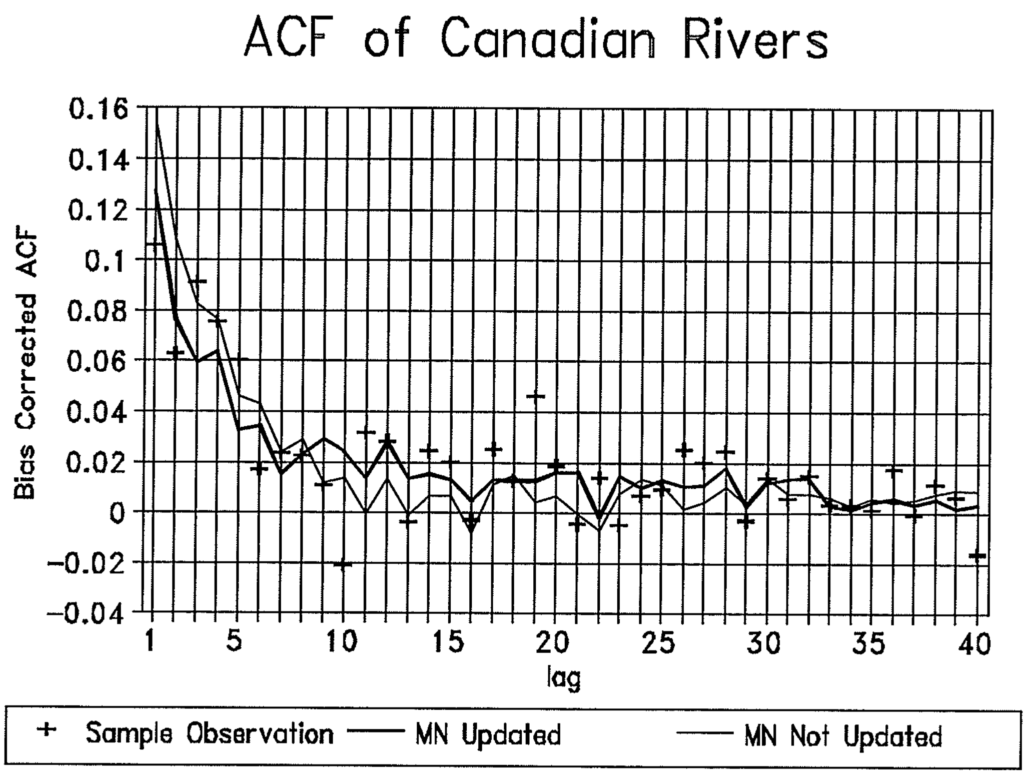


Figure 7-10. Comparison of Bias Corrected ACF Between Techniques With Updated and Observed Parameters

CHAPTER 8.

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER STUDIES

8.1. CONCLUSIONS

1. Annual peak flow series obtained from Canadian rivers show a strong Hurst phenomenon and a very low lag-one autocorrelation. Although Anderson's test suggests that hypothesis of time independent can not be rejected, these rivers cannot be seen as time independent because of the long term persistence. Only the models which can preserve short term as well as long term autocorrelation characteristics should be used for accurate modelling of peak flow series;

2. The Hurst estimation K given by H.E.Hurst has serious bias. This is partly caused by forcing N equal to 2 and rescaled range R/s equal to 1 simultaneously. By removing this constraint, a bias correction formula is derived based on Monte Carlo simulation;

3. The method given by Kendall (1954) for bias correction of autocorrelation coefficients of time series is not valid when the long term persistence is significant, that is when $h > 0.5$. The bias can be corrected by the formula given in chapter five, which contains h;

4. The sample observation of Hurst statistic is a random variable, which can be updated by means of Bayes' theorem when regional information becomes available. The updated h has less variability and is more reliable;

5. By using bias corrected and updated parameters in the mixed noise model, some incompatibilities between lag-one autocorrelation and model Hurst h can be eliminated. When

the mixed noise model is applied for modelling annual peak flows of Canadian rivers with these more reliable parameter combinations, better simulation results can be obtained.

8.2. RECOMMENDATIONS FOR FURTHER STUDIES

1. After bias correction of the ACF, we still occasionally observe negative lag-one autocorrelations or low lag-one autocorrelations with relatively high model Hurst, which is due to the variability of ACF and Hurst. Further studies may be required to investigate the uncertainty of ACF for observed time series.

2. Bayesian updating may also be used to better estimate the lag-one autocorrelation coefficient if information on this parameter can be obtained for a region, in which there are number of stations or rivers;

3. The ACF could possibly be estimated by means of regression line techniques. The parameters of the line could then replaced the model Hurst h and the lag-one autcorrelation coefficient. This regression line could use the following function

$$ACF(k) = a * k^{b} \qquad \textbf{(8-1)}$$

where k represents the lag.

REFERENCES

Bartsch, H.J. (1974). Handbook of Mathematical Formulas. Academic Press, 1974.

Benjamin,J.R. and C.A.Cornell (1970). Probability, Statistics and Decisions for Civil Engineers. Mcgraw Hill, New York. 684 p.

Booy, C. and Lye, L.M. (1985). The Use of Accumulated Basin Storage in Flood Peak Analysis. Proceedings of 7th canadian Hydrotechnical Conference, Saskatoon. 317-335.

Booy, C. and D.R.Morgan (1985). The Effect of Clustering of Flood Peaks on a Flood Risk Analysis for the Red River . Canadian Journal of Civil Engineering, 12(1):156-165.

Booy, C. and Lye, L.M. (1986). Accumulated Basin Storage as a Factor in the Correlation Structure of Annual Peak Flows on the Red River. Canadian Journal of Civil Engineering, 13(3), 365-374.

Booy, C. and Lye, L.M. (1987). Uncertainty in Flood Risk Analysis. Proceedings 8th Hydrotechnical Conference, Montreal, 401-418.

Booy, C. and L.M.Lye (1989). A New Look at Flood Risk Determination . Water Resources Bulletin, 25(5) 933-943.

Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, 553p.

Box, G.E.P. and Jenkins, G.M.(1976). Time Series Analysis-- Forecasting and Control. Holden-Day, San Francisco,575p.

Dawdy, D.R. and Matalas N.C. (1964). In: Handbook of Applied Hydrology, V.T. Chow Editor. Section 8, p85-87, McGraw Hill, New York.

Feller, W. (1951). The Asymptotic Distribution of the Range of Sums of Independent Random Variables. Annals of Mathematical Statistics, 22, 427-432.

Gracia, L.E., Dawdy, D.R. and Mejia, J.M. (1972). Long Memory Monthly Stream flow Simulation By a Broken Line Model. Water Resources Research 8(4),1100-1105.

Hald, A. (1951). Statistical Theory with Engineering Applications. John Wiley and Sons, New York, New York.

Hurst, E.H. (1951). Long Term Storage Capacity of Reservoirs . Transactions American Society of Civil Engineers ,116:770-808.

Hurst, E. H. (1956). Method of Using Long Term Storage in Reservoir, Proc. Inst. Civil Eng., Part 1,519.

Jenkins, G.M. and Watts, D.G. (1969). Spectral Analysis and Its Applications. Holden--Day Series in Time Series Analysis. San Francisco, California.

Kendall, M.G. and Stuart, H. (1968). The Advanced Theory of Statistics, Vol. 3, Design and Analysis and Time Series. 2nd Edition, Hafner, New York.

Klemes,V. (1974). The Hurst Phenomenon:A Puzzle? Water Resources Research,10(4): 675-687.

Lawrence, A.J. and Kottegoda, N.T. (1977). Stochastic Modelling of River Flow Time Series. Journal Royal Statistical Society, Series A, 140, Part 1, 1-47

Lee,P.M. (1989). Bayesian Statistics-----An Introduction. Oxford University Press, New York . 294 p.

Lettenmaier,D.P. and S.J.Burges (1977a). Operational Assessment of Hydrologic Models of Long-Term Persistence . Water Resources Research,13(1):113-124.

Lettenmaier,D.P. and S.J.Burges (1977b). An Operational Approach to Preserving Skew in Hydrological Models of Long Term Persistence. Water Resources Research, 13(2), 281-290.

Lye,L.M. (1988). Uncertainty in Flood Risk Analysis . Ph.D Thesis, University of Manitoba , Canada .

Mandelbrot,B.B. (1965). Une Classe de Processus Stochastiques Homothetiques á Soi; Application á la Loi Climatologique de H.E.Hurst., Comptes Rendus de L'Academie des Sciences de Praies, 260,3274-7.

Mandelbrot,B.B and Wallis,J.R. (1968). Noah,Joseph, and Operational Hydrology. Water Resources Research ,4(5) 909-917 .

Mandelbrot,B.B. and Van Ness, J.W. (1968). Fractional Brownian Motions, Fractional Noise and Applications. S.I.A.M. Review, 10(4), 422-437.

Mandelbrot,B.B. and Wallis,J.R. (1969a). Computer Experiments with Fractional Gaussian Noise . Water Resources Research ,5(1):229-267.

Mandelbrot, B.B. and Wallis,J.R. (1969b). Computer Experiments with Fractional Gaussian Noise. Part 2 : Rescaled Ranges and Spectra. Water Resources Research, 5(1), 242-259.

Mandelbrot,B.B. and Wallis,J.R. (1969c). Computer Experiments with Fractional Gaussian Noises. Part 3 - Mathematical Appendix. Water Resources Research, 5(1), 260-267.

Mandelbrot,B.B. and Wallis,J.R. (1969d). Some Long Run Properties of Geophysical Records. Water Resources Research,5(2),321-340.

Mandelbrot,B.B. and Wallis,J.R. (1969e). Robustness of the Rescaled Range R/S in the Measurement of Non-cyclic Long-Run Statistical Dependence. Water Resources Research, 5(5), 967-988.

Mandelbrot,B.B. (1971). A Fast Fractional Gaussian Noise Generator .Water Resources Research ,7(3):543-553 .

Mandelbrot,B.B. (1972). Broken Lines Process Derived as an Approximation to Fractional Noise . 8(5):1354-1357 .

Mejia, J.M. (1971). On the Generation of Multivariate sequences exhibiting the Hurst Phenomenon and Some Flood Frequency Analysis. Ph.D. Thesis, Colo. State Univ., Fort Collins.

Mejia, J.M., Dawdy, D.R. and Nordin, C.F. (1974). Streamflows Simulation 3. The Broken Line Process and Operational Hydrology. Water Resources Research, 10(2), 242-245.

Mejia, J.M., Rodriquez-Itrube, I. and Dawdy, D.R. (1972a). Streamflow Simulation 1. A New Look at Markovian Models , Fractional Gaussian Noise, and Cross Theory. Water Resources Research, 8(4),921-930.

Mejia, J.M., Rodriquez-Itrube, I. and Dawdy, D.R. (1972b). Streamflow Simulation 2. The Broken Line Process as a Potential Model for Hydrological Simulation. Water Resources Research, 8(4), 931-941.

O'Connell, P.E. (1971). A Simple Stochastic Modelling of Hurst's Law. Proc. Int. Symposium on Mathematical Models in Hydrology. Int. Association Hydrological Science, Warsaw, 169-187.

O'Connell, P.E. (1974). Stochastic Modelling of Long Term Persistence in Streamflow Sequences. Ph.D. Thesis, Imperial College, University of London,284 p.

Press, S.J. (1989). Bayesian Statistics: Principles, Models, and Applications. John Willey & Sons, U.S.A..237 p.

Salas, J.D. , Delleur, J.W., Yevjevich, V. and Lane, W.L. (1980). Applied Modelling of Hydrologic Time Series. Water Resources Publications, Colorado, U.S.A..

Schmitt,S.A. (1969). Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics. Addison-Wesley Publishing Company,Massachusetts.400 P.

Srikanthan , R. (1979). Stochastic Generation of Annual and Monthly Flow Volumes. Ph.D. Thesis, Monash University, Australia.

Thomas, H.A. and Fiering , M.B., (1962). Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation. In Design of Water Resources System. Cambridge, Massachusetts, Harvard University Press.

Vicens, G.J., Rodriquez-Iturbe, I. and Schaake, J.C. (1975). A Bayesian Framework for the Use of Regional Information in Hydrology. Water Resources Research 11(3):405-414.

Wall, D.J. and Englot M. (1985). Correlation of Annual Peak Flows for Pennsylvania Streams. Water resources Bulletin 21(3):459-464.

Wallis,J.R. and Matalas, N.C. (1971a). Small sample properties of H and K --- Estimators of the Hurst Coefficient h . Water Resources Research,6(6): 1583 - 1594.

Wallis ,J.R. and Matalas,N.C. (1971b). Correlogram Analysis Revisited . Water Resources Research ,7(6):1448-1459.

Wallis , J.R. and Matalas, N.C. (1972). Sensitivity of Reservoir Design to Generating Mechanism of Inflows. Water Resources Research,8(3), 634-641.

Wallis, J.R. and O'connell, P.E. (1972a). Small Sample Estimation of $\rho 1$. Water Resources Research, 8(3), 707-712.

Wood, E. and Rodriquez-Iturbe, I. (1975). Bayesian Inference and Decision Making for Extreme hydrologic Events. Water Resources Research 11(4):533-542.

Yevjevich, V.M. (1963). Fluctuations of Wet and Dry Years Part 1. Research Data Assembly and Mathematical Models. Hydrology Paper 1, Colorado State University, Fort Collins, Colorado.

Yevjevich,V.M. (1972a). Probability and Statistics in Hydrology. Water Resources Publications, Fort Collins, Colorado, U.S.A.

Yevjevich, V.M. (1972b). Stochastic Processes in Hydrology. Water Resources Publications, Fort Collins, Colorado, U.S.A.