UNIVERSITI PUTRA MALAYSIA

COMPARISON OF HYDROLOGICAL HOMOGENIZATION METHODS FOR DEVELOPMENT OF FLOOD REGIONAL MODELS

ATEFEH ABDOLHAY

COMPARISON OF HYDROLOGICAL HOMOGENIZATION METHODS FOR DEVELOPMENT OF FLOOD REGIONAL MODELS

BY

ATEFEH ABDOLHAY

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Partial Requirements for the Degree of Master of Science

November 2008

DEDICATION

For my loving family. Their never ending support and encouragement helped me to believe in myself and discover I can complete anything I put my heart in.

Abstract of the thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Master of Science

COMPARISON OF HYDROLOGICAL HOMOGENIZATION METHODS FOR DEVELOPMENT OF FLOOD REGIONAL MODELS

By

ATEFEH ABDOLHAY

November 2008

CHAIRMAN: Professor Mohd Amin Mohd Soom, PhD

FACULTY: Engineering

ABSTRACT

Water resource management, design of hydraulic structures such as bridges, dams, roads and railways, land use management and flood control depend on reliable estimates of flood with various risk of exceedance. Lack of funds and human resources in many developing countries such as Iran have resulted in limited climatological and hydrological data. Regional flood frequency analysis (RFFA) is used for characterization of flood in ungauged location with the support of gauged sites in areas of similar hydrological characteristics.

Gorganrood basin is one of the areas which are exposed to frequent floods. Due to insufficient hydrometric stations in this area direct prediction of flood is not applicable. In order to apply RFFA to study area parameters such as NDVI, curvature, area, mean slope of basin and etc were estimated for each subbasin.

For the purpose of homogenization, Hierarchical clustering, K-means clustering, Fuzzy clustering and Kohonen method were applied. The region was grouped into two homogenous sub regions. All of the clustering methods showed same results. L-moment heterogeneity values were greater than two (H>2) therefore some adjustments were done.

Regional models were developed for each homogenous region and the whole basin. The values of R-squared for homogenous groups were higher (Average R^2= 0.98) than for the whole region (Average R^2= 0.6). Besides, the standard error for homogenous sub region is much lesser than whole region which proves the necessity of identification of homogenous regions.

The controlling factors on AMF were area, mean annual rainfall, average elevation of basin, Gravilious coefficient, and percentage area of class4 NDVI. Therefore flood quantiles can be estimated by the mentioned parameters from the regression models.

Abstrak tesis untuk kelulusan Senat Universiti Putra Malaysia sebagai memenuhi keperluan ijazah Master of Science

PERBANDINGAN KAEDAH PENYERUBAN HIDROLOGI UNTUK PEMBANGUNAN MODEL BANJIR SERANTAU

Oleh

ATEFEH ABDOLHAY

November 2008

PENGERUSI: **PROFESOR MOHD AMIN MOHD SOOM, PHD PENG**

FAKULTI: **KEJURUTERAAN**

Amaran ancaman banjir yang boleh dipercayai adalah penting untuk kawalan banjir, pengurusan sumber air, pengurusan gunatanah, dan rekabentuk struktur hidraulik seperti jambatan, empangan, jalan dan rel keretapi. Kekurangan dana dan tenaga manusia di kebanyakan negara membangun biasanya menyebabkan kekurangan stesen mengumpul data iklim dan data hidrologi. Sebaliknya analisis kekerapan banjir serantau digunakan untuk mencirikan banjir di sesuatu kawasan yang tiada stesen menyukat dengan membandingkan data yang diperolehi dari kawasan yang mempunyai ciri hidrologi yang sama. Lembangan Gorganrood di Iran adalah salah satu daripada kawasan yang selalu terdedah kepada ancaman banjir. Oleh yang demikian terdapat keperluan mendesak untuk menganggar dan meramal air larian kawasan ini bagi persediaan yang lebih baik dalam menghadapi banjir. Ramalan terus banjir di kawasan ini adalah mustahil kerana tiada stesen menyukat, jadi RFFA digunakan. RFFA mempunyai tiga langkah iaitu mengenalpasti kawasan serupa, analisis kekerapan luahan maksimum semasa, dan pembangunan model anggaran aliran serantau bagi kawasan yang sama ciri hidrologinya. Kajian ini meliputi pemilihan stesen hidrologi dan pengumpulan data aliran puncak dari stesen-stesen tersebut. Analisis kekerapan luahan maksimum tahunan telah dijalankan. Rantau Gorganrood telah disamakan ciri hidrologinya mengikut parameter yang diekstrak. Satu dari parameter tesebut adalah Normalized Difference Vegetation Index (NDVI) yang diperolehi daripada imej MODIS. Lengkungan bumi adalah satu lagi parameter berkaitan ciri topografi. Dari analisis faktor, pembolehubah terpenting telah dikenalpasti. Kawasan tersebut telah dikelaskan kepada rantau homogen berdasarkan parameter seperti NDVI, lengkungan, luas kawasan dan cerun. Bagi tujuan penyerupaan, kaedah-kaedah Hierarchical clustering, K-means clustering, Fuzzy clustering dan Kohonen telah digunakan. Teknik L-moment telah digunakan untuk menyiasat keputusan kajian. Ketidaksamaan untuk kumpulan 1 melebihi dua, oleh itu beberapa modifikasi telah dilakukan. Model serantau telah dianggar bagi setiap kawasan yang sama ciri dan juga keseluruhan lembangan. Nilai R^2 bagi kumpulan yang sama ciri adalah lebih tinggi (R^2=0.98) berbanding bagi keseluruhan lembangan (R^2= 0.60). Tambahan pula ralat piawai untuk kawasan yang serupa adalah lebih rendah dari keseluruhan lembangan, ini membuktikan keperluan untuk mengenalpasti kawasan yang serupa. Lembangan tersebut telah dibahagi kepada dua kawasan yang mempunyai ciri hidrologi yang serupa. Semua kaedah penyatuan kluster menunjukkan keputusan yang sama. Model yang dibina menunjukkan Class4 NDVI mempengaruhi banjir dalam sesetengah kala kembali. Model yang terhasil dalam kajian ini boleh digunakan untuk kajian seterusnya bagi pelbagai aspek hidrologi gunaan.

ACKNOWLEDGMENTS

First and Foremost I would like to thank the Almighty God for his perpetual divine assistance without which nothing could have been achieved. With any work of this scope there are always many people to thank for their assistance and advice. Much of this work would not have been possible without the boundless energy of my advisor Professor Saghafian. His innovative thinking and love of learning was a great motivation for me to excel in my work.

I would like to express my gratitude to Professor Mohd Amin Mohd Soom who accepted the challenge of supervising a student who was going to spend most of the time away from UPM. Sincere appreciation is also due to Assoc.Prof.Dr. Abdul Halim Ghazali for his comments and suggestions.

I would like to thank my mom and dad. They have provided me with strength and encouragement in times when I needed it most. They are two of the greatest people I know, and I am truly blessed to have them as parents. Without financial support and guidance of my father I could not accomplished my studies. I thank my mother for preparing me to achieve anything I wish .Her constant words of encouragement and her ever-ready shoulder to cry on were a saving grace. I owe a great deal of gratitude to Ali, my dear brother who was patient during the time I was doing my research in Iran.

A special thank is also owed to my dear friend Samane who never left me alone in Malaysia. I would like to extend my thanks to my dear friend Ms Salamzade and my lovely sister Fateme for their moral and prayerful support.

I am grateful to help of Mr. Chavoshi for his help and support in Malaysia. I am also thankful for the comments of Mr. Ghalekhani through out this research. Finally, I would like to acknowledge all those who contributed in diverse ways towards the success of this research.

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

ATEFEH ABDOLHAY

DATE: November 2008

TABLE OF CONTENT

	Page
DEDICATION	**II**
ABSTRACT	**III**
ABSTRAK	ERROR! BOOKMARK NOT DEFINED.
ACKNOWLEDGMENTS	**VII**
APPROVAL SHEETS	**IX**
DECLARATION	**XI**
TABLE OF CONTENT	**XII**
LIST OF TABLES	**XIV**
LIST OF FIGURES	**XV**
CHAPTER 1	1.1
INTRODUCTION	1.1
1.1 General	1.1
1.1.1 Hydrology of Iran	1.4
1.1.2 Floods in Iran	1.6
1.2 Statement of the problem	1.7
1.3 Objectives	1.9
1.4 Scope of the study	1.9
CHAPTER 2	2.1
LITERATURE REVIEW	2.1
2.1 Hydrological Homogeneity	2.1
2.1.1 Geographical Proximity	2.2
2.1.2 Dalrymple Approach	2.2
2.1.3 Hierarchical Clustering	2.4
2.1.4 Andrews curve	2.10
2.1.5 L-moment	2.12
2.1.6 K-means clustering (Non-hierarchical clustering)	2.15
2.1.7 Artificial Neural Networks	2.17
2.1.7.1 Self Organizing Map	2.21
2.1.7.2 Simple example of how SOM operates	2.25
2.1.8 Fuzzy clustering	2.29
2.2 Regional flood frequency analysis	2.33
2.2.1 Square grid method	2.34
2.2.2 Hybrid method	2.36
2.2.3 Flood index method	2.36
2.2.4 L-moments	2.37
2.2.5 Multiple regression method	2.39
2.3 Normalized Difference Vegetation Index (NDVI)	2.45
2.4 Factor analysis	2.49
2.5 Curvature	2.52

2.5.1 Surface Interpolation 2.55

CHAPTER 3 3.1

METHODOLOGY 3.1

3.1 Study area 3.1

3.2 Data review 3.7

3.3 Statistical analysis of data 3.9

3.3.1 Statistical analysis 3.9

3.3.1.1 Randomness 3.9

3.4 Frequency analysis 3.10

3.5 Climatic and Physiographic Characteristics 3.16

3.6 Factor Analysis 3.18

3.7 Identification of homogenous regions 3.20

3.7.1 Hierarchical clustering 3.21

3.7.2 Fuzzy clustering 3.21

3.7.3 K-means clustering 3.22

3.7.4 Kohonen method 3.22

3.8 Determination of regional model 3.24

CHAPTER 4 4.1

RESULTS AND DISCUSSIONS 4.1

4.1 Climatic and physiographic characteristics 4.1

4.2 Frequency analysis 4.7

4.2.1 Result of *run test* analysis 4.9

4.2.2 Outlier detection 4.10

4.2.3 Frequency analysis 4.11

4.3 Identification of homogenous regions 4.18

4.3.1 Hierarchical clustering 4.18

4.3.2 K-means Clustering 4.24

4.3.3 Fuzzy clustering 4.25

4.4.4 Kohonen Clustering 4.26

4.5. Regional model 4.34

CHAPTER 5 5.1

CONCLUSION 5.1

REFRENCES

APPENDICE

BIODATA OF STUDENT

LIST OF TABLES

1.1: Natural disasters of Iran (EM-DAT,2008)	1.6
1.2. Loss of life and properties of past flood in Gorganrood basin	1.7
2.1. Input data set	2.26
2.2. Silhouette table	2.32
3.1. Characteristics of the selected hydrometric stations in Gorganrood basin	3.8
3.2. Curvature classified range	3.17
3.3. NDVI classified range	3.18
3.4 Variance explained by factor analysis	3.19
3.5. Selected variables of sub basins in Gorganrood basin	3.20
4.1. Mean annual precipitation in each sub basin	4.1
4.2. Run test analysis result	4.9
4.3. Result of outlier detection in hydrometric stations	4.10
4.4. Flow characteristics of selected stations in Gorganrood basin	4.13
4.5. Chi – square distribution comparison	4.14
4.6. Flood quantiles (m^3/s) based on station skew	4.15
4.7. Flood quantiles (m^3/s) based on generalized skew	4.16
4.8. Flood quantiles (m^3/s) based on weighted skew	4.17
4.9. Comparison of hierarchical methods	4.19
4.10. Cluster members	4.23
4.11. Cluster member of two clusters	4.24
4.12. Comparison of different number of clusters in fuzzy clustering	4.25
4.13. L-moment results of Total region	4.30
4.14. Heterogeneity values of total region	4.30
4.15. L-moment results of Group 1	4.31
4.16. Heterogeneity values of Group 1	4.31
4.17. L-moment results of Group 2	4.32
4.18: Heterogeneity values of group 2	4.32
4.19. L-moment results of group 1 with Galikesh	4.33
4.20. Heterogeneity values of group 1 with Galikesh	4.33
4.21. L-moment result of group 2 with Arazkose	4.33
4.22. Heterogeneity result of group 2 with Arazkose	4.34
4.23. Regression models for whole region	4.34
4.24. Regression models for Group 1	4.35
4.25. Regression models for Group2	4.35
4.26. RERR of Polordogah station	4.40

LIST OF FIGURES

2.1. Biological neuron cell 2.18
2.2. Artificial neuron 2.19
2.3. Kohonen network 2.21
2.4. Network structure of SOM example 2.26
2.5. Curvature analysis 2.53
3.1. Study area location of selected station 3.3
3.2 Classified slop map of Gorganrood basin 3.4
3.3. DEM of Gorganrood basin 3.5
3.4. The methodology chart 3.6
4.1. Isohyetal rainfall map of Gorganrood basin 4.2
4.2. River profile of Zaringol River 4.3
4.3. River profile of Zyarat River 4.3
4.4. Location of Zyarat River and Zaringol River 4.4
4.5. Curvature map of Gorganrood basin 4.5
4.6. NDVI map of Gorganrood basin 4.6
4.7. Relation between daily discharge and AMF of Zaringol station 4.7
4.8. Relation between daily discharge and AMF of Lazore station 4.8
4.9. Relation between daily discharge and AMF of Ramyan station 4.8
4.10. Skew map of Gorganrood basin 4.12
4.11. Group average dendrogram 4.20
4.12. Complete linkage dendrogram 4.20
4.13. Simple average dendrogram 4.21
4.14. Single linkage dendrogram 4.21
4.15. Centroid dendrogram 4.22
4.16. Median dendrogram 4.22
4.17. Ward's minimum variance dendrogram 4.23
4.18. U-matrix and component plane of map size [5 3] 4.26
4.19. U-matrix and labeled map ([5 3]) 4.27
4.20. Cluster members of map size [5 3] 4.28
4.21. U-matrix and component plane of map size [2 2] 4.28
4.22. U-matrix and labeled map ([2 2]) 4.29
4.23. Cluster members of map size [2 2] 4.29
4.24. Predicted versus fitted discharge of Group 2 for return period 2 year 4.36
4.25. Predicted versus fitted discharge of Group 2 for return period 5 year 4.37
4.26. Predicted versus fitted discharge of Group 2 for return period 10 year 4.37
4.27. Predicted versus fitted discharge of Group 2 for return period 20 year 4.37
4.28. Predicted versus fitted discharge of Group 2 for return period 50 year 4.38
4.29. Predicted versus fitted discharge of Group 2 for return period 100 year 4.38
4.30. Predicted versus fitted discharge of Group 1 for return period 2 year 4.38
4.31. Predicted versus fitted discharge of Group 1 for return period 10 year 4.39
4.32. Predicted versus fitted discharge of Group 1 for return period 20 year 4.39
4.33. Predicted versus fitted discharge of Group 1 for return period 50 year 4.39
4.34. Predicted versus fitted discharge of Group 1 for return period 100 year 4.40

CHAPTER 1

INTRODUCTION

1.1 General

Among the 17 main catastrophes in the world, flood is a serious natural hazard due to frequent loss of lives and properties. In spite of extensive research on various aspect of flood, yet mankind suffer seriously from this disaster which necessitates more investigation.

Flooding occurs when a river or stream overflows unexpectedly and inundate the surrounding area. This over flow may be caused by continuous rainfall over several days, intense rainfall over a short period of time, erosion of river banks, or an ice or debris jam. Flooding causes the destruction of valuable agricultural land, structures such as bridges, roads and dams. Therefore flood prevention is one of the necessary aspects of water resource development.

Water resource management, design of hydraulic structures such as bridges, dams, roads and railways, land use management and flood control depend on reliable estimates of flood with various probability of exceedance. The aim of hydrological studies is to estimate the amount of floodwater which the hydraulic structures confront during their life. Therefore accurate estimation of floodwater will be the most important aspect of hydraulic structure design. So, designing of dams and spillways requires estimation of flow with different probabilities of exceedance. There are different ways of design flood determination, some of which are based on statistical analysis and others based on empirical models.

In statistical analysis, if sufficient length of record exists, a suitable probability distribution is fitted and peak discharge with desired return period is estimated. But where sufficient records are not available or no stream gauge is available, it is difficult for engineers to estimate flood directly. Lack of funds and human resources in many developing countries resulted in limited climatological and hydrological data. Since availability of flood records is very important in designing hydraulic structures, various methods have been developed for estimation of peak discharges. Some of these techniques are mentioned below:

1. Empirical models

In case where there is no runoff information, it has been tried to develop models for run off estimation such as Dicken's formula, Creager formula, Mayer formula and etc.

Dicken's formula is as follows:

$$Q = CA^{0.75}$$

Where:

A: area (km^2)

Q: run off (m^3/s)

C coefficient would range from 14 to 28 for mountainous areas and 2.8 to 5.6 for plain areas.

The empirical models are developed in special situation, so they cannot be used or may not be suitable in other conditions.

2. Envelope curve

Sometimes peak discharge can be estimated from enveloping curve of all the existing peak discharge which is plotted in logarithmic paper based on area of basin. This method is not reliable for peak estimation and is just used for evaluation of other methods.

3. Hydrograph method

Design discharge is estimated from design storm by this method but it is not applicable for large basins with different hydrological characteristics (Salimi,1997).

4. Simulation method

These methods need lots of input so that if these data cannot be representative of the watershed, the output would not be reliable.

5. Regional flood frequency analysis (RFFA)

RFFA is a technique that estimates extreme events with specific return period. This method is used where the existing record length is short relative to return period. In other words it is effective in substituting an increased spatial characterization of the data for insufficient temporal characterization, although problems exist with the implementation of regional flood frequency analysis techniques.

Regional flood frequency analysis (RFFA) is also used for characterization of flood in ungauged location with the help of gauged sites based on similar hydrological characteristics. Whereas regional flood frequency analysis is applied for estimation of peak discharge in this research, more detail on this method is explained in chapter 2.

1.1.1 Hydrology of Iran

World population for the past 70 years has grown three times and water consumption increased six times. Annually, 75 million people is added to the world population and it is predicted that growth rate in undeveloped countries would increase 300 % in the next 50 years, (Abdolahi,2000).

Predictions of future population have been estimated at 7. 9 billion individuals by 2050. However, others have predicted population of the world to reach to 10.9 billion in 2050. An average value of 9.3 billion individuals has been also proposed. It has been also estimated that population of IRAN to be about 100 millions in year 2050, (Abdolahi,2000).

Mean annual precipitation of Iran is about one third of the world precipitation which should be noted that this amount is not well distributed. Most regions have dry climate, therefore, high evapo-transpiration would be another problem. Notwithstanding that Iran covered 1.1% area of the world's land; the country holds only 34 % of water supply. Seasonal and uneven distribution of precipitation is also not uniform, (Abdolahi,2000).

Mean annual volume of Iran's precipitation is 400 billion cubic meters based on Ministry of Energy studies. Three hundred and ten billion cubic meters of this precipitation occurs in mountainous regions (870,000 km^2) and 90 billion meter cube in plain regions (778 Km^2). Seventy one percent of precipitation is lost due to evapo-transpiration which includes 200 billion meter cube in mountainous regions and 84 billion cubic meters in plain regions. From the remainder volume, 59 billion meter cube of precipitation infiltrates in mountainous regions and two billion meter cube in plain zones. So runoff in mountainous zones would be 51 billion meter cube and four billion meter cube in plain zones. The next two categories are drinking water (6.67 %) and industrial sector; Agriculture has the largest Water consumption in Iran (88.8 %), (Abdolahi,2000).

In Table1.1 the top natural disasters of Iran are mentioned. It can be seen that flood is the second destructive hazard in Iran.

Table 1.1: Natural disasters of Iran (EM-DAT,2008)

Disaster	Date	Killed	Disaster	Date	Killed
Disaster	Date	Killed	Disaster	Date	Killed
Earthquake	23-Jan-1909	5000	Earthquake	16-sep-1978	20000
Earthquake	1929	3253	Flood	Jul-1980	120000
Earthquake	Jan-1939	23000	Flood	26-Aug-1988	150000
Earthquake	13-Dec-1957	3000	Earthquake	4-Apr1989	400000
Earthquake	Sep-1962	12000	Drought	1-Jun-1990	625000
flood	1964	9500000	Earthquake	1 Jun-1990	40000
Earthquake	31-Aug-1968	10000	Earthquake	8-Feb-1993	605000
Earthquake	10-Apr-1972	5057	Drought	2000	37000000
Windstorm	Feb-1972	4000	Drought	Jul-2001	37000000

1.1.2 Floods in Iran

Previous records showed 94 important floods in years upto 1951. Subsequent research indicates that 1890 flood events have occurred from year 1952 to 1991. This increasing trend of flood events (20 times) is worrying. Undoubtedly recording of flood events in the past was not considered precisely. But it can be concluded that some unnatural causes may caused increase in flood events. Destruction of 66% of pasture lands and half of the forests in Iran in the past half century and loss of 2 -2.5 billion tons of soil due to erosion is a reality, may attribute to excess runoff and floods (Khalili and Sedghi, 2000).

Irregular distribution of precipitation in time and space in north part of Iran results in frequent flooding. Two extreme floods have occurred in north eastern part of Iran in year 2001 and 2002 (August). These catastrophes resulted in loss of human (more than 300 death in total) and properties.

Some factors believed to be the causes of these damages were changing land usage, improper location of villages, improper design of bridges and other hydraulic structures, existence of natural parks alongside rivers, roads being built in the floodplains and no attention to safety alarms (Mosaedi, 2003). In spite of research and extensive works on floods, flooding still occurs in northern part of Iran. Therefore, flood disaster management and flood prevention must be revised in this region. ,

1.2 Statement of the problem

Mean annual precipitation in Iran is about one third of the world average. Precipitation in most part of Iran (about 96% of Iran) is less than 200 mm (Abdolahi,2000). However, this volume is not evenly distributed. High-intensity, short duration rainfall is the most important reason of flood occurrence in many regions. Gorganrood basin is one of the areas exposed to frequent floods which caused damages especially in year 1991, 2000 and 2001 as illustrated in Table 1.2.

Table 1.2. Loss of life and properties of past flood in Gorganrood basin

Year	Financial loss	No. Death
1991	180	5
2000	570	220
2001	220	43

The occurrence of floods, due to high-intensity rainfall of short duration is the dominant natural disaster in the study area, causing loss of human lives and properties. Due to insufficient number of hydrometric stations in this area, direct prediction of flood is not applicable. Regional Flood Frequency Analysis (RFFA) can be used in regions with insufficient hydrometric stations. The initial step in RFFA is to find the homogenous groups but Gorganrood basin has not been clustered into homogenous groups previously which necessitate the homogenization in the study area.

There are various methods for clustering and finding homogenous groups. Kohonen and Fuzzy technique which are two novel clustering techniques have been applied in advanced countries and the results were satisfactory. Therefore it appeared that these methods which have not been used before in Iran can be applicable to study area.

Innovation of remote sensing and its application in the hydrologic projects help to solve problems associated with data preparation. In order to develop the most accurate regional models the most suitable parameters should be selected. Land cover as represented by Normalized Difference Vegetation Index (NDVI) can be used as a representative parameter in Regional Flood Frequency Analysis which has not been considered in RFFA previously.

1.3 Objectives

The aim of the study was to develop regional flood models. The specific objectives were as follows:

1. To delineate hydrological homogenous regions using hierarchical and non-hierarchical algorithms, Kohonen method and fuzzy technique.
2. To evaluate the performance of the derived homogenous regions on flood quantiles.
3. To determine the effect of vegetation cover as represented by satellite images on regional flood models.

1.4 Scope of the study

The study involved selection of appropriate hydrometric stations and use of peak flow data measured at these stations. Frequency analysis on annual maximum discharges has been performed. The region was delineated into homogenous zones according to important factors. For each homogenous region, a flood flow model was developed. This model can be applied in future studies to determine flood discharges at ungauged locations.