A Data Based Flood Forecasting Model for the Mekong River

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Muhammad Khurram Shahzad, MSc.
aus Pakistan

Tag der mündlichen Prüfung: 25.02.2011

Hauptreferent: Prof. Dr.-Ing. Dr.-Ing E.h. Erich J. Plate
Korreferent: Prof. Dr.-Ing. Dr.h.c. mult. Franz Nestmann

Karlsruhe 2011

Prologue

The dissertation of Khurram Shahzad, which is before the reader, is the outcome of a contest, where a simple model applied by a research group, which considers only available data, is pinned against a ready-made, fully developed physically based model, which is adapted to a new situation by consultants. There is an interesting background story that led to it. The devastating floods of the year 2000 in the Mekong area brought about a number of international activities, in the course of which I was invited by the German Foreign Office to participate in the formative meeting of the Mekong River Commission´s flood management and mitigation program. My experience with this kind of activities included participation in an expert meeting in Mozambique after the big Limpopo flood of 1998, where a mixed group of experts had gathered to make recommendations for future flood control. To involve many different types of experts at the same time to discuss every aspect of flood management in a two day meeting did not seem to me to be very efficient, and I became convinced that the only way a group of experts can be effective is if they are concentrating on one aspect only. I thought that flood warning was of the first priority, and organized a workshop of international experts in cooperation with the Mekong River Commission Secretariat to make recommendations only on the issue of early warning and flood forecasting for the Mekong River. The workshop took place in 2002 in Phnom Penh, Cambodia, and was supported by the German Foreign Ministry. The recommendations of this meeting included setting up a center for early warning, and to go for a local flood forecast model on the basis of the available data, which should improve or replace the then existing model SSARR, which had been used since the early sixties. The MRC then went ahead, the Regional Flood Management and Mitigation Center was founded, and a program was generated which defined the duties of this center, with early warning and flood forecasting of highest priority. I was asked to draw up a program for the forecast activities, for which I spent two weeks in Phnom Penh. I developed specific for the Mekong, the concept of a data based forecast model, based on available information, along with the necessary data gathering activity. However, while I was working on drafting this concept, the center decided to use an existing model, the model URBS developed in Australia, and to put the model structuring and application planning into the hands of an Australian consultant. When I found out that this was decided, I went to find support for the data based study that I had in mind, and fortunately the German Ministry of Science and Technology agreed to support this as part of the program WISDOM, and the DAAD sponsored Khurram Shahzad to do the study under my supervision.

Against this background one has to see the work of Khurram Shahzad, who developed a concept that might well be used also for other rivers. Using the principle to go from the simple to the more complex – if the simple does not do the job – he investigated as a first approach how far one could go with classical time series analysis, i.e. by using regression models for interconnecting time series from all the gauging stations along the middle course of the Mekong River. Naturally, this approach could not provide information for what happens in the regions between gages, where the rainfall dominates what is coming down from the lateral hills – in particular from the Eastern mountains between Laos and Cambodia, where rainfalls from the Southwest monsoon cause the runoff. Using rainfall information in addition to river discharges improves the forecasts – if it is properly used. It is interesting to know that the pecu-

liarities the Mekong middle region, where the rainfall is converted into runoff in a very systematic and consistent manner due to the fact that the Southwest monsoon is raining on almost completely dry land, which in the first couple month of the flood season (starting in June) is gradually filled up to almost complete saturation in August, are making it possible to estimate the runoff coefficient without the help of any soil moisture accounting model, both during calibration and real time operation. This is a most significant result of the study. The errors from the forecasts are greatly reduced when this input is calculated, for which a simple gamma distribution as unit hydrograph is sufficient. If in the end the forecast is further cleaned by means of an error correction routine, then a very good forecast is obtained, as is made clear by looking at quality assessment parameters calculated from the data, such as the persistence index. In combining the simple regression model with this simple rainfall -runoff a model is obtained which in actual application to the discharges in the years 2009 and 2010 proved the value, if not the superiority of the data based approach.

Karlsruhe, June 2011 Erich J. Plate

Acknowledgement

All praises are to Almighty Allah, the creator of the universe, who blessed me with courage, strength and motivation in order to complete this research work.

I am deeply indebted to my reverend research-supervisor, Prof. Dr. Erich J. Plate, without his continuous guidance I would not have been able to complete this work. I had the privilege to get his unconditional help in the research work, in a write up and in testing of developed model in the study area. He spared his valuable time to accompany me in the field visits. A word of thank is clearly not sufficient to express my gratitude for his contributions to this study. I would like to extend my gratitude to Mrs. Plate, who always encouraged me during this study.

I am cordially thankful to my honorable co-supervisor, Prof. Dr. Franz Nestmann. It was him, who invited me to Germany on the first place and provided me the opportunity to start my research in this prestigious Institute of Water and River Basin Management (IWG). His deep commitment for research and development in Pakistan always kept me motivated to complete my studies quickly and return to my home land. Despite his busy hours, he always spared time for me in order to provide the required support.

I am grateful to the other members of my doctoral supervisory committee, Prof. Dr. Erwin Zehe and Prof. Dr. Sarah Jones, who reviewed my dissertation at the short notice and gave their valuable comments to improve it. I would like to extend my gratitude to the chair of doctoral examination commission, PD. Dr. Stefan Norra.

My deepest gratitude goes to Dr. J rgen Ihringer, who has not only integrated me in his research group, i.e. the hydrology section but also helped me out throughout my study period. I extend my gratitude to Dr. Charlotte K mpf, who provided me the opportunity to study different courses in the international study program of Resource Engineering.

I reserve my special thanks for my colleagues, without their support it was not possible to complete this work. I would like to thank Dr. Sandra Schneider, who not only guided me with the procedural details of doing doctorate in Karlsruhe Institute of Technology (KIT) but also helped me in every possible way to overcome different problems, which I have confronted during my studies. The accomplishment of this study would not have been possible in this limited time without the technical co-operation of my dear colleagues, Dr. Falk Lindenmaier and Mr. Martin Helms. Especially, the study lead by Dr. Lindenmaier on SSARR model was an insightful effort to determine the limitations of the SSARR in its application for forecast in the catchment of Middle Mekong River. The support of Mr. Patrick Preu and Oleg Evdakov in different matters, especially, regarding the installation of different software saved my lot of time and therefore they deserve my deep appreciation. I want to thank Mr. Preu and Mr. Helms for their valuable time which they spend in the proofreading of some important sections of this manuscript. Mr. Robert Mikovec deserves my heartiest gratitude for arranging short but nice recreational visits in the mountains, which proved to be healthy breaks from the work. I also owe special words of thanks for Mr. Johannes H fer, Mr. Joachim Liebert, Mrs. Olga Kiselova, Mr. Simon H llering and Dr. Frauke K nig, who always encouraged me to complete this study.

I would like to thank Mr. Thorsten Ulbrich, Dr. Falk Lindenmaier, Ms. Anastasia Golovina and Dr. Ulrika Scherer, who used to be my room partners from time to time in this institute – each of them took special care in order to provide me the healthy working environment required for this huge task.

I could not forget to mention the support of Mrs. Sandra Knoll, Mrs. Kathrina Remmler, Ms. Jutta Szabadics, Mrs. Raziye Fiden, Ms. Ines Ferring and Mrs. Karin Krix in administrative matters and in providing me an access to the libraries of IWG-IWK and IWG-Hyd.

This study would not have materialized without the financial support available to me through the scholarship scheme of HEC-DAAD. The WISDOM project deserves my gratitude for its support in travels to the study areas in Cambodia and Lao-PDR.

The special thanks go to the flood forecasting team of the Mekong River Commission (MRC) who not only made their data available in order to conduct this study but also provided me the working place and their personal coordination during my visits in Cambodia and Lao-PDR. I take this opportunity to thank Mr. Pradeep Rathore for his input in the analysis of typhoon tracks in the Mekong region while working on mini thesis of his BSc. civil engineering in this institute. It is an honor for me to acknowledge the suggestions of Prof. Dr. Peter Adamson in order to improve the quality of this work.

Last but not the least; I would like to express my gratefulness to each and every member of my family who has provided me the constant and unconditional moral support during my studies. Especially, my parents without whose prayers I could not have managed to finish my work successfully.

Karlsruhe, June 2011 Muhammad Khurram Shahzad

Abstract

The Mekong, 10th largest River of the world, is an important life line for South East Asian countries. The Lower Mekong is the major source of surface water flow in Lao-PDR and Cambodia. However, during the flood season, i.e. June to October, water overflows the banks of the river. The discharge increases as Mekong flows from North to South, due to continuous inflows from its Eastern tributaries. The SE monsoons and NW typhoons cause heavy rainfall in the mountains of northern Lao-PDR and along the border of Lao-PDR to Vietnam which results in exceptionally high floods, up to 30,000 -70,000 m^3/sec in the Lower Mekong. Owing to their low altitude, almost entire inhabited part of Lao-PDR, central Cambodia and Southern Vietnam are the sufferers of Mekong flooding.

On one hand, the flood overflow has been used for number of beneficial applications, i.e. agriculture, fishing, etc. But on the other hand, it is responsible for many human casualties, for example in the flood of the year 2000, the death tolls exceed 800. However, in financial terms, the flood benefits of 7100 (US$ millions) outnumber the flood damages, i.e. 76 (US$ millions). Therefore, the South East Asian countries have decided to live with the floods. However, a major effort is required to reduce the intangible costs in order to optimize the social benefits of Mekong floods. An appropriate flood risk zoning, and an efficient early warning system are possible non-structural means in order to reduce flood damages while keeping its benefits.

Flood forecasting is a most important element of an early warning system. However, in the case of Lower Mekong, the existing flood forecasting models, i.e. SSARR and URBS, fail to produce the flood forecast of required efficiency. Both of these models were developed in other regions (SSARR in USA and URBS in Australia) and were then fitted to the conditions of the Mekong. However, when some pre-build (semi-empirical) model is fitted to local conditions, the quality of model performance is subjected to certain restrictions. First of all, there should be understanding of application purpose, based on a study of the catchment, which should determine the structure of the model. Second, the application site should comply with basic data requirements of pre-build model. First condition is satisfied by defining proper model structure and calibrating the parameters of a model, and the second condition by furnishing the required data. The main problem observed in the fitting of SSARR and URBS was inability to fulfill these conditions. Therefore, objective of this study was to develop a better forecasting approach, which shall be goal oriented, based on local conditions, i.e. local hydrology, and shall be workable with the available data base. The application domain of this model was restricted to Middle Mekong, i.e. Luang Prabang to Stung Treng. Further downstream of Stung Treng, the nature of flooding is quite different from upstream Stung Treng because of specific effects of the large Tonle Sap Lake, which were not investigated in present work.

Owing to data scarcity of catchment's physical properties, i.e. geologic terrain, river cross sections, land use, land cover, soil moisture and its hydro-meteorologic gauging, i.e. 31 rainfall and 7 discharge gauges to represent the catchment area of 795000 km^2, a data based forecasting approach was preferred. In contrast to detailed and complex physical models, two simple data based modeling approaches were used, i.e. Metric (Type 1) and Hybrid Metric Conceptual (Type 2), which are data efficient and therefore can be used in the typical conditions of Mekong. In data based modeling approaches, the functional dependencies are structured from the empirical relationship of subsequent discharges (Type 1 Model) or of discharge with a rainfall (Type 2 Model) with the help of historical data.

But prior to model building, the objective function was selected in order to judge the quality of a model. The quality was taken in terms of accuracy and reliability. In order to gauge accuracy, instead of Nash Sutcliffe, the persistence Index (PI) was used. And reliability was measured by providing the probability distributions of the remaining errors.

In the development of data based models, the information from upstream discharge gauges was used with the help of routing and only the portion of discharge coming from lateral tributaries is forecasted. The average discharge travel time from one gauge to another was determined with the help of cross correlation analysis and by noting time delay of flood peak from one gauge to another.

Daily discharge data of 7 mainstream discharge gauges Mekong for year 1960 to 1990 and 1990 to 2005 have been used to develop and test the model respectively. The objective was to produce 1 to 5 days flood forecast at each gauge. In Type 1 Model, the lateral inflows were forecasted on the basis of multi-regression. These lateral inflows were added to the routed flow from upstream in order to generate future discharges. The correlation of subsequent errors suggested the use of error correction by forecasting errors. The forecasting efficiency after error correction, as expressed by PI, ranges from 0.7 to 0.5 in first 3 days lead time, however it was reduced to 0.2 to 0.4 for 4 and 5 days lead time.

In Type 2 Model, the lateral inflows were forecasted on the basis of conceptual rainfall-runoff models. In total, 6 number of conceptual rainfall runoff models were constructed, one for each sub-basin. In addition to discharge data, daily rainfall data of 31 rainfall gauges was used for year 1990 to 2005. The period 1990 to 2000 was taken as simulation period and that one of 2000 to 2005 as validation period for model development. The basic 2 parameter Nash cascade was used in order to convert effective rainfall into runoff. The effective rainfall was computed from areal average rainfall data with the help of a seasonal adjustment coefficient KN, which includes, among other factors, the runoff coefficient. The lateral inflows were computed by the best combinations of KN and n & k parameters, which are updated by correcting error using the linear regression between subsequent errors.

The forecasting efficiency of a model with perfect rainfall data (assuming future rainfall is known exactly) for 1 to 5 day forecast, as expressed by PI, ranges from 0.6 to 0.7, except at Nakhon Phanom where it remains 0.1 to 0.5. However, forecasting efficiency is reduced if one replaces the perfect input data with a forecasted rainfall for future days, and then range from 0.5 to 0.6 in first 3 days lead time and 0.2 to 0.4 in 4 and 5 days lead time. A special combination of n and k parameter was introduced in order to cope with the poor rainfall forecast situation. However, the resultant unit hydrograph shapes of these n and k variant did not seem to be realistic, although they improved the forecasting efficiency, as expressed by PI, to ranges from 0.6 to 0.7 in first 4 days lead time, and 0.5 to 0.6 in 5 days lead time. Again the forecasting efficiency at Nakhon Phanom is smaller, i.e. ranges from 0.1 to 0.4. But if Type 2 Model is used in the forecasting mode the efficiency of forecasts for longer lead times is very much reduced.

The typical compensating pattern of forecasting errors in Type 1 Model and Type 2 Model suggested the use of weighted average of these two forecasts. Optimal weights were determined with the help of standard regression. The forecasting efficiency of the mixed model in the analysis mode, as expressed by PI, ranges from 0.6 to 0.7 in first 4 days lead time, and is reduced to 0.4 to 0.5 after 5 days lead time.

In summary, it has been demonstrated that simple data based models, i.e. Metric (Type 1) and Hybrid Metric Conceptual (Type 2) can be successfully used in data scarce catchments like that of the Mekong. To combine the advantages of the two methods weighted averages of the forecasts from the two models are used in an approach that is called Mix Model. But it became evident that when improvements for longer lead times are sought, an improvement of rainfall forecasts will be needed.

The finding of this study was that in large rivers flood forecasting, owing to physical continuity of large rivers system, simple data based models can perform equally well or even better than complex physical models. The best model for forecasting should not be the one which best captures the (complete) physical reality, but the one that delivers the most accurate and reliable flood forecasts to be used in the practice of flood management.

Zusammenfassung

Der Mekong, der zehntgrößte Fluss der Welt, ist eine wichtige Lebensader für die südostasiatischen Länder. Der untere Mekong ist das bedeutendste Oberflächengewässer in Laos und Kambodscha. Während der Hochwassersaison zwischen Juni und Oktober kommt es jedoch entlang dieses Flusses zu Überflutungen. Auf seinem Weg von Norden nach Süden steigt der Abfluss des Mekongs durch perennierende linksseitige Zuflüsse stetig an. Der Südost-Monsun und Nordwest-Taifune bringen hohe Niederschläge in den Gebirgsbereichen von Nord-Laos und entlang der Grenze zwischen Laos und Vietnam, welche im unteren Mekong ausgesprochen große Hochwasserabflüsse zwischen 30.000 und 70.000 m³/s auslösen. Hauptsächlich Laos, Zentral-Kambodscha und der Süden Vietnams leiden bedingt durch ihren Tieflandcharakter unter diesen Hochwassern.

Einerseits haben die Überschwemmungen für die Bevölkerung eine ganze Reihe nützlicher Effekte, z.B. im Ackerbau und in der Fischerei. Auf der anderen Seite sind sie für viele Todesopfer verantwortlich, im Jahr 2005 beispielsweise wurden 800 Tote in Folge der Hochwasser gezählt. In der ökonomischen Bilanz überwiegt jedoch die positive Wirkung der Hochwasser (7100 Millionen US-Dollar) deutlich gegenüber den verursachten Schäden (76 Millionen US-Dollar). Aus diesen Gründen haben sich die südostasiatischen Länder entschieden, mit den Überschwemmungen zu leben. Allerdings sind große Anstrengungen notwendig, um die immateriellen Schäden zu reduzieren und damit den sozialen Nutzen der Mekong-Hochwasser zu verbessern. Eine geeignete Ausweisung von Hochwasserflächen und ein effizientes Frühwarnsystem können dazu beitragen, die Hochwasserschäden zu reduzieren und den Nutzen der Überschwemmungen dennoch auszuschöpfen.

Die Hochwasservorhersage ist ein wesentliches Element eines Frühwarnsystems. Im Fall des Mekongs sind bestehende Vorhersagemodelle, wie SSARR und URBS, nicht in der Lage, Hochwasservorhersagen der notwendigen Güte zu erbringen. Beide Modelle wurden in anderen Gebieten entwickelt (SSARR in den USA und URBS in Australien) und erst dann an die Bedingungen am Mekong angepasst. Wenn jedoch vorgefertigte (semi-empirische) Modelle an lokale Bedingungen angepasst werden, sind sie in ihrer Einsatzfähigkeit begrenzt. Zunächst sollte basierend auf einer Einzugsgebietsstudie der Anwendungszweck klar definiert werden, um eine geeignete Modellstruktur festzulegen. Weiterhin sollten in dem Anwendungsgebiet die notwendigen Datengrundlagen für die Modellanwendung gegeben sein. Die erstgenannte Bedingung kann durch die Kalibrierung der Parameter eines Modells erreicht werden, die zweite indem die geforderten Daten zur Verfügung gestellt werden. Dies konnte bei der Anpassung der Modelle SSARR und URBS nicht geleistet werden.

Daher war es das Ziel dieser Studie, einen besseren Vorhersageansatz zu entwickeln. Dieser sollte zielorientiert sein, auf den lokalen (v.a. hydrologischen) Gegebenheiten basieren und mit der verfügbaren Datenbasis anwendbar sein. Das Anwendungsgebiet dieses Modells wurde auf den mittleren Mekong beschränkt, von ‚Luang Prabang' bis ‚Stung Treng', weil der Abschnitt unterhalb durch den Einfluss des ‚Tonle Se Up' Sees einer gesonderten Methode unter Zuhilfenahme eines hydraulischen Modells bedarf.

Im betrachteten Einzugsgebiet ist nur eine knappe Datengrundlage im Hinblick auf die physikalischen Eigenschaften des Einzugsgebietes, insbesondere mit seinen geologischen Verhältnissen, Gewässerprofilen, Bodennutzungen, seiner Vegetation, seinen Bodenfeuchteverhältnissen, gegeben. Auch sind nur wenige hydrometrologische Messstationen verfügbar: 31 Niederschlagsstationen und 7 Abflusspegel, die ein Einzugsgebiet mit einer Größe von 795.000 km² repräsentieren. Vor diesem Hintergrund ist ein Modell, das einen datenbasierten Vorhersageansatz impliziert, vorzuziehen.

Im Gegensatz zu detaillierten, komplex-physikalischen Modellen gibt es zwei einfache datenbasierte Ansätze: der metrische (Typ 1) und der hydrid-metrische konzeptuelle (Typ 2). Beide sind dateneffizient und können deswegen für die Verhältnisse am Mekong genutzt werden. Beim datenbasierten Modellansatz werden die funktionellen Abhängigkeiten aus der empirischen Beziehung von

aufeinanderfolgenden Abflüssen (Typ-1 Modell) oder von Abflüssen im Zusammenhang mit dem Niederschlag (Typ-2) anhand historischer Daten ermittelt.

Um die Güte des Modells bewerten zu können, wurde vor der Modellbildung die Zielfunktion gewählt. Die Qualität wurde dabei hinsichtlich Genauigkeit und Verlässlichkeit beurteilt. Für die Abflussbewertung wurde statt des Nash-Sutcliffe-Faktors der Persistence-Index PI verwendet. Die Verlässlichkeit wurde über die Wahrscheinlichkeitsverteilung der Restfehler gemessen.
Bei der Entwicklung datenbasierender Modelle wurden Abflüsse oberstromiger Pegel geroutet und nur der Anteil aus seitlichen Zuflüssen über Vorhersagen einbezogen. Die mittlere Transportzeit zwischen den einzelnen Pegeln wurde mit Hilfe von Kreuzkorrelationsanalysen und durch die Bestimmung von zeitlichen Abständen der Hochwasserspitzen an den Pegeln identifiziert.

Tägliche Abflusswerte der Jahre 1960 bis 1990 und der Jahre 1990 bis 2005 an 7 Hauptpegeln am Mekong wurden verwendet, um die Modelle zu entwickeln und zu testen. Ziel war es, an jedem der genannten Pegel eine 5-Tages-Vorhersage zu liefern. Beim Typ-1 Modell wurden die seitlichen Zuflüsse auf Basis einer multiplen Regression vorhergesagt. Diese lateralen Zuflüsse wurden dem Hauptfluss von Oberstrom aufgeprägt und geroutet, um zukünftige Abflüsse zu generieren. Die Korrelation resultierender Fehler legte die Anwendung einer Fehlerkorrektur durch eine Fehlervorhersage nahe. Die Vorhersageeffizienz nach der Fehlerkorrektur, dargestellt durch PI, liegt im Bereich zwischen 0.5 und 0.7 bei einer Vorhersagezeit von 3 Tagen und im Bereich 0.2 bis 0.4 bei 4 bis 5 Tagen Vorhersagezeit.

Beim Typ-2 Modell wurden die seitlichen Zuflüsse auf Basis eines konzeptuellen Niederschlag-Abfluss-Modells vorhergesagt. Dafür wurden 6 unterschiedliche N-A-Modelle aufgebaut, für jedes Teileinzugsgebiet eines. Zusätzlich zu den Abflusswerten konnte auf tägliche Niederschlagsmessungen von 31 Stationen der Jahre 1990 bis 2005 zurückgegriffen werden. Das Zeitfenster 1990 bis 2000 wurde dabei als Simulationsreihe verwendet, im Anschluss erfolgte für die Modellentwicklung eine Validierung mit den Jahren 2000 bis 2005. Die einfache 2-parametrige Nash-Kaskade wurde eingesetzt um den effektiven Niederschlag in Abflüsse zu transformieren. Der effektive Niederschlag konnte aus dem mittlerem Gebietniederschlag der Messungen mit Hilfe des Korrekturfaktors KN, der u.a. auch den Abflussbeiwert beinhaltet, errechnet werden. Die seitlichen Zuflüsse wurden aus der geeignetsten Kombination vom Faktor KN und der Parameter n und k berechnet, die durch eine Fehlerkorrektur unter Zuhilfenahme der linearen Regression der Folgefehler aktualisiert wurden.

Die Vorhersageeffizienz eines Modells mit ‚perfekten' Niederschlagsdaten (unter Annahme dass der Niederschlag genau bekannt ist) liegt für die Vorhersagetage 1 bis 5, ausgedrückt durch PI, zwischen 0.6 und 0.7, mit Ausnahme von ‚Nakhon Phanom' wo der Wert 0.1 bis 0.5 erreicht. Wenn die perfekten Vorhersagedaten durch die tatsächliche Vorhersage ersetzt werden so erreichen die Werte 0.5–0.6 für die ersten 3 Tage und 0.2–0.4 für die Vorhersagetage 4 und 5. Eine besondere Kombination aus den n und k Parametern wurde eingeführt, um Defizite in der Niederschlagsvorhersage auszugleichen. Die resultierende Einheitsganglinie aus diesen n und k Parametern erschien nicht realistisch, obwohl die Vorhersagegüte, wieder ausgedrückt durch PI, im Bereich von 0.6 bis 0.7 (1.-4. Tag) und 0.5 bis 0.6 (5. Tag) lag. Auch hier war der Wert in ‚Nakhon Phanom' schlechter, er lag zwischen 0.1 und 0.4. Wenn das Typ-2 Modell für größere Vorhersagezeiträume angewandt wird, fällt die Güte der Vorhersage stark ab.

Die typischen Ausgleichsschemen der Vorhersagefehler der beiden Modelltypen suggerierten die Mittelbildung der beiden Vorhersagen. Die optimalen Wichtungen wurden mit Hilfe der Standardregression bestimmt. Die Vorhersagegüte der gemischten Modelle erreichte wieder als PI ausgedrückt 0.6 bis 0.7 für die ersten 4 Tage und 0.4 bis 0.5 für den 5. Vorhersagetag.

Zusammenfassend wurde aufgezeigt dass einfache datenbasierende Modelle, das metrische (Typ-1) und das hydrid-metrische konzeptuelle (Typ-2) Modell, erfolgreich in datenarmen Einzugsgebieten, wie das des Mekong, eingesetzt werden können. Um die Vorteile der beiden Methoden in einem kombinierten

Modell einzusetzen werden gewichtete Mittelwerte erzeugt. Gleichzeitig wurde ersichtlich, dass wenn bessere Abflussvorhersagen erreicht werden sollen, auch die Niederschlagsvorhersagen für längere Dauern verbessert werden müssen.

Die Erkenntnisse der vorliegenden Studie äußern sich wie folgt, die Vorhersage an großen Gewässern kann mit einfach datenbasierten Modellen ebenso gut oder sogar besser geleistet werden wie mit komplexen physikalischen Modellen. Das beste Modell für eine Vorhersage sollte nicht das Modell sein, welches die physikalischen Gegebenheiten am besten abbildet, sondern jenes, welches die präziseste und verlässlichste Hochwasservorhersage liefert.

Note that some of the texts and equations were taken without detailed referencing from the papers by the writer and his co-authors, which are however listed in the list of references

Contents

1. Introduction

The River Mekong is one of the greatest rivers of South East Asia. Among the world´s largest rivers, it is ranked 10th according to outflow at mouth. From its source in Tibetan plateau glaciers at about 4500 m amsl (above mean sea level), River Mekong flows from north to south, draining South East Asia. For centuries the lower River Mekong (downstream of China) has been the life line in the socio-economics of the riparian countries. A study of the Siem Reap civilization indicated the high dependence of local life on River Mekong throughout history. Direct and indirect benefits of River Mekong to agriculture, fishery farming etc. reached 7100 (US$ millions) in year 2008. However, this benefit did not come without cost in the shape of huge damages caused by River Mekong floods. Inflow into River Mekong is unevenly distributed over the year and mainly concentrated in summer flood seasons, i.e. June to October, due to heavy rains caused by the south west monsoon, further enhanced through superposition of rain from typhoon storm surges. The vulnerability of major flood victims in the riparian countries, i.e., in Lao-PDR, Central Cambodia, and Southern Vietnam has increased in the last few decades because of new settlement along the river and increase in the population density in existing cities along River Mekong, for example, the cities Phnom Penh in Cambodia and Vientiane in Lao-PDR.

Latest research on integrated water management suggest that development of any river should be based on the concept of "living with floods", which has been adopted by the riparian countries of the Mekong river basin. This implies to avoid tampering with the river, and puts emphasis on using non-structural measures, i.e. flood forecast, flood risk management, early warning system etc. to reduce risks to life and property of people, while keeping the traditional benefits of River Mekong.

There is a tradition of flood forecasting in the Mekong River Basin (MRB) as a non-structural means for reducing vulnerability against harmful floods. The former Mekong Committee (replaced in 1995 by the Mekong River Commission (MRC)) initiated a flood forecasting program for the lower MRB in response to severe flooding in 1966. Normally flood forecasts of 1 to 5 days lead time are produced for 12 gauges along the main stem of the Mekong starting from Chiang Sean in Lao-PDR to Kompang Cham in Cambodia. However, hydrologic models for flood forecasting - the formerly used model SSARR, and the model URBS introduced in 2009, fail to produce flood forecasts and warnings of desired accuracy. The main flaw of the adopted URBS forecasting model is its dependency on satellite rainfall input which first is not regularly available in time, and secondly, representativeness of satellite rainfall estimates (SRE) and their forecasts to actual rainfall occurrence's depth and distribution is questionable. Yet another problem of adopted models is their general distributed structure, which requires to model the hydrological characteristics of a watershed for conversion of rainfall to runoff. The absence of detailed physiographic data is a major constraint in efficient simulation and calibration of semi-distributed models. In order to improve flood forecast quality, a better forecasting model is

needed. But the scarcity of available hydro-meteorological and physiographic data limits the applicability of available sophisticated flood forecasting models to the Mekong region. A review of contemporary flood forecasting approaches revealed the feasibility of using data based stochastic models, combined with Nash cascade based semi-conceptual models with variable runoff coefficients for conversion of rainfall into runoff in the reaches between gauging stations, as appropriate in view of the Mekong catchment data scarcity.

In this study the design of such a data based flood forecasting model is described, which can be used for large rivers in general, and for the River Mekong in particular. The efficiency of this flood forecasting model is tested for the River Mekong. To start with, it was found necessary to find a measure by means of which the quality of flood forecasts can be measured in an objective manner. A suitable objective quality criterion is selected, based on a literature survey of different quality criteria. In order to establish benchmarks, this defined quality criterion is used to judge not only the quality of forecasts from the new model, but also of the forecasts produced by existing forecasting models for the Mekong.

Before starting on model building, the available hydro- meteorological data base for River Mekong is analyzed and checked for consistency, in order to facilitate development of a data based model. Keeping in mind both the scarcity of available data and the problems associated with the lack of a dense network of rain gauges, the data based flood forecasting model is designed with minimum data requirement. By adopting a model developing approach moving from simple to complex, a model structure, with limited data requirement was selected in the beginning. With these data a first model, based solely on conventional regression analyses was developed, to assess the potential of hydro-meteorological data in producing flood forecast. Efficiency of each further addition of complexity, with increased data requirement for each step, is weighed against quality improvement furnished by each next step. Moving from a model for flood forecasting that is based only on discharges to a conceptual rainfall – runoff model in the first two steps, the two models are then used in combination using both rainfall and discharge data in a third step. The reliability of the forecasts is documented by quality criteria and probability density functions of the resulting forecasting errors. This allows to identify different confidence intervals for each crisp forecasted value.

2. Description of Study Area

2.1. Mekong Description

Mekong is one of the world's greatest rivers, ranked 10th largest as per length and flow at the mouth (Liu et al. 2009, flood report 2005). The 100 km wide strip of eastern Tibetan Plateau at about 4500 meters above sea level is source area to the Mekong and also of Salween and Yangtze rivers. Starting their journey few hundred km apart, the courses of these rivers separate at the point of origin of Irrawaddy. Salween moves to south west and falls in Indian Ocean after draining part of Myanmar; Yangtze drains central China and flows east. And Mekong in the middle flows toward south east in between the catchments' of Salween and Red rivers. From China downstream it keeps its journey southward and empties into the South China Sea after draining part of Myanmar, Lao PDR, Cambodia, and Vietnam (Clark et al. 2004; Twidale 2004). The long stretched drainage patterns of Salween, Mekong and upper Yangtze are unique in comparison to other large river system that drain the interiors of continents, such as the rivers Amazon, Congo, and Mississippi, which have broad dendritic tributary network. The tributary networks of Salween, Irriwaddy, Yangtze and particularly Mekong River exhibit different drainage patterns due to heterogeneous underlying geological structure. These underlying geological factors control the course of rivers and the landscape they carve out (Twidale 2004; Tandon & Sinha 2007; state of the basin report 2010)

The greater Mekong can be divided into three parts; Lancang Jiang, as the river is called in China has its upper basin in China, the middle Mekong, from downstream of Yunnan province of China to Northern Cambodia, and the Mekong delta, formed by the river before reaching the South China Sea. Middle Mekong and delta are usually combined into the Lower Mekong, which is the area covered by the Mekong River Commission. The upper basin makes up 24% of the total area and contributes on the average15% to 20% of discharge that flows into the Mekong River. Upper Mekong flows for almost 2,200 km from its source before it enters the Lower Basin where the borders of Thailand, Lao PDR, China and Myanmar come together in the Golden Triangle (Overview of the hydrology of Mekong Basin, 2005). Soil erosion has been a major problem here and approximately 50 % of the sediment in the river comes from the Upper Basin. As per statistics of 1998, up to 28 % of the Mekong basin in Yunnan was classified as "erosion prone". Therefore cultivation was restricted in favor of reforestation. In the south of Yunnan, the river slows down as the valley opens out and the flood plain becomes wider.

After leaving China, the Lower Mekong forms the border between Myanmar and Lao-PDR for about 100 km then turns southwest to form briefly the border of Lao-PDR with Thailand. After flowing some 400 km in Lao-PDR from the Lao-Thailand border again for about 850 km as it moves east and then turns south through central south east Asia, passing through the capital of Lao PDR, Vientiane. Lao PDRA lies almost entirely within the Lower Mekong Basin. The Mun River's confluence with the Mekong occurs right before it crosses into Cambodia. Mekong

interacts with the complex Tonle Sap River system as it flows through Phnom Penh, the capital of Cambodia. The seasonal cycle of changing water levels in the Mekong lead to changes in the direction of flow toward and from the Tonle Sap River. High water levels in the Mekong and low water levels in the Tonle Sap lake turn the direction of flow towards the lake, whereas for low water levels in the Mekong in low flow season cause the Tonle Sap river to flow from the lake towards the Mekong. Over all, the length of the Lower Mekong is about 2,600 km from golden triangle to the South China Sea through the complex delta system of Vietnam (State of the basin report, 2010). This complex Mekong delta actually begins in Phnom Penh, Cambodia, where the Mekong divides into Lower Mekong and Bassac. Lower Mekong and Bassac then divide into six and three main channels respectively to form the "Nine Dragons" in Vietnam before entering into South China Sea. The main delta is flooded almost every year because it is less than five meters high above sea level (MRC Annual Mekong Flood Report 2005, 2006).

2.2. Geology of Mekong River Basin

Geology generally defines the catchment shape, boundaries and outlet together with River slope. Mekong originates from Guyong-Pudigao creek at 5160 m above sea level near the foot of Mt Jifu on the Tibetan Plateau (Liu et al. 2007). In its journey from Guong-Pudigao creek to South china Sea it descend about 5 km in 4600 km (Fig.2.1 and 2.2). The slope does not remain constant but changes along the longitudinal section of River (Fig.2.1). The Tibetan plateau is the most densely glaciated area of the world, of which 316 Km^2 glacier area is attributed to Mekong Basin (Eastham et al. 2008). The course of the Mekong on this plateau is influenced by the tectonic fabric of the Tibetan Plateau. Here mainstream and major tributaries run parallel in NNW to SSE direction (State of the Basin report, 2010)

Fig. 2.1: Elevation and Geological formations of Mekong (Source, Encyclopedia Britannica, 1994)

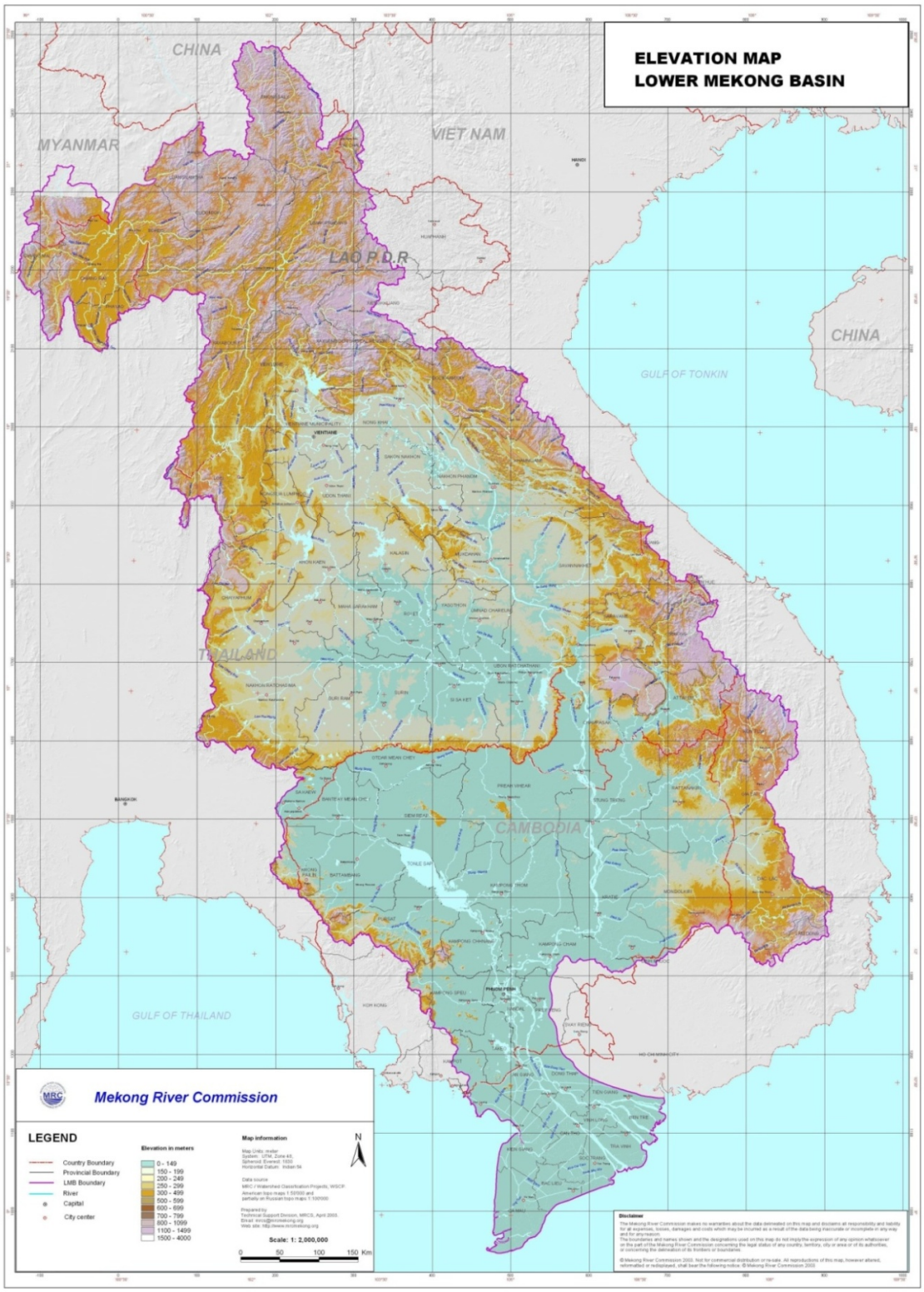


Fig. 2.2: Elevation map of lower Mekong (Source, MRC)

In three rivers area, Mekong is confined in deep narrow sections where the depth of the river goes up to 2500 m at some places as it passes through broad north-south trending arc for about 500 km. The Upper Mekong basin broadens at the point where the Salween and Yangtze diverges to west and east respectively. This area is called Lancang Basin.

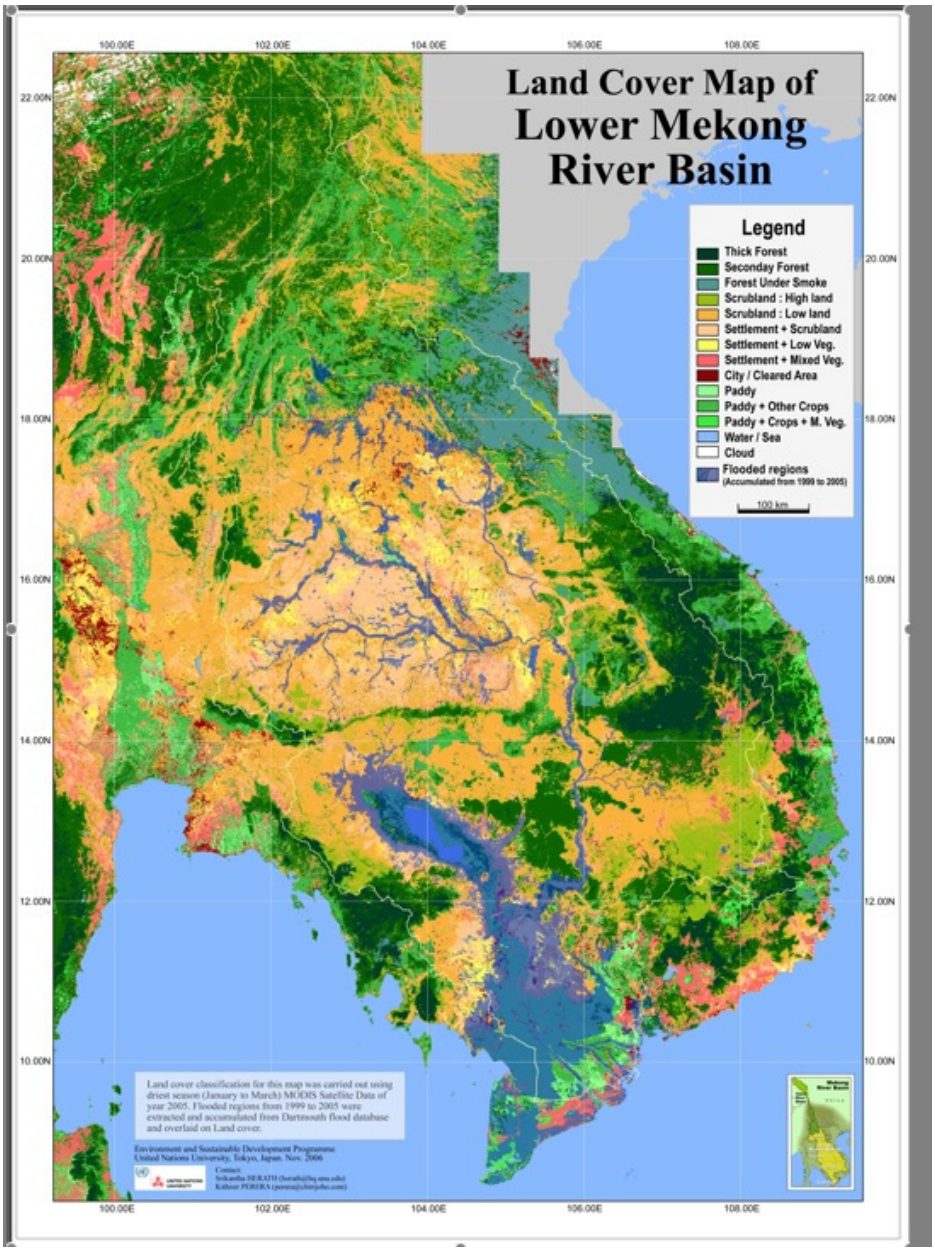


Fig. 2.3: land cover map of lower Mekong (Source, Saito et. al. (2007))

After the river passed through the golden triangle, the northern Highlands of northeast Myanmar, northern Thailand and northern Lao-PDR restrict the Mekong into steep-sided bedrock channels. The river continues its journey from Northern Highlands to the saucer-shaped Khorat Plateau at

an elevation of 300 meter above sea level, forming a border between Lao-PDR and Thailand running from north to south. The rim of the basin is comprised of sharp cuesta formed by the highly resistant sandstones of the Khorat Group. The Plateau is bounded by the Loei – Petchabun fold belt from the west and by the Annamite Mountains in the east. In this plateau, the low range hills which run from NNW to SSE, known as "Phuphan uplift" divides the basin in to two sections, Savankhat basin to the north and the Mun/Chi Basin in the south (State of the Basin report, 2010).

The Mekong enters the Tonle Sap basin just north of Pakse and flows in between the Khorat mountains in the west and the Boloven Plateau in the east. At the southern end of this stretch, the mainstream breaks up into a complex network of branching and reconnecting, with islands in between. The Mekong in its passage from southern Lao-PDR to the alluvial plains of northern Cambodia, passes through a series of cataracts at Khone Falls (Gupta and Liew, 2006, State of the Basin report, 2010). The river flows further southwards until Kratie, where it takes a right angle turn towards west due to the upland formed by the extensive basaltic lava flows near Ho Chi Minh City. The mountainous terrain that runs from north to south in the shape of a broad arc from the Boloven Plateau in Lao PDR to the volcanic uplands of southern Vietnam forms the extreme border to the sub basins of the tributaries which are entering from the eastern side into the Mekong mainstream (State of the Basin report, 2010).

The Mekong river delta forms downstream of Phnom Penh. The delta plain can be divided into two parts, the low- laying inner delta which is close to sea level, yet has its topography dominated by fluvial processes, and the outer delta, which is built by coastal plain deposits, and is dominated by marine processes (Te et al., 2002 a).

2.3. Land Cover and Land Use

"Land cover is the observed (bio) physical cover on the earth's surface. In a very pure sense it should be confined to describe vegetation and man-made features. Consequently, areas where the surface consists of bare rock or bare soil are describing *land* itself rather than land *cover*. Also, it is disputable whether water surfaces are real land cover. However, in practice, the scientific community usually describes those aspects under the term land cover. (http://www.fao.org/docrep/003/x0596e/X0596e01e.htm#P213_18188). From the hydrologic point of view, both land and land cover are important because they, along with soil type, define the catchment response to rainfall, both by magnitude and its temporal distribution.

The Mekong River basin is mainly comprised of ever green forest, deciduous forest and grass land. Although the boundary between evergreen, mixed deciduous and deciduous forest ascertained by satellite imagery is more or less fuzzy, it still gives the general features of land cover with a certain degree of confidence. The study conducted by Saito et. al. (2007) mentioned that evergreen forests predominate the land cover of the Mekong Basin. The upper and middle

Mekong together with the eastern portion of the lower Mekong is mainly covered with evergreen forests. Deciduous forests are located mainly in the northern and the central Thailand area. Grass lands together with crop fields cover the south western portion of Mekong river catchment. These areas are mainly the low lands of Thailand and portions of northern Cambodia (Saito et. al., 2007). Along with the forest cover there is a substantial portion of Mekong catchment covered with permanent and temporary water bodies that includes lakes, ponds, wetlands and marshes etc. The biggest of these is the Tonle Sap lake in Cambodia. Land cover maps as produced by ESDP (2006) indicate the presence of thick and secondary forests in the catchments of Sekong, Sen San and Sre Prok, extending further northward from the border of Cambodia to Lao-PDR (Fig.2.3). This heterogeneous land cover condition in different sub-catchments of Mekong is one of the main reasons behind the runoff coefficient variability in different areas of the River.

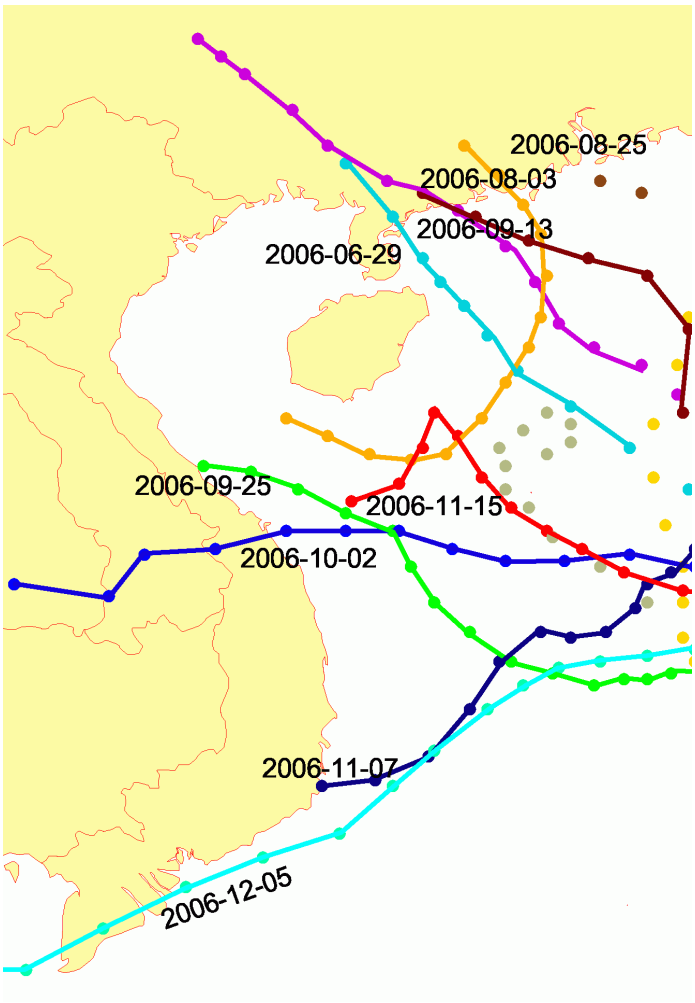


Fig. 2.4: Typhoon tracks along lower Mekong

2.4. Climate

The climate of the Lower Mekong Basin (LMB) is dominated by the Southwest Monsoon, which generates wet and dry seasons of more or less equal length (Overview of the hydrology of Mekong Basin, 2005). In the wet season, there are usually heavy rains in most parts of the basin. The flood period in the LMB is mainly caused by the Southwest monsoon season which usually lasts from May until late September or early October. July, August and September are generally the months of highest rainfall, although there is

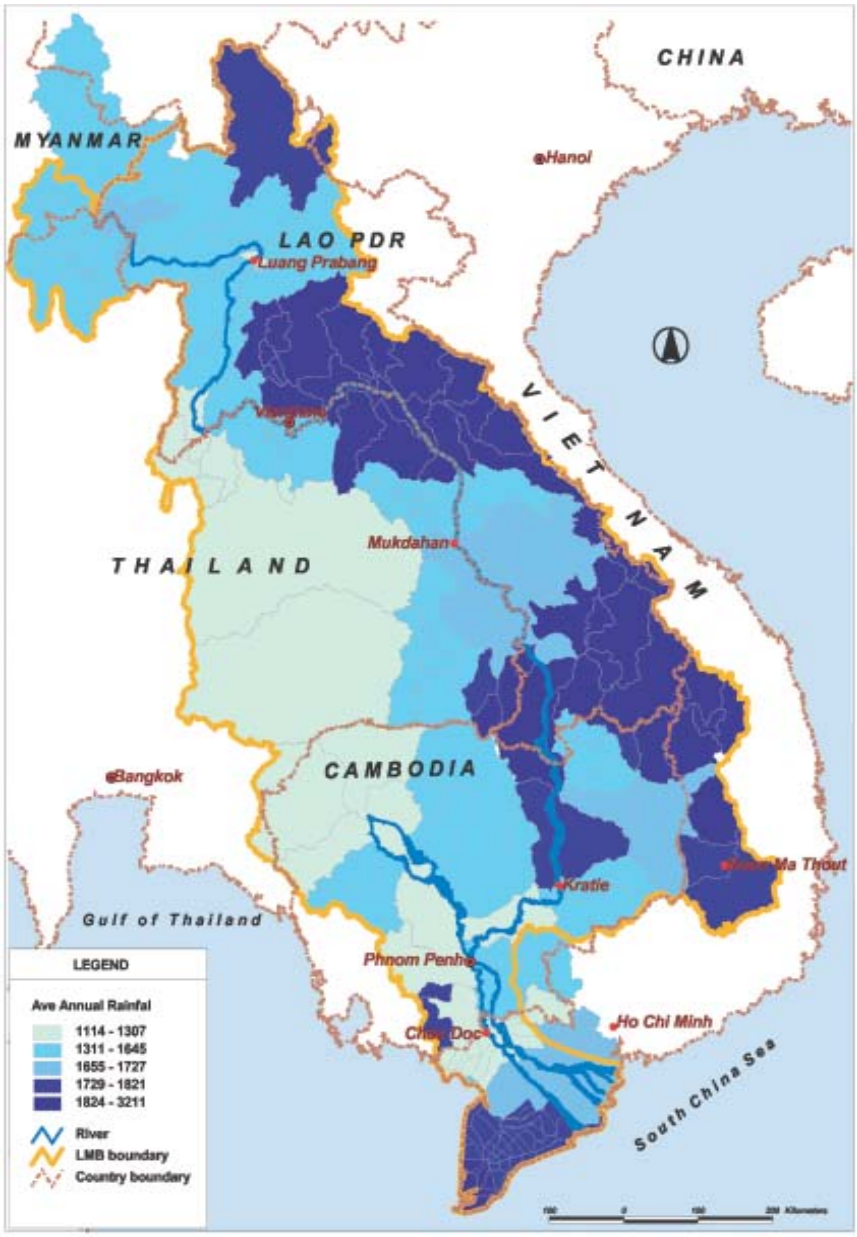


Fig. 2.5: Rainfall distribution in Lower Mekong Catchment (Source, hydrology report-05, MRC)

evidence of a shift later in the season in Cambodia and in the deltaic region, where more rain falls in September and October. Later in the season, tropical cyclones are responsible for making August, September and sometimes even October the wettest months of the year in the delta region as shown in Fig.2.4 (MRC Annual Mekong Flood Report 2005, 2006).

Annual average rainfalls over the Cambodian floodplain and the Mekong delta are equally low, which is less than 1,500 mm. The highest rainfall is in the Central Highlands and within the mainstream valley around Pakse. The distribution of mean annual rainfall over the Lower Basin is given in Fig.2.5 as reproduced from (Overview of the hydrology of the Mekong Basin,2005). This Fig. shows that left bank tributaries are receiving much higher rainfalls than right hand tributaries of Thailand. This trend makes the sub-catchments of Central Highlands of Lao PDR between Vientiane and Nakhon Phanom, along with Pakse and Kratie in Cambodia, the major contributing catchments by receiving about 1800 mm of rainfall per annum. The range of rainfall in the Thai sub-basin starts from less than 1,500 mm and goes up to 1700mm, and only a very small fraction of the area receives over 1800 mm. The map clearly shows that the left bank tributaries of Lao PDR receive much more rainfall than right bank tributaries.

Tab. 2.1: Mean monthly rainfall in different sub-regions of lower Mekong (Source, hydrology report-05, MRC)

Month	Northern Region	Central Region	Korat Plateau	Central Highlands	Cambodian Floodplain	Vietnam Delta
	Chiang Rai	Pakse	Khon Kaen	Pleiku	Phnom Penh	Chau Doc
Jan	13	2	5	6	8	8
Feb	10	7	15	6	3	3
Mar	20	20	35	25	15	15
Apr	85	70	60	85	65	75
May	190	220	170	225	115	165
Jun	210	380	180	350	125	110
Jul	310	390	160	360	160	140
Aug	390	500	185	460	160	170
Sep	280	320	260	360	265	160
Oct	140	100	120	220	255	250
Nov	60	20	10	75	130	160
Dec	20	3	3	20	20	40
ANNUAL	1,730	2,050	1,210	2,200	1,320	1,300

The Lower Mekong Basin has been classified into six sub regions according to spatio-temporal variation of rainfall in (Overview of the hydrology report of the Mekong Basin, 2005). Table 2.1 compares long-term averages of rainfall of these six sub regions. According to this table, the area of maximum rainfall are the central region and the central high land, where rainfall exceeds 2000

mm annually with monthly averages of almost 500 mm in the month of August. On the other hand the annual average rainfalls over the Cambodian floodplain, Korat Plateau and the Vietnamese delta is less than 1,500 mm. Rainfall in Northern region is little lower than central region's rainfall with the yearly average of 1700 mm. Amount of rainfall approximately doubles from April to May in each of the region,. The average rainfall keeps on increasing from May to August, where August receives highest rainfall. Downward trends start from August to September and rainfall amounts decrease drastically in the month of November.

Temporal pattern of average monthly rainfall on the left bank sub-catchments of Mekong for the years 1990-2005 are plotted in Fig.2.6. On the average, rainfall remains high from June to September, with maxima in July or August. The maximum average rainfall occurs in the contributing sub-basins between Nakhon Phanom and Mukdahan (NM), and between Pakse and Stung Treng (PS) whereas the minimum average rainfall occurs in the contributing sub- basin between Luang Prabang and Vientiane (LV). The range of rainfall from minimum maxima to maximum maxima among Mekong left bank catchments is as high as 300 mm.

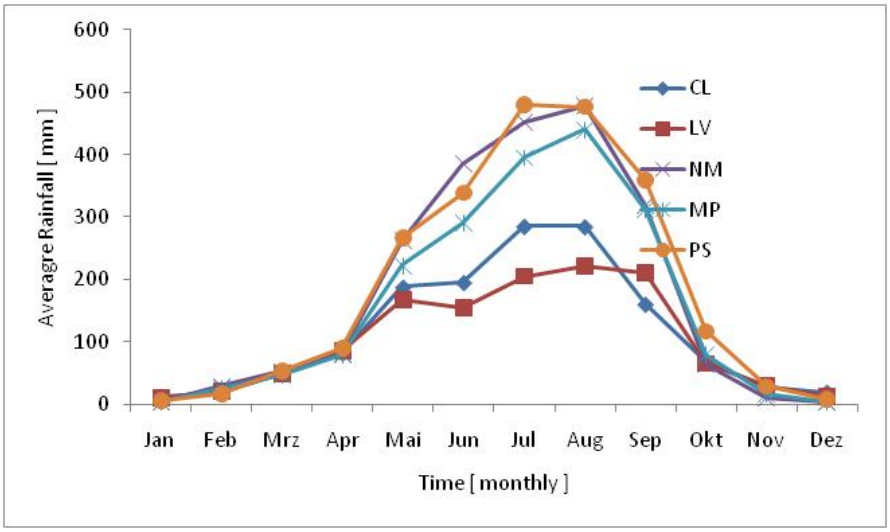


Fig. 2.6: Spatio-temporal rainfall distribution

As per (Overview of the hydrology report of the Mekong Basin, 2005), tropical storms and cyclones have a strong effect on the rainfall climate of the Basin. This effect shows up as a double peak in the rainfall distribution over most of the Lower Basin during wet years, and the concentration of maximum rainfalls during the last quarter of the year 2005 in Cambodia and Viet Nam. The influence of cyclones is not widely recognized. Data on the wider regional occurrence of tropical cyclones show that over Central and Southern Viet Nam they are most frequent between September and November and are largely responsible for the higher rainfalls occurring in these later months.

The above average cyclone activity has wider impacts on the rainfall climate of north of Viet Nam and Cambodia. Cyclones are further responsible for late wet season rainfall peaks. During tropical cyclone periods, the probability that any given day will be wet (> 1 mm) reaches a maximum over most of the Lower Basin. This happens even in the relatively dry parts of Northeast Thailand. The Southwest Monsoon, combined with severe tropical storms, has been the cause of flood disasters in the Lower Mekong, particularly in Cambodia and the delta. But the major impacts of such weather systems are not confined to the southern part of the Basin. In 1966, Typhoon Phyllis was responsible for the most extreme flood recorded at Vientiane since 1913 (Overview of the hydrology report of the Mekong Basin, 2005).

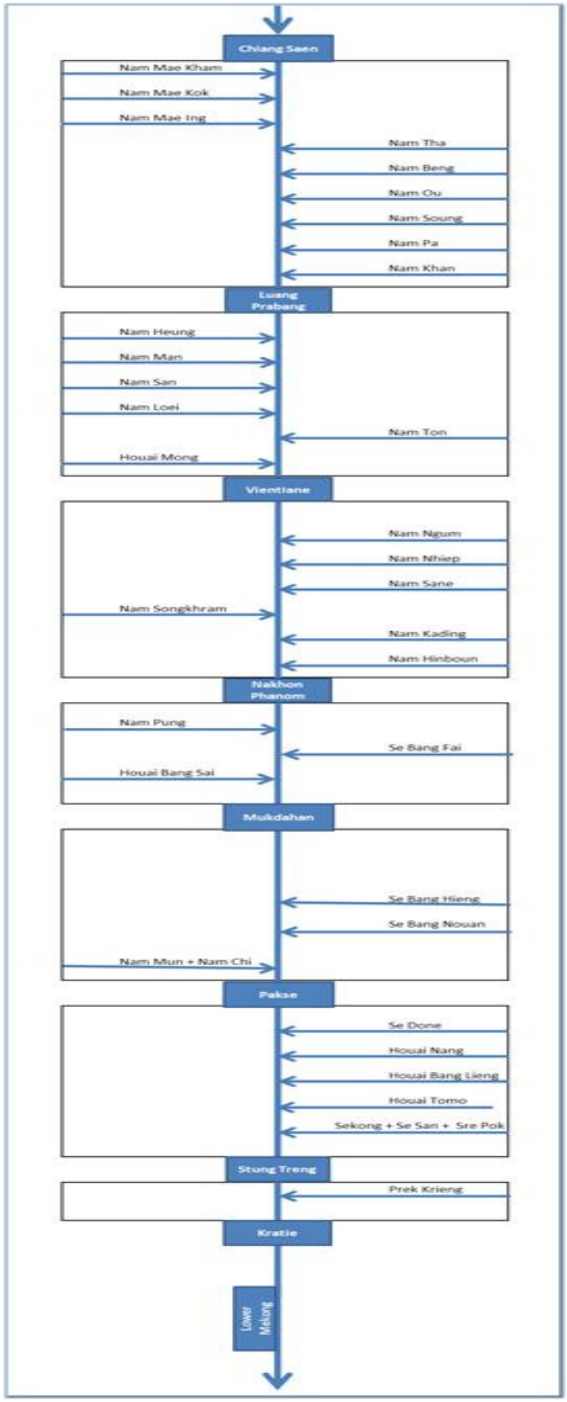


Fig. 2.8: Schematic diagram of lower Mekong with laterals

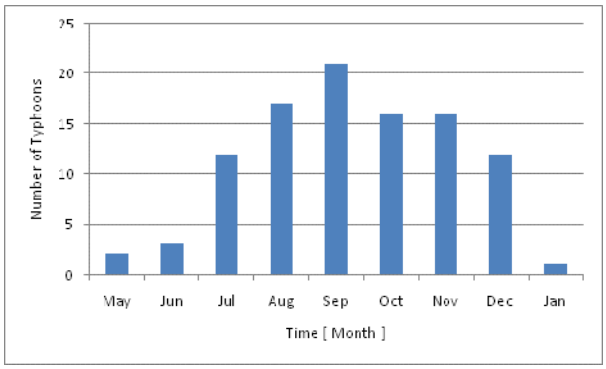


Fig. 2.7: Typhoon storms frequency in Lower Mekong

The analysis of typhoon frequency for the period of 1991 to 2008 conducted by Pradeep et. al. (2010) indicates that June receives the maximum number of typhoons followed by May and July. This does not support the findings of hydrology report (2005) according to which the maximum typhoon intensity occurs between September and November (Fig.2.7)

There is not much difference of mean temperatures between lowlands and river valleys of the Lower Basin. However, there are significant changes, both from season to season and from day to night at increasing altitudes and in the more temperate climates to the north.

Mean summer, i.e. March to October

temperatures are similar within the Lower Basin from Phnom Penh, Cambodia to as far north as Luang Prabang in Lao PDR and Chiang Rai in Thailand. The summer temperatures are a little lower at altitudes above 500 masl. At almost 2,500 masl in the Upper Se San sub basin in Viet Nam, station Pleiku has mean summer temperatures that are only 2° to 3° C lower than those typical of the Mekong lowlands. Winter mean temperatures decrease significantly from south to north, from 26° - 27° C in Phnom Penh to 21° - 23° C in Chiang Rai.

In the entire lower Mekong Region the annual rates of evapo-transpiration, remain constant inter-annually within the range of 1 and 2 meters. This range of evapo-transpiration makes the Korat Plateau in Northeast Thailand, mainly the Mun and Chi Basins, with annual rainfall of 1200 mm, one of the driest areas in Southeast Asia. In many climate classification systems this region is defined as semi-arid. Lack of soil moisture in the area becomes critical during the late dry season from February to April. Further south in the Cambodian and Vietnamese parts of the basin, the annual evaporation rates are 1,500 to 1,700 mm. To the north at Chiang Rai, the rate is around 1,400 mm.

2.5. Hydrology of the individual reaches of the Mekong

The topography defines the hydrology and flow pattern of Mekong. Most of the inflow comes from the sub-catchments located in the Northern Mountains in Lao-PDR and in the mountainous regime along the border of Lao-PDR and Vietnam (Fig.2.10). The schematic diagram of the lateral tributaries joining Lower Mekong in between discharge gauging stations is shown in Fig.2.8. And the Fig.2.9 shows the longitudinal elevation profile of Lower Mekong.

2.5.1. Chiang Saen to Luang Prabang

The lateral sub-catchments of this reach are almost entirely mountainous and covered with natural forests. This reach consists of a complex of right and left bank tributaries. Among these laterals, Nam Mae Kok with an area of 10,870 km^2 from right and Nam Ou with an area of 25,810 km^2 from left are significant.

From Chiang Saen to Luang Prabang, the Mekong covers the distance of 353 km and drops a height of 90 m (Fig.2.9). The total right and left bank tributaries of this reach contribute 10% to the total Mekong river flow at gauge Stung Treng. Mean annual rainfall in this reach ranges from 1500 to 2000 mm.

2.5.2. Luang Prabang to Vientiane

With 5 right and one left bank tributary this reach has a length of 427 km with a slope of 0.00025 from Luang Prabang to Vientiane. The proportion of the total Mekong flow up to Stung Treng of

the flow from this reach is 4%. Unlike all other reaches of Mekong, the hydrology of this reach is dominated by right bank tributaries. Flash flooding in the lateral tributaries of this reach are attributed to deforestation coupled with the culture of slash and burn agriculture.

Up to Vientiane, the flow of the main Mekong River remains dominated by the Yunnan component of flow from China. Between Chiang Saen to Vientiane, no big lateral tributary enters into the Main River to alter this dominance.

Flash floods are most common in Northern Thailand and northern Lao-PDR with catastrophic events particularly in year 2000 and 2001. Two distinguished weather processes typical of this area, have been indicated in (MRC AMFR 2008, 2009): first, typical monsoonal storms in combination with average antecedent moist catchment, and secondly extreme weather systems such as intense tropical depression, or severe tropical storms in combination with saturated antecedent moisture conditions. However, the flood peaks due to the latter conditions are rapid and of short duration, less than a day.

2.5.3. Vientiane to Nakhon Phanom

There is a significant discontinuity between the hydrological sub-regions upstream and downstream of Vientiane (MRC AMFR 2007, 2008). Downstream of Vientiane, the impact of the Yunnan flow component of the flood discharges is highly reduced by the large left bank tributaries of Lao PDR. Downstream Vientiane, hydrologic conditions change because of contributions from major left bank tributaries as result of heavy rainfall in the central high land of Lao-PDR. In this 368 km long reach almost 20% of the total Mekong flow is produced, bringing it to the top with respect to flow contributions. Tributaries Nam Ngum 50 km downstream of Vientiane in combination with Nam Kaeding brings the largest part of total lateral inflow. The dam on Nam Ngum is another distinctive feature of this reach, making the natural lateral flow subject to reservoir operation. In this reach, these left bank tributaries mainly exert their influence on the flood hydrology of Mekong.

In contrast to the reach from Chiang Saen to Vientiane, flash floods do not dominate the flood hydrology of this stretch. In recent years the events of 2000, 2001 and 2002 were larger floods both in terms of peak and volume in this reach (MRC AMFR 2007, 2008).

2.5.4. Nakhon Phanom to Mukdahan

Se Bang Fai from the left and Houai Bang Sai from the right with respective catchment areas of 10,240 and 3500 km^2, are the major lateral tributaries in this reach of 91 km. Total contribution of this reach to the Mekong river flow at Stung Treng is small i.e. 4% , mainly due to low rainfall on both right and left side areas that range from 1250 to 1500 mm per annum. This reach may have an additional problem of negative flow in the winter, i.e. seepage as base- or ground water flow from the lateral catchments and evaporation losses in the winter exceed the local gain.

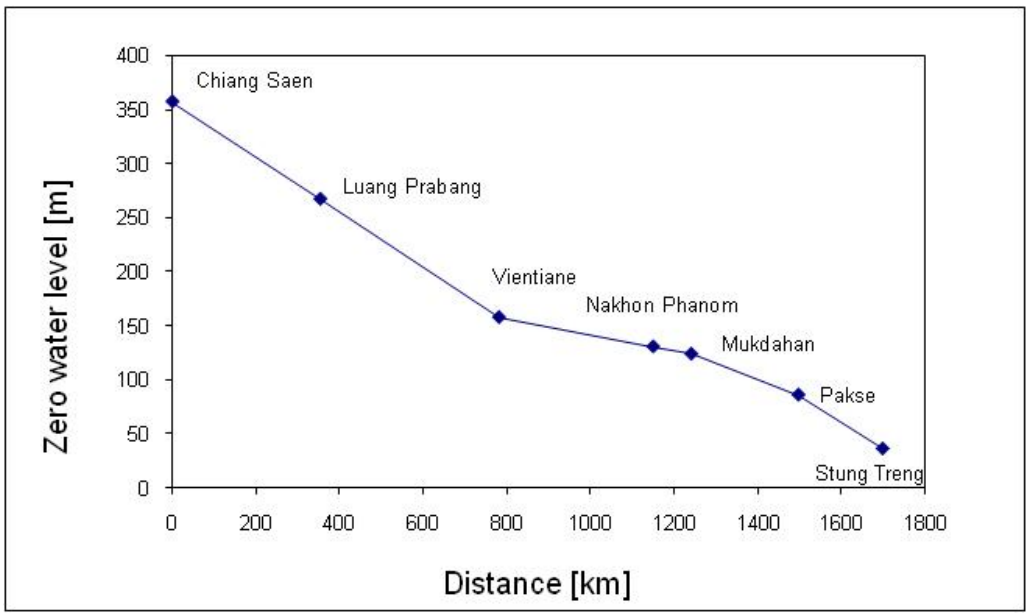


Fig. 2.9: Longitudinal elevation profile of Lower Mekong

2.5.5. Mukdahan to Pakse

Mukdahan to Pakse reach is comparatively short with a total length of 256 km. This reach is unique in the sense that on the right side,the huge Nam Mun/ Nam Chi system with a catchment area of 120,000 km^2enter just above Pakse, after draining southern Thailand, however with very small contribution in terms of lateral inflow to the total flow of Mekong. On the left side the main contributor is Se Bang Hieng with a catchment area of 19,300 km^2.

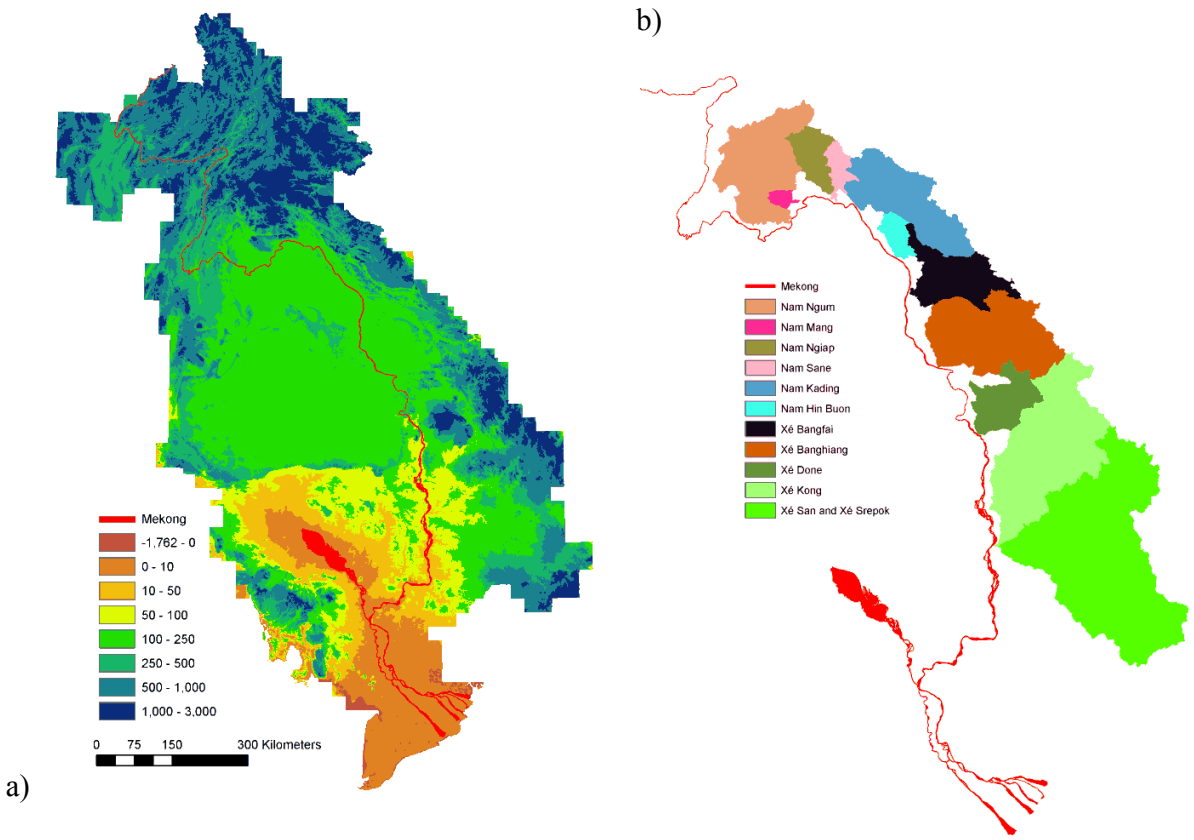


Fig. 2.10: a) Topography and Mekong River with Tonle Sap and Major delta branches. b) Left sided tributaries which have major contribution to the Mekong due to the border mountains to Vietnam.

2.5.6. Pakse to Stung Treng

The large river systems of Se Kong, Se San, and Sre Pok on the left bank control the hydrology of the Mekong in this reach. A detailed hydrological analysis of the flood hydrology of these tributaries is constrained by the lack of availability of data. However, the contribution of these Rivers to the mainstream flood is significant as can be seen for the example of the year 2000 flood. In this 201 km long reach a discharge of 3000 m^3/sec on the average is generated which adds 23 % to the total average Mekong flow. The main reason of this inflow is again heavy rainfall in Se Kong and Se San catchment which goes up to 2500 mm per annum. This region is also influenced by typhoon storms in the late summer months. The overall impact of this reach in defining the average flow characteristics of Main Mekong, which does not change much from Stung Treng to Phnom Penh, because on the average there is only very little lateral inflow into Mekong downstream of Stung Treng.

2.5.7. Stung Treng to Kratie

Comparatively small inflow is contributed by tributaries downstream of Stung Treng. Only Prek Krieng enters from left side in the main River. But the impact of Prek Krieng is hard to ignore in summer flooding months when local storm outpours may result in lateral inflows, that are as much as10,000 to 15,000 m^3/sec.

2.5.8. Downstream of Kratie

The gauge site at Kratie is the point where the hydrological system develops into an almost completely hydraulics dominated river system. Until then, about 90 % of the water inflow has occurred, and outflow starts to decrease the runoff volume (AFMR 2007, 2008).. From there on, overland flow and the reversing hydrodynamic system of the Tonle Sap has to be accounted for, so that hydrological forecasting for this region has to provide the inflow at Kratie and hydraulic models need to be used for forecasting further downstream. The seasonal cycle of changing water levels at Phnom Penh results in "flow reversal" of water into and out of the Great Lake via the Tonle Sap River. During the flood months, water flows up the Tonle Sap from the Mekong mainstream into the Lake. As the water level decreases in the mainstream in late September, water flows out of the Tonle Sap down into the Mekong mainstream (AFMR 2007, 2008). The main contributing sub-basins of River Mekong with their topographic profile are shown in Fig.2.10. The cross-section of River at Kratie is narrow in comparison to the River cross-sections at upstream gauges of Mukdahan and Pakse (Fig.2.11). This is unique feature because normally the river section normally widened as it enters in the flat lands.

2.6. Mekong River Basin Floods

All river floods are natural phenomena that occur recurrently within hydrological time scales. They are characterized by increased water levels in river channels leading to overspill of natural banks or artificial embankments and subsequent inundation of the surrounding flood plains (White, 2000). River floods in tropical region are characterized by rapid increases of water levels, due to rainstorms of high intensity.

Floods in the lower Mekong River occur in the summer month between May to October. They are caused by extensive orographic rainfall occurrences as a result of south-west monsoon, amplified by the downpour of tropical storms moving from the east over Gulf of Tonkin to the northwestern Mekong catchment. Most of these tropical storms occur in the summer months of June to October on the central mountain range, which run from north to south and form the border of Lao-PDR and Vietnam. Therefore, mainly the left bank tributaries contribute to the Mekong main stream flood.

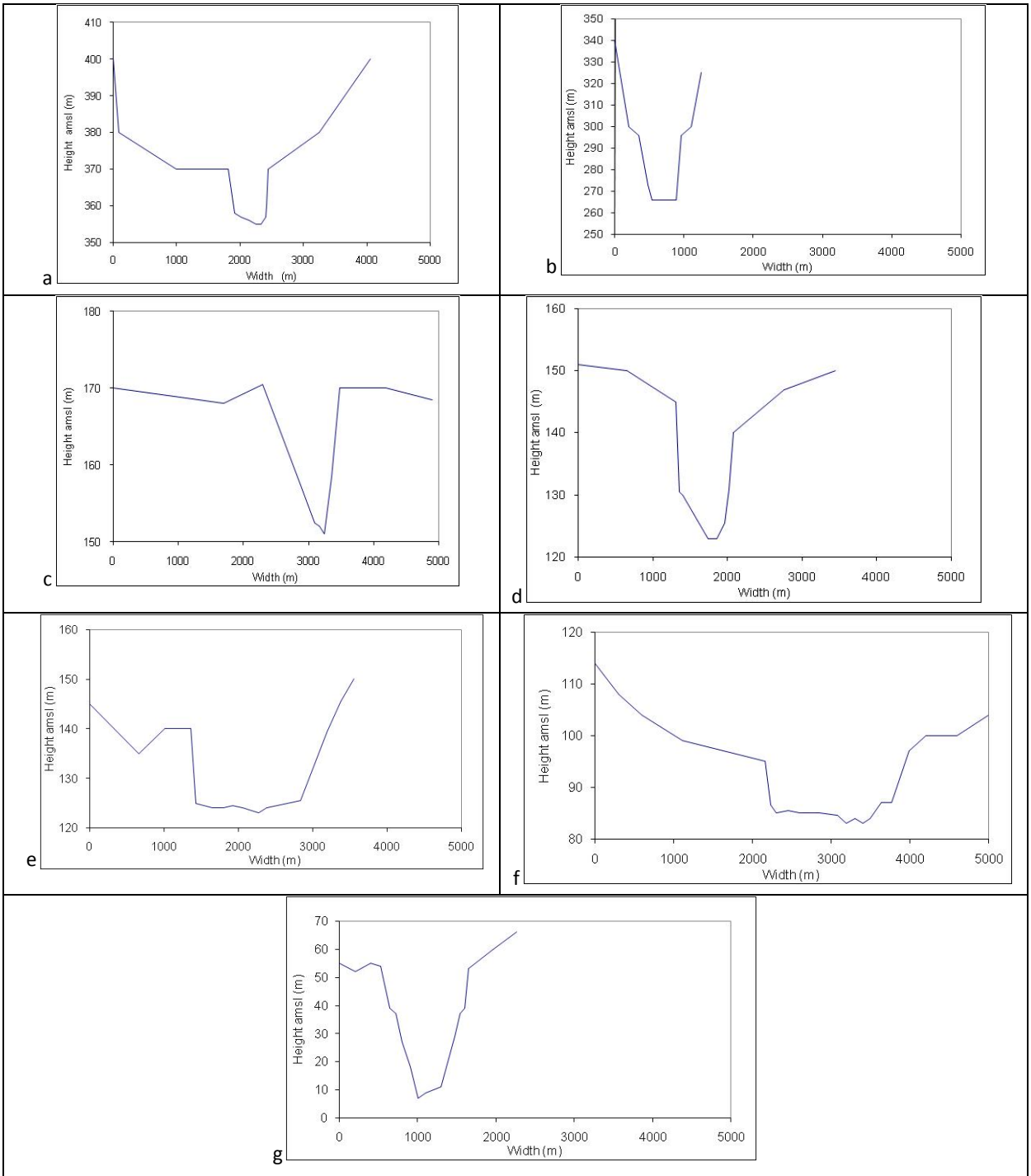


Fig. 2.11: Cross sections of lower Mekong at gauging stations; a) Chiang Saen, b) Luang Prabang; c) Vientiane; d) Nakhon Phanom; e) Mukdahan; f) Pakse and g) Stung Treng

The floods in Mekong basin can be subdivided according to place and mode of occurrence into five categories; flash floods in the tributaries, floods in the main channel of the Mekong River,

combined Tonle Sap and Mekong floods, rainfall floods on Cambodian plains, and floods in the Mekong delta (MRC Annual Flood Report, 2005).

2.6.1. Flash floods in side tributaries

Short duration, few hours to few days, flash flooding mainly occurs in the tributaries of Mekong River due to intense and/ or long periods of rain - in particular in the mountainous areas of LMB. The rapid rise of discharge in the lateral narrow streams sometimes also affects the water level of the main River. These floods are largely harmful to people, cattle and infrastructure in the impacted area. Mitigation measures to limit damages mainly should include regulating land use/ cover and effective flood forecasts, based on local rainfall runoff modeling supported by good rainfall forecast. Dams and reservoirs built upstream of the flash flood risk areas may contribute to the reduction of flash floods.

2.6.2. Floods in the Main Mekong River

The main River floods are caused by large scale rainfalls over large parts of the river basin leading to high water levels in the mainstream resulting in overflows of lower sections of the flood plain. The flooding time when water levels are higher than critical flood levels normally lasts from one to two weeks. In the flood months from June to October, the Yunnan component flow from China is superimposed by lateral inflows from left bank tributaries of the lower Mekong that result into overflow of the river in number of different sections. The cities located on the Mekong river banks, for example Vientiane in Laos and cities in Cambodia, suffer directly from this over bank flow. Normally, some overflow is considered beneficial for watering crops. However, long duration high water levels of flood waves affect crops adversely. Even small overtoppings are considered harmful in the densely population regions along Mekong, if they are not protected by flood embankments. Apart from human life loss, damages due to these type of floods are caused to infrastructure, crops, live stock and private homes (MRC AMFR 2008, 2009)

Casualties and loss of cattle can be reduced by efficient flood forecasts. Mitigation measures for limiting damage from such floods would consist mainly of a well functioning early warning system, regulated land use, diversion of flood waters to wetland areas where possible, and of building embankments for protection of urban areas or areas with important assets.

2.6.3. Floods in Eastern Lao-PDR

The low land of the Lao-PDR is particularly affected adversely by damage caused by Main Mekong overflow, although the overtopping frequency of flood wave at Vientiane and

Mukdahan is less than 10 years. Same is the case for Chiang Saen which is further upstream at the border of Myanmar to Lao-PDR.

2.6.4. Floods in the Cambodian flood plain

There are three very different flood regions in Cambodia: the first is the region from Pakse to Phnom Penh, which is dominated by floods from the upper Mekong. The second is the region of floods in the Tonle Sap area, and the flow downstream of the confluence of Tonle Sap and Mekong and the splitting up of the Mekong into Bassac and lower Mekong which are very much influenced by the Tonle Sap. Downstream lies the third region,the delta, with an entirely different regime of the floods .

2.6.5. Floods in the Mekong Delta

The areas inundated by water from the Bassac River, the Mekong River and the numerous natural and artificial canals linking the two rivers that flow laterally to low lands from Phnom Penh onwards mainly characterize the Mekong delta floods, together with the floods due to tidal motion of the sea. Sometimes, high water levels in the Bassac / Mekong systems may not drain easily to the sea due to tidal backwater. This was the case during the severe 2000 flood in the downstream provinces of the Mekong Delta (MRC Annual Flood Report 2005, 2006).

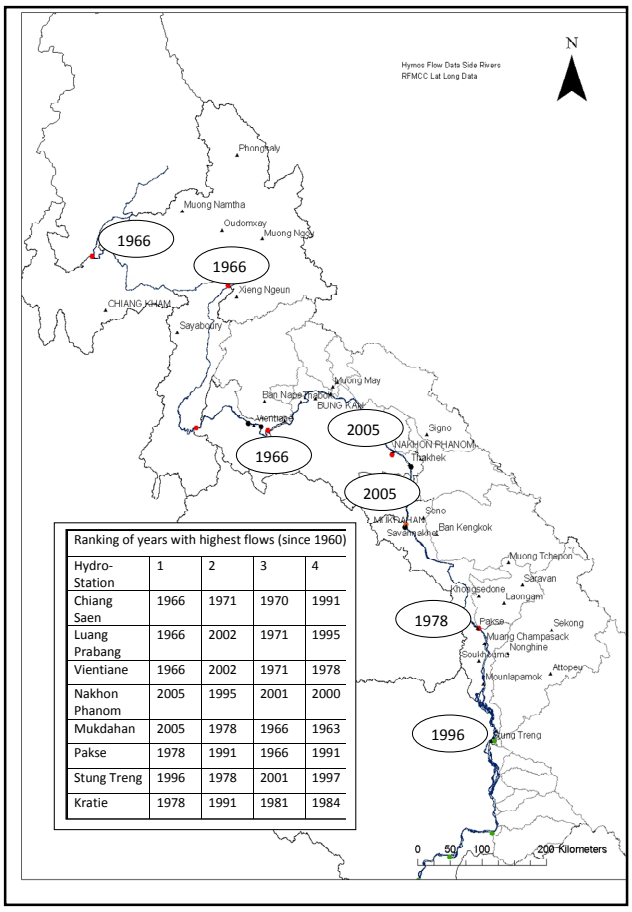
2.7. Spatio-temporal Analysis of Mekong Flooding

Because the lower Mekong is a very long river, affected by variable incidences of weather systems such as south western monsoons and eastern typhoons of different severity level along its length, it is not necessary that it would cause flooding simultaneously along the full stretch from Chiang Saen to Kratie. This is the reason that historical flood at different forecasting stations were recorded in different years (Fig.2.12). Flood waves move from upstream to downstream causing critical situations at some points, whereas no flood occurs elsewhere. This selective flooding of different Mekong river stretches occurs due to a number of reasons, such as variable conveyances of different sections along the river, synchronization of lateral tributary flows, or flood wave attenuation because of over flow into the flood plains.

A historical analysis of flooding shows that there were some years with serious flooding in the upper part of Lower Mekong with low to medium floods further downstream. For instance, the year 1966 was the worst year in the upper part of the LMB, with its effects extending to Pakse. Year 1978 was the year with the most serious floods for the area south of Mukdahan up to the confluence with the Tonle Sap River in the flood plain of Cambodia. Year 1996 a severe flood was recorded, but it was limited to the area of Stung Treng at the confluence with the Sekong River. Years 1961, 1966 and 2000 were nearly equally devastating years for the Mekong Delta,

although 2000 was seen as particularly damaging and remains memorable to all the people living in the floodplains. Year 2005 was the most severe flood for the central area of Lao PDR and Thailand (Nakhon Phanom, Thakek, Mukdahan and Savannakhet). In second ranking: years 1971, 1974, 1984, 1991, 1995 and 2002 may also be considered as severe for one or another section of the LMB (MRC Annual Flood Report 2005, 2006).

Apart from the flood peak, the magnitude of flood volume is a useful indicator of flood intensity, especially in large rivers like Mekong. The historical geography of the annual flood regime, as reflected by flood magnitude of the Mekong (1960-2000) is presented in terms of matrix in a Fig.2.13 (MRC Annual Flood Report 2005, 2006). This Fig. reiterates the evidence of variability of flooding along the length of Mekong River.



Ranking of years with highest flows (since 1960)				
Hydro-Station	1	2	3	4
Chiang Saen	1966	1971	1970	1991
Luang Prabang	1966	2002	1971	1995
Vientiane	1966	2002	1971	1978
Nakhon Phanom	2005	1995	2001	2000
Mukdahan	2005	1978	1966	1963
Pakse	1978	1991	1966	1991
Stung Treng	1996	1978	2001	1997
Kratie	1978	1991	1981	1984

Fig. 2.12: Historical floods along lower Mekong

2.8. Mekong River Floods in a Global Prospective

In contrast to the flood hydrology of Temperate Zone Rivers, such as the Chenab in Pakistan, Rio Uruguay in South America, and River Thames in the UK etc, with seemingly random flood pulses throughout the year, tropical rivers like Brahmaputra, Ganges, Yangtze, Huang He and Mekong produce immense volumes of runoff in the shape of single annual flood hydrograph.

In large River basins like that of the Mekong the relatively rapid variation in flows are smoothed out as the catchment scale increases as the longer duration responses due to each storm episode coalesce, resulting in the highly coherent hydrograph in a flood season. This convergence, and the associated accumulation of monsoonal flood runoff into a single seasonal hydrograph places the Mekong amongst the global river systems within which the largest floods have been recorded (MRC, 2006).

Fig.2.13. shows the location of these large river basins in latitude and the unit runoff produced in each basin. In the list of river basins exceeding 500,000 km^2 (O' Connor, 2004), the Mekong falls into the category of global extreme, as far as per peak discharge per unit area is concerned. This is due to the historical maximum discharge event which occurred on the 3rd September,1939 at Kratie (Fig.2.13). However, this event seems to be not unique, as almost the same maximum was repeated in 1978 and 2000.

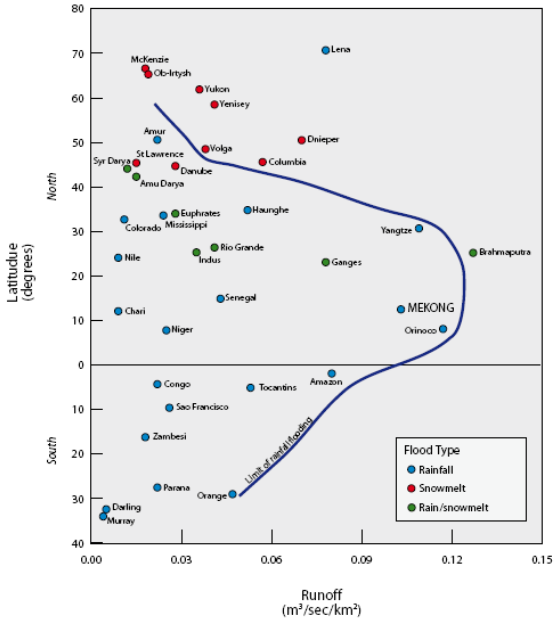


Fig. 2.13: The largest meteorological floods observed for global River basins exceeding 500,000 km2, (Source reproduced in MRC Report, 2007 from O' Connor, 2004)

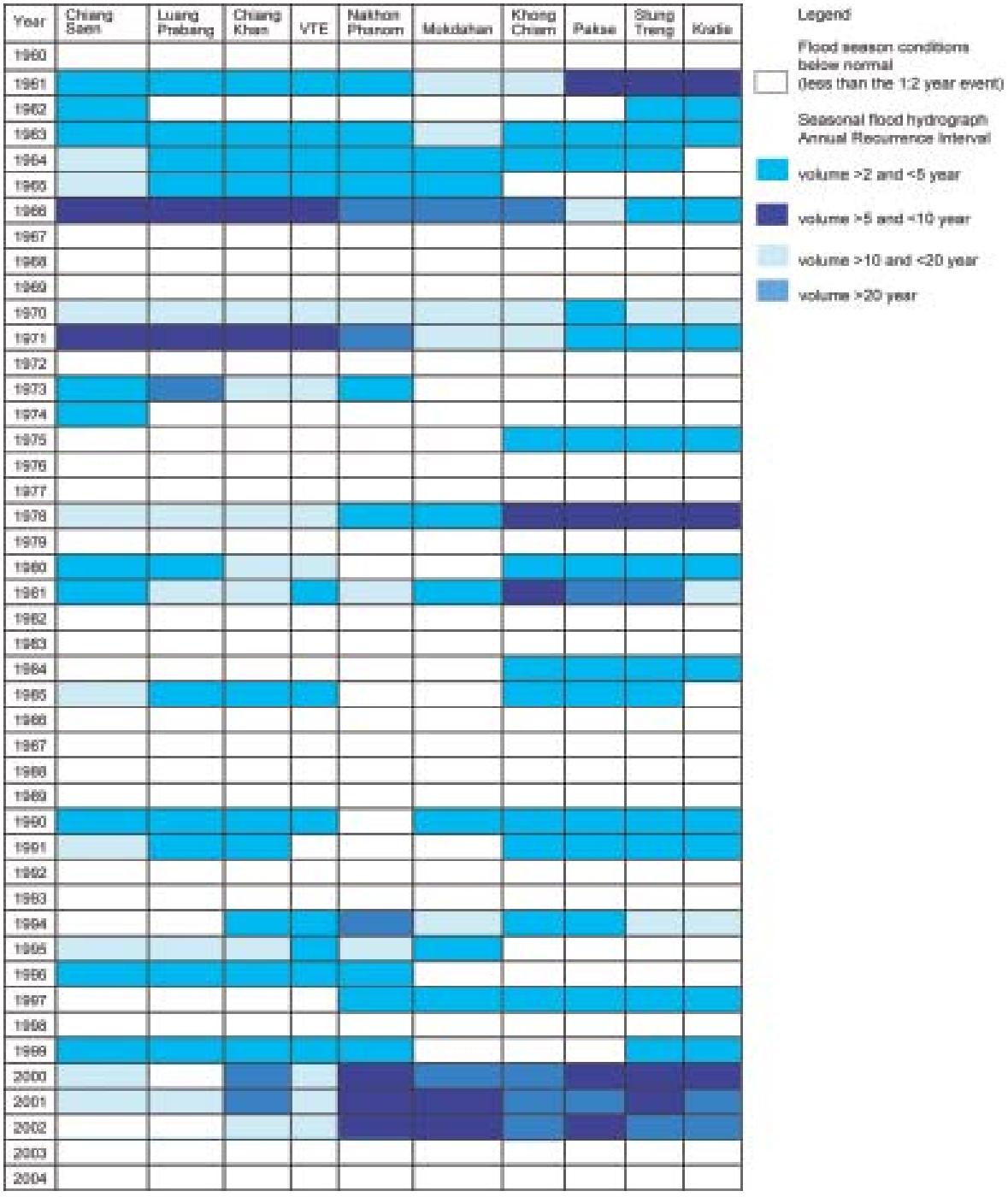


Fig. 2.14: dry and wet years frequency along lower Mekong (Source, MRC)

2.9. Flooding Frequency

Fig.2.14 shows the reoccurrence volume of flood water between 2 to 20 year return periods at different forecasting stations along Mekong. As mentioned in Section.2.7, that maximum historical flood occurs in different year at different forecasting stations. Similar behavior is reported in volume of flood water. The analysis of yearly flood volume at each gauge along Mekong, as shown in Fig.14, confirm this pattern of the variability of flood volume at different gauges.

A flood frequency analysis plot of flood peaks is shown in Fig.2.15. This plotting is based on data from (MRC, 2006). The floods are actually multivariate events, characterized by peak flow, flood volume and flood duration. Peak flows are related to inundation depth, whereas, flood volume and duration decide the area under inundation and the duration of inundation (MRC, 2006). Therefore, an exhaustive frequency analysis should incorporate flood peak, flood volume and flood duration.

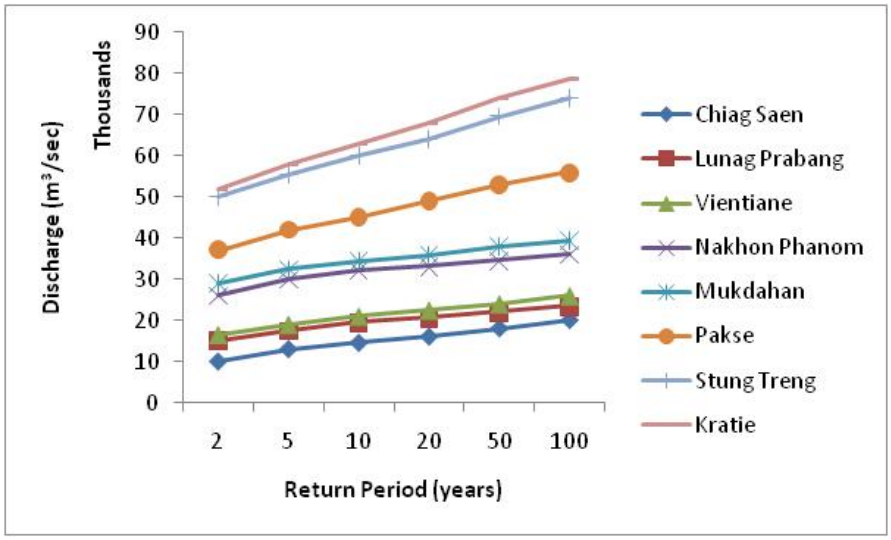


Fig. 2.15 : Flooding return periods along lower Mekong (data for this fig has been taken from MRC AMFR 2006, 2007)

A distribution of the joint probabilities of annual flood peaks and volumes as presented in (MRC, 2007) and reproduced in Fig.2.16 shows the reoccurrence of flood with 2 to 50 year return period at Chiang Saen and Vientiane.

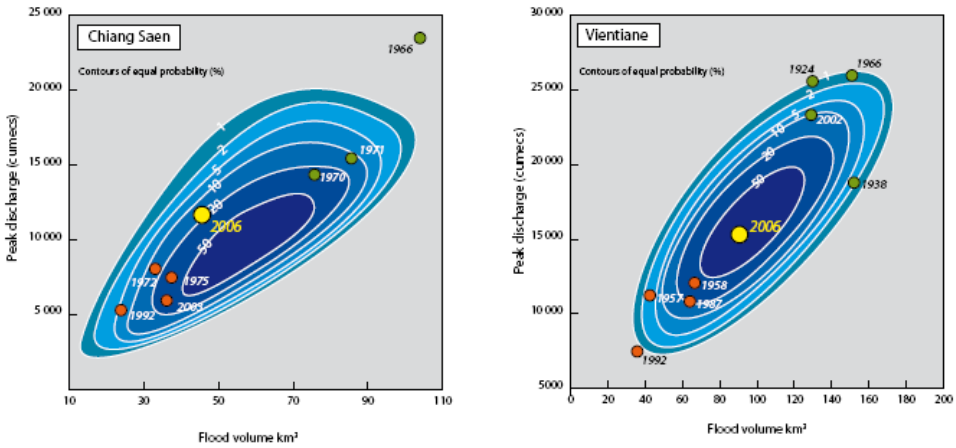
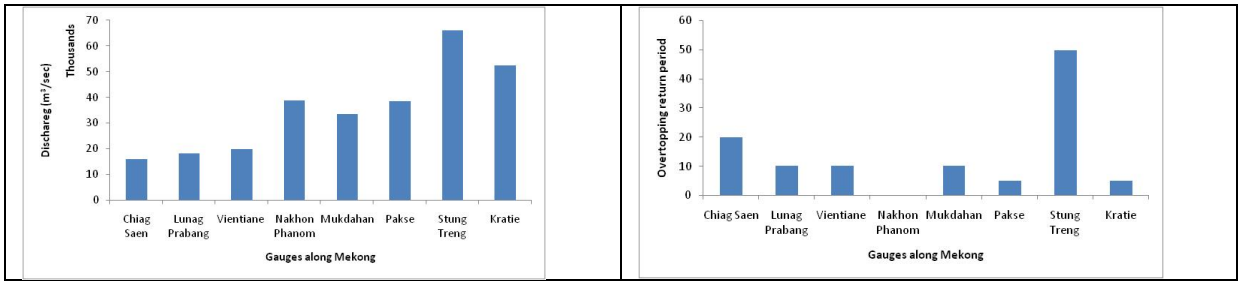


Fig. 2.16: Bivariate probabilities of the joint distribution of annual flood peak and volume at Chiang Saen and Vientiane (Source: MRC AMFR 2006, 2007)

The statistics of flood discharges occurrence is meaningless, if it is not seen in the perspective of overtopping flood levels or flood discharges. The water carrying capacity of the Mekong River changes along its longitudinal section as shown in Fig.2.16a. The discharge carrying capacity varies along longitudinal section of Mekong. The discharge carrying capacity at Mukdahan and Pakse is lower than Nakhon Phanam, which is the next station upstream of Mukdahan. The critical flood levels / discharges are plotted against the discharge carrying capacity at various cross sections along Mekong in Fig.2.17a. The return period of flood overtopping at various sections has been worked out by comparing magnitudes of floods of certain period against overtopping flood discharge as presented in Fig.2.17b.



a) b)

Fig. 2.17: a) Critical flood discharges along Mekong; b) flood overtopping return periods along lower Mekong (Flood return period data have been taken from AMFR-2006, 2007)

2.10. Impact of Flooding in Mekong basin

Floods generally are perceived as natural disasters, however in the Mekong area, floods are a mixed blessing. During the Monsoon season the river regularly overflows its banks in the lower Mekong area, usually with beneficial effects (agriculture and fisheries) but once every 6 to 10 years river flood level exceed the critical beneficial level causing extensive damage to life and property. Such a flood caused 800 human causalities in the year 2000 flood (Plate, 2007). The major sufferers to these harmful floods are the residents of the Lower Mekong area who are living along the river in the Central & Southern Lao-PDR, Central Cambodia and the deltaic regions of South-Vietnam.

2.10.1. Flood benefits in Mekong basin

Agriculture in the lower Mekong basin mainly depends on irrigation from water overflow through and over river banks that are caused by the flood wave. Farmers prepare land and sow the seeds prior to the flood season and wait for the onset of the flood for irrigation. Paddy rice is the main crop of the area. The benefits of floods in terms of agriculture, fisheries, other aquatic animals, and for creation & maintenance of wetlands are tremendous for the riparian countries Lao-PDR, Cambodia, Thailand and Vietnam. Only some of these benefits can be quantified in financial terms as given in Table 2.2 (MRC AMFR 2008, 2009). Intangible benefits derived from eco-system and wetland maintenance can be inferred from the fact that the lower Mekong countries contain 16 WWF Global 2000 eco-regions, and that 1068 new species, without counting invertebrates, have been discovered along the Mekong between 1997 to 2007 (Thompson, 2008).

Tab. 2.2- Flood benefits in Mekong basin

Country	Estimated annual agricultural value accruing from the annual Mekong Flood(US$ millions)	Value of national inland fishery based on year 2000 estimates (US$ millions)
Cambodia	1000	608
Lao PDR	Not Significant	212
Thailand	Not Significant	900
Vietnam	3500	880
Total Lower Mekong Basin	4500	2600

2.10.2. Flood damages in Mekong basin

The damage caused by flood water overflow in the flood plains, depends on flood peak and flood peak duration which in turn decide depth and duration of inundation. However, like other geo-physical hazards the flood damage depends on the area where it hits. The magnitude of damage (Shrestha, 2005) depends on the settlement density, and the infrastructure of the flood prone area, as well as on the type and frequency of crops and other agricultural uses.

Some flood damage data have been collected, by means of surveys, interviews and questionnaires, by the national disaster management organizations of the riparian countries. Damages listed in Table 2.3 (MRC AMFR 2008, 2009) include only the annual financial flood damage of the riparian countries of the lower Mekong. However, the total damages must include also social costs. For expressing social damages in financial terms there is the problem of the inability to quantify loss of human lives and other intangible social consequences including distress and health issues.

2.10.3. Cost to benefit optimization

If we take the damage as the cost of the floods, the financial benefits of the floods in lower Mekong region far exceed their costs. However, if one includes intangible costs, the situation may be changed.

Tab. 2.3- Flood damages in Mekong basin

Country	Estimated annual agricultural value accruing from the annual Mekong Flood (US$ millions)
Cambodia	25
Lao PDR	10
Thailand	16
Vietnam	25
Total Lower Mekong Basin	76

Management of the flood risk should also include intangible factors, and thus it is a multi-objective decision process. A major effort is required to reduce the intangible costs in order to optimize the social benefits of Mekong floods. Non-structural measures need to be used to maximum extent, using techniques as mentioned earlier in this article, to optimize the social benefit. Above all, losses of human lives must be avoided. The population density along the Mekong normally ranges from 30 to 50 persons per km^2 as per population density map of year 2000 (State of the Basin Report: 2003) which is further higher in big cities like Vientiane and

Phnom Penh which are also located along Mekong Banks (Fig.2.19). One of the most effective ways to reduce human vulnerability to extreme floods is a good early warning system, supported by reliable and accurate flood forecasts

The riparian countries have to come up with strategies for reducing flood damages while keeping the beneficial impacts of floods. Along with flood risk zoning they include long, medium, and short term flood forecasting systems for the purpose of planning irrigation and to give early warnings for vacating settlements and saving lives when the flood exceed the critical beneficial level. One of the most effective ways to reduce human vulnerability to severe floods is effective early warning supported by reliable and accurate flood forecasts.

2.11. Mekong River Flood Forecast

The settlement density is high along the Mekong River (Fig.2.18). The capitals of Lao-PDR and Cambodia are also located along River Bank. This high settlement density requires good flood forecast to evacuate the region before the hitting of flood wave. There is a tradition of flood forecasting in the Mekong River Basin (MRB) as a non-structural means for reducing vulnerability against harmful floods. The former Mekong Committee (replaced in 1995 by the Mekong River Commission (MRC)) initiated a flood forecasting program for the lower MRB in response to severe flooding in 1966. A hydrologic model (SSARR) and a hydrodynamic model (DELTA) were adapted for flood forecasting. In the nineties the SSARR model forecast has been enhanced by the implementation of a regression model for the Tonle Sap region (MRC, 1999; Tanaka, 1999). Forecast products, including water level forecast bulletins, are published on the MRC website (www.mrcmekong.org).

An expert meeting in 2002 on early warning for the Mekong River recommended the improvement of flood forecasts and other aspects of the early warning process (Plate and Hewitt, 2002). The recommendations supported the creation of a Regional Flood Management and Mitigation Center (RFMMC) within the MRC for flood studies on the Mekong. The center was established in 2004. As a first step, the outdated SSARR model was to be replaced. The Australian URBS (Carrol, 2007) hydrological model was selected by the RFMMC as a model to be implemented (Pengel *et al.*, 2008). The URBS model is a semi-distributed rainfall-runoff model with a built-in flood routing capability which is combined with FEWS. FEWS is an envelope that takes care of all data capturing, pre-validation and processing, model runs and output post processing; enabling easy access to graphs and tables of input and output data, both for recent and historical situations. Since June 2009 the new forecasting model based on the URBS-Code (http://www.sunwater.com.au) is in use (Pengel *et al.*, 2008).

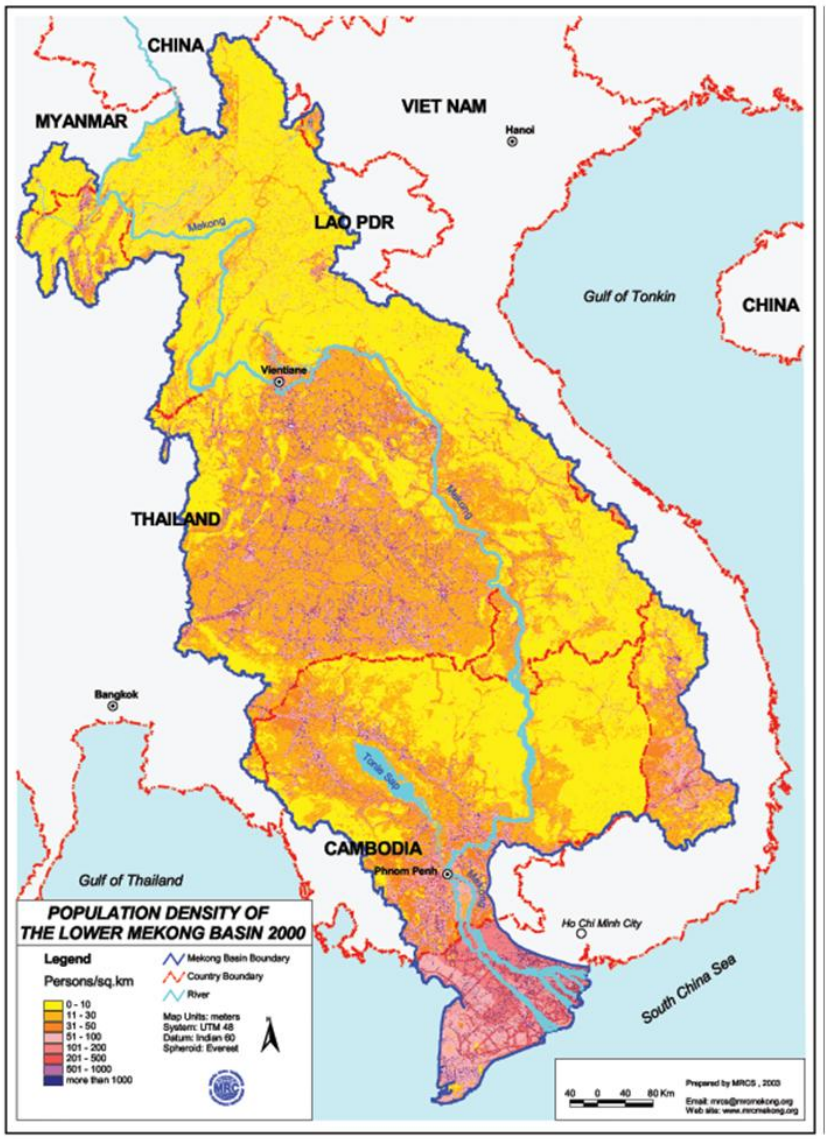


Fig. 2.18: Population density of the lower Mekong basin (State of the Basin Report: 2003)

3. Literature Review

In this chapter, the flood forecasting problem of large rivers is explored in a purpose oriented manner, in order to choose the appropriate tools for an optimum solution to flood forecasting on the Mekong. After a brief introductory section on general concepts for forecasting (nature of problem), the requirement of flood forecasting models will be presented. In the next section the feasibility of contemporary hydrologic models will be analyzed in order to serve the purpose of flood forecasting of River Mekong. The findings of this comparative analysis will be used towards the decision support for a new flood forecasting model for Mekong. In the end, the potential efficiency of a new model to serve the required objective (accuracy of forecast and sustainability of proposed model) will be compared with the efficiency of previous flood forecasting approach used in the Lower Mekong River.

3.1. Flood forecasting

Forecasting is the process of making statements about events whose actual outcomes (typically) has not yet been observed. A commonplace example might be estimation of the expected value for some variable of interest at some specified future date. Prediction is a similar, but more general term. In hydrology, the term "forecast" is reserved for estimates of values at certain specific future times based on initial conditions, while the term "prediction" is used for more general estimates, when the initial conditions become irrelevant, such as probability based statistics of the number of times floods will occur over a long period (Plate, 2007).

Forecasting is required in hydrology in order to estimate future incoming floods on a river. These incoming flood events may occur in near or far future. However, there is a lack of consensus regarding the classification of forecasts with respect to lead time, such as division of forecast lead time into short, medium and long term, and their respective spans. The differences could be attributed to the basis of classification, i.e. time, purpose, catchment size, rainfall field type and forecasting approach. The forecast lead time, as in the case of Mekong River basin, is based on purpose oriented fixed time markings, i.e. short term, medium term and long term. A short term flood forecast for a large river, such as the Mekong ranges to 5 days. The medium term forecast range from 5 to 10 days (Malone, 2006). On the other hand, Lettenmaier and Wood (1992) have divided forecasts into two categories, i.e., short (less than 7 days) and long term (up to several months) forecast. Plate´s (2007) classification of forecast lead times with their respective spans is more relevant because of its context to system stability and continuity of physical processes. He stated that short times are times when the variation of value from present to future is small and thus can be foretold by simple models. The span of short term forecasts is further explained as depending on catchment size and extent of dominant rainfall fields - as in large catchments forecasting times of a few days may come under the definition of short time, in comparison to

small catchments of a few km^2, where the definition of short time may only apply to forecast times of a few minutes.

The dominating factors affecting the flow dynamics of large rivers like Mekong are essentially different from those of small rivers. The flows in small rivers are mainly driven by catchment response rather channel flow dynamics - as the hydrologic response time, T_f (the time of concentration of the catchment response at the forecast point) is dominated by hydrologic response time of the catchment, T_c in comparison to channel routing time T_r. In small catchments, the catchment runoff follows a typical sequence of land surface – runoff - channel system, which means that catchment response turns into channel flow only after overland flow from the land surface of the catchment has occurred. In this case we have four possibilities;

1) $T_f < T_c + T_r$ and $T_c << T_r$, Tc overland flow, Tr river flow Tf forecast time

2) $T_f < T_c + T_r$ and $T_r << T_c$,

3) $T_f > T_c + T_r$ and $T_c << T_r$,

4) $T_f > T_c + T_r$ and $T_r << T_c$.

The flow dynamics, in the first case, is driven by the catchment physics in comparison to second case where it is driven by the channel behavior. Unlike the first two cases, in the third and fourth cases the required forecast lead time T_f exceeds the available total response time T_C. In these situations the extra lead time is acquired through rainfall forecasts, but at the cost of additional meteorological uncertainties. But in actual practice, overland and channel flow of the large river doesn't follow this typical response – rather its flood discharge is generated on a number of lateral sub-catchments which add to the main river in its course from upstream to downstream. Thus, the flow dynamics at the forecast point is the combined result of catchment response of a number of parallel lateral sub-catchments and the flow dynamic of the Main River.

For modeling the flood generating system of large rivers, it will be assumed, for the sake of simplicity, that the lateral sub-catchments drain directly at the river gauges into the main river as over land flow, to which one has to add the flow in the main river. Actually of course, the flow from the sub-catchments is also a combination of overland flow and tributary flow.

Flow dynamics of large rivers is complex, as it is controlled by many different catchment characteristics, such as catchment shape, overland and groundwater response, and channel flow characteristics. For any particular application, one may base a flood forecast on channel routing in the main river, rainfall-runoff models for the lateral subcatchments, with precipitation inputs obtained from rainfall measurements or forecasts (for example by means of numerical weather forecasts) or a combination of these, depending on user requirements and data availability. The model component selection of a given forecasting problem depends on the relation between required forecast lead time, T_f to the time of concentration of the catchment response at the

forecast point. The combination of components of models for flood forecasting of large river is a function of the river basin structure and varies from case to case. For example, in the case of Mekong, one has to account for the fact that the components work together, but for the sake of clarity each of the components is discussed separately.

3.1.1. Runoff Routing Models

Before starting the modeling of discharge routing in the main river, the persistence of discharges at each of the gauges should be checked, by studying the changes in discharge at two subsequent gauging times. In large rivers, owing to physical continuity of the river flow, the change in two subsequently measured discharge values is small and therefore could be used as potential for future forecast. In the selection of a runoff routing approach, one may move from time delayed simple linear regression of downstream to upstream discharges. As described by Plate (2007) on many large rivers flood forecasting is based on this sort of regression. However, if regression fails to produce flood forecasts of required accuracy the option of hydrologic and hydraulic runoff routing could be explored.

Routing of water down a river channel is generally described by the one-dimensional hydrodynamic equations of unsteady flow known as the St. Venant equations. There are various simplified version of this equation used in river forecasting, as listed and described by Lettenmaier and Wood (1992); such as diffusion equation method, kinematic wave method, Muskingum method, and also impulse response function methods, such as linear reservoir models, cascade of linear reservoirs, lag and route method and others. These methods could be mainly categorized into hydraulic and hydrologic models for flow routing.

However, for the purpose of flood forecasting, it seems appropriate, at least for the large Mekong River (Plate, 2007; Apirumanekul, 2006), to use the time delayed discharge at the next upstream gauge as approximation for the downstream discharge without considering the attenuation of flood peak. One should move from simple to a more complex routing approach, under the constraints of available data, if additional accuracy is gained by the added level of complexity.

3.1.2. Rainfall Runoff Modeling

Lettenmaier and Wood (1992) state that there is a range of forecasting models for stream flow based on rainfall inputs. However, comprehensive rainfall runoff models inherently involve, and are consequently classified on the basis of water storages; interception, soil moisture, surface storage and process components on the basis of fluxes: infiltration, evapo-transpiration, snowmelt, interflow, ground water base flow and surface runoff. The classification of rainfall runoff models could be based on the specific combinations of various storages and flux, as well as on the level of modeling complexity. The U.S National Weather Service River Forecast system is an example of conceptual storage models. An even more comprehensive list of watershed hydrologic models was provided by Singh and Woolhiser (2002)

3.1.3. Rainfall Prediction Models

There are various methods of rainfall forecasting for hydrological applications; however, only a few of them are listed here. For short period spatial rainfall forecasts, radar can be used by virtue of its ability to measure both quantitative rainfall and short time changes in rainfall. For lead times less than 2h, rainfall predictions are often based on radar rainfall data by adopting precipitation projection or now-casting procedures (Smith, 1992).

Georgakos and Bras (1984) have developed a physically based prediction model for rainfall at a site. Smith (1992) quoted that U.S. National Weather Service produce 6-h quantitative weather forecast (QPF) products based on probability of heavy precipitation. QPF products are developed using statistical procedures in which model output variables are input to regression models for the precipitation variables of interest. This procedure is described by Glahn and Lowry (1972)

For real time flood forecasting in the Mediterranean catchment of the Gardon d' Anduze Lardet and Obled (1993) employed a stochastic rainfall generator which is based on renewal processes. They found reliable forecasts for very short lead time of 4h ahead.

For the Anas catchment of Northwest India Zehe et al. (2005) predicted the long term monsoonal rainfall (monthly) by a stochastic approach. The rainfall time series used was generated on the basis of applying observed meso-scale circulation patterns.

Currently, in many countries numerical weather prediction is normally based on Global Circulation Models (GCM). In weather forecasting meteorologists make 'ensemble' runs with different initial conditions to test the sensitivity of the predictions. Apart from NOAA´s Satellite based rainfall forecast estimates; there are available other global numerical weather products to produce rainfall prediction (Beven, 2001).

The tracking of typhoon or tropical storm movement proved important in the qualitative prediction of rainfall in many Asian countries (Mekong region, Indo-Pak subcontinent), which receive heavy rainfall from these storm systems. The quantitative use of typhoon information has not yet been developed very far. For a statistical approach, some of the results from Camargo et al. (2006) may be useful, who describe the tropical storm trajectories by probabilistic clustering, in the western north pacific region. The analysis acknowledged the usefulness of track identification towards the improved prediction of typhoon landfall, several days in advance.

3.2. Requirements of flood forecasting model

3.2.1. Application oriented

Flow forecast requirements for floods are different from other typical applications such as navigation or reservoir operation. Unlike flow forecast for floods, the flow forecast for navigation becomes relevant not only during critically high flows near to overtopping,

particularly at critical sections (less discharge carrying capacity sections, thus potential to be overtopped) but also for low flows along the river. Therefore, flow forecasts for the purpose of navigation require sophisticated hydraulic modeling which provides forecasts of water levels along the channel at fine spatial-temporal resolution, especially during low flows. On the other hand, reservoir operations require reliable forecasts of flow volumes within discrete time intervals. Unlike discharge volume reliability in the case of flow forecast for reservoir operation, the time to peak and total flood volume is important in the context of flow forecasts for floods.

3.2.2. Object Oriented

There are number of mathematical models available in hydrology which can be potentially used and have been used for flood forecasting. For example Plate (2007) distinguished the application of models for flood forecasting from models for other typical applications such as; synthesis of past hydrologic events, effects of anthropogenic and climate change on hydrologic response, predicting future hydrologic events for design (Freeze and Harlan, 1969). Although the structure of both forecasting and design models include discharge time series, runoff routing, rainfall runoff modeling, they both serve quite different purposes. The purpose of design models is to predict extreme value statistics of certain return period, such as 100 year flood. Unlike time to actual peak, as in the case of flood forecasting models, only the order of magnitude of the extreme flood is important in design models. The inherent and accepted uncertainty, primarily due to input data, model, parameter, distribution choice, in the prediction of extreme value statistics, is covered in design of protection work by suitable margins of safety. In contrast the accepted accuracy in forecasting is much smaller in comparison and should be expressed by an auxiliary error band together with crisp forecast or with different forecast ensembles.

Flood forecasting models primarily use simulation and trend extrapolation modeling. However because of their specific objectives, they can be clearly differentiated from other categories of watershed and trend extrapolation models. For example, Plate (2009) distinguishes between real time forecast models and planning models in the domain of flood management hydrological models. He said that each model has a different objective, consequently the structure of a model should be a function of its application. Similarly in time series modeling simple decomposition or prediction models can be differentiated from real time flood forecasting models with respect to their application.

Maximum possible achievable accuracy and reliability are the primary goals of any forecasting model. The accuracy should be read as the minimum possible disagreement between the timing and magnitude of predicted and actual outcome. As there is no forecast with absolute agreement of actual to predicted outcomes due to input, parametric, modeling and future uncertainty, consequently a quantified statistics of possible disagreement is required with each outcome to ensure passing on reliable information to the decision maker. The reliability of expected outcomes depends on many factors. In some cases it is possible to attribute uncertainty of

forecasts to its sources, by a Bayesian reliability analysis. Thus, it is possible to split up the conversion problem from rainfall to river discharge into meteorological component and hydrologic-hydraulic components, whose uncertainty can be determined individually, as done by Krzysztofowicz (2001), who introduced appropriate hydrologic uncertainty processors (HUP) and meteorological uncertainty processors (MUP). These can be treated separately, as illustrated by Krzysztofowicz (2001) or collectively, as will be done for the work of this thesis.

3.2.3. Top to bottom versus bottom to top modeling

Young (2003) discussed the issue of top to bottom (moving from channel to catchment) and bottom to top (moving from catchment to channel) modeling approach in the context of forecasting models. He rightly argued in favor of top to bottom models, because of the uncertainty involved in the process of bottom to top modeling. The bottom to top modeling, essentially, requires physical/ conceptual modeling of catchment response, starting from remotest catchment and coming down to the forecast point in the Main River gradually, which may not be necessary for flood forecasting. Top to bottom models start from the next point upstream of the foresting point in the River. The frequent use of regression models for hydrologic routing reflects the popularity of top to bottom models in flood forecasting.

3.2.4. Physical Representativeness versus functional accuracy

Among the requirements of forecasting models, the accuracy of a model in producing flood forecasts is more valuable in comparison to its physical representativeness of the catchment process (Nash et al., 1977; Beven, 2000, 2001; Young, 2002; Plate, 2007; Kachroo, 1991).

Nash et. al. (1977) supported the findings of W.M.O (1975) on the advantages of simple conceptual models in comparison to more elaborate forecasting models. Kitanides (1980a) stated that ideally, the best model for real time forecasting would be a deterministic description of the complete rainfall runoff process based on well known physical laws. He states, however, that application of these laws is hindered in practice by the complexity of natural catchments.

Beven (2000) argued that there seems to be a tendency to think that the more physically based a model is the more accurate will be its predictions, and to ignore the fact that all models should depend on field data to define the characteristics of each unique catchment. Though process based models may be richer scientifically than, say, transfer function models, yet that does not mean they are more accurate in reproducing the data. Consequently the important thing in forecasting is that the outcomes should be correct. (Kachroo, 1993; Plate, 2007)

3.2.5. Catchment Type and scale

Landscape, land use, hydrological scale and climate greatly influence the choice of a flood forecasting model. Plate (2009) described the effects of landscape and scale in the model

selection. The issue of scale is highly relevant to the selection of appropriate forecasting model. Singh and Woolhiser (2002) described the analysis of Seyfried and Wilcox (1995) on how the nature of spatial variability affects the hydrologic response over a range of scales. The need of complexity of a hydrologic model is influenced by the scale for which it is built. The appropriate selection of the correct scale may save one from the development or use of an unnecessary complex model.

Plate (2009) stressed the dependence of rainfall runoff model selection on the basis of the hydrological scale of the catchment. He explained the model choice with respect to four types of hydrological scales: point scale, micro scale, meso scale and macro scale. The dominant variables vary with scales, and models should be selected accordingly. For the point scale local soil characteristics and plant cover determine the runoff. The flow for micro scale processes is can be represented by fundamental laws of continuum physics. The meso scale is better described by conceptual models based on system functions. Typical of this scale are unit hydrographs. There is no clear limit of the size of the catchment area in the application of unit hydrograph models, which depends on catchment characteristics and available data. With increase of catchment size to macro scale flow in the river network becomes increasingly more important than the rainfall runoff process

Large rivers are normally fed by macro scale sub-catchments, such as the sub-catchments of the Mekong River. The macro scale refers to catchment sizes ranging from 1000 to several 10,000 s km^2. On this scale, instead of modeling surface runoff in detail, determination of the runoff coefficient is most important for all sub areas. Furthermore, models combining sub-catchment runoff with river networks are particularly advantageous for situations in which the geological or topographic properties are very inhomogeneous, causing different runoff formation processes in each sub-catchment. For example, the geological and topographic properties of the Mekong River sub-catchments are unique to each sub-unit.

Climate and land use of a catchment also affect the choice of a forecasting model. The climate dictated by geography required special kinds of model to suite the local condition. For example, humid and arid catchments require different treatments. Similarly, forecasting models in typhoon induced flash flooding are more dependent on storm system tracking and rainfall forecast in comparison to orographic rainfall based or snowmelt based floods. Also, land use, i.e., cities, crops, barren land etc, dictate the required spatio-temporal sensitivity and accuracy of flood forecasts. (Plate, 2009)

3.2.6. Uncertainty Analysis

Kanning et. al. (2004) described uncertainties with reference to reliability of a flood defense system. They classified the uncertainties into natural variability and knowledge uncertainty where the latter can be reduced and the former has to be endured. Natural variability represents the randomness or variation in nature in both time and space (Van Gelder, 2000). Knowledge

uncertainty caused by basic lack of understanding and modeling the physical phenomena, or by lack of sufficient data.

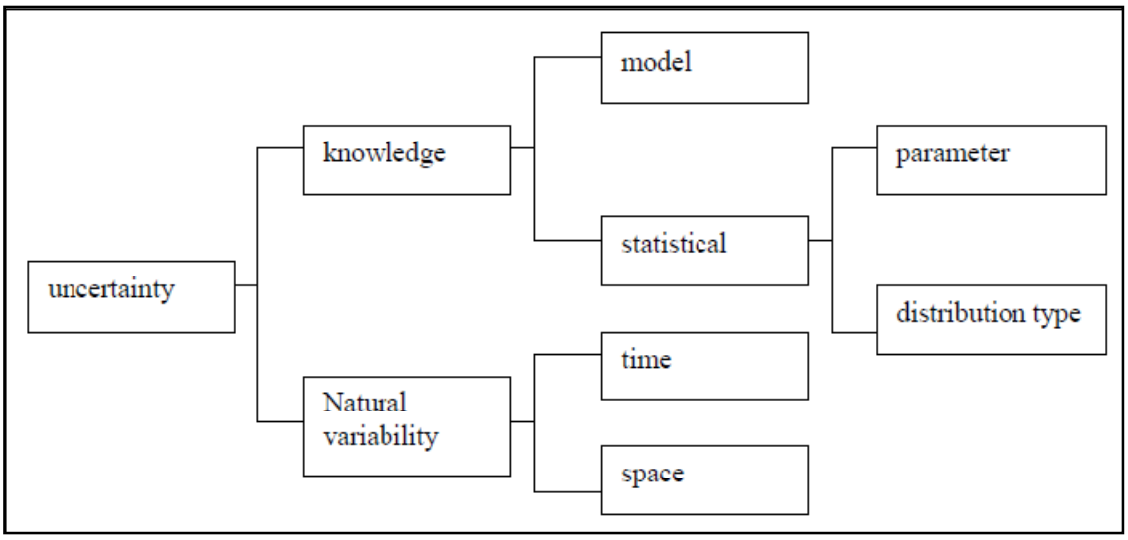


Fig. 3.1; Classification of uncertainty by Kanning (2007)

Input uncertainty: the input data is either based on point and areal measurement or come from quantitative precipitation and weather forecasts. In the first case there are considerable measurement preprocessing errors, while in the second there are considerable forecast errors.

Initial state of the system uncertainty: imperfect knowledge of the initial state introduces these errors. However, since hydrologic forecasting is performed in a continuous way, this source of error is not important.

Model uncertainty: in modeling the various components of the rainfall-runoff process, several simplifications have to be made. The various components of the hydrologic cycle separately and the interactions between elementary processes are described by simplified functional relations. These imperfections of the model introduce model errors.

Parameter uncertainty: Some of the model parameters have no exact physical meaning and have to be chosen through calibration, on the basis of their reproduction efficiency of the input to output behavior of the system. If more than one parameter has to be determined in this way, the problem of equi-finality (Beven, 1993) arises, i.e. compensating effects of different sets of models may yield the same result in application. Other parameters are found from the physical characteristics of the basin, which cannot be determined with accuracy.

Kitanidis et al. (1980a, 1980b) discuss in detail the issue of uncertainty analysis with respect to data and model structure for real time forecasting. They mention four major sources of

uncertainty i.e., model uncertainty, input uncertainty, parameter uncertainty and uncertainty of the initial state of the system, Once the model is available, the deterministic formulation assumes that input information (precipitation and temperature) is sufficient to describe the condition of the system, so that measurement of output (river flow) is actually redundant information except for the cumulative effect of input, model, parametric and initial system state errors. Therefore, they have suggested to modify output information in order to reduce these uncertainties.

To sum up: apart from natural variability, type and extent of uncertainty have to be considered in flood forecasts, which mainly depends on forecasting input data, availability and accuracy of calibration data, selection/availability, choice of model/ approach for forecast and user requirement for the output. In a forecasting model, it is not useful to identify uncertainties of each component of the model, or for each of the processors. Instead, it is better to lump the results of all uncertainties, which make up the forecast error, and use the error statistics, i.e. the difference of forecast from real discharges or water levels as measure of the uncertainty, as will be done in this work.

3.2.7. Input Data Requirement for flood forecasting

Data requirement is normally seen in the context of an a priori chosen forecasting approach, i.e. fitting the pre-built model on some catchment. However, it is more relevant that data availability should dictate the choice of a forecasting method, especially in the context of hydrologic modeling which requires tremendous amount of time series and physiographic data that is not usually available and cannot be obtained at short notice. The method of fitting a model by assuming values or parameters estimated from similar catchments does not serve the purpose well in most of the cases.

In addition to basic physical laws prior information on the system in the shape of catchment characteristics (physiographic data) and measured data time series (input and output) should be used as basis for forecasting model construction. Generally, the requirements vary with the selected type (metric, pure physical, conceptual and hybrid) of forecasting model. Metric and conceptually empirical modeling normally requires input-output data series only. In contrast, pure physical and hybrid models require not only modeling of the appropriate physical laws for describing each of the processes of the discharge formation process, but also physiographic data (atmospheric and geosphere) and input data in the form of ordinal gauging of water – as per requirement - at various stages of water cycle, i.e., rainfall, evaporation, transpiration, infiltration, interflow and surface flow (overland and channel flow).

In the case of large river flood forecasting by data based modeling; the availability of required data dictates the choice of the model. These data have to be analyzed in detail with respect to its spatial discretization, temporal discretization, length of record and data quality. Normally the watersheds are categorized into un-gauged, sparsely gauged and densely gauged catchments with respect to data availability.

3.2.7.1. Spatial discretization

For detailed models of the rainfall – runoff process the areal distribution of rainfall may be important. The major distinction in precipitation input is between point measurements and areal gauging. Although the situation varies from case to case, time series from point measurements are the normally available data form in most catchments. But the issues of density of point measurements and their representativeness are to be considered. Point measurements could also be used for areal approximation by some sort of areal averaging, i.e., mean areal average, weighted areal average, isohyets and Thessian's polygons, as are normally used in conceptual, lumped or semi distributed models. Satellite Rainfall Estimates (SRE) are now globally available, but the problem of downscaling and validation of their representativeness by means of ground based observed reality is an issue to be resolved prior to using them in any rainfall runoff modeling. In the case of Mekong River, large differences have been found between ground observations of rainfall and SRE (Malone, 2009).

There were, though, attempts to base physically distributed models on areal averaged inputs. A study by Freeze and Harlan (1993) however, yielded limited success, (Freeze and Harlan, 1969). It was rightly argued by Kachroo (1992) that the response to spatially averaged values of input is not necessarily the average of responses to the corresponding distributed values. Consequently, the spatial discretization, i.e., lumped, semi distributed and distributed, of physical or empirical models should be oriented on the available spatial discretization of the input time series, i.e., rainfall and snow.

3.2.7.2. Temporal discretization

The availability of input data could be further analyzed with respect to its temporal gauging frequency that may vary from few minutes, hourly, 6 hourly to daily resolution. Although temporal gauging frequency of both, point and areal measurement in the case of rainfall, and discharge measurement place the limit to the output interval of the forecasting model. However, in principle output intervals smaller than measurement intervals could be approximated by interpolation, but not without efficiency loss.

3.2.7.3. Data length

Unlike the high dependence on the frequency of spatio-temporal gauging as in distributed models, the length of data is a decisive factor in the choice of the data driven model. For example, artificial neural network (ANN) and empirical models require sufficient length of data record for the purpose of training, parameterization, simulation and finally validation. Therefore, time series and empirical models, cannot be used in the catchments with short historical data records. The data should be used in the split sampling mode, by calibrating the model on a part

of the record, the using the second part for verification, and a final validation takes place by means of real time application of the model.

3.3. Review of Contemporary Flood Forecasting Approaches

Before the design/ selection of a flood forecasting model for large rivers, it is beneficial to analyze the merits and demerits of existing flood forecasting techniques/ models. There is a large variety of different forecasting models, for example (SSARR, URBS, Mike 11), based on different principles, consequently with different input, modeling and output data requirements. A literature survey revealed that various classifications of flood simulation models based on different criterion are already part of different hydrologic papers and books. For example ASCE (1996) reviewed and categorized flood analysis models into event based precipitation models, continuous precipitation models, steady flow routing models, unsteady flow routing model, reservoir regulation models, and flood frequency analysis models. Plate´s (2009) classification distinguishes operation and design models.

A classification of flood forecasting models by Plate (2007), has placed the family of models into five groups; a,b,c,d and e. Group a comprises deterministic models for forecasting a single value. Group b models involve the stochastic nature of the catchment, by incorporating the uncertainty of variation in rainfall distribution and hydraulic process. Normally, Monte Carlo methods are employed for calculating from large numbers of observed hydrographs, deviation of actual value from forecasted value. The sample of deviation is analyzed statistically and error bands are derived (Krzysztofwicz, 2001). Group c models use historical data to determine an empirical set of response hydrographs for many different initial conditions. Group d models are just the improved variant of group b models for large number of input conditions. Finally group e represents traditional method of regression analysis of upstream with downstream gauges. The initial four groups are essentially based on rainfall runoff modeling with possible distinction in the process of whether or not and accounting random component. The last group is principally based on empirical statistics of discharges or water levels.

But none of these classifications serve the required purpose of comparison of different flood forecasting approaches with respect to their input requirements, the forecasting technique, modeling type, modeling discretization, modeling optimization and type of output. Consequently the available flood forecasting models' plethora must be sorted and classified on an elementary level into certain categories, and the practical applicability to the problem of large rivers' flood forecasting of each class must be assessed. This classification, unlike general model comparison, will provide the opportunity to analyze the merits of these forecasting approaches to serve the requirement of flood forecasting as outlined in Section-3.2.

3.3.1. Forecasting Techniques

There are various categories of forecasting techniques such as time series methods, causal methods, judgmental methods, artificial intelligence methods, however, trend extrapolation and simulation methods are relevant to the problem of flood forecasting for the Mekong.

3.3.1.1. Trend Extrapolation

Trend extrapolation, commonly known as statistical forecasting methods, is based on time series analysis, such as, examination of trends and cycles in historical data, and their extrapolation into the future. The essence of statistical forecasting is identification of structure in the data, fitting a time series model to the situation and then extrapolation of that model into the future. However, the model structure should be based on the structure found in the data, and it is further assumed that the structure will remain stable over the period being forecasted. There are many mathematical models for forecasting trends and cycles. These time series methods are essentially deterministic models, i.e. constant mean model, linear trend model, simple additive seasonal model, more general linear model, growth model, regression models or essentially stochastic models; autoregressive- moving average models, stochastic component model, models defined by moment etc (Gilchrist, 1974).

For a very short term forecast, when the data (discharge) persistence is high (high correlation of subsequent values) over the period being forecasted, trend extrapolation methods like simple auto-linear regression, curve fitting can be useful. But if subsequent value persistence, over the period being forecasted, gets disturbed by atmospheric storm forcing (rainfall), simulation (modeling) methods are to be preferred, because time series or trend extrapolation methods are useful only for stationary conditions.

3.3.1.2. Simulation Methods

Simulation methods use analogs to model complex systems. These analogs can take on several forms, for example, mechanical or mathematical analogs. Mathematical analogs ranging from pure empiricism, to simple physics based and conceptual simulations (unit hydrograph, transfer function) are used in hydrologic modeling, i.e. for rainfall-runoff modeling, or runoff routing. In general, a hydrologic model is defined (see for example Haan et al. (1982)) as a mathematical representation of hydrological processes of a catchment in a simplified form.

The difference between trend extrapolation and empirical modeling should be noted. Though both depend on statistical time series, in trend extrapolation the data structure is modeled and replicated, whereas in a mathematical analog the physical system is imitated by means of mathematical equations.

For large rivers, stream flow forecast are invariably based on simulation of observations or forecasted rainfall on the upper catchment, or of river flows at upstream points on the main river or tributaries, often supplemented by rainfall measurement on the intervening catchment, to the flow at the forecasting point (Kachroo, 1993). Channel simulation is required if variation in persistence is caused by strong short term changes of upstream discharges, and rainfall-runoff watershed simulation is needed for modeling change in persistence of discharges due to strong rainfall forcing in the lateral catchment between two subsequent gauging points. Normally both channel and watershed simulations are required in the flood forecast of large rivers, because of consecutive sub-catchment runoff (overland flow to channel flow) and concurrent incoming river flows (sub-catchment runoff addition to upstream main river flow). Sometimes, forecast of rainfall is additionally required for flood forecast.

3.3.2. Modeling Type

The level of complexity of the mathematical analogs for watershed processes divides simulation methods into various types. The components of the hydrological cycle can be divided into structural elements, process and storage elements (ESA, 1997). Similarly, simulation methods can be classified according to which element of hydrologic cycle is modeled with how much complexity. But each simulation approach needs a mathematical equation, or a set of interconnected equations, to model real world behavior. As regular mathematical abstractions of the real world physics of conversion of rainfall into runoff are not possible without assumptions, consequently, each approach has its own set of empirical conditions.

Hydrologic simulation modeling could be broadly categorized into four different prominent approaches, which developed parallel, in order to answer the Penman (1961) question of what happens to the rain? The answer of this question was given depending on the intended applications, i.e. discharge data synthesis, design models, forecasting models, etc. These approaches are abstracted by Young (2002) according to Wheather et al. (1993) as: pure physics approach, metric approach, conceptual approach, and hybrid metric conceptual approach. These different approaches have been used in the last 150 years in a quest to achieve better watershed modeling. The following section will define and review these hydrologic models from a flood forecasting perspective.

3.3.2.1. Pure physics approach

Simulation of water flow from precipitation to catchment overland and tributary flow with its auxiliary components, inter- and ground water flow require complex modeling for physically meaningful real time flood forecast. Basically, the structure and functional dynamics of a catchment is to be modeled. The structure is represented by catchment parameters and the description of the catchment initial conditions. The functional dynamics is normally characterized by the interaction of the catchment structural elements with input data from

atmospheric forcing, i.e. it describes the response of the catchment to the precipitation and heat input from the atmosphere. The spatial-temporal distribution of these forcings, combined with different initial state conditions of the catchment, result in a variety of different outcomes of runoff and water levels within the catchment. The initial condition of the catchment is characterized by small scale processes and water level variation, induced by spatio-temporal distribution of water within the catchment.

The imitation of natural catchment processes by means of pure physical simulation through mathematical equations requires comparatively fixed/static structural model elements of a catchment, i.e., size and shape of the basin, soil properties, land use and cover, topography, geological and geomorphologic formation, location of tributary network of the river, slope, cross-sectional and longitudinal sectional details of each tributary. And the dynamic part of the model, i.e., storages and processes of water flow on, within, and below the catchment and in the tributaries. These processes are needed to be defined by mathematical equations based on physical laws.

If structure and structural dynamic of a catchment is properly defined, ideally, by pure physical models, then the variety of different outcomes could be predicted simply as a function of rainfall input and structural network in a deterministic environment. But it would require an enormous effort of measurements to completely describe the natural variability of the structural elements of process and storage elements. Consequently, the lack of detailed measurements along with other constraints as described by Beven (2000, 2001, 2002) hinder the development of such pure physical models on a hill slope scale, and even less on a catchment scale.

The “pure physics approach” is mostly based on partial differential equations. The blueprint for physically-based digitally-simulated hydrologic response model was given by Freeze and Harlan (1969). The component process, i.e. precipitation and evaporation, infiltration and soil moisture flow, ground water flow and overland and channel flow were tried to be represented by well established physical laws with exact mathematical representation. The model was considered a composite boundary value problem described by partial differential equation and potential theory. The proposed model was non-unique with respect to both time and space and applicable over a wide range of hydrologic and geographic conditions. This type of distributed model was capable to predict the local hydrologic responses for points within the catchment. The first application of this kind of model was by Freeze (1972, 1974) followed by the SHE model of Abbott et al.(1986). SHETRAN, MIKE SHE, (Bathurst et al., 1995; Parkin et al., 1996; Refsgaard and Storm, 1995; Resgaard, 1997), IHMD (Calver and Wood, 1995) and HILLFLOW (Bronstert and Plate, 1997) are a few examples of distributed physically based models (Singh and Woolhiser, 2002). Most of these models followed the blue print of Freeze and Harlan (1969) with different ways to discretise and solve process equations (Beven, 2002a). But only a few (such as WATFLOOD and MIKE SHE) of these pure physical models were used to address the problem of flood forecasting.

A wide application of MIKE SHE in large river flood forecasting is constrained by unavailability of required physiographic data coupled with issues of scaling (Plate, 2009, Singh, 2002) and efficient output replication (Beven,2002). In macro scale catchments, complex physical modeling is not required, when the purpose is to define variation in discharge persistence resulting from large scale averaging processes.

3.3.2.2. Metric approach

Unlike physical models block box/ pure empirical models, simulate input-output relations empirically without complex catchment process imitation. The "metric approach", based primarily on observational data, seek to characterize the flow response using some form of statistical estimation or optimization (e.g. Wood & O' Connell, 1985). These include purely black box- or time series models such as discrete and continuous time transfer function, artificial neural network (ANN) and fuzzy representations (e.g. Tokar & Johnson, 1999). Metric approaches in the form of regression models have been widely used for flood forecasting as quoted by Plate (2007). The wide range applicability of the metric approach in flood forecasting is due to their limited data requirement, i.e. discharge time series of subsequent gauges or catchment rainfall and discharge time series. However, these methods have also distinct short-comings. Regression models, although proved to be valuable in modeling comparatively linear relations of upstream to downstream discharges, fail to simulate the time delayed behavior of rainfall to runoff conversion, and ANN application is problematic due to their unknown complex random parameterization of hidden layers in training process.

Simple linear regression models can be used to estimate the dependent variable from the independent variable. If "Y" is to be estimated from "X" by means of some equation then it is called a regression equation of "Y" on "X". The simplest linear model for this case is as follows:

$$\hat{Y}=b_0+b_1*X \qquad (3.1)$$

Constant "b_0" represents the intercept of straight line defined by the model or value of "$\hat{Y}$" when "X" is zero and "b_1" represents slope of line. Shahzad (2004) quoted that forecast results from simple linear regression might be improved by means of linear multi-regression to evaluate correlations inferred from knowledge of the physical environment. The resulting equations are in the form (Acreman, 1985):

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 \ldots\ldots\ldots\ldots b_n X_n \qquad (3.2)$$

where "b_0" is constant coefficient or intercept.
Usually "n" observations are available for the variable "$\hat{Y}$" resulting in "n" numbers of equations, one for each observation. Therefore, if these "n" equations have to be solved for the "p' unknown parameters (regression coefficients) then "n" must be greater than "p" (Haan, 1979). An example of "n" equations may be:

$Y_1 = B_1X_{1,1}+B_2X_{1,2}+B_3X_{1,3}+\ldots\ldots\ldots\ldots B_pX_{1,p}$ (3.3)

$Y_2 = B_1X_{2,1}+B_2X_{2,2}+B_3X_{2,3}+\ldots\ldots\ldots\ldots B_pX_{2,p}$ (3.4)

$Y_3 = B_1X_{3,1}+B_2X_{3,2}+B_3X_{3,3}+\ldots\ldots\ldots\ldots B_pX_{3,p}$ (3.5)

$Y_n = B_1X_{n,1}+B_2X_{n,2}+B_3X_{n,3}+\ldots\ldots\ldots\ldots B_pX_{n,p}$ (3.6)

The regression analyses are used to identify the mathematical dependence between the observed values of physically related variables and thus can account for the additional information contained in the correlated sequence of events. There are two accepted ways of solving Eq.3.1. The first consists of least squares optimization of the system of equations, by one of the standard optimization methods. The second consists of a combination of linear regressions separately for each independent variable and determination of the variable with maximum effect on the root mean square error. This regression is then used to obtain a first estimate for the multi-regression relationship. The error after using this relation is used on the additional variables in sequence of their importance. By means of this method, the effect of spurious correlations or of cross correlation dependency between apparent independent variables is avoided.

3.3.2.3. Conceptual approach

Conceptual models provide a logical description of simple conceptual elements that simulate processes occurring in the catchment. However, addition of a number of conceptual elements in series and/or parallel in this semi empirical process, results in complex parameterization and optimization schemes.

Normally, a conceptual approach" may be based on empirical equations which represent the effect of internal storages, expressed through the Instantaneous Unit hydrograph (IUH). Todini (2007) pointed out that since its first formulation this conceptual modeling approach, evolved in two different directions: physically meaning full IUH modeling, and data driven hybrid models for the IUH.

The unit hydrograph application, based on observed hydrographs of rainfall and runoff was proposed and used by Nash (1970), Kachroo (1991) and many others. Kachroo and Liang (1991) stated that the unit hydrograph based on the assumptions of proportionality and superposition of time invariant responses, expresses the operation of a system in converting the precipitation excess x(t) to direct storm runoff y(t) by means of convolution integrals.

$$y(t) = \int_{\tau=0}^{\tau=1} x(\tau)h(t-\tau)d\tau \quad (3.7)$$

where τ is the time variable of integration and h(t) is the unit impulse response function or instantaneous unit hydrograph ordinate at time t.

In the physically meaningful approach the shape of IUH is defined a priori by modeler as the integral solution to a set of linear differential equations, and parameters are computed as

functions of physical characteristics of the phenomenon. The Nash cascade (1958, 1960) parameter estimation by virtue of Froude number, the bed slope, velocity, etc., by Dooge (1973) is an example to this end.

3.3.2.4. Hybrid metric conceptual approach

In the "hybrid metric conceptual approach", the conceptual models are estimated from the available data and used to test hypotheses about the structure of hydrological storages and processes on a catchment scale. Kitanidis and Bras (1980a) proposed a methodology of conceptual model structure verification by output data. But their emphasis was on describing the nature of the catchment response, i.e. essentially deterministic or stochastic. They reported that the prior conception of catchment response, i.e. essentially deterministic or stochastic suggests different types of modeling. The initial conceptual rainfall runoff modeling was based on deterministic formulation. They suggested replacing it with inherently catchment stochastic behavior. This modification resulted in a probability density function of system state.

Irrespective of a catchment´s deterministic or stochastic behavior, one can test the hypothetical conceptual model of a catchment against available data. Kachroo (1991) described a theory and applied hybrid metric conceptual models in flood forecast. In data driven hybrid approach, the Nash parameters are estimated by input to output historical data as proposed by Natale and Todini (1976a, b).

But there is agreement among hydrologists that neither the rainfall-runoff nor the runoff routing process is really linear. In the case of rainfall-runoff, the effect of rainfall on a catchment is greatly altered by wetness state/ antecedent moisture of a catchment. The rainfall may be completely or partially absorbed depending on soil wetness state. Seasonal variations in the runoff coefficient, i.e. the ratio of total stream flow volume to the total precipitation over a certain area and time, is analyzed by Kadoglu (2001) by the use of polygons instead of fixed ratio precipitation-runoff round the year. If one drops the assumption of time invariance in catchment response by introducing varying runoff coefficients makes it possible to use the basically linear convolution integrals, for a good replication of the non-linear rainfall-runoff process. The concept of variable runoff coefficient is tested, in this study, for the design of conceptual rainfall-runoff modeling routine of flood forecasting.

Application of conceptual approach in flood forecasting is of limited use in the absence of updating (Kachroo, 1991) by regular recalibration. In addition, adoption of an error updating in order to improve the estimated runoff from the conceptual approach, helps to overcomes the inherent limitation of general rainfall-runoff modeling in the context of flood forecasting.

In the light of above discussion, metric, and data driven hybrid metric conceptual models seem viable approaches for simulating runoff routing and rainfall runoff processes for the purpose of large rivers' flood forecast. Therefore, these two approaches for flood forecasting have been

tested on the Mekong River. From this point onward in this text, the metric model will be called Type-1 model, and data hybrid metric conceptual model will be called Type-2 Model.

3.3.3.1. Spatial discretization of forecasting models

In terms of spatial discretization, hydrologic models vary from simple lumped models to semi distributed models to more complex distributed models. This differentiation is based on the units of heterogeneity considered in the catchment. Lumped models consider the catchment as one homogeneous unit; consequently the parameters of lumped models often do not represent physical features of a hydrologic process and usually involve a certain degree of empiricism.

In lumped models, the impact of spatial variability of parameters is accounted by effective value computation for entire basin such as by area weighted average (Haan et al., 1982). In the case of discharge prediction, these models can provide as good a simulation as complex physical based distributed models (Beven, 2000).

Cunderlik (2003) stated that the parameters of semi distributed models are partially allowed to vary in space by dividing the basin into number of smaller basins. There are two main types of semi-distributed models: Kinematic wave theory models such as HEC HMS, and probability distribution models such as TOPMODEL (Beven, 1997).

Plate (2009) classified the plethora of available distributed models into three types; models based on rectangular grid, models based on sub-catchments, and models based on response units. This classification is used to describe the geological characteristics of the basin, trace its river networks, and identify surface and ground water interaction. The rectangular grid models utilize a digital terrain model (DTMs) or regular grid format. Grid based models are frequently applied for flood forecasting (i.e. Todini, 1996). But in actuality catchment based models should better be vector oriented. The response unit models divide the catchment into units of equal response.

For distributed models the catchment is divided into elementary heterogeneous units and flow is passed from one node to another as water drains the basin (Singh, 1988). The distributed modeling approach attempts to incorporate data concerning the spatial distribution of parameter variations together with computational algorithms to evaluate the influence of this distribution on simulated precipitation-runoff behavior. These models generally require large amount of (often unavailable) data for parameterization in each grid cell (Cunderlik, 2003).

The question remains, what should be the scale of heterogeneity in flood forecasting of large rivers. No doubt, the issue of heterogeneity should be referred to if it leads to an increase in output efficiency with each detailed level of heterogeneity consideration in moving from coarser to fine resolution. In the absence of practical pure physical models, efficiency of output in physically meaningful models at catchment scale is reduced, if emphasis of parameterization is on the point or process scale. The heterogeneity consideration in metric and hybrid models

should be based on moving from simple (lumped) to complex (semi distributed) models guided by increase of efficiency at each subsequent level as proposed by Nash and Sutcliffe (1970).

3.3.3.2. Temporal continuum of forecasting models

Cunderlik (2003) stated that the hydrological models could be divided into event driven models, continuous process models, or models capable of simulating both short term events and continuous hydrographs. Event driven models are designed to simulate individual precipitation-runoff events with emphasis on direct runoff, consequently they are required for flash flood forecasting approaches. Continuous – process models take into account the direct and indirect runoff by considering moisture recovery during the period of no precipitation. They are suited for long term volume forecasting (Ponce, 1989). In the case of large river flood forecasting, where the flood wave is a complex function of time delayed effective rainfall aggregation of a number of small and large scale events in different contributing sub-catchments, the continuous process model is appears to be best suited, as used in this work.

3.3.4. Parameterization methods

Unlike pure physical models, parameterization is an issue to be resolved in empirical, semi-empirical and conceptual modeling. Parameterization, as stated by Hochschild, (1999), is the so called quantification of model parameters (measured constant or variables) describing the system response. It is a basic component of model building.

Identification and optimization of parameters are two basic tasks in parameterization process. Different methods of identification and optimization lead to different types of modeling. Normally the identification and optimization are object driven, for example either to imitate the functional relationship to make useful predictions, or to reproduce realistic behavior of the process, which are two different objectives which lead to different sort of parameterization

As per Nash and Sutcliffe (1970) in empirical or analytical model parameterization, one should move from simple to complex models. Optimization of parameters should be based on minimizing the root mean square error, by suitable methods, for example by steepest descent method, or a by conducting search in the possible parameter space by moving parallel to the parameter axes. The optimization of a first set of parameters with a suitable objective function, the error variance R^2 in their case, should be stabilized before moving to the next set of parameters with increased order of complexity.

The same approach should be used in feed forward stepwise multiple regression unlike, feed backward stepwise multiple regression. In feed backward stepwise multiple regression one start from maximum parameter set and reduces the number of parameters stepwise to come to the optimum number of parameters.

In addition to model accuracy, Kachroo (1993) stressed the need of model consistency, whereby the level of accuracy and the estimates of the parameter values persist through different sets of data. The examination of parameter stability was assessed by Kachroo (1993) by 'split sampling', i.e., the division of available record into two periods, in one of which the model is calibrated and in the other, the verification period, it is tested and validated.They further proposed to check appropriateness of parameter sets for the respective model, and to assess its inadequacy i.e., usage of a linear model for a highly nonlinear system.

In general, the complexity of the parameterization process increases with the number of sets of parameters, their interdependence, and the non-linearity of the system. This makes complex models difficult to parameterize. The degree of freedom in calibration increases by involving more parameters, out of necessity to represent more process reality in the modeling. That is an effort, in recent practices of hydrological modeling, to make model realistic in terms of representing all the details of all the processes along with its primary useful predictive capabilities. The quest of realistic representation of process raised the issues of non-identifiability (Beck (1987, 1990)) and of equi-finality (Beven (1993, 1996a)), which emerged due to complex interplay of different parameter sets in the parameter space. Non-identifiability prevents to find some "true" description of the system, and equi-finality implies that many different sets of parameters might lead to the same hydrological system response, so that it is unlikely ever to be able to say that one has the true set (Beven, 2001). This issue is particular important if complex parameterization brings up the problem of optimization of different parameter sets. There are approaches like global optimum and multi-criterion optimization that are labeled as Pareto optimal set of models (Gupta et. al., 1998), which have been developed over time to resolve the issue of optimization with respect to different parameter sets. However, the issue of complex parameterization is irrelevant if the flood forecasting of large rivers is based on lumped or semi distributed catchments with rainfall runoff conceptualization by unit hydrograph.

3.3.5. Objective function

An objective function is needed as criterion for model parameter optimization. Most of the rainfall runoff model use Nash and Sutcliffe (NS) (1970) efficiency criteria as objective functions. This criterion is based on the proportion of improvement in the stage of "no model forecast" to "model forecast". In no model forecast, the mean of observed discharges was taken as best prediction estimate. The ratio of residual sum of squared errors of model forecasts to no model forecasts essentially reflects the model efficiency and consequently is used as objective function.

In flow forecasting this criterion, however, is subject to criticism due to firstly, unnecessarily primitiveness of the no model forecast and secondly, due to failure to draw distinction between uncorrelated random and correlated errors.

Nash et al. (1978) addressed this critique by proposing modifications to the original NS efficiency criterion. The first modification was to replace mean discharges in no model forecast by taking the mean of observed discharges on particular dates in the calibration period:

$$q_d = (1/n)\ (q_{d,1} + q_{d,2} + \ldots . q_{d,n}) \quad (3.8)$$

where $q_{d,1}$, etc., refer to the discharges on date d in the first, second years, etc., of the calibration period.

The second modification was proposed in the context of models which use an updating procedure in real time forecasting. If persistence in the errors (e) is observed during the calibration period, error corrections could be estimated by use of the autocorrelation function, and added to each no model forecast estimates as described in Nash et al. (1978),.

Kitanidis and Bras (1980b), however, proposed a list of different efficiency criteria Coefficient of efficiency, coefficient of determination, coefficient of persistence, coefficient of extrapolation, etc. In particular, the coefficient of persistence (PI), expresses the benefit achieved in real time model forecast through updating by using latest measurement instead of no model forecast. The coefficient is expressed through:

$$PI(k) = 1 - \frac{\sum_{i=0}^{imax}[(x_a(i+t) - x_f(i+t)]^2}{\sum_{i=0}^{imax}[(x_a(i) - x_a(i+t)]^2} \quad (3.9)$$

where x_a and x_f are observed and forecasted values and i is time index,

3.3.6. Forecasting model output

Rainfall runoff models for flood forecast are different from rainfall runoff models for design in the sense that the former can employ the advantage of real time updating, i.e., improvement in each estimated result on the basis of subsequent observations of actual values and thereby making available errors of previous estimates. Kitanidis and Bras (1980a) quoted that operational forecasting requires, in addition to a rainfall runoff model, a method for the continuous correction of forecasts based on observed errors in earlier forecasts (Nash and Sutcliffe, 1970). This feedback information proved valuable in improving real time forecasting performance as illustrated by among many others, Rodriguez-Iturbe et al. (1978) and, Kachroo and Liang (1992).

Updating is a kind of filtering, because the errors in measurement and model are filtered out through special algorithms. Frequently, Kalman filters are applied in linear models for updating (Lettenmaier and Wood, 1992). But other methods are also used: for example Kachroo and Liang (1992) have proposed autoregressive updating in real time flow forecasting.

The output of a flood forecasting model can be a single value of discharge or water level,, a trajectory of future outcomes as ensemble members, or ensemble average with various occurrence probability bands. The type of output, i.e. single or multiple results depends on the

nature of the forecasting model. For example, deterministic models inherently do not consider random components, consequently assume that a specific input results in one fixed output (Chow et al., 1988). On the other hand, stochastic models consider random variables, consequently produce various ensemble members based on random inputs and random input to output trajectories-.

Output and input may have various temporal discretizations. For example hourly, 6 hourly, daily etc, depending on the requirement. The spatial discretization on the other hand depends on the nature of flow routing, i.e., 1D, 2D or 3D.

3.3.7. Selection of flood forecasting model for the Mekong

There is a range of different forecasting schemes (as discussed), however, each rests on input data, model and output data. The nature of input data, i.e., data type, spatio-temporal discretization, length of time series influence the choice of the forecasting scheme. Data spatio-temporal discretization and gauging length in combination with required forecasting objective i.e. flood control, navigation, etc. influence the choice of modeling type, i.e. metric, conceptual, pure- physics based and hybrid. For large river forecasting models based on data spatio-temporal discretization are more appropriate than pure- physics based and physically meaningful conceptual models. However, these require a reasonable length of input and output data record, which is more important for metric and hybrid modeling approaches.

As per requirement of flood forecasting as explained in section-3.2, trend extrapolation and metric forecasting models (Type-1 Model) are more likely to be successful in flood forecasting, provided the empirical relation between input to output data is revealed correctly, and that dependency of output to input is largely to be detected by correlation and covariance statistics. Uncomplicated simple linear relationships can be used in flood forecasting, provided that no significant changes in the relation of output to input by anthropogenic or climatic changes occur.

On the other hand, pure physical and conceptual models are not a good choice for flood forecasting in large rivers because these models rely on the explanation of catchment physics by partial differential equations, and the appropriate conceptualization of the catchment model primarily is the key to the relation of output to input. The degree of realistic representation of catchment physics by pure physical laws or by conceptual abstraction is seen in the accuracy of the output. However, the problem of producing a realistic physical representation of the structural, process and storage elements is limited because of insufficient data, and thus hinder the application of pure physical models to any practical flood forecasting problem on macro scale catchments. Application of pure conceptual models in flood forecasting is limited by the problems of catchment representativeness, scaling, non-identifiability and equi-finality. The physically meaningful IUH conceptual models as illustrated by Dooge (1973) are also not free

from the problems of complex parameterization.

It is concluded that (Type-2 Models) can be successfully used in flood forecasting. Suitable conceptualizations can be tried on and verified against the available data, and hypotheses about the structure of catchment scale hydrological storages and processes can be tested with more or less emphasis on theory (physics) and data (empiricism). But in this approach, physical interpretation must be coupled with subjective choices, i.e., separation of storm runoff from base flow. Another problem is the estimation of effective rainfall and derivation of actual IUH/TUH shape. The advantage of better physical interpretation usually is lost, however, when parameter estimation is solely subjected to minimization of a suitable objective function.

From this discussion it is evident that one can use both Type-1 and Type-2 Models in flood forecasting of large rivers – and therefore both will be used in flood forecasting for the Mekong. However, the problem of subjective choices in Type-2 Model has to be solved. Therefore, in this study, instead of separating the flow components, they will be taken as lumped quantities.. For estimation of effective rainfall from total rainfall, a special methodology will be developed (described in chapter-7). In this methodology the empirical behavior of catchment processes is used in a prior analysis and later verified by input to output data validation, which indirectly preserves the physics of the catchment in the data.

The benefit of updating will be used in the forecasting model to be developed, and three criteria, i.e., Coefficient of determination (RSQ), NS and Persistence index will be used. The former two will serve the purpose of basic rainfall-runoff model optimization and later for updating model optimization.

In line with required objectives of flood forecast accuracy and quality, the effect of uncertainty due to natural variability and knowledge uncertainty will be covered by generating predictive uncertainty bands. This approach combines the results from both the hydrologic uncertainty processors (HUP) and meteorological uncertainty processors (MUP) of Krzysztofowicz (2001) into one description of predictive uncertainty, without identifying their origin from input data, model building, parameter estimation. In order to clarify, the term predictive uncertainty used here is different from its use by Todini (2007), who defined uncertainty based on covariate and parameter sets.

3.4. Review of Flood Forecasting Models Applied in LMB

In this section the existing flood forecasting models, i.e. SSARR and URBS of Mekong will be presented. The emphasis will be on the structure and performance of these models.

3.4.1. SSARR Model

The SSARR model has been used by RFMMC for flood forecast of Mekong from 1980 to 2005. The Stream flow Synthesis and Reservoir Regulation Model (SSARR) is owned by the U.S. Army Corps of Engineers but is available via Internet at sites of the U.S. Geological Survey and the Columbia Basin Water Management Division (http://www.nwd-wc.usace.army.mil/). The SSARR Model is a watershed simulation model developed around 1960. It is a simple linear storage model which includes the major hydrological processes: rainfall as input, snow accumulation and melt (not implemented in MRB), evapo-transpiration, a fast interflow (soil) and slow interflow (groundwater) and river channel routing. Included is a reservoir model to calculate the effects of dam control and natural reservoirs.

3.4.1.1. Basic input and structure of SSARR model

The SSARR model is divided into two major model concepts: the watershed model and the river system and reservoir regulation model (Fig.3.2). The watershed model has two options, which are based on different approaches for snow cover incorporation. The watershed model includes snow accumulation and melt, evapo-transpiration after Thornthwaite's formula, and an interception routine as well as a cascade of linear storage approaches. It satisfies the law of continuity through the storage equation with inputs from the other components of the hydrological cycle, such as surface runoff routing. This is coupled with the soil moisture storage (also named subsurface) and the evapo-transpiration module (Rockwood, 1972; US Army Corps of Engineers, 1987). Base flow components as well as a lower zone components are included, which are - as also other components- determined as a percentage value of the faster subsurface storage components (Lindenmaier et. al., 2010).

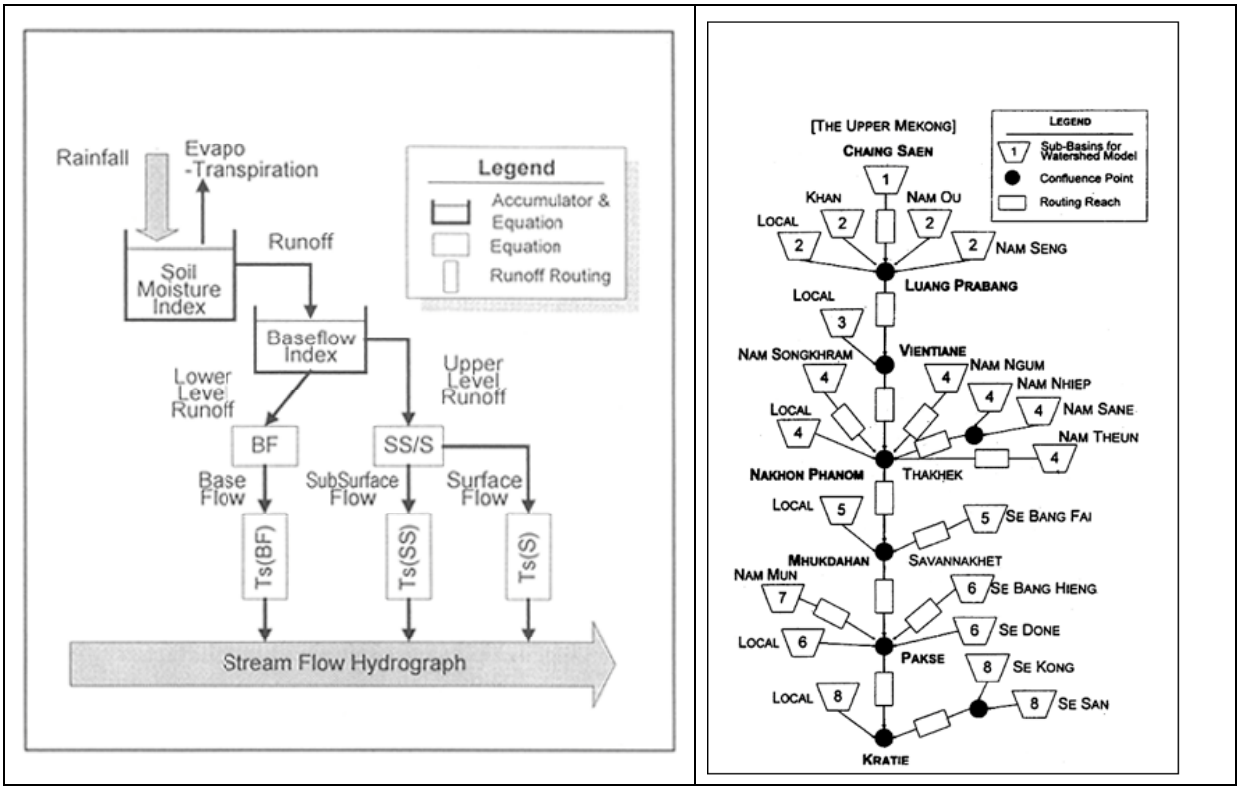


Fig. 3.2: a) basic principle of SSARR watershed model b) structure of SSARR model as it was in 1997, both graphs are from (MRC, 1999).

In SSARR, the river system and reservoir regulation is taken as a linear storage approach like the watershed model, the reach is separated in segments and each of these segments is again represented by several storages which are based on following equation:

$$T_S = \frac{KTS}{Q^n} \tag{3.10}$$

where

TS = the time of storage per increment in hours

KTS = A constant to regulate the outflow hydrograph's shape

Q = Discharge [volume/time]

n = a coefficient usually between -1 and 1

KTS affects the time of storage linearly, larger values of KTS result in a longer time of storage for a given number of phases. KTS is rather responsible for the translation, the n value for the peak height of a runoff curve (Lindenmaier et. al., 2010).

To ease the pre- and post-processing for the acting hydrologist at the RFMMC several shells have been designed. The pre-processing/ processing shell is a Visual Basic routine that helps to re-format the input data for the use of the SSARR executable. In the post-processing, data from the SSARR output file is imported to an Excel file (Pich, 2006).

The SSARR executable of the MRC does not work on Windows XP and Windows 2000 system software due to the closure of a security gap by Microsoft Inc, and thus it cannot be revised from its present condition for adaptation to the present situation (Lindenmaier et. al., 2010).

3.4.1.2. SSARR model performance

The poor performance of SSARR in recent year floods of 2000 and 2005, forced the stakeholders to revisit the forecasting model by either updating SSARR or replacing it with a new system. The study conducted by (Lindenmaier et. al., 2010) on improving flood early warning system of Mekong, stressed the need of adopting a new forecasting model by quoting that the SSARR model as it is used in the RFMMC has the limitation that the parameter setting has been done a long time ago and that the knowledge about parameter change and model justifications has been lost after the long time hydrological forecaster retired. Reaches were added and subtracted in the model depending on data availability which is not possible nowadays any more (MRC, 1999). In addition to this, the model as it is used at the RFMMC lacks proper description and is very user unfriendly.

3.4.2. URBS Model

In order to replace model SSARR by a better approach, the model URBS was selected as trial model by RFMMC. URBS is a runoff-routing networked model of sub-catchments based on centroid inflows. Pengel et al. (2008) described the application of URBS for flood forecast of for the Lower Mekong basin (LMB). URBS is a semi-distributed non-linear conceptual runoff routing model, (Carroll, 2004). It is a computer based, hydrologic modeling program that enables the simulation of catchment storages and runoff responses by a network of conceptual storages representing the stream network and reservoirs. URBS combines two hydrological modeling processes into one model: rainfall runoff modeling, which converts the gross rainfall into net or excess rainfall and runoff routing modeling, which takes the excess rainfall as input and converts it into flow (Carroll, 2007).

The excess rainfall is determined by accounting for losses by two different options. First mode is to determine loss as event model and second is by accounting for continuous losses. For the event model, separate loss calculations are done for pervious and impervious areas. In

impervious areas the URBs model by default assumes no loss. However, in pervious areas the loss is computed by continuous loss model (rainfall is lost on all parts of the catchment), proportionate runoff model (only part of the catchment contributes to runoff), and Manley Philips loss model (Rainfall is lost on all parts of catchment up to end infiltration) (Carroll, 2007).

After the excess rainfall has been determined, the runoff routing component of the model routs the excess rainfall through a series of conceptual non-linear storages to determine the distribution of flow in the catchment. The runoff routing component can be applied in either the basic or the split mode. In the basic mode, the effect of sub-catchment and channel storage is treated as a lumped storage at the centre of each sub-catchment. This basic model is a simple RORB-like model (Laursenson & Mein, 1990) in which stream lengths are assumed to represent both catchment and channel storage. Each conceptual non-linear reservoir is represented by the storage-discharge relationship (Carroll, 2007):

$$S = K_c Q^m \qquad (3.11)$$

where K_c is the non-linear routing constant for a single reservoir and is a function of the sub-catchment and channel storage characteristics, empirically determined to read (Carroll, 2007):

$$K_c = \alpha f Ln(1 + F)^2 / \sqrt{Sc}\ (1+U)^2 \qquad (3.12)$$

where
S = catchment and channel storage [m3h/s]
α = storage lag parameter
f = reach length factor
L = length of reach [km]
U = fraction urbanization of sub-catchment
F = fraction of sub-catchment forested
n = channel roughness or Manning's n
Sc= channel slope [m/m]
Q = outflow [m3/s]
m = catchment non-linearity parameter

When stream length alone is used to represent catchment and channel storage, the default values of α and m adopted by URBS are 1.2 and 0.8 respectively. These values have been adopted from typical catchments of South- East Queensland (McMahon and Muller, 1986).

In the split mode, the effects of the sub-catchment and channel routing are calculated separately.

Firstly, the excess rainfall on a sub-catchment is routed through a conceptual storage at the centre of the sub-catchment to the creek channel. The lag of the sub-catchment storage is assumed proportional to the square root of the sub-catchment area. Next, the channel inflow is routed along a reach using a linear or non-linear Muskingum method, whose lag time is assumed proportional to the length of the reach. The split mode was preferred and adopted by RFMMC. The split mode model is similar to Watershed Bounded Network Model or WBNM (Boyd, 1987), except assuming channel storage is proportional to channel length instead of catchment area as in WBNM. In catchment routing, the travel time (T) in hrs from sub-catchment perimeter to the centroid is computed as (Carroll, 2007):

$$T = \sqrt{A/\pi} \ / \ v \qquad (3.13)$$

where A is the area of sub-catchment and v is the velocity of flow in km per hr (Carroll, 2007).

Once the rainfall has been routed using the time-area diagram, it is routed again through a nonlinear reservoir. The storage-discharge relationship for this reservoir is (Carroll, 2007):

$$S_{catch} = (\beta\sqrt{A}(1 + F)^2 / (1+U)^2) Q^m \qquad (3.14)$$

where
S_{catch} = catchment storage [m^3h/s]
β = catchment lag parameter
A = area of sub-catchment [km2]
U = fraction urbanization of sub-catchment
F = fraction of sub-catchment forested.
m = catchment non-linearity parameter

3.4.2.1. URBS model performance

The performance of URBS model in Mekong River flood forecasting is documented by Malone (2009). He stated that the quality of URBS forecast is a function of the input, and that the poor output of URBS forecast, as seen in flood season 2008, could be attributed to poor input of rainfall by SRE. The issue of inconsistencies in the parameters adopted for URBS with reference to different sub-catchments is further discussed as reason for poor flood forecast. Further, the possibility of adding base flow modeling is under consideration to improve model performance.

3.4.3. Remarks on Contemporary Flood Forecasting Approaches of Mekong

MRC has used SSARR and URBS model in order to produce 1 to 5 days flood forecast for LMB. However, there are number of reservations on the performance of these models. The poor performance of these models has different reasons but is due mainly because both of these models do not satisfy the requirement of flood forecasting model as discussed in section-3.2.

The estimated parameters of SSARR to define soil moisture index, subsurface and surface flow are not serving the purpose of efficient forecast anymore as reflected in forecasting performance of year 2000 and 2005 floods. The improved semi-distributed non-linear conceptual runoff routing model URBS is under trial by RFMMC, however, with limited success at number of stations upstream in Lao-PDR and downstream in Cambodia of Lower Middle Mekong. The basic problem in application of both SSARR and URBS is because of their approach of top to bottom modeling by moving from catchment to channel. In this process of moving from catchment to channel, there are lot of uncertainties involved. Part of this is due to the fact, that the physiographic data required to parameterize efficiently the model in calibration stage is not available. Secondly the high dependence on SRE as rainfall input, where SRE has never been validated by ground based data till today. Thirdly, the semi-distributed URBS Model requires a lot of physiographic and hydrologic data in order to optimize the model and its subsequent operation. But data of required quantity and accuracy is not available in the data scarce catchment of Mekong. Fourthly, both SSARR and URBS do not have any special routine which can account for the possible uncertainty of flood forecast. All these factors result in the failure of SSARR and URBS to produce flood forecasts of desired accuracy and reliability. Consequently, in the absence of success in flood forecasting by SSARR and semi distributed model URBS, the need emerges for a better forecast model. The metric (Type-1 Model) and hybrid conceptual Data Based Modeling (Type-2 Model) approach will be tried in this thesis to improve flood forecasting quality in terms of accuracy and reliability.

4. Data Review

4.1. Data Collection

The quality of any engineering analysis and/ or design is dependent on the input data. Initial screening of available data suggested that the problem of Middle Mekong flood forecasting will have to be addressed under the constraints of limited data availability. The former Colonial Governments in Southeast Asia started to implement rainfall and runoff observations in the early 60's in the Mekong River Basin (MRB). However, the civil unrest from the 1970 up to 1990 led to huge data gaps, especially in the hinterland of Lao PDR and Cambodia. On the Thailand side, observation of water level at gauges was going on throughout the years, but the water stage relationships were not measured, as the Mekong was the border to Lao PDR. Furthermore the availability of rainfall data is poor because only a small number of rain-gauges existed, and secondly, because most of the gauges are concentrated in plain areas, near the main river and hence they do not very well represent climatic conditions of whole sub-catchments. The data availability sheet by MRC includes a total of 480 rainfall and 243 discharge gauges in the Lower Mekong basin which were operational of and on since 1900. But the length of data at most of the gauges is limited to few years.

4.1.1. Sources for water level and runoff data

Data collection in trans-boundary rivers like river Mekong is a complex issue. Data are collected and kept by any number of different local, regional and national agencies in several countries. In the absence of uniform quality standard methods of data collection quality of data also vary from agency to agency. Furthermore the data banks are scattered over different agencies. MRC, being the central coordinating commission, solved the problem of gathering data in one form. Unfortunately, all the required data couldn't be obtained at the start of this study. Rather data were supplied in four increments from different sources.

One major source of historical runoff data for this study was access to the HYMOS database of MRC, which was located at the MRC headquarters in Vientiane. It contains discharge and rainfall data of all stations in the member countries of the MRC. However, data gathering and implementation was temporary stopped in 2002 for political reasons. A second data set for major discharge and rainfall stations was available from the RFMMC. Observed water levels and precipitation of 20 to 40 stations were obtained from the line agencies of member countries. This data set was unfortunately neither complete nor checked on mistakes. In addition, the number of recorded stations changed from year to year, and finally the records contained only water level data and no discharges. This data set was available from 2001 up to 2007. The MRC headquarter started another consolidated effort to collect data from different line agencies and to compile them in a single data base. The data were compiled by the British consultant firm Halcrow, who prepared an initial report to Water Utilization Project (WUP) of the MRC in November, 2001.

Another set of data for this study has been provided by the MRC headquarter in 2008, which was a first revision of the HYMOS database (third data set). On comparing this data set with previously collected (2007) data from RFMMC and in the HYMOS data base (initial version), certain discrepancies were found and brought to the attention of both RFMMC and MRC personnel. Finally, a revised set of data was made available to IWG in early 2010 (fourth data set). All four data sets were thoroughly examined for quality and corrected, firstly in 2007-2008 and then rechecked in early 2010.

One consistent data bank has been created and used in this study for further analysis. Only two, the revised HYMOS and MRC's 2010 data out of the four mentioned data sets, i.e. HYMOS, RFMMC, Revised HYMOS and MRC's 2010 data came with some sort of supporting information in the shape of Meta data and rating curves. Therefore, these two data sets were preferred, also because they were complete in recording in comparison to the other two data sets. In the end, the revised HYMOS data set was taken, which is augmented by recent data collected in early 2010.

4.1.2. Main Mekong River Data

The 2600 km long stretch of Lower Mekong from the Golden Triangle (at the borders of Thailand, Lao PDR, China and Myanmar) to South China Sea is gauged at 17 locations. However, only for 7 gauges (Table-4.1) are the water levels and discharge data for the main Mekong River fairly complete and of higher quality in comparison with other gauges.

4.2. Method used in Hydro-meteorological data collection

4.2.1. Water Level Data

The water level data have been mostly taken by vertical staff gauges. Quality of water level data collected by this method is heavily affected by bank erosion and by damage to gauge installation by floating logs and trash. Accuracy of the water level depends on maintenance of these installations. Though staff gauges were in recent years replaced by less vulnerable slope gauges yet these are also prone to damage in lowest and highest water levels.

Bubble type automatic water level recorders have been used in the Mekong region from 1960 to around mid seventies, but due to lack of maintenance these gauges were abandoned. There was a new attempt by the Australian AHNIP project in early 2000 to equip 18 stations with automatic water level recorders and telemetry system for real time water level data sharing among the participating nations, including 2 stations in China. Furthermore, a new generation of electronic automatic water level recorders has been introduced into the region by the Japanese WUP-JICA project, and other Hydro-meteorological Network Improvement projects. However, maintenance

problems are limiting regular gauging at each station even by this new system. Therefore, main set of water level data used in this study were collected by vertical and slope staff gauges.

Tab. 4.1: Gauging stations along Lower Mekong

Sr. No.	Gauging Station	Latitude	Longitude	Elevation of Zero gauge (meter amsl)	Catchment Area ($10^3 km^2$)	Distance from Mouth km	Consistent Data Record
1	Chiang Saen	20.273	100.083	357.11	189	2363	1960-2007
2	Luang Prabang	19.892	102.137	267.195	268	2010	1960-2007
3	Viantiene	17.928	102.620	158.04	299	1583	1960-2007
4	Nakhon Phanom	17.398	104.803	130.961	373	1215	1960-2007
5	Mukdahan	16.540	104.737	124.219	391	1124	1960-2007
6	Pakse	15.117	105.800	86.49	545	868	1960-2007
7	Stung Treng	13.545	106.017	36.79	635	667	1960-2007
8	Kratie	12.240	105.987	-1.08	646		1960-2007

4.2.2. Discharge Gauging

In discharge gauging, the measurements are made by boat at the measuring cross section, which is usually fixed due to fixed infrastructure for positioning. Error in the boat position is sometimes high, due to difficulties in maintaining the boat in a fixed position against strong currents of water. A cup type current meter was initially used which was replaced by a propeller type meter, which is a more accurate instrument for estimation of water velocities, in particular if propellers can be changed according to the magnitude of the velocity to be measured. But they are more vulnerable to floating logs and rough treatment. A number of manufactured brand name current meter and related equipment were introduced into the Mekong namely: GURLEY (a cup type), NEYERPIC, A.OTT, SEBA, OSS B1, VILLEPORT (propeller types) etc. (MRC Hydro-meteorological data base review report, 2004)

4.2.3. Rating Curves

The establishment of a reliable relationship between the monitored variable stage and the corresponding discharge is essential at all river gauging stations when continuous-flow data is required from the continuous stage records. This calibration of the gauges is dependent on the nature of the channel section and of the length of the channel between the site of the staff gauge and the discharge measuring cross-section. All continuous estimates of discharge derived from a continuous stage record depend on the accuracy of the stage values and the rating curves. However, the river cross sections change over the years because of silting and scouring, therefore, updating of this stage discharge relation is required at regular intervals, especially, in rivers where floods are frequent.

4.3. Data Availability

4.3.1. Availability of Water level and Discharge data of Main River

The water level of main stream gauges has been recorded fairly regular from year 1960; however, discharge gauging was not conducted regularly. As it is presented in Table-4.2, there are certain gaps in discharge gauging. Normally, regular discharge measurements are required to update ratings curves for computing realistic discharges. But in the cases of gaps in discharge measurements, the stage-discharge relations of neighboring years have been used to produce discharge data. For example, in the case of Chiang Saen, the missing year discharge data is produced by neighboring years rating curves, which are divided into five groups from year 1960 to 2000 as given below (MRC Hydro-meteorological data base review report, 2004):

.

- Group of 1960, 1962, 1968, 1969 (black), and 1970 in light blue.
- Group of 1998,1998 , 2000 (red)
- Group of 1971,1872, 1973,1974, 1975, 1995, 1996 (pink), and
- Group of 1997, 1998, 1999 and 2000.

Discharges computed by using ratings of neighboring years add additional uncertainty to the quality of data, which however is inevitable. A similar procedure is repeated for other gauges also to generate discharges of the years where measurements were not available.

Tab. 4.2: Gaps in discharge gauging along Lower Mekong

Nr.	Gauge Name	Gauge Type	Gauging initiation	Daily Gauging frequency	Discharge gauging gaps	Maximum recorded water level (m)	Date of maximum water level
1	Chiang Saen	Bubble Gauge	1957	3-13	1975-1994	13.82	3-09-1966
2	Luang Prabang	Vertical staff gauge, slope gauge, Bubble gauge	1950	2-11	1962-66, 1969-70 ,1974-85	22.38	02-09-1966
3	Vientiane	-	1895	-	1962-66, 69-73, 74-86, 89-90	12.71	04-09-1966
4	Nakhon Phanom	-	1960	3-5	1966, 1976-93	13.34	11-09-1966
5	Mukdahan	Slope gauge, Staff gauge	1959	-	1975-82, 1984,86,87	14.24	19-08-1978
6	Pakse	-	1960	-	1962-66	14.48	17-08-1978
7	Stung Treng	Staff gauge, Slope gauge	1900	-	1900-1954,57 1960-04	13.00	02-09-1939
8	Kratie	Slope gauge	1933	-	1933-1959, 1970-2001	24.28	03-09-1939

4.3.2. Data availability of Lateral Tributaries of Lower Mekong

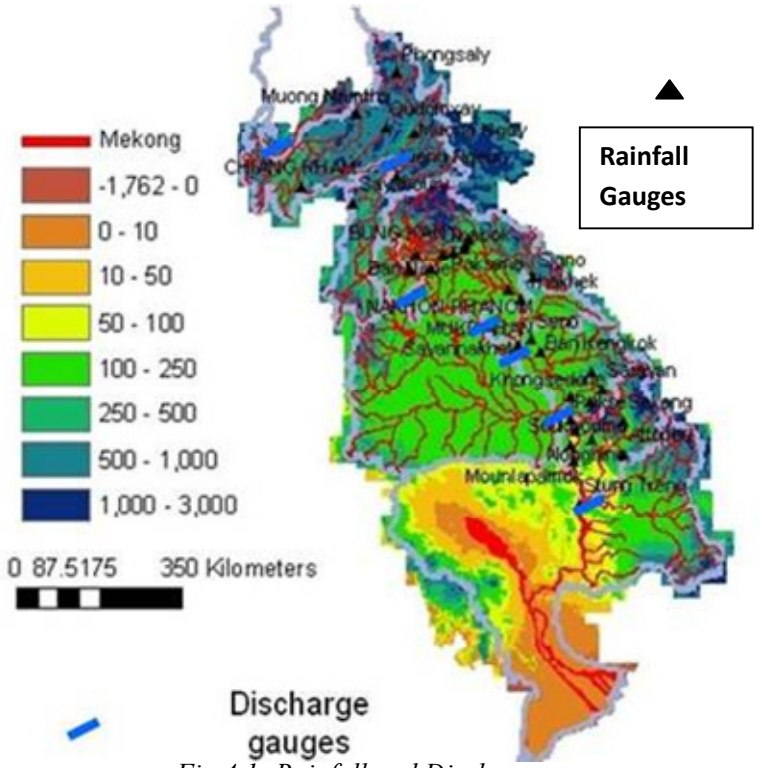


Fig.4.1: Rainfall and Discharge gauges

As the Mekong flows from north to south, from China to South China Sea, a number of tributaries from both banks discharge into the main Mekong. Water level and discharge gauging on these lateral tributaries were conductedduring the last 40 years but not regularly. Furthermore, only some of the lateral rivers have been completely gauged. For example, between Chiang Saen to Luang Prabang, only 65% area of the total lateral tributaries' catchments is gauged. With a total tributary catchment area of 79000 km^2, this reach covers 14% of the area of the lower Mekong basin. Similarly, gauged lateral tributaries drain only 21% of the area between Luang Prabang and Vientiane. The condition of discharge gauging is alsonot different in other reaches.

Rainfall has been gauged at a number of stations in different sub-catchments of lower Mekong, but most of these gauges are not truly representative of the sub-basins. Furthermore, most of these gauges were not operational for a number of years. Only 34 rainfall stations in left bank tributaries of the Mekong have fairly regular rainfall gauging records (Fig.4.1 and Table-4.3).

4.4. Quality of data

4.4.1. Water stage relation curves, river cross sections and warning levels

The quality of data, especially of the rating curves, poses a big problem for the understanding of the flow regime of the Mekong. First of all, the water stage - discharge relation has to be measured each year, because the morphological activity of the riverbed is substantial. In the upper reaches erosion and river bank displacement lead to ever changing runoff relationships. A report on the water stage relationships (RFMMC, 2004) shows that almost all gauges of main Mekong show large variability, especially when high water levels are measured (Fig.4.2) , which

is due to changes of the riverbed. Lower river gauges are greatly influenced by whether measurements were taken during a rising or falling limb of the hydrograph. For example, Kompong Cham shows three to four relationships depending on the seasonal time. It is not clear which rating curves were used for the HYMOS data or for its revised versions by MRC for the

Tab. 4.3: Availability of rainfall data

No.	St.Code in:		Station names	Coordinates (Indian 1960 geodetic datum)		Country	Altitude (m)	Data Availability
	Map	HYMOS		Latitude	Longitude			
1	229	140504	Mounlapamok	14.333	105.867	Laos	94	65-70;79-87;89-01;05
2	256	140506	Soukhouma	14.650	105.800	Laos	95	79-92;94-05
3	259	140507	Muang Champasack	14.900	105.883	Laos	95	79-80;82-05
4	266	140705	Attopeu	14.467	106.833	Laos		29-30;32-33;35-38;43-44;88-05
5	225	150504	Pakse	15.117	105.783	Laos	93	29-39;50-05
6	224	150506	Khongsedone	15.567	105.800	Laos	122	63-70;72;79;83;88-05
7	227	150602	Saravan	15.717	106.433	Laos	170	29-33;35-39;42;64;70;87-05
8	270	150604	Laongam	15.467	106.167	Laos	451	30-33;35-39;89-05
9	260	150605	Nonghine	14.750	106.217	Laos		80-05
10	223	160405	Savannakhet	16.550	104.750	Laos	155	27-29;31-40;56;65-92;94-05
11	221	160502	Seno	16.667	105.000	Laos	184	50-73;75;76;78-81;84-90;95;99,01-04
12	220	160505	Ban Kengkok	16.433	105.200	Laos	126	31;35-39;65-67;89-05
13	265	160601	Muong Tchepon	16.033	106.233	Laos		23-25;27;30-32;35-38;90-05
14	218	170404	Thakhek	17.417	104.800	Laos	146	29-32;35-39;42;56;61-64;80-82;84-92;94-05
15	262	170501	Signo	17.833	105.050	Laos		87;89-05
16	216	180203	Ban Nasone(Maknao)	18.017	102.967	Laos	161	63;65-76;79;82;90-05
17	208	180303	Paksane	18.400	103.633	Laos	157	24;30;31;33;36-39;41-43;65-80;83;87-91;93-05
18	209	180304	Thabok	18.283	103.200	Laos	159	65-73;89-95;02;04-05
19	254	180307	Muong Kao(Borikhane)	18.567	103.733	Laos		29-33;36-39;41;43;78;79;88-05
20	253	180308	Muong May	18.500	103.667	Laos		78-80;85;88-05
21	285	180501	Ban Nape	18.283	102.667	Laos	100	22-24;27-32;34-45;93-94;96-05
22	233	190103	Sayaboury	19.233	101.367	Laos	323	64-75;78;80-89;91-05
23	243	190205	Xieng Ngeun	19.750	102.233	Laos	304	75;88-05
24	252	200101	Muong Namtha	20.930	101.400	Laos	600	29-33;36-38;41-43;92-05
25	251	200201	Muong Ngoy	20.567	102.600	Laos		96-05
26	273	200204	Oudomxay	20.680	102.000	Laos	550	91-05
27	250	210201	Phongsaly	21.733	102.200	Laos		22-25;27-44;88;90-05
28	428	160401	MUKDAHAN	16.533	104.733	Thailand	138	50-04
30	343	170403	NAKHON PHANOM	17.500	104.333	Thailand	140	53-04
31	342	180302	BUNG KAN	18.333	103.417	Thailand	164	79-04
32	307	190002	CHIANG KHAM	19.517	100.300	Thailand	394	83-85; 87-04
33	072	120502	Stung Trang	12.250	105.540	Cambodia		85-87, 92, 2001-02
34	226	150609	Sekong	15.083	106.850	Laos	126	92-93,95-05

lower Mekong River gauges, in particular starting from Kratie and further downstream. Therefore, the analysis of stage discharge data to produce forecasts is conducted only up to Stung Treng, which is the next gauge upstream of Kratie.

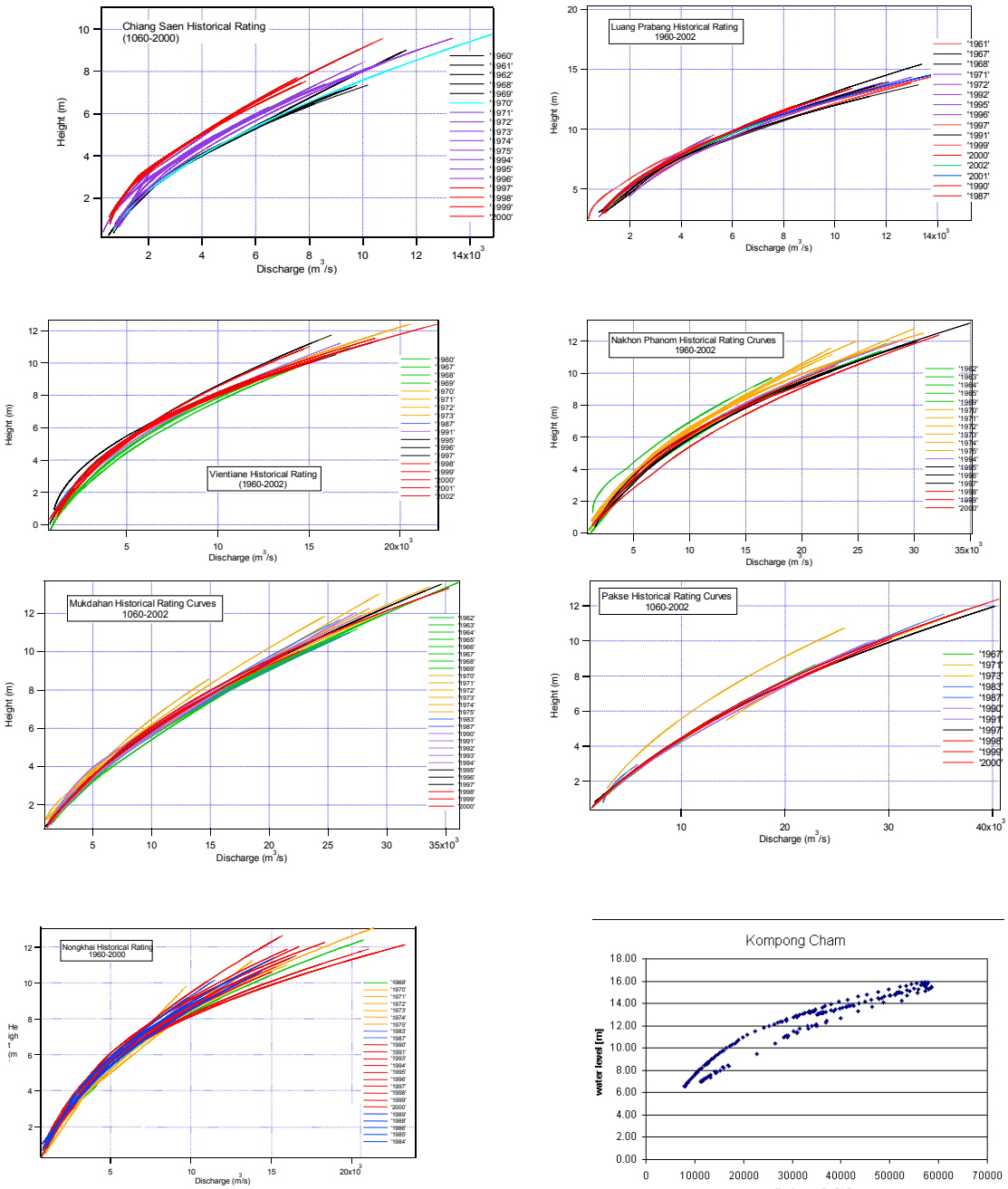


Fig. 4.2 a) Historical rating curves of Chiang Saen, Luang Prabang, Viantiane, Nakhon Phanom, Nong, Mukdahan, Pakse and Nong Khai which show discrepancy especially at high water levels (MRC Hydro-meteorological data base review report, 2004)

b) rating curve of Kompong Cham which shows several rising and falling limbs

There were four different versions of rating curves, i.e. SSARR, URBS, MRC old and MRC NEW, available to the writer.

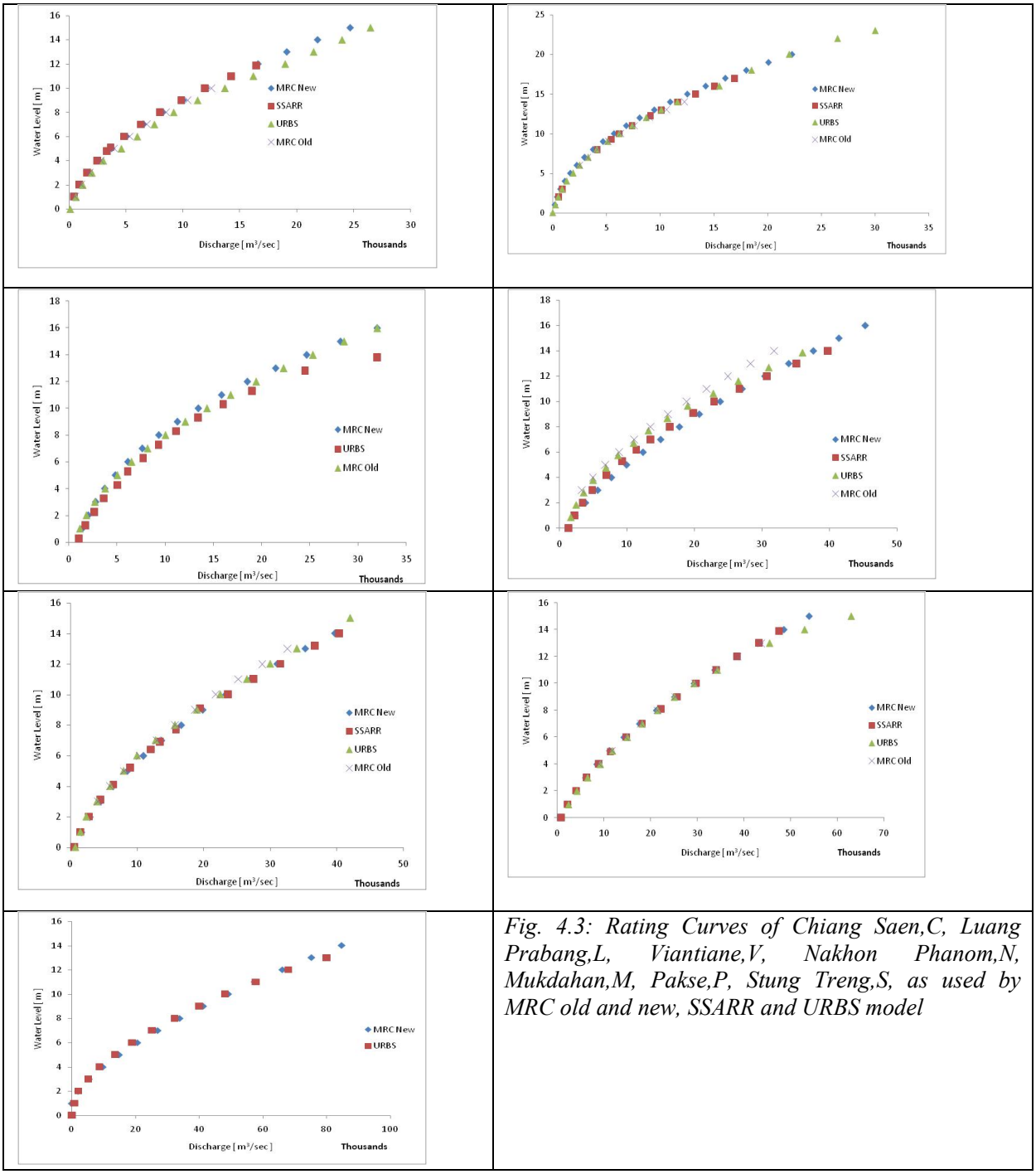


Fig. 4.3: Rating Curves of Chiang Saen,C, Luang Prabang,L, Viantiane,V, Nakhon Phanom,N, Mukdahan,M, Pakse,P, Stung Treng,S, as used by MRC old and new, SSARR and URBS model

It is certain that the rating curves used for the SSARR model are different from the ones that are used today by MRC in the URBS model. And both types of rating curves do not agree with

official MRC old and new rating curves. Curves used in SSARR are very old and do not represent the stage discharge relation of today. The set of URBS rating curves was especially prepared in the context of URBS model application. Rating curves for the URBS model were produced so that the discharges produced by URBS model under given conditions of rainfall input yielded measured water levels of those dates. And on the basis of the assumption that the URBS model is performing with 100% accuracy, the measured water levels were fixed against discharge output produced by URBS. Therefore, these rating curves are not actual water level - discharge rating curves, but discharge - water level calibration curves for the URBS model. Consequently, only the old and new sets of MRC rating curves are left to be adopted for water level to discharge conversion. Therefore, in this study the MRC new set of rating curves is used for all gauges except Nakhon Phanom (Fig.4.3).

A review of the data quality was conducted by plotting the time series of water levels and discharges of each gauge separately. A problem in reviewing the data quality of water levels and discharges was, as indicated earlier, that there existed four different available data sets, each based on different sets of rating curves with incomplete documentation. Documentation of Meta data on discharge measurement dates and types for each data set is seldom available. Comparison of these different data sets revealed that discrepancies in data were both random and regular. Regular differences were due to different ratings, whereas random difference may be due to typing errors.

The yearly plots of water level against discharge supported the findings of (MRC review of data base report, 2004) that cross section morphology changes annually. Plotting of yearly discharge of consecutive gauges in the same Fig. reflect discharge contributions of lateral tributaries (Fig.4.4). Lateral discharge contributions were determined indirectly by subtracting upstream discharge from downstream discharge with time lag because gauges on lateral rivers are mostly not available. In some cases these inferred discharges show trends, for example the lateral contributions between Viantiane and Nakhon Phanom, and between Nakhon Phanom and Mukdahan show upward and downward trends respectively (Fig.4.5a).

This trend in discharge starting in year 1993 onwards seems questionable, therefore they were cross-checked against trends in rainfall time series. Those discharge trends which seemingly were not confirmed by a trend in rainfall, were corrected by adjusting the rating curves of the respective years, i.e. 1993 to 2007. That lateral discharges in these reaches could not be correct was also emphasized by negative contributions in the flood season in the reach from Nakhon Phanom to Mukdahan, which is not physically possible unless there is outflow, which was not observed. There seems to be a problem in the stage – discharge relation of Nakhon Phnom from year 1993 onwards. However, there is no such trend in discharges of year 1991 to 1993. Therefore, the stage –discharge relation of year 1991 to 1993 was used to compute discharges of Nakhon Phanom for year 1993 to 2007. The lateral inflow between Vientiane and Nakhon Phanom (VN) and between Nakhon Phanom and Mukdahan (NM), computed by using previous stage – discharge relation and revised stage – discharge relation,

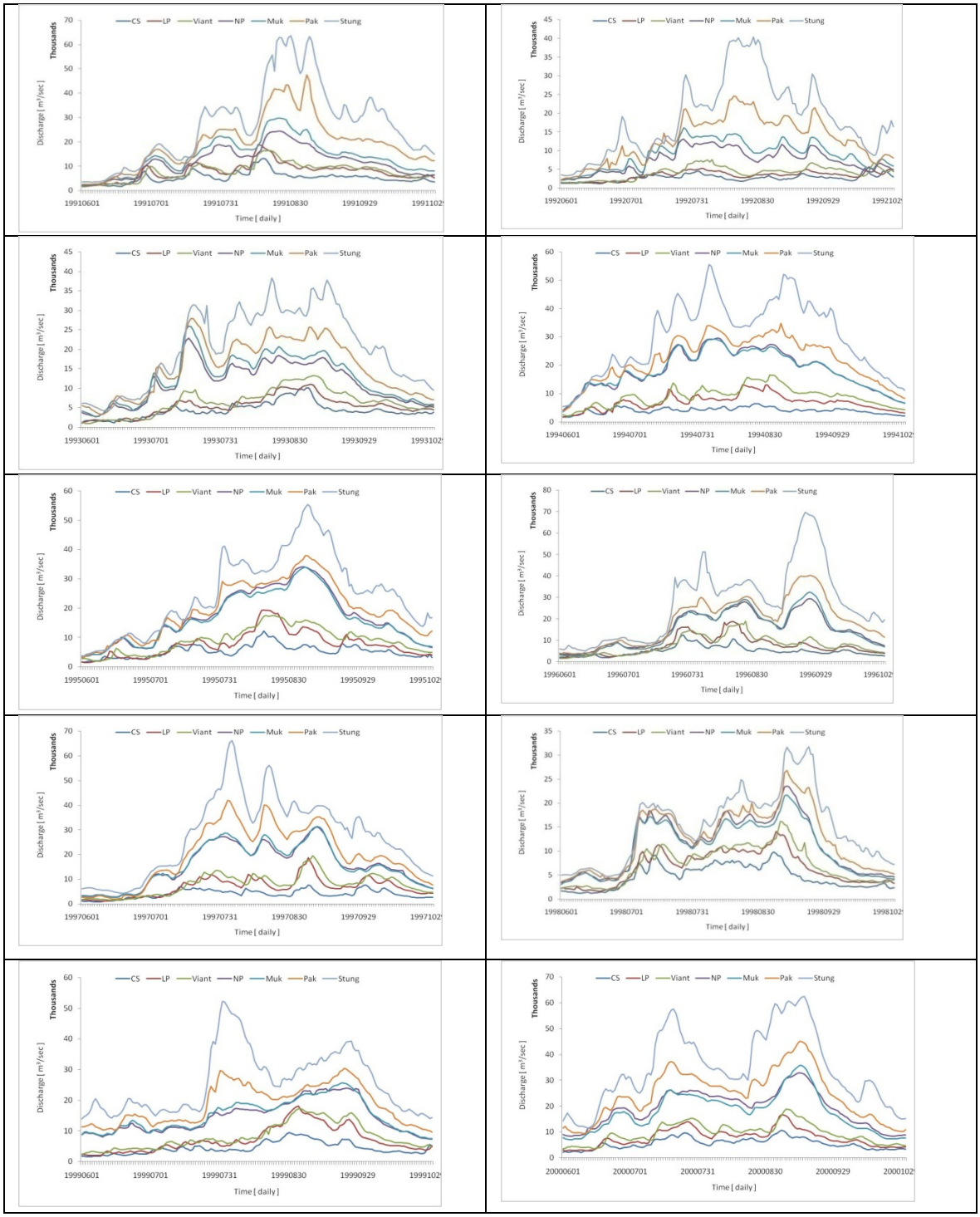
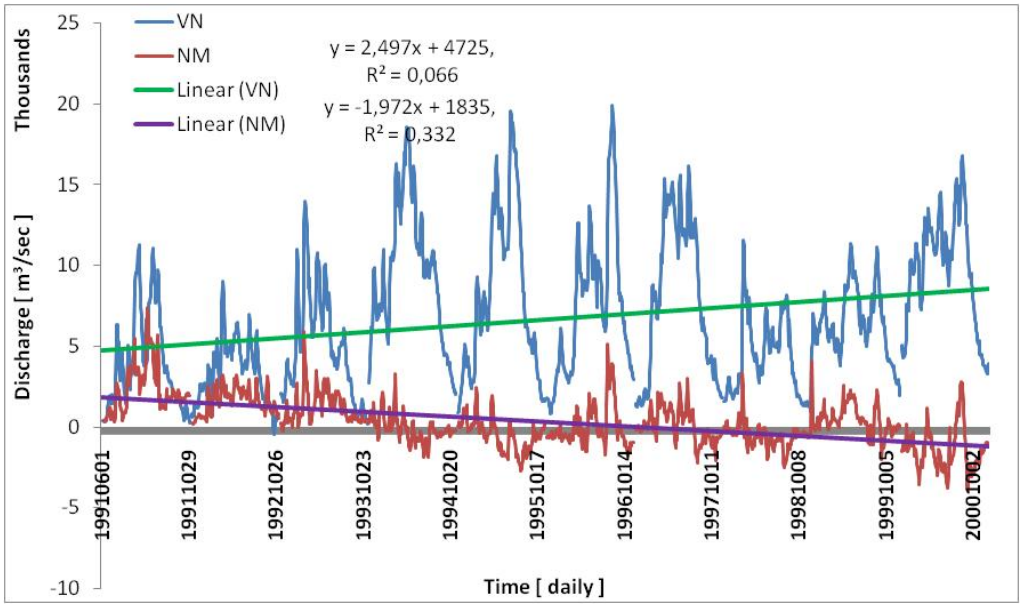
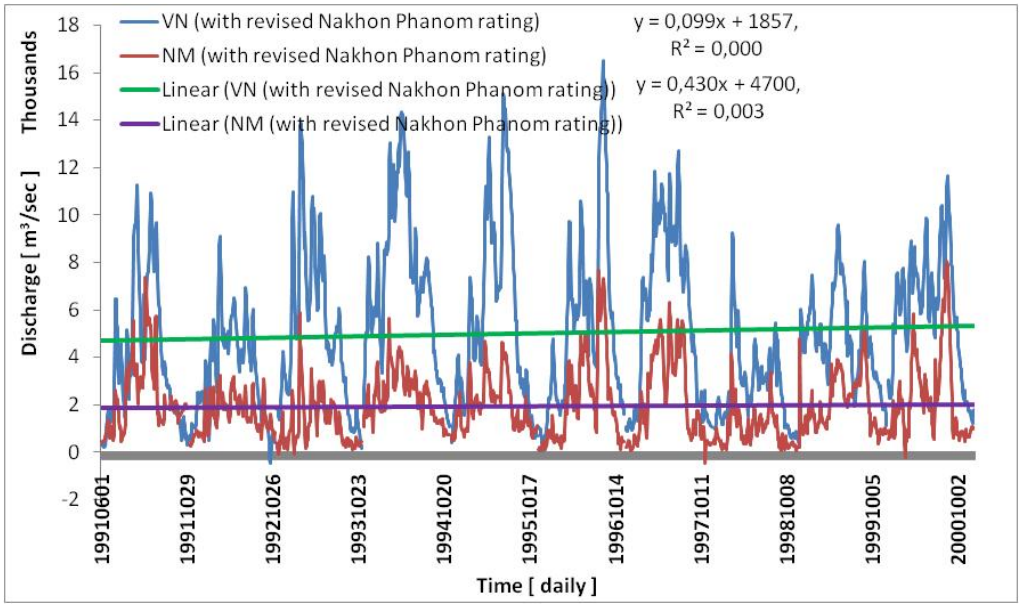


Fig. 4.4: Daily discharge plotting along middle Mekong gauges (1991 to 2000)

are shown in Fig.4.5a and 4.5b respectively. And Fig.4.5b shows that the trend in time series of VN and NM has been removed by using revised stage – discharge relationship of Nakhon Phanom.



a)



b)

Fig. 4.5: Viantiane to Nakhon Phanom and Nakhon Phanom to Mukdahan lateral discharges computation by using data of Nakhon Phanom, a) produced by MRC new rating and, b) revised rating based on year 1991-1993

The availability and quality of river cross sections is similar to those of the water stage relation curves. The RFMMC provides river cross sections in their forecast pages in the internet, but these are quite coarse and also produced with varying spacing for the width of the rivers (Fig.4.6) which gives a false impression of the river cross section. More precise cross sections were generated by using the information available at www.mrcmekong.org as given in Fig.2.12.

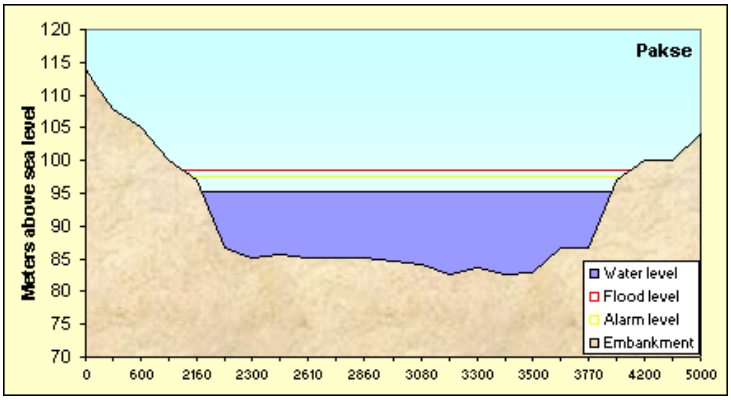


Fig. 4.6: a) River cross section at Pakse, reproduced from the internet page of MRC. Note the varying spacing of the width, water level is of 30th September 2009. b) Flood and alarm level of major Mekong gauges.

b) Gauge station	flood level (m)	alarm level (m)
Chiang Saen	11.8	11.5
Luang Prabang	18.0	17.5
Chiang Khan	17.4	17.3
Vientiane	12.5	11.5
Nong Khai	12.2	11.4
Paksane	14.5	13.5
Nakhon	12.7	12.6
Mukdahan	12.6	12.5
Pakse	12.0	11.0
Stung Treng	12.0	10.7
Kratie	23.0	22.0
Phnom Penh	11.0	9.5
Tan Chau	4.2	3.0

4.5. Conclusion

The data of seven discharge gauges, i.e. Chiang Saen, Luang Prabang, Vientiane, Nakhon Phanom, Mukdahan, Pakse, Stung Treng and 34 rainfall gauges was considered of acceptable quality for use in this study.

After analyzing the four available data sets, the revised HYMOS data set is used. These discharge data iare based on different stage – discharge relations for different years. The quality of these discharge data seems acceptable except for Nakhon Phanom, because the historical discharges (produced on the basis of variable stage – discharge relations) of Nakhon Phanom show trends from year 1993 onwards. Therefore, discharge data from year 1993 onwards are adjusted by using the stage – discharge relation of year 1991 to 1993 because there was no observable trend in the discharge data for years 1991 to 1993.

After comparing four different rating curves, i.e. SSARR, URBS, MRC Old and MRC New, for current data, the MRC new rating curves were preferred for the inter-conversion of stages to discharges and vice-versa except for Nakhon Phanom, where the revised rating curve is used (produced on the basis on 1991-93 stage – discharge relation).

5. Quality Criterion for Flood Forecasting

Prior to assess the performance quality of a forecast by any forecasting model, an objective quality criterion to detect performance quality has to be selected. The criterion or set of criteria should be sufficient to judge the model in terms of accuracy and reliability as given in section 1.3 of 3rd chapter. There are many quality criteria which are used as objective functions in hydrology. To name a few: root man square error, mean absolute error, accumulated volume error, correlation coefficient, coefficient of determination, Nash and Sutcliffe coefficient, coefficient of extrapolation, coefficient of persistence are normally used to assess the performance of hydrologic models. However, root mean square error, mean absolute error, and accumulated volume error are more or less subjective criteria because one cannot judge the model performance by these criteria, which should be quantified by some standard scale, like a Nash and Sutcliffe coefficient between 0 and 1.

In the other mentioned criteria, the squared error is normalized by some standard value. For example, in Nash and Sutcliffe (1970) coefficient, the squared error is divided by standard deviation of predicted values from the average value. This criterion is often used in hydrology to determine the quality of model performance. If adapted to the forecasting case and denoted as CNS (t) this criterion reads:

$$NS(t) = 1 - CNS\,(t) = \frac{\sum_{i=0}^{imax}[Xob(i+t)-Xf(i+t)]^2}{\sum_{i=0}^{imax}[Xob(i+t)-\bar{X}]^2} \qquad (5.1)$$

where $\bar{X}$x is the average value of the quantity x over the time from i=i_0 to i=i_{max}. It is obtained from:

$$\bar{X} = \frac{1}{n}\sum_{i=0}^{i=imax} Xob\,(i) \qquad (5.2)$$

Note that this quantity depends on the forecasting time, expressed through parameter t. The original Nash-Sutcliffe criterion was designed for the whole time series and essentially is a measure of the contribution of forecast error (the numerator of Eq.5. 1) to the variance (the denominator of Eq.5.1) of the total record. For a perfect fit, the numerator is zero, whereas for a fit, which is no better than the average value the ratio, NS(t) becomes 1 or larger. By subtracting NS(t) from 1 the direction of the criterion is reversed, and a good fit, in Nash - Sutcliffe terminology, is a value close to 1, a poor fit will lead to a criterion CNS(t) =0, whereas for QNS(t) the opposite is the case. As adapted to forecasting the Nash-Sutcliffe criterion implies that we start the forecast with no information except that obtained from ranking all measured values of forecasts. In that case, the best estimate of the future value is the mean value $\bar{X}$x. But the problem with the Nash and Sutcliffe criterion in measuring the quality of forecast lies in its comparison index denominator D, which does not take account of the difference of model

estimate $\bar{X}$x from future observed values $x_{ob}(i+t)$. This no model estimate is fairly primitive as discussed by Nash et. al. (1978) and given in Section 3.3.5. No model approximate of future outcome could be much better than $\bar{X}$x, by virtue of real time data availability. In particular, the last observed value $x_{obs}(i)$ is a much better estimate of future $x_{ob}(i+t)$ than $\bar{X}$x. Other problems of Nash and Sutcliffe coefficient are; that a constant over/ underestimation of the predicting model cannot be pointed out, the same error yields different Nash and Sutcliffe coefficient values depending on the difference of future observed value from the average value $\bar{X}$x. Furthermore, the criterion becomes meaningless when observed and predicted values are near to average $\bar{X}$x and vice versa because when the difference between $x_{obs}(i)$ and $x_f(i+t)$ is divided by difference between $x_{obs}(i)$ and $\bar{X}$x, it depends on the magnitude of difference between $x_{obs}(i)$ and $\bar{X}$x in addition to difference between $x_{obs}(i)$ and $x_f(i+t)$. And if $x_{obs}(i)$ is located far from $\bar{X}$x, then D becomes high making high NS and it is vice versa if $x_{obs}(i)$ is located near $\bar{X}$x.

The concern that the Nash and Sutcliffe (1970) criterion does not reflect the true quality of a model, has been raised by many hydrologist, i.e. Krause et al.(2005). Therefore an alternate criterion is required to better assess the performance of a forecasting model.

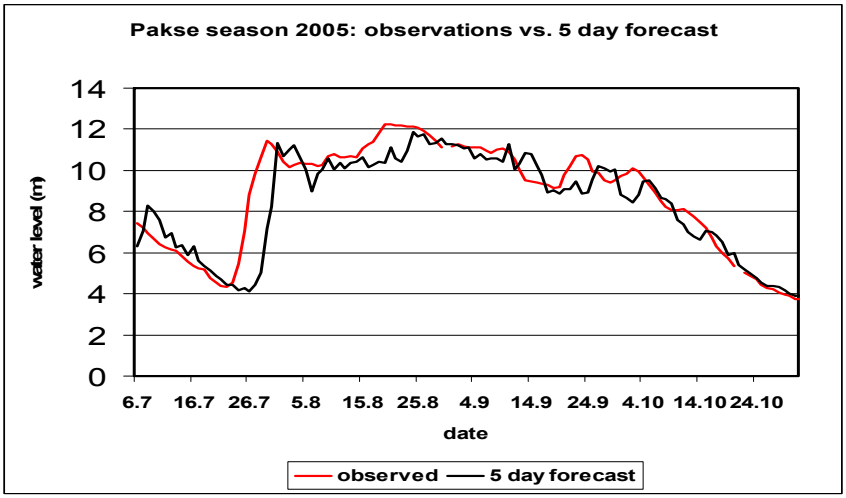


Fig. 5.3: Observed water level and 5 day forecast at Pakse gauge in 2005(Plate and Lindenmaier, 2008)

The study by (Plate and Lindenmaier, 2008) emphasized the need of better quality criterion. For illustration, they have given an example of a 5-day forecast as function of time (Fig.5.3). For this example forecasts are made by means of the model SSARR for station Pakse, for 2005. D(t) is the numerator and E(t) the denominator of CNS(t), as expressed in Equ. 5.1. The sum is extended over the whole season. From Fig.5.3 it is seen that when the season starts, i.e. when the Monsoon rains start in early summer, the initial steep rise of the water surface is not correctly predicted by

model SSARR, and yet one gets a CNS(t) value of 0.8 despite of poor performance of the model (Plate and Lindenmaier, 2008, Lindenmaier et. al., 2010).

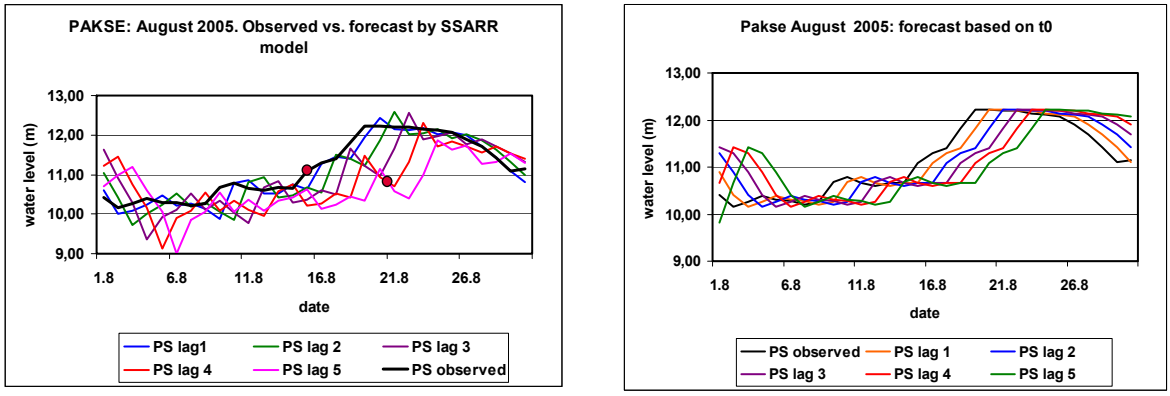
Tab. 5.1: Nash-Sutcliffe criterion applied to Pakse gauge readings for the 2005 season

	D(t)	E(t)
sum	180.5	898.9
CNS(t)	0.80	

5.2. Better quality criterion for forecasting model

A better quality criterion is obtained if one uses the fact that at time t_0 one already has some information available, at least the value of the time series to be forecasted at the present time $t = t_0$, as well as for earlier times. If there is no other information available but the present day value, then the best forecast for the near future is to assume the value at time t_0+T_F, i.e. at i + t, as being equal to the value at t_0 (Plate and Lindenmaier, 2008, Lindenmaier et. al., 2010)..

In Fig.5.4 the forecast using the SSARR model and using the no information model with forecast equal to $x_o(i)$ are shown for Pakse, using the data of August only. The 1 to 5 days forecasts are shown as family of curves. The figures must be interpreted from one point $x_o(i)$ on the observed gauge where a forecast is made for different forecasting times, represented by the other curves. For example, on August 16 we observe a value of 11.30 m, indicated by the red dot on the left in Fig.5.4a on the observed curve. A five days ahead forecast is made with the SSARR model, which yields a value of 10.58 m, shown by the right red dot on the lag 5 curve in Fig.5.4a. This value has to be compared with the actual value on the observed curve, which has a value of 12.19 m. On the right side of Fig.5.4, the same data are shown with a forecast based on present day value $x_0(i) = x_0(i+t)$ (Plate and Lindenmaier, 2008, Lindenmaier et. al., 2010).



a) b)

Fig. 5.4: Gauge Pakse, August 2005: a) forecast calculated with the SSARR model, b) taking $x_0(i)$ as forecast vs. observed data (Plate and Lindenmaier, 2008)

The comparison of Fig. 5.4a and 5.4b shows that the forecast using $x_0(i)$ is not a bad first approximation. The performance of a forecast model, expressed through a forecasted value $x_0(i) = x_f(i+t)$ after time $T_F = t$ should be judged relative to this value. This implies that the deviation of the observed value $x_{ob}(i+t)$ from the present value $x_0(i)$ should be larger than the deviation of the observed value from the forecasted value $x_f(i+t)$, if the model is to provide a better forecast than what one can estimate by just knowing the value $x_0(i)$ at the time of the forecast (Plate and Lindenmaier, 2008). This condition is expressed quantitatively by a persistence index PI, which is an index similar in structure to the NS criterion (Berthet *et al.*, 2009; Kitanidis and Bras, 1980a; Kitanidis and Bras, 1980b) and defined as:

$$PI_0(t) = 1 - \left[\frac{\sum_{i=io}^{i\max} [x_{ob}(i+t) - x_f(i+t)]^2}{\sum_{i=0}^{i\max} [x_{ob}(i+t) - x_0(i)]^2} \right] \qquad (5.3)$$

for which during calibration the sums have to be taken at each time $t = i \cdot \Delta t, \quad i = 1, 2, \ldots i_{max}$, for each t, i.e. for each $T_F = t \cdot \Delta t$, where Δt is the time increment. A small value of $PI_0(t)$ indicates poor, a large value good performance, i.e. if $PI_0(t)$ is close to 0, performance of the forecast is not better than taking the value of today (at time t) as forecast for the value at T_F.

Table 5.2: Pakse season 2005: quality criterion Equation 3 applied to SSARR forecasts (Plate and Lindenmaier, 2008)

Pakse flood season 2005: comparison of SSARR model with equation (5.3)										
	D(1)	E(1)	D(2)	E(2)	D(3)	E(3)	D(4)	E(4)	D(5)	E(5)
sum	7.60	16.1	27.8	57.1	65	113	119	178	179	243.8
quality	0.53		0.51		0.43		0.33		0.26	

An application of this criterion to the Mekong for station Pakse is shown in Table 5.2, where D(t) is the numerator and E(t) is the denominator for the t-day ahead forecast. The data are taken for the whole season from 1 July to 31 October. As presented in Table 5.2, the performance of SSARR is realistically assessed by $PI_0(t)$ by declaring it poor – as it was shown in Fig. 5.3(Plate and Lindenmaier, 2008). But this quality criterion also has some positive and negative points. The criterion loses its significance in situations where the difference between current and future observed value is small. On the other hand it has the advantage of reflecting the gain from having this particular forecasting model as compared to a no model forecast. However, PI is a relative criterion. It therefore does not reflect the magnitudes of the forecasting errors, which are described by the standard deviation as criterion, or the range of the forecasting errors. These are best described by probability distributions of errors.

5.3. Performance of SSARR Model

The quality of SSARR model performance as reflected by the quality criterion PI is shown in Fig.5.5. This Fig. presents the quality of forecasts for Pakse. PI ranges from 0.58 to 0.62 in the cases of 1 to 4 days lead time forecast.

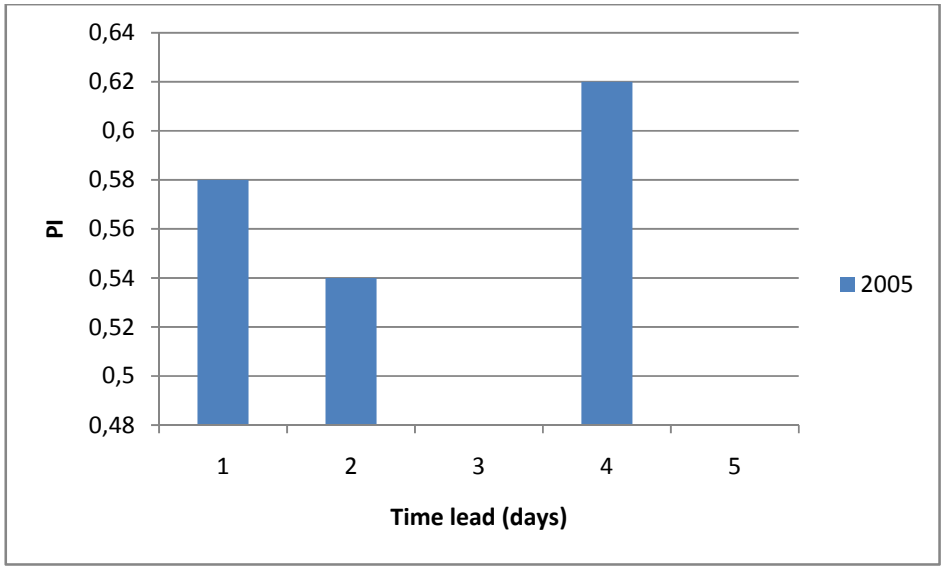
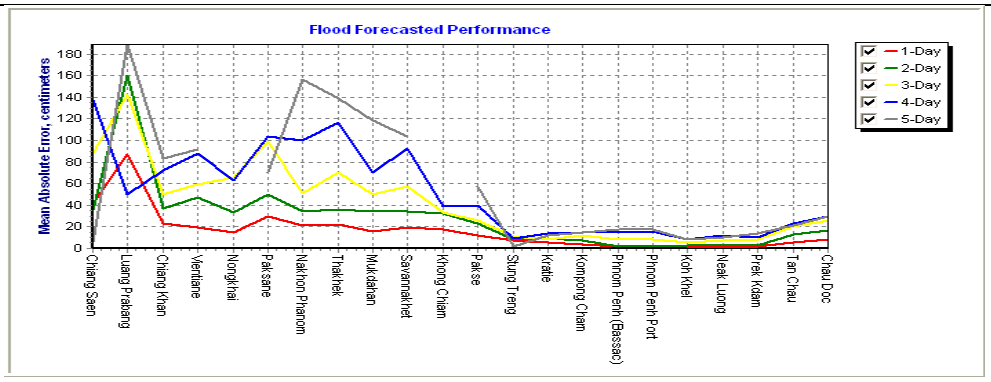


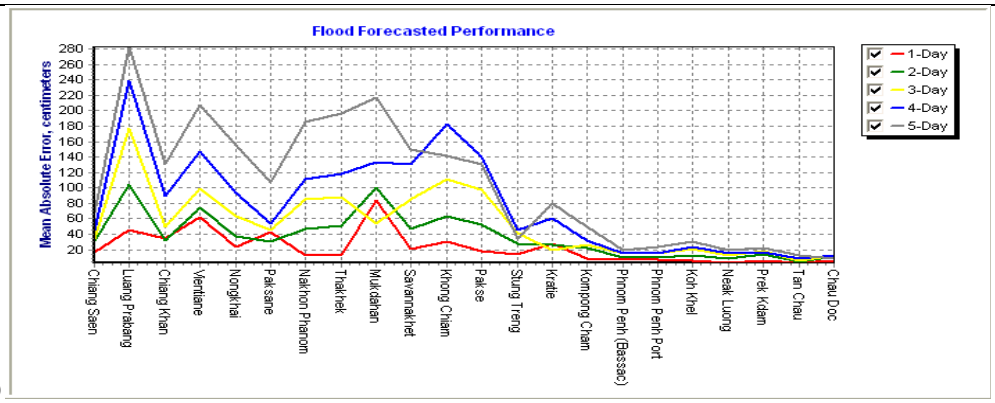
Fig. 5.5: Coefficient of efficiency distribution for the whole forecasting season at Pakse for the forecasting year 2005 (June to October).

5.4. Performance of URBS Model

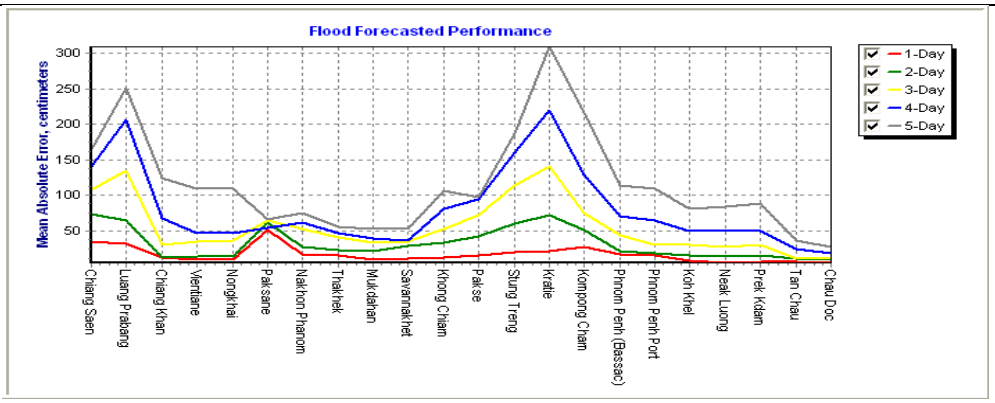
The mean absolute error of URBS model forecasts for first to 5th week of June, 2009 are given in Fig.5.6a, 5.6b, 5.6c, 5.6d, and 5.6e, respectively. Mean absolute error actually is not a good quality criterion to reflect daily forecast model performance but unfortunately, the author did not have an access to daily forecasts by



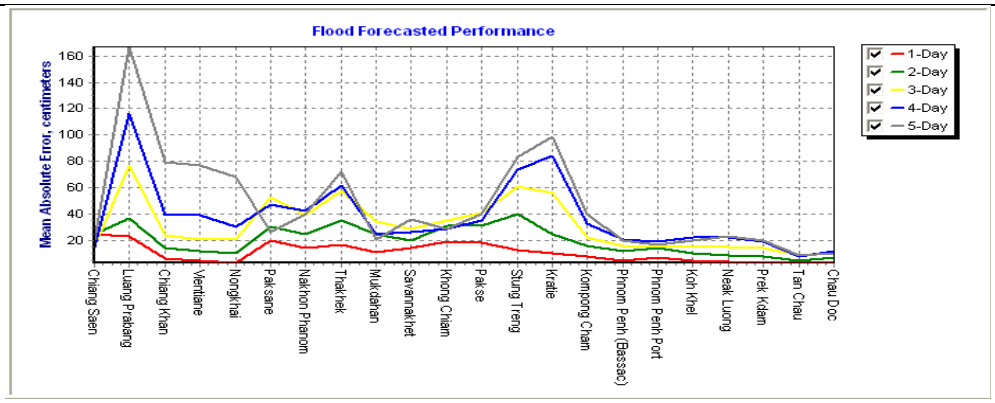
5.6a


Flood Forecasted Performance
Mean Absolute Error, centimeters
280
260
240
220
200
180
160
140
120
100
80
60
40
20
1-Day
2-Day
3-Day
4-Day
5-Day
Chiang Saen
Luang Prabang
Chiang Khan
Vientiane
Nongkhai
Paksane
Nakhon Phanom
Thakhek
Mukdahan
Savannakhet
Khong Chiam
Pakse
Stung Treng
Kratie
Kompong Cham
Phnom Penh (Bassac)
Phnom Penh Port
Koh Khel
Neak Luong
Prek Kdam
Tan Chau
Chau Doc

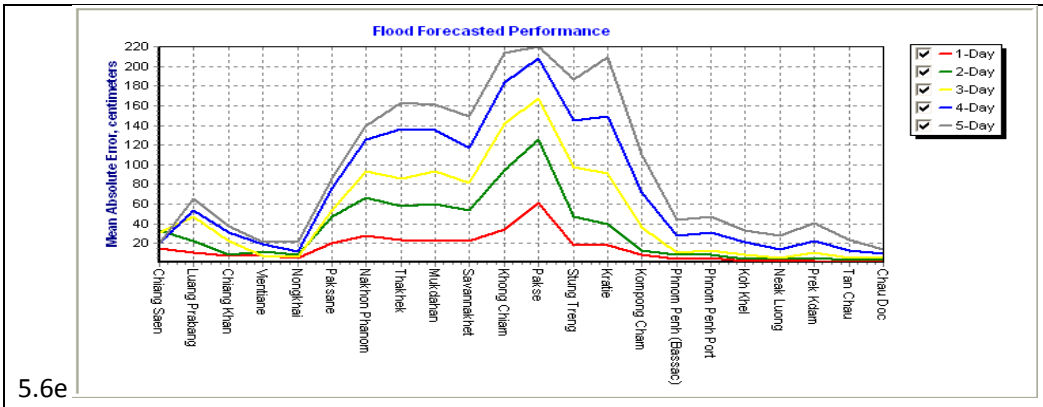
5.6b


Flood Forecasted Performance
Mean Absolute Error, centimeters
300
250
200
150
100
50
1-Day
2-Day
3-Day
4-Day
5-Day
Chiang Saen
Luang Prabang
Chiang Khan
Vientiane
Nongkhai
Paksane
Nakhon Phanom
Thakhek
Mukdahan
Savannakhet
Khong Chiam
Pakse
Stung Treng
Kratie
Kompong Cham
Phnom Penh (Bassac)
Phnom Penh Port
Koh Khel
Neak Luong
Prek Kdam
Tan Chau
Chau Doc

5.6c


Flood Forecasted Performance
Mean Absolute Error, centimeters
160
140
120
100
80
60
40
20
1-Day
2-Day
3-Day
4-Day
5-Day
Chiang Saen
Luang Prabang
Chiang Khan
Vientiane
Nongkhai
Paksane
Nakhon Phanom
Thakhek
Mukdahan
Savannakhet
Khong Chiam
Pakse
Stung Treng
Kratie
Kompong Cham
Phnom Penh (Bassac)
Phnom Penh Port
Koh Khel
Neak Luong
Prek Kdam
Tan Chau
Chau Doc

5.6d



5.6e

Fig. 5.6: Mean absolute error of URBS Model forecast for 1st to last week of July-2009 (source, www.mrcmekong.org)

URBS model. According to this Fig. the mean absolute errors ranges from 20 cm to 300 cm. The errors for the upstream gauges, i.e. C to S are higher than for the gauges downstream of S.

At present, the success of the forecasting system is comparatively modest: it is claimed that during flood seasons the one day forecasts are quite good, but that three day forecasts are not good enough. In view of the large area of many 100000 km² upstream of the Mekong delta this is surprising, although it is realized that the hydraulic conditions are quite complex, in particular due to the damping role of the large lake Tonle Sap in Cambodia. The reason for this has numerous roots: obvious are lack of data, or incomplete or even erroneous data, and a model base which either no longer is up to date or it is not appropriate for the present application.

Therefore in this study, it is tried to use comparatively better data and adopt alternative modelling approach in order to improve the flood forecast in particular at upstream gauges.

6. Discharge Data Based forecasting (Type 1 Model)

6.1. Data Based Flood Forecasting

When physically based models are used, the functional dependencies are structured from the channel and catchment physics of flood routing and rainfall – runoff models, and only the parameters of the model are determined from the data base (Shahzad et. al, 2010). For a data based model, the present to future discharge change is quantified and directly correlated with other data (known discharge and rainfall), with only conceptual models for the runoff from sub-catchments between gauges[1]. To develop appropriate data based model components for this purpose, two different approaches are tested: a regression model based on discharge data, and a rainfall – runoff model based on a unit hydrograph. A data based model which is structured from the available data base may give results with similar error bounds as physically based models. It was the objective of this study to explore the forecast capability of such data based models for the Mekong.

Forecasts for large rivers are traditionally based on fitting existing models to the data base of measured discharges and/or stages at different gauges. Consequently, the quality of the forecast first of all is a function of the model used. The earliest types of models used were based on discharge trend and regressions of discharges from past or/and upstream gauges against time delayed downstream discharges (Model Type 1). More sophisticated models use rainfall – runoff modeling and runoff routing components (Model Type 2).

The writer preferred to approach the forecasting problem not from the use of any existing model, but to develop a forecast model from the data base, where it is less important that the model is physically complete, than that the forecasts are as accurate as possible. Because a perfect forecast is not possible in view of the uncertainty of the many factors contributing to the discharge formation process, an error band must be expected. The quality of the forecast must be based on two factors: on the mean value of the forecast, which should be as bias-free as possible, and on the spread of the error band, which should be as narrow as possible, where the former is indicator of model reliability and the latter of model accuracy. In contrast to physical models based on analytical descriptions of all discharge forming processes, which have to consider uncertainties in all parameters of the individual processes, a data based model lumps all errors into one probability distribution (pdf) for the forecast, whose standard deviation is to be minimized

[1] Here the hydrological distinction is made between physical models and conceptual models. Physical models describe the physical processes of the hydrological cycle as closely as possible. In contrast, conceptual models are simple empirical models, such as the unit hydrograph model to fit the rainfall – runoff process.

regardless of physical meaning of the forecast model used. However, use of a data based model doesn't mean that system physics is ignored completely; rather the imprints of system physics are captured in the structure of data time series whose structure is a function of catchment and channel physics.

In large rivers like the Mekong, a change of discharge from present to short time in the future is mostly small in comparison to total present discharge. The discharge change is caused by rainfall storms on the sub-catchments of the river. Very large hydro-meteorological forcings are required to generate abrupt changes. Such storms do indeed occur in the flood season, and abrupt and significant discharge changes can occur. The magnitudes of these discharge changes vary with lead times and season. However, one can make an estimate of occurrence, of these discharge changes by a prior discharge and rainfall time series analysis.

In the absence of a priori information on discharge change behavior, the range of discharge change from the present to the future could be any value between 0 and infinity which leads to absolute uncertainty. However, these discharge changes are outcomes of physical processes and hence would be ideally predictable, were it not for the difficulty in quantifying the contribution of each process to the generation of future discharge changes. This is due to model, measurement, and other uncertainties as explained in chapter-3.. Irrespective of the cause, these uncertain discharge changes could be classified into two major groups, i.e. conversion of rainfall into runoff and river discharge, embodied in the hydrologic data processor, and the forecast of the rainfall field, as expressed through the meteorological (rainfall) processor (Fig.6.1a). By means of such processors, the hydrologic and meteorological uncertainty bands could be narrowed down by modeling respective historical time series information. For short lead times of large rivers, it is pragmatic to only use the discharge processor, whereas for long lead times rainfall contributions from the sub-catchments may dominate the uncertainty, and reduction of the uncertainty by means of a suitable rainfall processor is most important for a good forecast. However, temporal short and long forecast lead times are relative terms, as explained in chapter.3, therefore the effectiveness of each of the processors must be determined in each case by a statistical analysis of discharge time series. Consequently, an iterative procedure should be followed, by first analyzing the time series of the discharges only, and to develop a suitable discharge processor based on the statistical structure of the discharge time series of adjacent gauges which reduces the error band as much as possible (Model Type 0 and Model Type 1). Only if this does not prove to be acceptable, then the error band should be further narrowed by introducing rainfall information and developing a rainfall processor, (as needed for a Model Type 2).

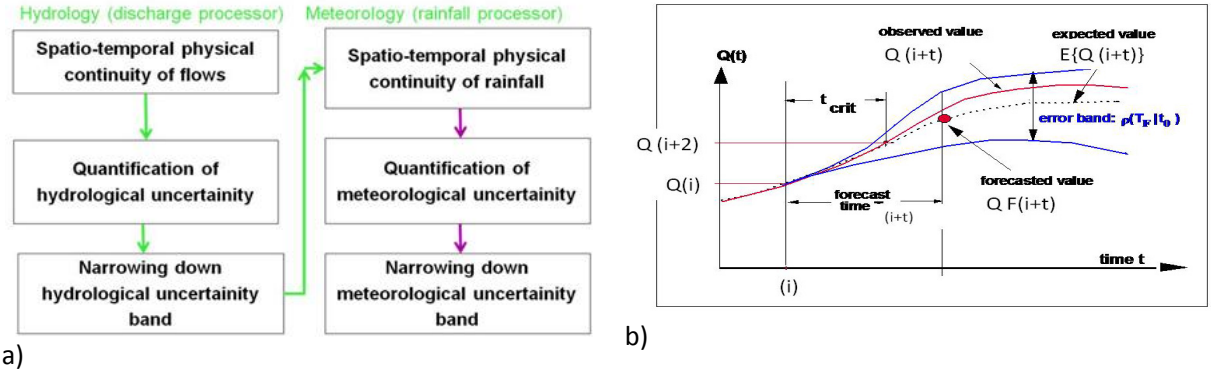


Fig. 6.1: Flood Forecasting Scheme

Basic characteristics of the process of flood forecasting in flood seasons are illustrated in Fig.6.1b. Assume that one is located at some station j, at a time identified by index, (i) and wants to do forecast for some future time, say (i+t), where t = 0, 1, 2 ...m. The expected (forecast) discharge is $Q_j(i+t)$ where j is the station index with j = 0, 1, 2 ...m. It could have any value between maximum and minimum discharge limits as represented by blue lines in Fig. 6.1. These extreme limits can be taken as limits of the uncertainty band, which is computed by a prior discharge time series analysis. The observed future discharge Q is shown by the red line and estimated/ expected values QF are given by the dotted line. Normally, the uncertainty band increases with forecast lead time, and so do the potential deviations between expected and observed values, because with increase in lead time, discharge information available at the time of the forecast loses its potential to help in estimating future discharges. For very long lead times, only the historical discharge probability distributions can be used to predict the span of possible discharges for the required lead time.

6.2. Methodology

In this chapter, as described in (Shahzad et. al, 2010), the flows of Mekong River are forecasted with 1 to 5 days lead time – as a first step, by using the information of forecasting gauge only (moving back in time). And in the second step, the information of required upstream gauges is used (moving back in time and upstream in space simultaneously). In each of these two steps Type 0 and Type 1 Models are used. Type 0 Model gives a prior span of future discharges and Type 1 Model gives a discharge forecast in the hind cast mode. Finally, the quality of these lumped models is tested against the distributed model (URBS) in producing discharge from the rainfall for flood forecasting at mainstream gauges.

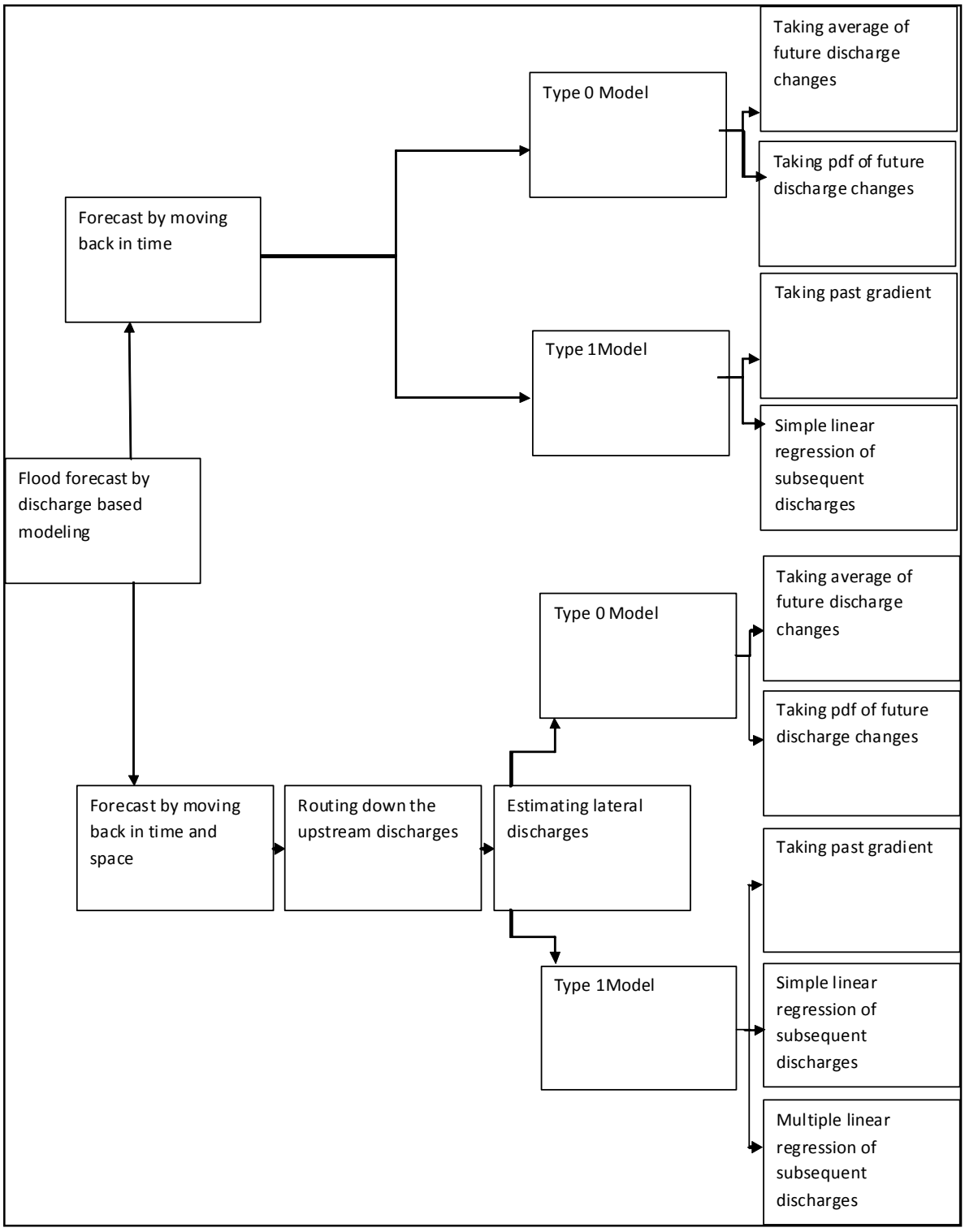


Fig. 6.2: Flood forecasting flow charts by different methods tested in Model Type 0 and Model Type 1domain

6.3. Forecast by moving back in time

In the first step, the problem of forecasting is discussed by using the discharge time series of forecasting gauge only – hence without using the discharge information of the adjacent upstream gauges. $Q_j(i+t) = Q_j(i) + \Delta Q_j(i+t)$ (6.1)

$$QF_j(i+t) = Q_j(i) + \Delta QF_j(i+t) + \varepsilon_{j0}(i+t) \quad (6.2)$$

where ε_{j0} is the forecasting error.

Eq.6.1 defines the change of discharges from time i to i+t in the analysis mode – where $\Delta Q_i(i+t)$ is known and therefore $Q_j(i+t)$ can be calculated. But in the forecasting mode, one needs to estimate $\Delta Q_i(i+t)$. Eq.6.2 presents the case of forecasting mode, where $Q_j(i+t)$ and $\Delta Q_i(i+t)$ of Eq.6.1 is replaced by $QF_j(i+t)$ and $\Delta QF_j(i+t)$ respectively – which are the quantities to be forecasted. The first logical choice to approximate the future discharge would be to assume ΔQ_i persistence, i.e., to assume that the discharge does not change from one unit step to the next, each of interval time t. Then, $\Delta Q_i(i)$ can be taken as estimate for future discharge $\Delta Q_i(i+t)$ provided that the variation of discharge over forecast time $(i+t)$ is negligible. However, even in large rivers; this assumption may lead to large errors especially in flood seasons, because discharges may change by large amounts, even for time intervals of 1 day, which is the gauging frequency of the Mekong River. Evidently the discharge has three possible future trajectories, i.e.

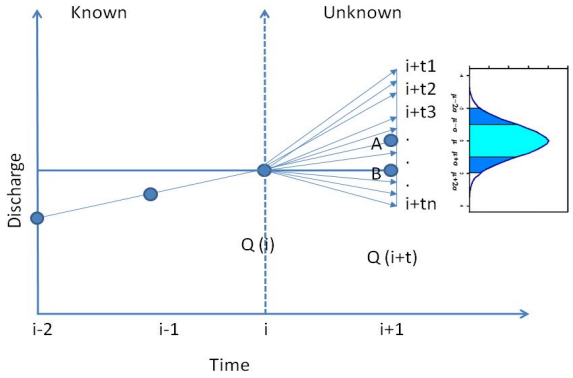


Fig. 6.3: Potential flood discharge extrapolations

it can remain constant, decrease or increase. The amount by which Q_j increases or decreases may or may not follow the previous Q_j slopes. There are countless possible trajectories, i.e. as indicated by indices i+t1, i+t2....i+tn in Fig.6.3. Consequently, the change of discharge Q_j (i) over time steps (i+t) must be improved for forecasting future discharges.

6.3.1. Type 0 Model (by moving back in time)

Prior to actual flood forecasting the hydrologic uncertainty, i.e. ΔQ_j in unit time $(i+t)$, where, t varies from one to five days, can be quantified empirically by fitting a suitable probability distribution to $\Delta Q_j(i+t)$. The quantification can consist of giving historical mean, maxima and minima, or of determining its probability distribution. The dependency of the $\Delta Q_j(i+t)$ distribution could be further refined by analyzing each month's $\Delta Q_j(i+t)$ time series separately. Model development steps for Type 0 and 1 models are shown in Fig.6.2. The first approximation of $\Delta Q_j(i+t)$ can be the average discharge change $\overline{\Delta Q_j\,(\imath,\imath+t)}$ during the unit time step from time i to $(i+t)$. The $\overline{\Delta Q_j\,(\imath,\imath+t)}$ can be taken as a function of season. This requires determination of the average discharge change $\overline{\Delta Q_j}$ for each season. If one assign an index s to each month of the flood season, i.e. s = 1, 2 ...12, then Eq.6.2 and 6.3will change to:

$$\overline{\Delta Q_{j\,s}(\imath,\imath+t)} = \frac{1}{ny}\left[\sum_1^y \sum_1^n \Delta Q_{j\,s}(i,i+t)\right] \qquad (6.3)$$

$$QF_{js}(i+t) = Q_{js}(i) + \overline{\Delta Q_{j\,s}(\imath,\imath+t)} + \varepsilon_{j0}(i+t) \qquad (6.4)$$

In Eq.6.3 n is the number of discharge changes inunit time Δt for each season s and y is the number of years.

One can argue that the main problem in forecast by Eq.6.4 is the possibility of the cancellation of negative and positive $\Delta Q_{j\,s}(i,i+t)$. However, it is not true because the frequency of positive and negative $\Delta Q_{j\,s}(i,i+t)$ of large rivers like Mekong depends on season. This is due to strong seasonality of flood occurrence, as in Mekong region, where the onset, rise and recession timing of floods are repeated over same the time with +/- of few 10-dailies. Consequently, in the rising part of the hydrograph, the change in discharge is dominated by big positive discharge differences and vice versa in a flood recession.

However, the positive $\overline{+\Delta Q_{j\,s}(\imath,\imath+t)}$ and negative $\overline{-\Delta Q_{j\,s}(\imath,\imath+t)}$discharge changes can also be lumped into two separate groups in each season to estimate the range of future discharges.

$$\overline{+\Delta Q_{j\,s}(\iota, \iota + t)} = \frac{1}{my}\left[\sum_1^y \sum_1^m \Delta Q_{j\,s}(i, i + t)\right] \quad (6.5)$$

m is number of positive discharge changes per season over unit time Δt, i.e.,

$\Delta Q_{j\,s}(i + t) > \Delta Q_{j\,s}(i)$

$$\overline{-\Delta Q_{j\,s}(\iota, \iota + t)} = \frac{1}{py}\left[\sum_1^y \sum_1^{np} \Delta Q_{j\,s}(i, i + t)\right] \quad (6.6)$$

p is number of positive discharge changes per season over unit time Δt, i.e.,

$\Delta Q_{j\,s}(i + t) < \Delta Q_{j\,s}(i)$

With the help of this approach, the range of future discharges can be approximated by adding pre-estimated positive and negative $\overline{\Delta Q_{j\,s}(\iota, \iota + t)}$to the current discharge $\Delta Q_{j\,s}(i)$ by using Eq.6.7 and 6.8.

$$QF_{js}(i + t) = Q_{js}(i) + (+\overline{\Delta Q_{j\,s}(\iota, \iota + t)}) + \varepsilon_{j0}(i + t) \quad (6.7)$$

$$QF_{js}(i + t) = Q_{js}(i) + (-\overline{\Delta Q_{j\,s}(\iota, \iota + t)}) + \varepsilon_{j0}(i + t) \quad (6.8)$$

This results into two possible future maximum and minimum approximations, which describes the possible range of future discharges. This range can be successfully used as future forecast, provided it is narrow, and the magnitude of deviations of positive and negative $\Delta Q_{j\,s}(i, i + t)$ time series is small. However, large deviations result in large error terms $\varepsilon_{j0}(i + t)$ because the spread of $\Delta Q_{j\,s}(i, i + t)$ around the mean discharge changes is wide. The coefficient of variation Cv, i.e. ratio of standard deviation to averages gives the approximation of possible discharge changes with respect to averages. Therefore, Eq. 6.7 and 6.8 can be successfully used in forecasting, when the magnitude of span, standard deviation and coefficient of variation Cv of positive and negative $\Delta Q_{j\,s}(i, i + t)$ is small.

The approach using average discharge changes assumes that in each season, increases or decreases in discharge over certain time period remains close to constant, i.e. equal to the average increase or decrease, which is less likely to be correct when there are large local rainfall forcings. However, one can estimate the future discharge span in between 5 and 95% occurrence probability by replacing the $+\overline{\Delta Q_{j\,s}(\iota, \iota + t)}$ and $-\overline{\Delta Q_{j\,s}(\iota, \iota + t)}$ in Eq.6.7-8 by $\mathrm{P}(\Delta \mathrm{Q}_{js5})(i, i + t)$ and $\mathrm{P}(\Delta \mathrm{Q}_{js95})(i, i + t)$ as given in Eq.6.9-10. Although the flood forecast span by Eqs.6.9

and 6.10 will be wider than by using Eqs.6.7 and 68 but with the high probability of future discharges within the estimated span.

$$QF_{js}(i+t) = Q_{js}(i) + \mathrm{P}\left(\Delta \mathrm{Q}_{js5}\right)(i, i+t) + \varepsilon_{j0}(i+t) \quad (6.9)$$

$$QF_{js}(i+t) = Q_{js}(i) + \mathrm{P}\left(\Delta \mathrm{Q}_{js95}\right)(i, i+t) + \varepsilon_{j0}(i+t) \quad (6.10)$$

$\mathrm{P}(\Delta \mathrm{Q}_5)$ =5% and $\mathrm{P}(\Delta \mathrm{Q}_{95})$ =95%

6.3.2. Type 1 Model (by moving back in time)

The hydrologic uncertainty band of the forecast can be narrowed down by identifying functional dependencies of future discharges $Q_j(i+t)$ on current discharges $Q_j(i)$, or of future $\Delta Q_j(i+t)$ on current $\Delta Q_j(i)$, because discharge $Q_j$and discharge change ΔQ_j must be continuous. Therefore, instead of using $\overline{\Delta Q_{J\,s}(\imath, \imath+t)}$ one can use the previous discharge gradient as approximation for the future discharge gradient as given by Eq. 6.11.

$$QF_j(i+t) = Q_j(i) + \Delta Q_j(i-t) + \varepsilon_{j0}(i+t) \quad (6.11)$$

A further reduction in error is obtained by using some weighted relation between $\Delta Q_j(i-t, i)$ and $\Delta Q_j(i)$ instead of equality, i.e. Eq. 6.11could be rewritten as regression equations as given in Eq.6.12 & 6.13;

$$QF_j(i+t) = Q_j(i) + b_j(i+t)[\Delta Q_j(i-t) + C_j\,(i+t)] + \varepsilon_{j0}(i+t) \quad (6.12)$$

where b is the regression coefficient, t is the lead time, and $C_j\,(i+t)$ is the regression constant or intercept.

Yet another approximation is obtained by going only one step into the past, i.e., to i-1 to forecast $Q_j\,(i+t)$ with t = 1,2..m. The advantage of this approach as given in Eq.6.13 is that it uses only recent discharge information instead of going t times into the past to forecast t time intervals into the future, analogous to the cases in Eq.6.11-12;

$$QF_j(i+t) = Q_j(i) + b_j(i+t)[\Delta Q_j(i, i-1) + C_j\,(i+t)] + \varepsilon_{j0}(i+t) \quad (6.13)$$

Another variation of Eq.6.13 could be multiplying $\Delta Q_j(i, i-1)$ with t as given in Eq.6.14. For this approach it is assumed that all future time steps $(i, i+1)$; of discharge change are the same as the one for times i-1 and i, in which case one gets:.

$$QF_j(i+t) = Q_j(i) + b_j(i+t).t.[\Delta Q_j(i, i-1) + C_j(i+t)] + \varepsilon_{j0}(i+t) \quad (6.14)$$

Yet another approach is to use simple linear auto regression between $Q_j(i)$ and $Q_j(i-t)$ to forecast $Q_j(i+t)$ in forecast mode, where the coefficients of the regression equation are determined in the analysis mode, using historical data:

$$QF_j(i+t) = b_j(i+t)Q_j(i) + C_j(i+t) + \varepsilon_{j0}(i+t) \quad (6.15)$$

Eq.6.11-15 can be used to produce forecasts, but the main shortcoming of these approaches is that the forecast depends entirely on the last observed discharge changes from time i to i-t or from i to i-1. Consequently, the quality of such a forecast is affected by short time discharge variations caused by local rainfall storm forcings.

6.4. Forecast by moving back in time and space

So far, the discharge change with time, of each discharge gauge, j has been discussed by only using the time series of station j, i.e. not including the data of upstream discharge gauges. However, the structure of the forecast model on a large river is determined by the combination of flows in the main river channel, which are measured at stream gauges, and lateral inflows from the sub-catchments between the gauges. The latter have to be inferred from local information: tributary discharges and runoff from the sub-basin. Consequently, a model based on this situation has to have the following components: an initial input of discharges from upstream stations, and an estimate for the lateral inflows between stations. This is shown in Fig.6.4, where the geometric notation used in this approach is presented. The point at which a forecast is to be made has index 0, and stations upstream are identified by index j, where j goes from 0 to m. Consequently, the station directly upstream of station 0 is station 1. The stations are supposed to be located in flow time 1 day apart, i.e. Δt = 1 day: it takes one day for the discharge at point 1 to reach point 0. In particular, the station with index m is located exactly m time intervals upstream

of station zero, where m is the number of intervals in the forecast time. If the flow time between adjacent gauges is larger than one day, then a virtual station has to be introduced, so that the number of gauges for each forecast time is exactly the same as the number of time steps for the forecast.

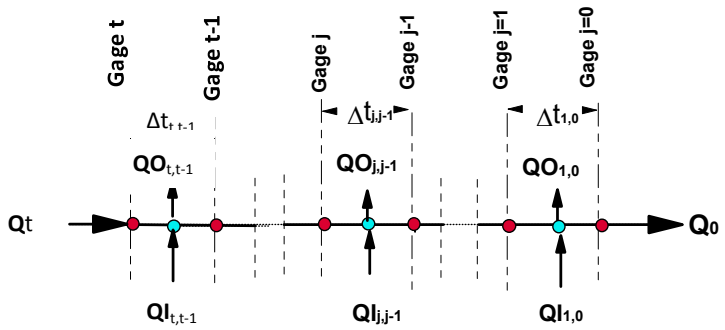


Fig. 6.4: Notation definitions.

Let the index i denote the time, as referred to station 0. That is, the real time at the time of forecast has the index i, which implies that at time t = i days a forecast $QF_0(i+t)$ is to be made for t days ahead. In this situation it is evident that the discharges at station j= t and all stations further upstream of station t are know exactly (within the error of measurement).

To describe the forecast situation, Fig.6.4 is redrawn for the forecast mode as in Fig.6.5.In this figure, the quantity:

$$DQ_j(i+1) = QI_{j,j-1}(i, i+1) - QO_{j,j-1}(i, i+1) \qquad (6.16)$$

is the net lateral inflow between stations j and j-1 in any time interval i,i+1. It is generated by rainfall on the basin, and inflows from tributaries minus such lateral outflows QO as may occur. With this notation, the continuity equation applied to the stretch between gauge 0 (the gauge for which discharges are to be forecast) and station t upstream yields:

$$Q_0(i+t) = Q_t(i) + \sum_{j=1}^{t-1} DQ_j(i+t-j) \qquad (6.17)$$

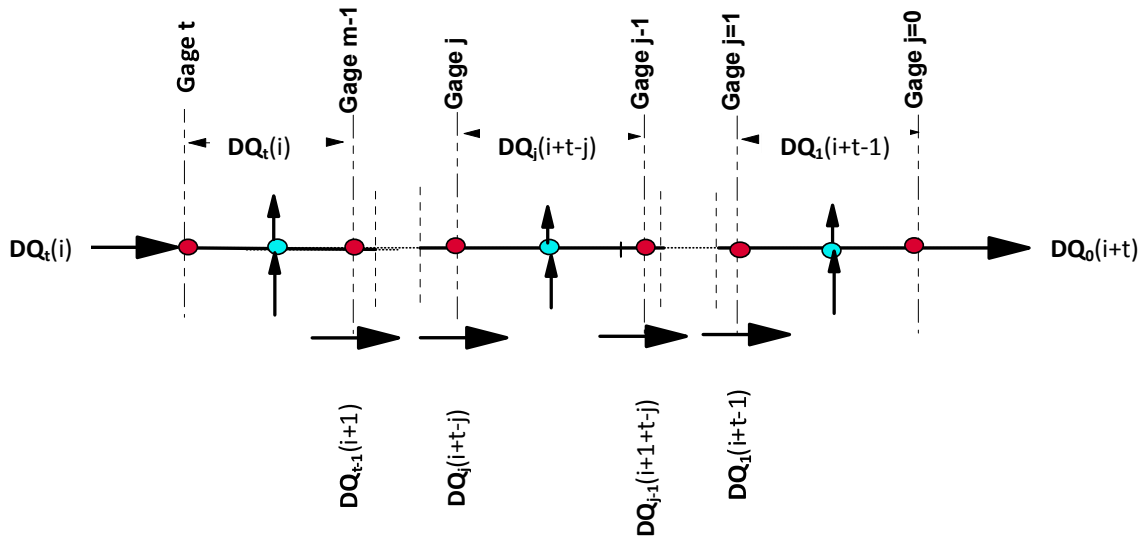


Fig. 6.5:. redrawn to reflect forecast conditions

This equation expresses the condition of continuity and is generally valid, independently of whether forecasts are considered or not. Q_t is the discharge at point t, and ε_0 is the error due to measurement uncertainties. For applications, Eq. 6.17 has to be considered both in the analysis mode and in the forecast mode. In the analysis mode, the equation is used to set up the forecast model, and in the forecast mode it is used to make actual forecasts.

a. <u>The analysis mode.</u>

The quantities in Eq. 6.17 cannot be measured directly. In the analysis mode they are determined by taking the difference:

$$DQ_j(i+t-j) = Q_j(i+t-j) - Q_{j-L}(i+t-j-L) \quad (6.18)$$

which yields a family of time functions $DQ_j(i + t - j)$, where j = 1, 2…t, for each station j, and L is the discharge travel time from station Q_{j-L} to subsequent downstream station Q_j. As a first step in the data analysis, this series of time functions has to be checked for consistency, by observing differences between three adjacent stations – consistency implies that when there is no overflow of the river banks then there cannot be a lower flow at the downstream gauge than at the upstream one. Other features, such as finding errors yielding strange outliers and other irregularities can be spotted and corrected in the parallel inspectional analysis of these time functions. This inspectional analysis is a time consuming but important and integral part of the data analysis, in particular in the Mekong area, where as mentioned before war and civil unrest has substantially interfered with correct data collection and determination of rating curves.

b. <u>The forecast mode</u>

In the forecast mode Eq.17 is written:

$$Q_jF_0(i + t) = Q_t(i) + \sum_{j=0}^{t-1} DQ_{j+1}F_0(i + t - j) + \varepsilon_{F0}\,(i + t) \qquad (6.19)$$

where the index F denotes forecasts, whereas quantities without this index - i.e. $Q_t(i)$ - are real time observed values. The error terms ε_{F0} are stochastically independent random variables. For example, for t = 1 and 2 one obtains:

$$Q_1F_0(i + 1) = Q_1(i) + DQ_1F_0(i + 1) + \varepsilon_{F0}(i+1)$$

$$Q_1F_0(i + 2) = Q_2(i) + DQ_1F_0(i + 1) + DQ_2F_0(i + 2) + \varepsilon_{F0}(i+2) \qquad \text{etc.}$$

It is worth noting that the use of spatio-temporal discharge information from upstream and lateral gauges in forecasting is more useful than the use of temporal discharge change ΔQ , which is obtained the help of only oneforecasting gauge j. The reason of this improvement is the separation of known quantities from unknown quantities in a better way. For example in Eq.6.2, ΔQ_j (i+t) is because of the lateral inflow DQ_j between gauge 0 & 1, 1 & 2…t-1 & t and t & m. Therefore, by using the discharge of upstream gauge, i.e. $Q_t(i)$ one can avoid forecasting the portion of inflows coming in between gauge t & m. Further, the lateral inflows between gauge 0

& 1, 1 & 2…t-1 & t are forecasted for different lead times, i.e. i+t, i+t-1, ….i+1 respectively as given in Eq.6.19. By using this approach, one can avoid forecasting the whole ΔQ_j (i+t) rather only that portion of this quantity is forecasted which is really unknown and hence avoiding the forecast of lateral inflow quantities which are already known.

The terms DQF in Eq.6.19 have to be determined ahead of time. It is assumed that $DQF_j(i)$ =function of known quantities + error term. This is written for the terms of Eq. 6.20:

$$DQ_jF(i+t-j) = DQ1_j(i+t-j) + DQ2_j(i+t-j) \quad (6.20)$$

where DQ1 is a deterministic part, exactly known at time i because of already occurred meteorological events and hydrologic processors, and DQ2 is the unknown part because of future rainfall occurrences that has to be forecast. Note that forType-0 and Type-1 Model, DQ_jF is not split into $DQ1_j$ and $DQ2_j$.

6.4.1. Type 0 Model (by moving back in time and space)

If one replaces the time series of Q_j of Eq. 6.4 to 6.10 with DQ_j,,the time series of lateral discharges or their potential spans could be approximated with the help of average discharge difference $\overline{\Delta DQ_{js}(\iota, \iota+t)}$) or by means of probability distribution. The methods used for Q can be applied to DQ, and can be expressed through Eq.6.21 to 6.24 by replacing $Q_j(i)$ with $DQ_j(i)$;

$$DQ_jF(i+t) = DQ_j(i) + (+\overline{\Delta DQ_j(\iota, \iota+t)}) + \varepsilon_{j0}(i+t) \quad (6.21)$$

$$DQ_jF(i+t) = DQ_{js}(i) + (-\overline{\Delta DQ_{j\,s}(\iota, \iota+t)}) + \varepsilon_{js0}(i+t) \quad (6.22)$$

$$DQ_jF(i+t) = DQ_{js}(i) + P(\Delta DQ_{js5})(i, i+t) + \varepsilon_{j0}(i+t) \quad (6.23)$$

$$DQ_jF(i+t) = DQ_{js}(i) + \mathrm{P}\left(\Delta \mathrm{DQ}_{js95}\right)(i, i+t) + \varepsilon_{j0}(i+t) \qquad (6.24)$$

Once the span of $DQ_jF(i+t)$ is estimated with the help of Eq.6.21 to 6.24, the $Q_jF(i+t)$ span could be determined with the help of Eq.6.19.

6.4.2. Type 1 Model (by moving back in time and space)

If one replaces the time series of Q_j of Eq. 6.11 to 6.15 with DQ_j, the time series of lateral discharges could be forecasted with the help of extrapolation or by simple linear Regression. The methods used for Q can be applied to DQ, and can be expressed through Eq.6.25 to 6.29 by replacing $Q_j(i)$ with $DQ_j(i)$.

$$DQ_jF(i+t) = DQ_j(i) + \Delta Q_i(i-t, i) + \varepsilon_{j0}(i+t) \qquad (6.25)$$

$$DQ_jF(i+t) = DQ_j(i) + b_j(i+t)[\Delta DQ_i(i, i-t)] + C_j(i+t) + \varepsilon_{j0}(i+t) \qquad (6.26)$$

$$DQ_jF(i+t) = DQ_j(i) + b_j(i+t)[\Delta DQ_i(i, i-1)] + C_j(i+t) + \varepsilon_{j0}(i+t) \qquad (6.27)$$

$$DQ_jF(i+t) = DQ_j(i) + t[b_j(i+t).\Delta DQ_i(i, i-t) + C_j(i+t)] + \varepsilon_{j0}(i+t) \qquad (6.28)$$

$$DQ_jF(i+t) = b_j(i+t)[\Delta DQ_i(i)] + C_j(i+t) + \varepsilon_{j0}(i+t) \qquad (6.29)$$

These forecasted $DQ_jF(i+t)$ could be used in Eq.6.19 in order to forecast $Q_jF(i+t)$.

6.4.2.1. Regression analysis mode

One can also estimate $DQ_jF(i+t)$ with the help of multi-linear regression. In the multi regression model, the first term of Eq.6.20 is zero, and the second one becomes:

$$DQ_jF(i+t) = \alpha_j(i+t).Q_k(i) + \sum_j^{j=k}(i+t)\cdot DQ_j(i) + \sum_j^{j=k}\cdot C_j(i+t) + \varepsilon_j(i+t) \quad (6.30)$$

where k is the upstream most available gauge. In the analysis mode, the empirical coefficients α, C and γ are found by least squares optimization, i.e. minimizing the standard deviation of ε_j. These coefficients then are used as known inputs into the forecasting model.

Eq.6.30 is based on the assumption that linear correlations exist, valid for all i, of $DQ_j(i)$ with known parts of the time series of upstream gauges. Inspection of an x-y plot of $DQ_j(i)$ to $DQ_j(i+1)$ for data of the Mekong River suggests the use of $DQ_j(i)$ as approximations for $DQF_j(i+1)$ for each station j. However, for longer forecasting times the nearest station alone is not sufficient, one wants to include also effects of stations which are further upstream. Consequently, the terms $DQF_j(i+t-j)$ could be regressed against all differences DQ_j upstream between station 0 and station k by means of a linear multi-regression analysis, yielding:

$$DQ_jF(i+t) = \alpha_j(i+t).Q_k(i) + \sum_j^{j=k}\gamma_j(i+t)\cdot DQ_j(i) + \sum_j^{j=k}\cdot C_j(i+t) + \varepsilon_j(i+t) \quad (6.31)$$

For example, the forecast of $DQF_7(i+5)$, i.e. 5 day lead time forecast for lateral discharge station j index 7, could be produced by using Eq. 6.32

$$DQ_7F(i+5) = \alpha_7(i+5).Q_9(i) + \gamma_7(i+5)\cdot DQ_7(i) + \gamma_9(i+5)\cdot DQ_9(i) + C_7(i+5) + C_9(i+5) + \varepsilon_0(i+5) \quad (6.32)$$

The difference between the forecasted value and the actual value is the error term given as $\varepsilon_{f0}(i+t)$ in Eq. 6.32.

When it is found that the error term time series is not entirely random, but a negative error is frequently followed by a negative value and vice versa, a positive value is followed by a positive value, then tthese errors $\varepsilon_{f0}(i+t)$ can be further reduced by regressing the future error term against current errors. In this case further improvements could be obtained by means of a lag one Marcov chain, i.e. as a first approximation to the time series structure of the error term one can write the lag-one auto-regression. However, advantages can only be gained from this approach, if the lead time of the error term is more than one day, and when significant correlations exist over the forecast lead time.

$$\varepsilon_{fj0}(i+t) = \beta j(i+t).\varepsilon_{f0}(i+t-t) + \varphi j(i+t) + \varepsilon_{f1}(i+t) \quad (6.33)$$

Since $\varepsilon_{f0}(i+t)$ is the estimate of the forecasting error term and determined in the analysis mode, it will be renamed as $\varepsilon_{fe}(i+t)$. Introducing these pre-estimated errors into Eq.6.19 yields:

$$\mathrm{Q_jF_0(i+t)} = \mathrm{Q_t(i)} + \sum_{\mathrm{j=0}}^{\mathrm{t-1}} \mathrm{DQ_jF(i+t-j)} + \varepsilon_{\mathrm{fje}}\,\mathrm{(i+t)} + \varepsilon_{fj1}(i+t) \quad (6.34)$$

The error term $\varepsilon_{fj1}(i+t)$ is analyzed further by finding its empirical probability distribution. In this way one can, through a frequency analysis of the error terms, obtain bands about the forecasted value, which are specified by the probability distribution P(ε). For example, the Probability of P(ε_{80}) =80% means that 80% of all values are lower than the forecast plus ε_{80}.

6.5. Data Based Modeling Application to the Mekong River (by moving back in time)

Daily discharges of 7 mainstream gauges of middle Mekong from Chiang Saen to Stung Treng for the period of 1960 to 2004 are used to quantify uncertainty bands, and to develop and apply

flood forecasting methods as described by Eq.6.1 to 6.15 in order to produce flood forecast at 6 gauges, i.e. Luang Prabang, Vientiane, Nakhon Phanom, Mukdahan, Pakse, and Stung Treng, with 1 to 5 day lead times. For the analysis, the data time series are divided into two parts, one for the time 1960 to 1990, and the other for the time from 1991 to 2004, for developing and testing of forecasting models respectively (using the "split sampling" technique). Correlation coefficients and Nash and Sutcliffe coefficients are used as preliminary criterion to assess each model performance. The quality is then judged by means of the Persistence Index (PI) in the final assessment of the forecasting model. However, a selection was made in as much as the approaches which were giving very low Nash and Sutcliffe coefficient values in the preliminary stages of model development were discarded and hence not verified by PI.

When models are applied to the data in the analysis mode, errors remain, due to the uncertainty associated with model, data and measurement uncertainties. The statistics of these error terms determine the limits of forecast accuracy of possible discharge changes. Application to actual forecasts introduces additional forecast errors. The quality of the forecast can be assessed by comparison of the historical error statistics – called here the hydrological uncertainty, with the forecast error statistics – called here the forecast uncertainty.

In this step, Qj(i,i+t) are forecasted by moving back in time by using only the discharge time series of Qj, hence not utilizing the information from upstream gauges or upstream reaches. The performance of Type 0 Model and Type 1 Model as described in Eq. 1 to 15 and in Eq. 21 to 27 respectively is tested to do flood forecast in the middle reach of Mekong between Chiang Saen and Stung Treng at C,L,V,N,M,P and S for 1 to 5 day lead times.

6.5.1. Application of Type-0 Model

At first the limits of the ΔQ_j are quantified by analyzing the discharge statistics of the Mekong. The spatio-temporal patterns of daily discharges at each forecasting gauge starting from Chiang Saen to Stung Treng are presented in Fig.6.6. The forecasting stations are identified by the initial letter of their name, i.e. Chiang Saen = C, Luang Prabang = L, Vientiane = V, Nakhon Phanom = N, Mukdahan = M, Pakse = P and Stung Treng = S. The whole middle reach of the Mekong between C and S is divided into 6 reaches, where each reach between adjacent gauging stations is represented by the initials of the stations, such as CL, LV, VN, NM, MP and PS reach. Similarly the lateral contribution in between two forecasting stations will also be represented by the initials of terminal stations, for example, the lateral inflow between Chiang Saen to Luang Prabang will be indicated by CL and so on.

Fig.6.6 shows the maximum and minimum recorded discharges at gauging stations along the middle reach of the Mekong in the period of 1960 to 1990. A plot of the maximum discharges of

the main river gauges as function of distance indicates that maximum recorded flow increases from upstream to downstream. The maximum discharge at upstream gauge C is about 23,500 m^3/sec in comparison to 68,880 m^3/sec at gauge S, which is the last downstream gauge in this analysis. Furthermore, maximum and minimum changes in main stream discharge $\Delta Q_j(i+t)$ and lateral discharge $\Delta DQ_j(i+t)$ are also determined, with t = 1,2…5, however only the results of t=1 are presented in Fig.6.6. The maxima ΔQ_j as shown by green bars illustrate that they also increase from upstream to downstream. It means that the potential of maximum discharge change in unit time is higher downstream than upstream.

Similarly, the statistics of the extremes, i.e., maxima and minima, of lateral inflows DQ_j and ΔDQ_j indicate that normally the maximum ΔQ-values in 1 day unit time are smaller than maximum temporal changes of the ΔQ-values of the main gauges. This information can be used to approximate the forecasted uncertainty bands without using real time information, by simply adding the span of maximum and minimum lateral ΔDQ to the upstream gauge discharge. The analysis of the historical ΔQ – time series has given the approximate range of discharge changes on either side in unit time, for example in the case of C it is +/- 4000 m^3/sec. Along with maximum and minimum ΔQ-values, one can also provide the probability distribution of the changes ΔQ per unit time.

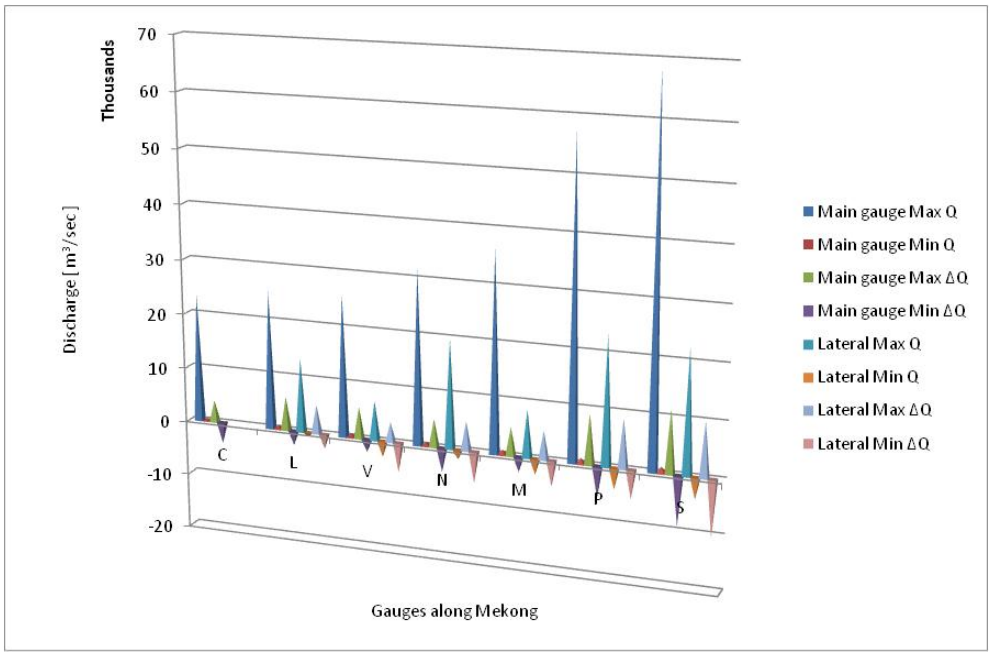


Fig. 6.6: Terminal discharge statistics

To start with, the maximum and minimum statistics of $\Delta Q_j(i+t)$ is determined as given in Fig.6.7, in order to fix lower and upper limits of the error band as described in Section-6.1. The ΔQ-values increase along the river - from 2500 to 10,000 m^3/sec in 1 to 5 days at Chiang Saen, and to 12000 to 32000 m^3/sec at Stung Treng.

But the maximum and minimum statistics of $\Delta Q_j(i+t)$ belong to extreme events, and most of the time $\Delta Q_j(i+t)$ remains smaller than these extremes.

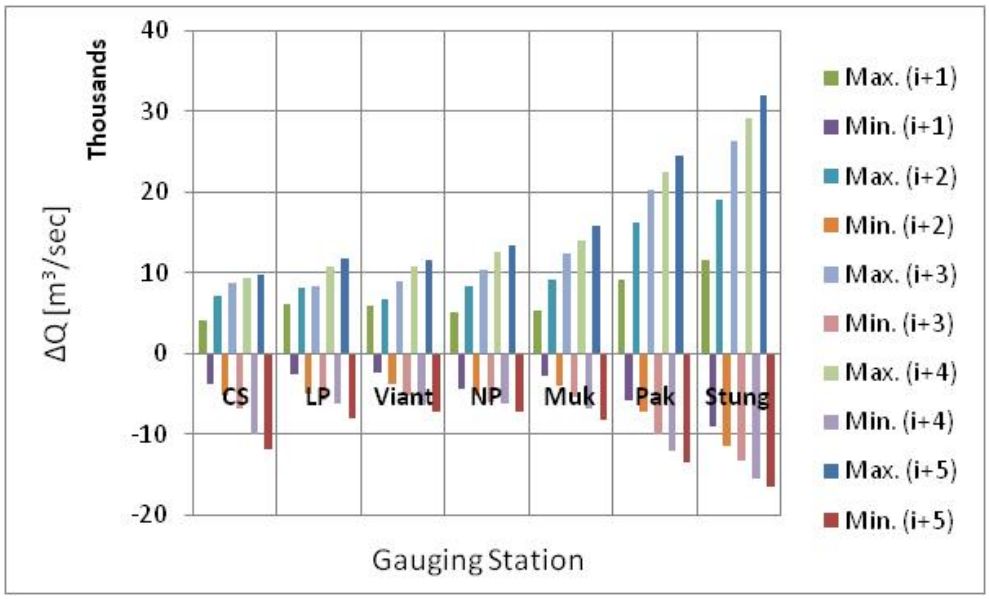


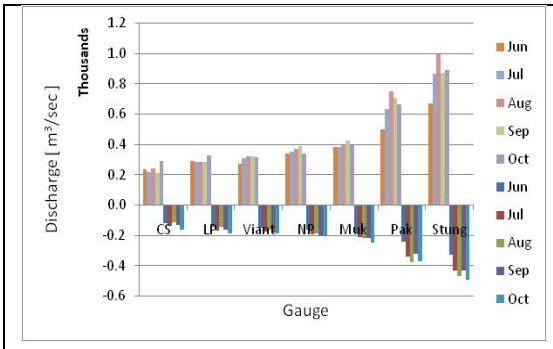
Fig. 6.7: Maximum and minimum discharge change in 1 to 5 days along Mekong

6.5.1.1. Forecast based on Average Differences

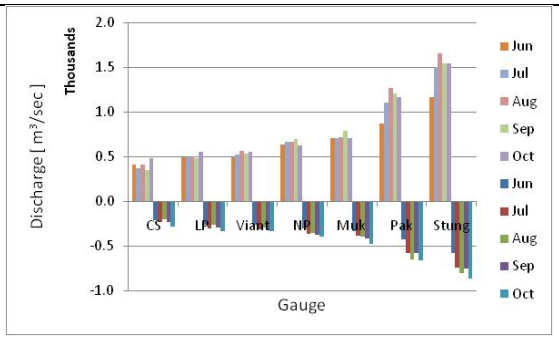
Both positive and negative $\overline{\Delta Q_j(1,1+t)}$ vary along the season and also from reach to reach. Plotting of $\overline{\Delta Q_j(1,1+t)}$ for t =1 to 5 days for the month of June to October shows the variation of ΔQ during the flood season (Fig.6.8). The positive and negative $\overline{\Delta Q_j(1,1+t)}$ values increase along the river and during the season from June to August and then decrease in September and October. However, positive and negative ΔQ-values are not of equal magnitude in different flood months.

6.5.1.2. Results and Discussion

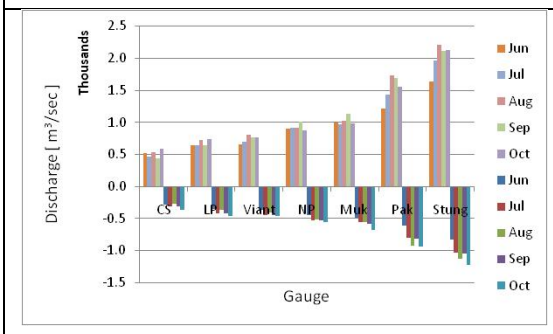
This information of positive and negative $\overline{\Delta Q_j(1,1+t)}$ could be used in Eq.6.7-6.8 in order to do estimate an a priori flood magnitude span. But since the range of $\overline{\Delta Q_j(1,1+t)}$ is based on average differences, future discharges are likely to exceed these upper and lower limits. Again, the span in between upper and lower limits obtained by adding $\overline{\Delta Q_j(1,1+t)}$ into Qj(i) increases with lead time from 1 o 5 days is given in Fig.6.8 . For example in the case of S, it increase from -400 to 1000 m^3/sec in 1 day lead time (Fig.6.8a) to -2000 to +3000 m^3/sec in five days lead time lead time (Fig.6.8e).



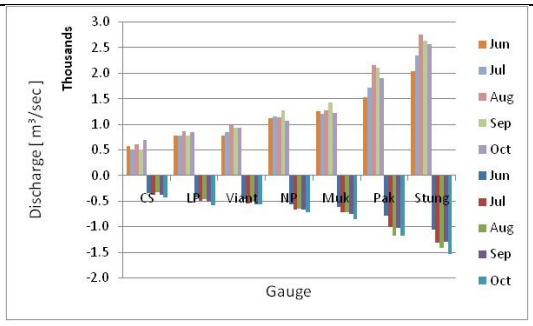
a)



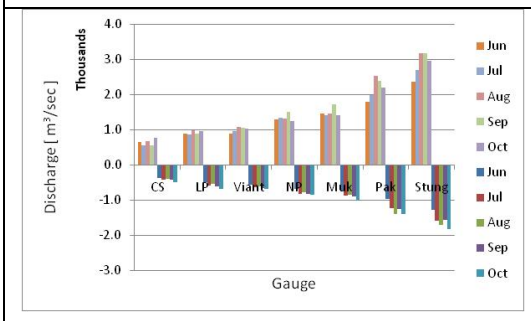
b)



c)



d)



e)

Fig. 6.8: Average discharge differences +/- (ΔQ_(j s) (i,i+t))¯ of main gauges a) t = 1, b) t = 2 , c) t = 3 , d) t = 4, e) t = 5 for flood season along Middle Mekong

6.5.2. Application of Type 1 Model (by moving back in time)

In the second step, different variants of Type 1 Model are used in order to do flood forecasts at C,L,V,N,M,P and S for 1 to 5 days lead times. The application potential of linear regression in forecasting future discharges is presented here.

6.5.2.1. Forecast based on simple linear regression

In the analysis mode the correlation of subsequent discharge differences , i.e. $\Delta Qj(i,i-1)$ vs. $\Delta Qj(i,i+t)$ and $\Delta DQj(i,i-1)$ vs. $\Delta DQj(i,i+t)$ for Mainstream discharges $Qj(i,i+t)$ and lateral inflows $DQj,j-1(i,i+t)$ are determined for each gauge j with lead time of 1 to 5 days. The high correlation value suggests the feasibility of Eqs.6.13 and 6.27 for flood forecast. It is observed that the correlation of subsequent discharge differences decrease with lead time, as shown in Fig.6.9.

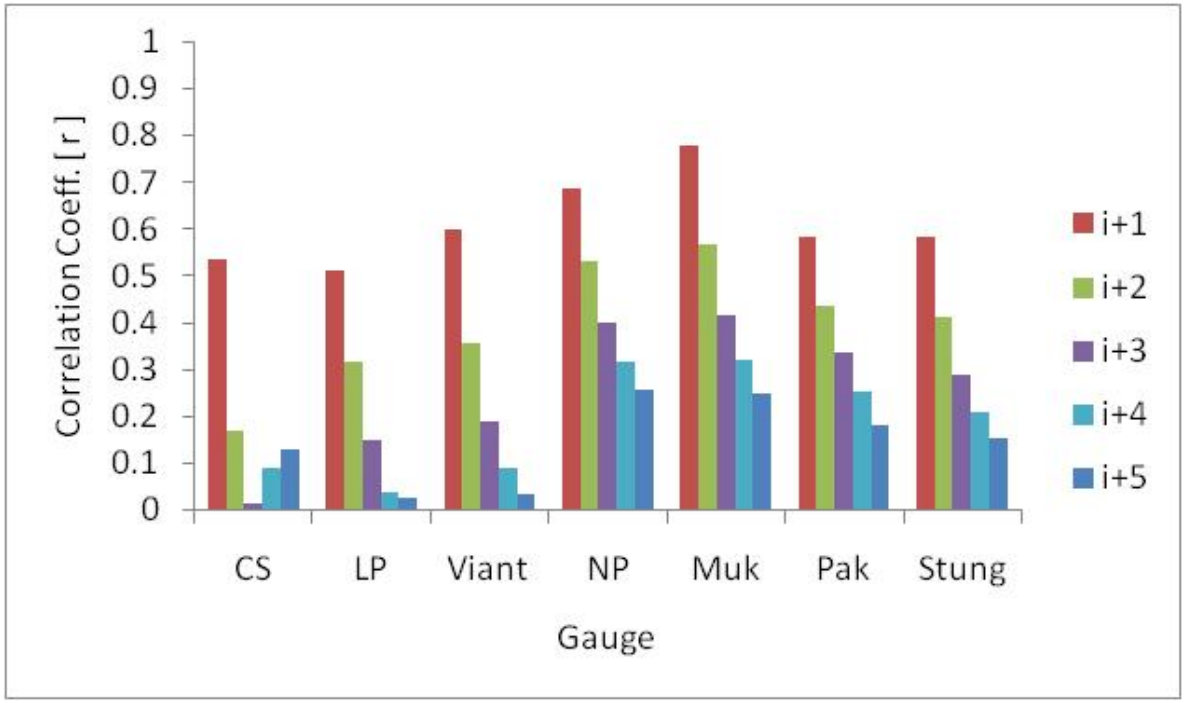


Fig. 6.9: Correlation coefficient of subsequent ΔQ

6.5.2.2. Results and Discussion

Prior to do forecasts by simple linear regression, the potential of this method is explored by further analysing the x-y plots $Q_j(i)$ vs $Q_j(i+t)$ for C,L,V,N,M,P and S. The forecasting results obtained by this approach are useful for one day forecasts because the magnitude of ΔQ is small in comparison to that of the total discharge, and also because the discharge correlation is high for 1 day lead time, as shown in Fig.6.9. But this approach cannot be successfully used for lead times more than 1 day as illustrated by the scatter plot of S for Q(i) vs Q(i+t) with t = 1 and t = 5 as shown in Fig.6.10. These x-y plots show for each gauge only small scatter for t =1, which increase for higher values of t reducing the usefulness of this approach for forecasts for more than 1 day lead time:as shown in Fig.6.10, the discharge spread ranges up to 30,000 m^3/sec? above the mainstream discharge of 40,000 m^3/secat S.

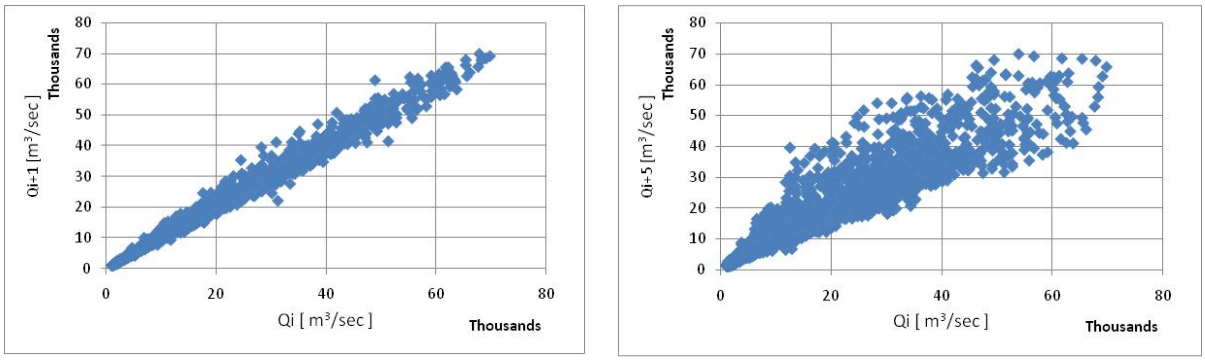


Fig. 6.10: x-y plots, right) Qj(i) vs Qj(i+1) and left) Qj(i) vs Qj(i+5) for S

6.6. Data Based Modeling Application on Mekong (by moving back in time and space)

6.6.1. Time lag calculation in between Major River gauges

The flow time between different stations can be determined by means of the cross correlation curve, or it can be estimated by means of velocity measurements in the reaches. The average travel time corresponds to the lag of the maximum of the cross correlation function. A time lag calculation helps to obtain the characteristic flood travel times of the river. The Lower Mekong

Basin is a large system where major stations are apart 100 to more than 300 km, yielding flow times between stations of the order of one or two days. These flow times correspond roughly to the time of measurement: the water level readings are made usually one or two times a day, depending on the policy of the specific country. Consequently, the analysis is made on daily values, and the forecast time is given in days.

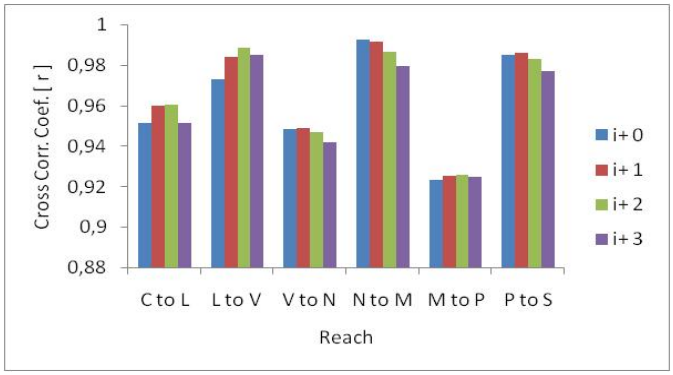


Fig. 6.11: Time lag computation by cross correlation analysis

Tab. 6.1: Time lag computation by a graphical approach

Reach		Length	Approx. Timelag	Approx. Timelag	Velocity	Velocity
		Km	Day	Hrs	Km/hr	m/sec
Chiang Saen	L.Prabang	353	2	48	7	2,0
L.Prabang	Vientiane	427	2	48	9	2,5
Vientiane	N.Phanom	368	2	48	8	2,1
N.Phanom	Mukdahan	91	1	24	4	1,1
Mukdahan	Pakse	256	1	24	11	3,0
Pakse	Stung Treng	201	1	24	8	2,3

Tables-6.1 & 6.2 give an overview of the results of computations of time lags between the major gauges. Fig.-6.11 gives time lags obtained from cross correlations of discharge time series, andTable-6.5 summarizes the time lag obtained by graphical plotting of storm peaks as shown in fig.6.12.The results for the upper reaches are better defined, because cross correlation peaks are sharp, whereas from Pakse downstream the cross correlation curves are spread more widely and identification of the maximum becomes more difficult. A similar problem occurs in locating the time delay of peaks between subsequent gauges in graphical plots of storm peaks of downstream gauges.

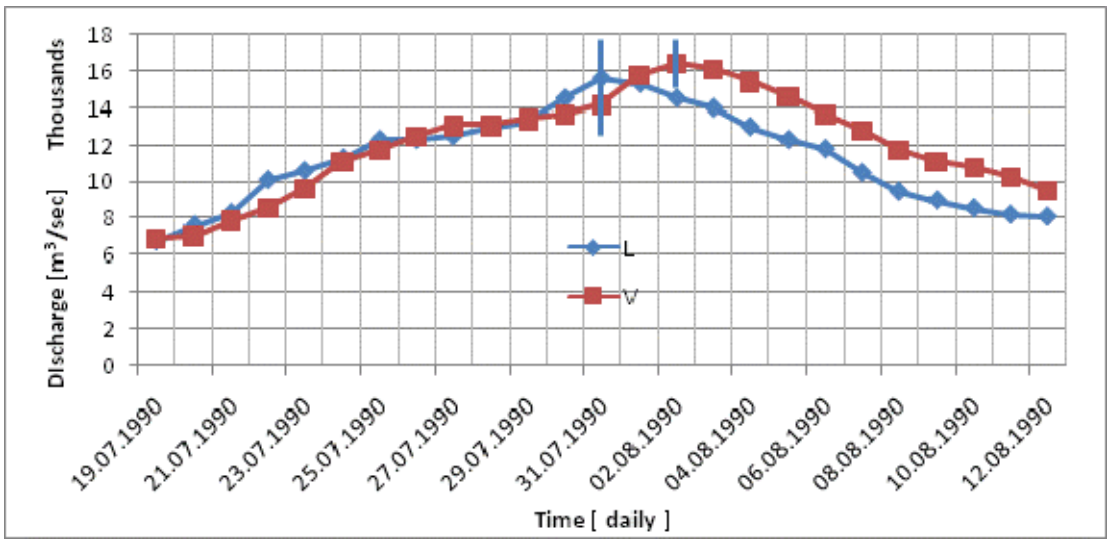


Fig. 6.12: Time lag computation by graphical approach

Results for Table-6.2 were calculated from given velocity data (RFMMC, 2004). The range between minimum and maximum values is due to velocity changes during different water stages. The data also show that time lags are about one day or smaller, which points out that in flood seasons a higher measurement interval would be helpful for near time forecasting.

Tab. 6.2: Time lag computations using velocities given by RFMMC

Guaging Station	Velocity (m/sec)			Reach		Length	Time Lag	Time Lag
	Min	Avg	Max			Km	Hrs	Days
Chiang Saen	0,7	1,5	2,6	Chiang Saen	L.Prabang	353	42	1,7
L.Prabang	0,2	1,1	2,1	L.Prabang	Vientiane	427	49	2,1
Vientiane	0,5	1,4	2,7	Vientiane	N.Phanom	368	41	1,7
N.Phanom	0,2	1,0	2,3	N.Phanom	Mukdahan	91	12	0,5
Mukdahan	0,5	1,0	1,9	Mukdahan	Pakse	256	38	1,6
Pakse	0,3	1,0	1,8					

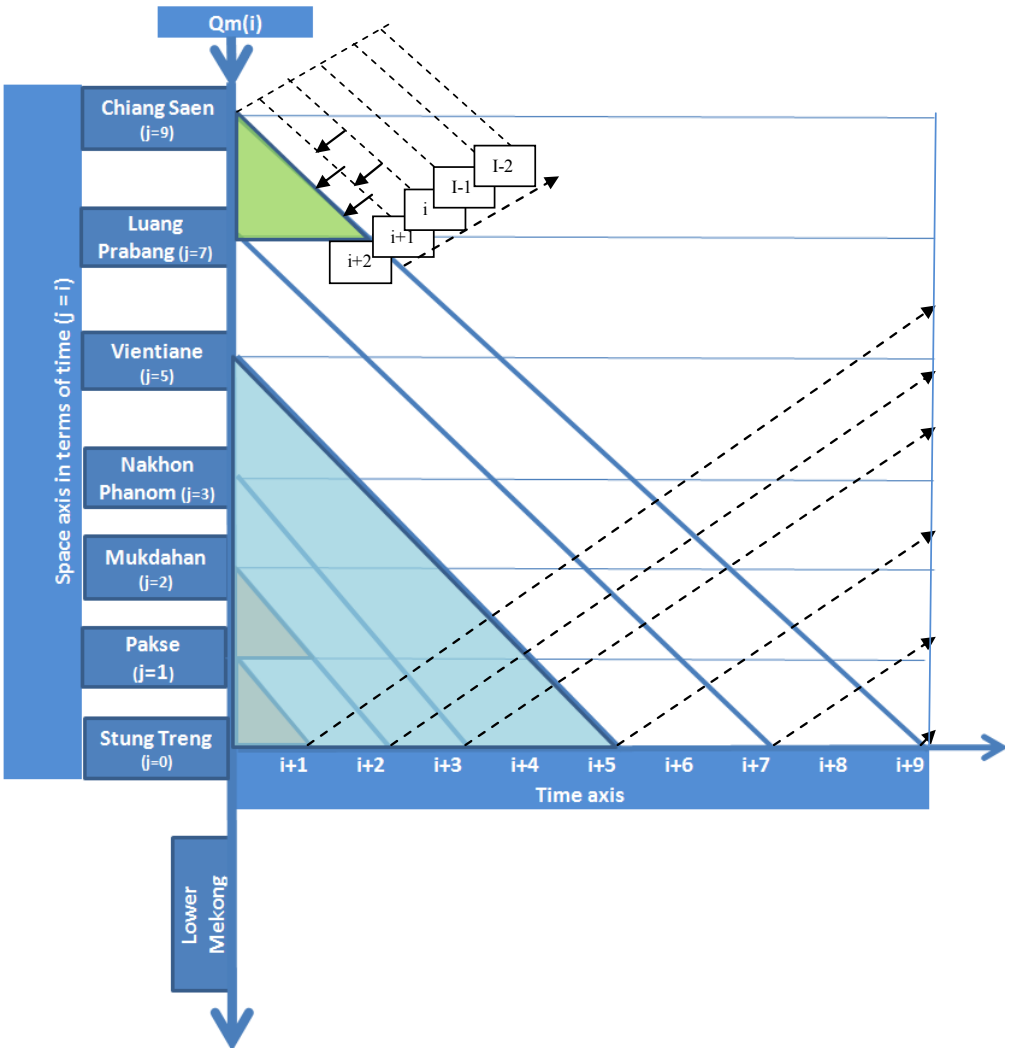


Fig. 6.13: Schematic diagram of middle section of Lower Mekong

The Mekong flood wave takes approximately 9 days to travel from C to S after passing through the intermediate flow gauging stations L, V, N, M and P. In the schematic diagram (Fig.6.13) of the middle Mekong, each gauge is assigned space index j, as indicated in Figs.6.4 & 6.5, which ranges from 0 to 9 increasing upstream from S to C. The travel time of flood wave is normally one day from one gauging station to another except for C to L, L to V, and V to N, where it is 2 days as given in Tables 6.1 & 6.2. In the course of travel from Chiang Saen to Stung Treng, the flood wave increases by contributions from lateral rivers which mainly drain the mountain range between Lao PDR and Vietnam.

In the flood forecasting procedure of moving back in time and space simultaneously, the information of upstream current discharges Qj(i) is required as indicated in Eq. 6.31. It is convenient to take space index j in terms of time t as indicated in Fig.6.13. The indices j are defined on the basis of time lag from one gauge to the next. For example, from gauge j = 0 to gauge j = 1, there is a time lag of one day and from gauge j = 3 to gauge j = 5, there is a time lag of 2 days and so on. In this situation of labelling space in terms of time, the forecasting lead time t at gauge j dictates the number of steps for moving upstream from the forecasting station. For example, for 1 day forecast at Stung Treng, where t = 1, upstream flow Qj(i) has to be considered taking j = 1 i.e., $Q_1(i)$, i.e., Pakse, and lateral inflows $\sum_{j=0}^{t-1} DQ_j(i + t - j)$ i.e., $DQ_0(i+1)$ which is lateral inflow between Pakse and Stung Treng. Similarly for 2 day forecast at Stung Treng, where t = 2 one needs to consider Qj(i), with j = 2 that is Mukdahan and lateral inflows $\sum_{j=0}^{t-1} DQ_j(i + t)$ - i.e., $DQ_0(i+2) + DQ_1(i+1)$ which is the lateral inflow between Pakse and Stung Treng, and Mukdahan and Pakse respectively. Therefore, the number of steps j movement upstream in space is equal to forecasting lead lead timet.

In order to do forecasts by using Eq.6.31, the lateral discharges $DQF_0(i+t)$ and $\varepsilon_{fe}(i + t)$ have to be estimated first in the analysis mode. Since lateral discharge data is not available for Mekong they are indirectly computed by subtracting upstream discharges from downstream discharges with time lags as given in tables-6.1 & 6.2, by using Eq.6.18.

6.6.2. Application of Type-0 Model (moving back in time and space)

The temporal structure of lateral inflows DQ_j-values is analyzed in order to determine the seasonality in lateral inflows. The box plot of mean monthly daily inflows, i.e. ($\frac{1}{n}, \frac{2}{n} \ldots\ldots \frac{12}{n}$), where n is number of years, of each gauge j of Main Mekong River has been plotted against time, i.e. 1, 2…12, in order to assess the mean monthly discharge temporal pattern with respect to empirical potential upper and lower limits of lateral inflows together with 25, 50 and 75 % quartiles. The maximum discharges are observed in the month of August or September. The box plots also indicate the deviations of discharges from the mean discharges in each month. The observed seasonality in discharges emphasizes the need of analyzing lateral ΔDQ separately for each month because the ΔDQ may be a function of the magnitude of DQ.

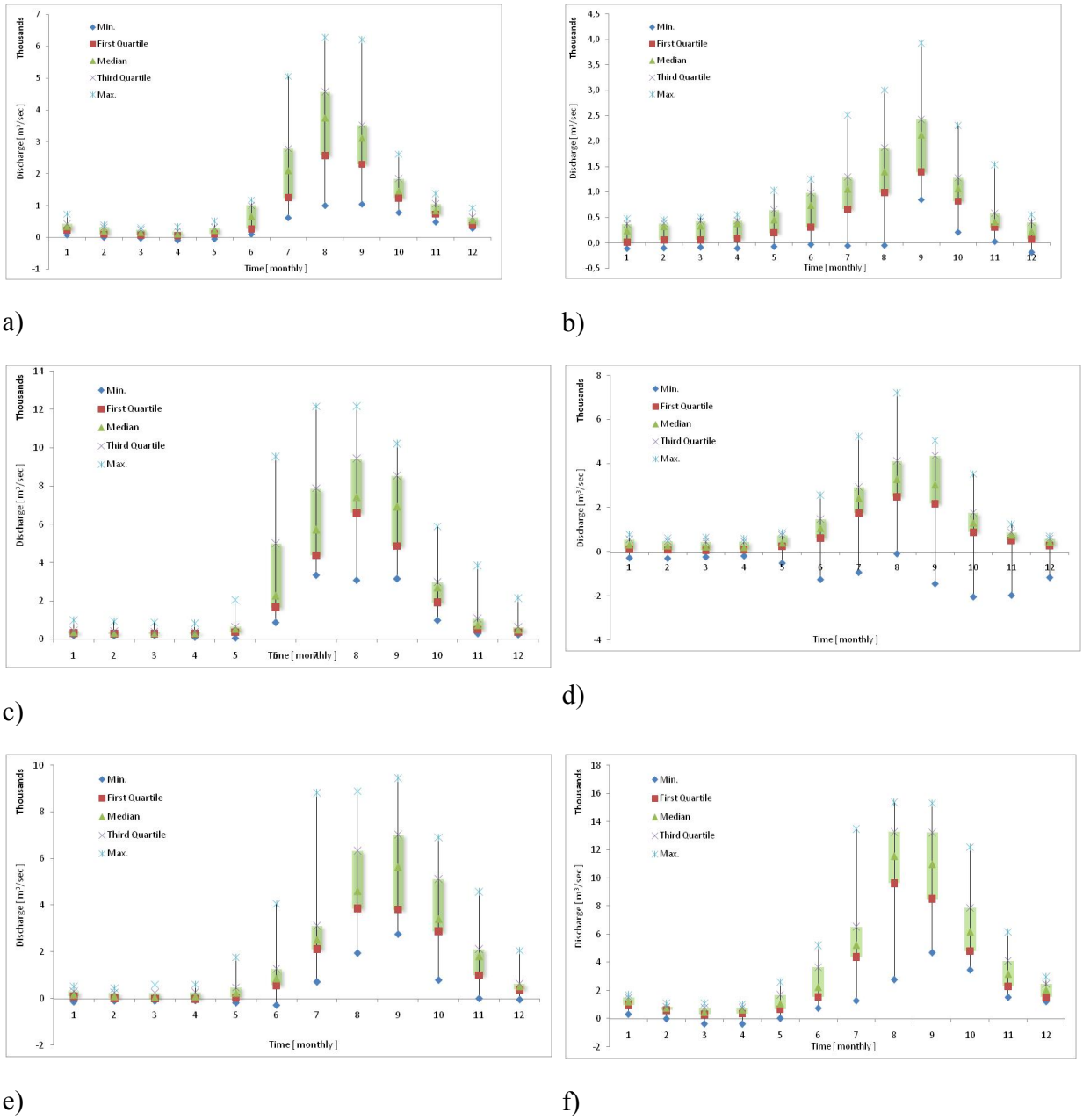
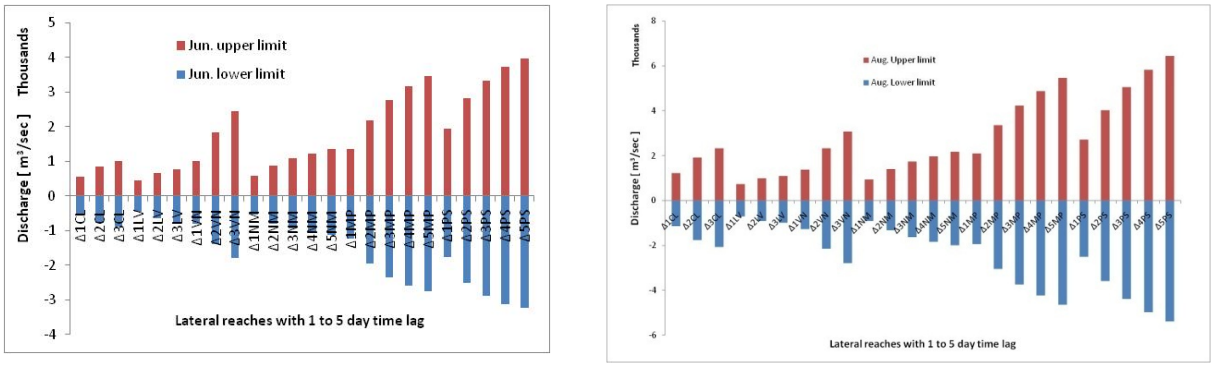


Fig. 6.14: Box plot of lateral inflows for CL, LV, VN, NM, MP and PS in a, b, c, d, e and f respectively

Therefore, $\Delta DQ_j(i+t)$ at $P(\Delta Q_5) = 5\%$ and $P(\Delta Q_{95}) = 95\%$ probability of occurrence are determined separately for each month in order to establish second order limits of the uncertainty bands. For example, the range of $\Delta DQ_j(i+t)$ for CL, LV, VN, NM, MP and PS reach between 0.05 and 0.95 probability limits per unit time of 1 to 5 days for the months of June and August

are given in Fig.6.15. It can be seen in Fig.6.15 that $\Delta DQ_j(i+t)$ vary with month and lead time. However, the variation of $\Delta DQ_j(i+t)$ does not increase linearly with lead time t.



a) b)

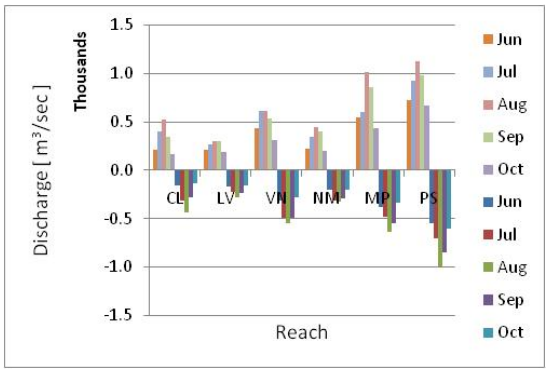
Fig. 6.15: Discharge change probability at P(ε_05) & ⟦P(ε)⟧_95) for June and August

6.6.2.1. Flood forecast by Type 0 Model

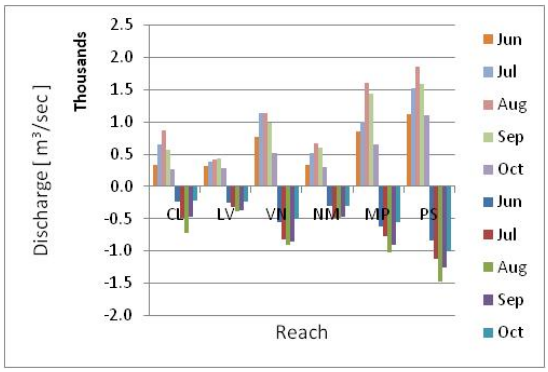
The $\overline{\Delta DQ_{J\,S}(\iota, \iota + t)}$ are of different magnitude in different reaches, being maximum in reach PS and minimum in reach LV. The maximum $\overline{\Delta DQ_{J\,S}(\iota, \iota + t)}$ of LV is in the month of September, in contrast to all other reaches where it is in August (Fig. 6.16).

6.6.2.2. Results and Discussion

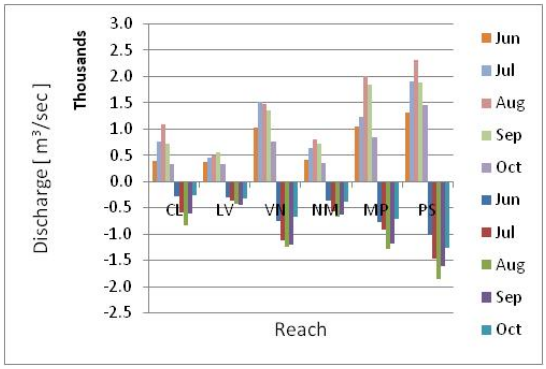
This information could be used as a priori information for estimating the flood magnitude span by means of Eqs.6.21-6.22 in combination with Eq.6.19. Fora1 to 2 days forecast, this approach can be successfully used, because the average change in discharge in short lead times is small. For example, for a two days forecast of S, one needs to add ΔPS (i+2) and ΔMP (i+1) into (S(i) +PS(i+2) + MP(i)), where the span of ΔPS (i+2) and ΔMP (i+1) ranges from -1500 to +1500 m^3/sec and -500 to +1000 m^3/sec respectively. So the total expected span of S(i+2) would become S(i) plus -2000 to +2500 m^3/sec (adding the negative and positive averages of ΔPS(1+2) and ΔMP(i+1) respectively). If most of the discharge changes remain near the average discharge change, then this approach gives a good initial guess of future discharges. However, $\Delta DQ_{j\,S}(i, i + t)$ do not always remain close to average. Therefore, better ways are needed in order forecast the discharge changes from current day to a future day.



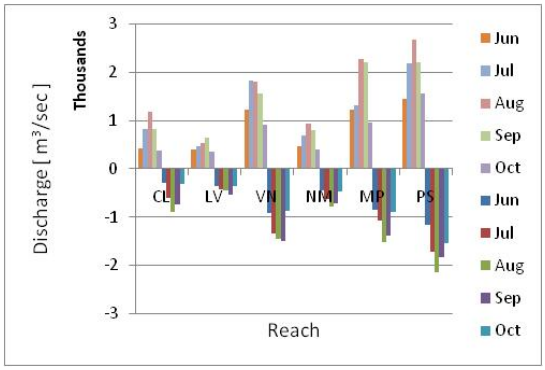
a)



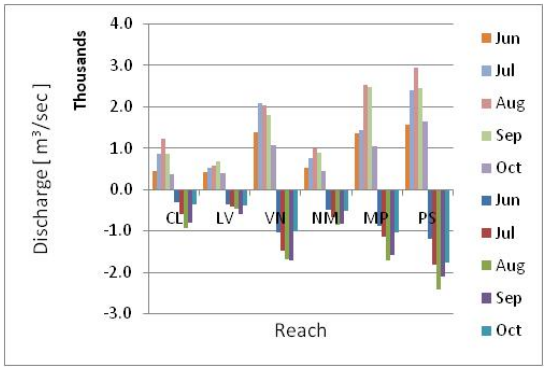
b)



c)



d)



e)

Fig. 6.16: Average discharge differences of lateral inflows +/- (Δ(DQ)_(j s) (i,i+t))¯ a) t = 1, b) t = 2 , c) t = 3 , d) t = 4, e) t = 5 for flood season along Middle Mekong

6.6.2.3. Forecast based on probability distribution

The $\Delta DQ_j(i+t)$ probability range between the limits of 0.05 to 0.95 of all the inflows of Middle Mekong are given in Fig.17 in different months of flood season from June to October. It is interesting to note that positive ΔDQ at 0.95 probability increase from June to August but decrease from September to October. In many reaches negative ΔDQ-values at 0.05 probabilities reach their maximum in September, for example in the MP reach.

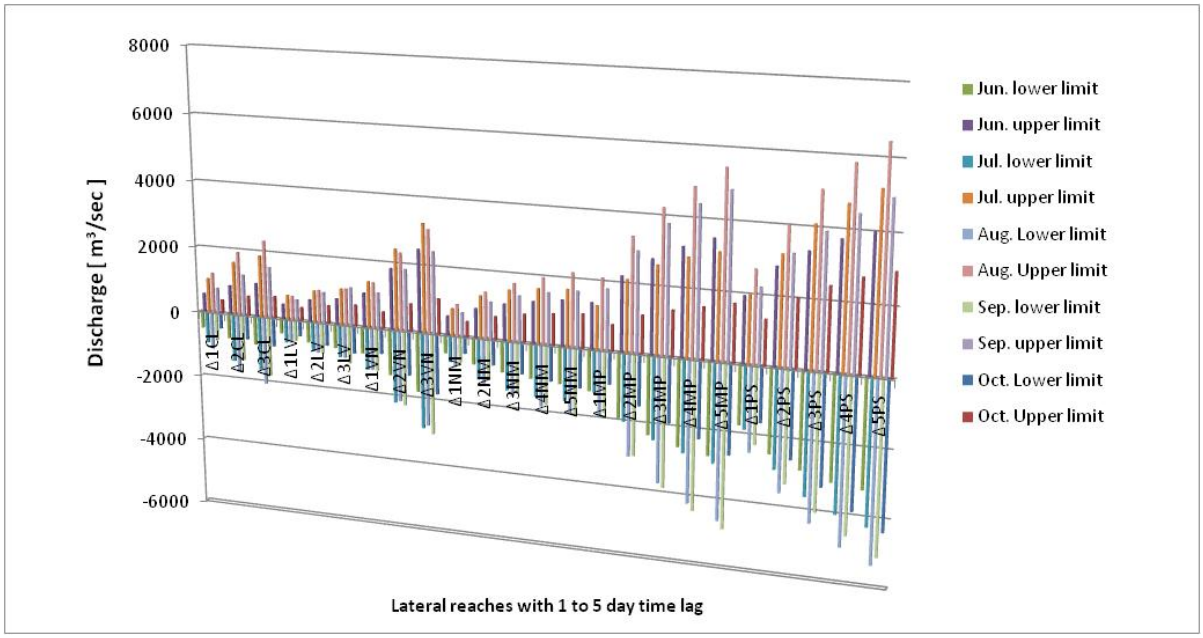


Fig. 6.17: Discharge variation Probability at P(ε_05) & (P(ε)_95) for flood season

Normalized probability density distributions of $\Delta DQ_j(i+t)$ of PS reach for the months of July, August, September and October are given in Fig.6.18, where, t = 1 to 5. The standard deviations and means of these pdfs are given in Table-6.3. The shape of the pdfs are Gaussian, however there is a positive bias in the pdf of $\Delta DQ_j(i+t)$ in July, whereas this bias becomes negative in the month of October, which indicates the $\Delta DQ_j(i+t)$ variation pattern during the flood season. The bias in the pdf is due to the rising and falling limbs of long flood hydrographs, where the increasing number of positive $\Delta DQ_j(i+t)$ in July and August pushes the mean of pdf towards right from the zero mean with the positive bias and increasing number of negative $\Delta DQ_j(i+t)$ in October push it toward left from zero mean. Similar distribution analysis is also conducted for other reaches for each month of flood season, i.e. June, July, August, September and October.

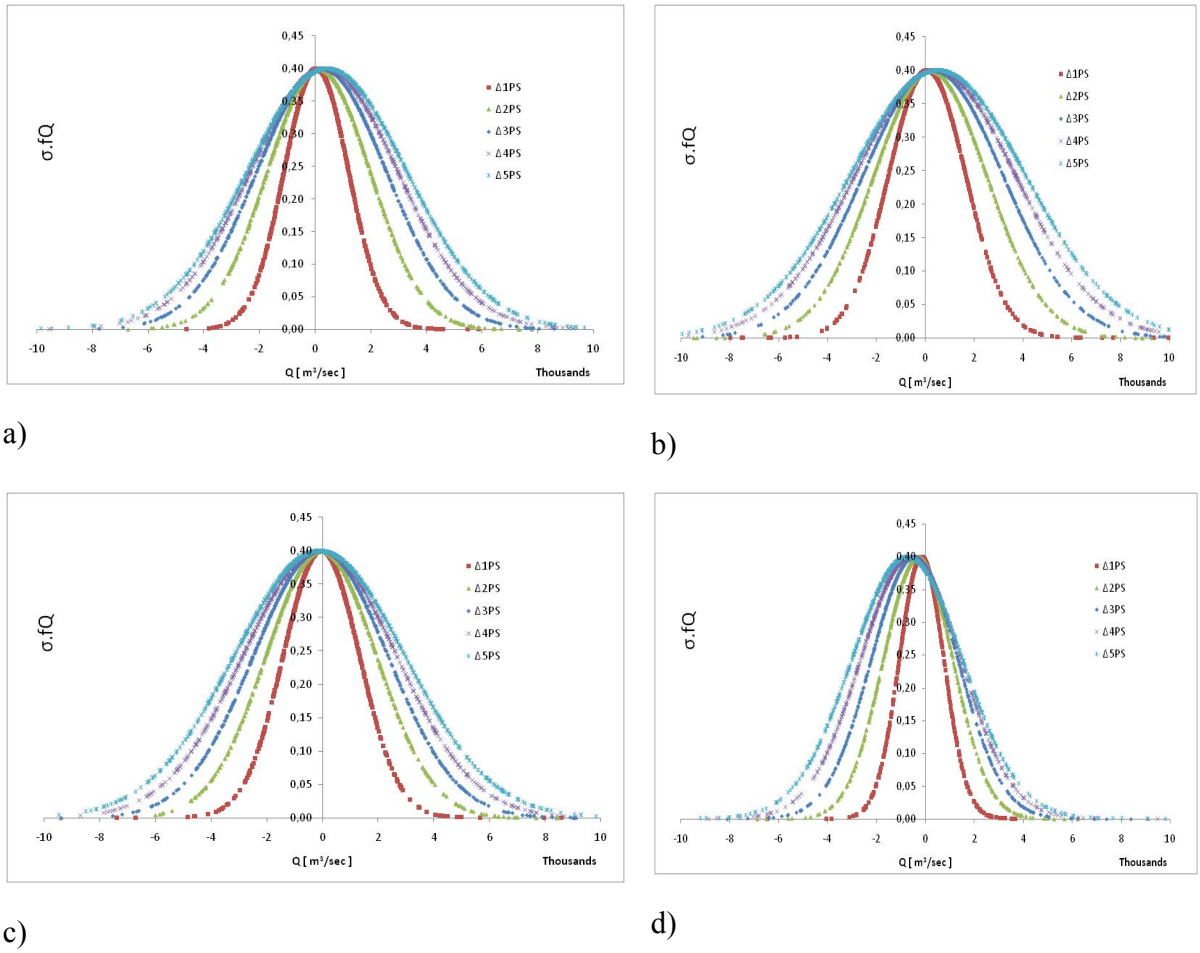


Fig. 6.18: Normal distribution plots of ΔQ of PS reach for the months of July, August, September and October in a, b, c and d respectively

6.6.2.4. Results and Discussion

The $\Delta DQ_j(i+t)$ probability range between the limits of 0.05 to 0.95 can be used in Eq.6.23-6.24 in order to estimate the span of $DQ_j(i+t)$ – which can further be used in Eq.6.19 in order to estimate the a prior span of future discharges $Q_j(i+t)$. But this span cannot be used as a forecast because the range between 0.05 to 0.95 % occurrence probabilities is too wide, as shown in Fig.6.18.

6.6.3. Flood forecast by Type 1 Model

The simple and multi-linear regressions are tried to determine lateral inflows in the application of Type 1 Models.

6.6.3.1. Forecast based on Simple linear Regression

In the analysis mode the correlation of subsequent discharge differences, i.e. $\Delta DQ_j(i,i-t)$ vs. $\Delta DQ_j(i,i+t)$ for lateral inflows $DQ_{j,j-1}(i,i+t)$ are determined for each gauge j with lead lead timeof 1 to 5 days. It is observed that the correlation of subsequent discharge differences decrease with lead time, as shown in Fig.6.19.

Tab. 6.3: Statistical parameters for normal distributions of ΔDQj(i+t) for PS

		i+1	i+2	i+3	i+4	i+5
Jul	Average	72	79	60	135	237
	St. dev.	1719	1849	1171	1865	2343
Aug	Average	330	421	100	215	328
	St. dev.	2741	3036	1587	2324	2879
Sep	Average	6	16	-13	-17	-26
	St. dev.	2679	2946	1381	2019	2455
Oct	Average	-361	-459	-143	-296	-441
	St. dev.	1159	1328	916	1433	1780

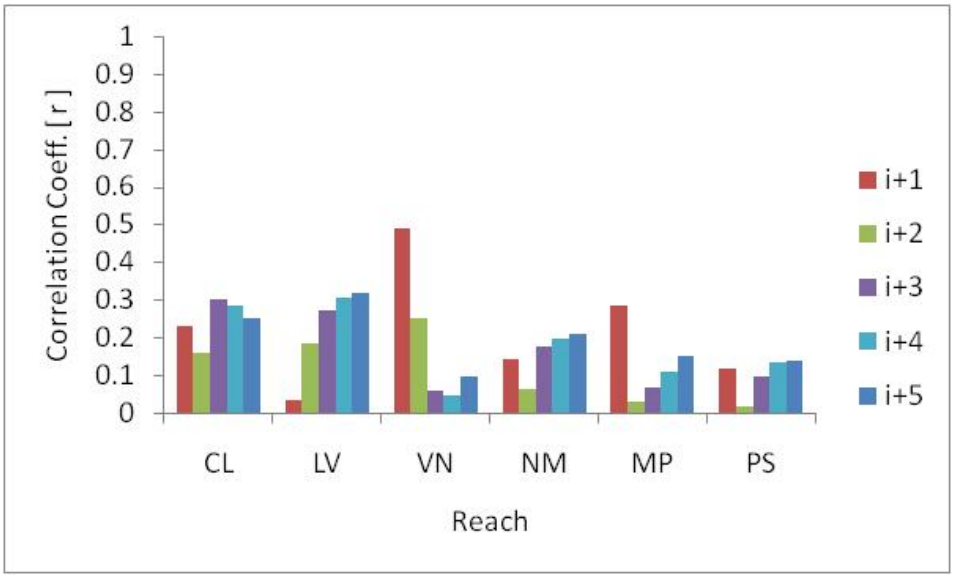


Fig. 6.19: Correlation coefficient of subsequent ΔDQ

6.6.3.2. Results and Discussion

The poor correlation between subsequent lateral inflows suggests the limited potential of this approach as augmented by the Scatter plot of DQ(i) vs DQ(i+t) for reach PS (Fig.6.20). Therefore, simple linear regression of subsequent discharges is not used for further computations.

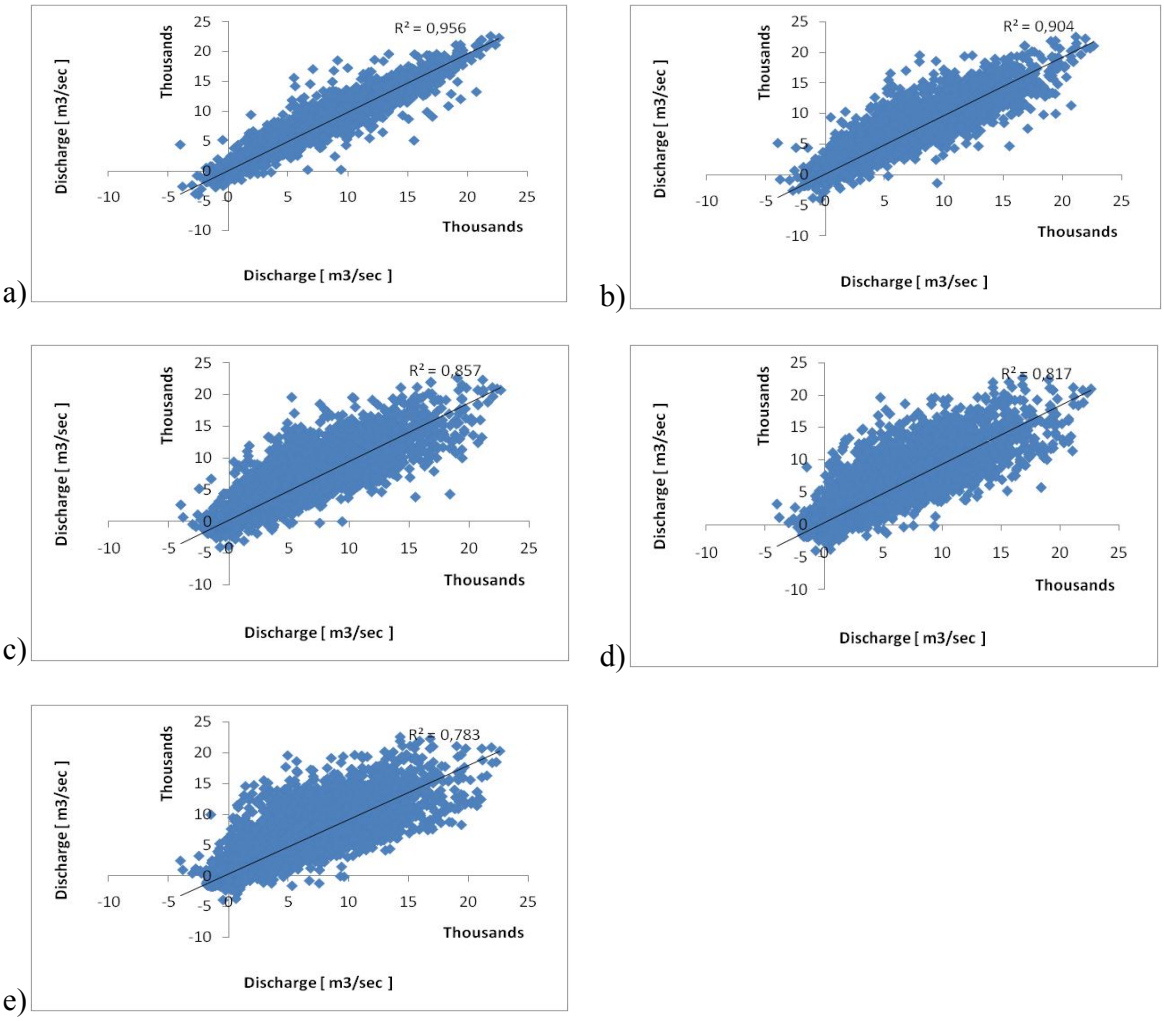


Fig. 6.20: scatter plot of PS reach DQ(i) vs DQ(i+t) with t is1 to 5 in a, b, c, d and e respectively

6.6.3.3. Forecast based on Multi linear Regression

The multi-regression parameter $\alpha j(i+t)$, $Cj(i+t)$ and $\gamma_j(i+t)$ of Eq.6.30 are determined in the analysis mode for approximating $DQF_0(i+t)$. The coefficients $\beta j(i+t)$ and $\varphi j(i+t)$ of Eq.6.33 are also pre-determined. These regression parameters are given in Table-6.4.

Tab. 6.4: Regression parameters to compute lateral inflows DQj(i+t)

Substantial Model										Updating	
DQ_j (i+t)	Q_k	DQ_9 (i+t)	DQ_7 (i+t)	DQ_5 (i+t)	DQ_3 (i+t)	DQ_2 (i+t)	DQ_1 (i+t)			$\beta j(i+t)$	$\varphi j(i+t)$
	CS	CL	LV	VN	NM	MP	PS				
$\gamma_j(i+t)$	$\gamma_j(i+t)$	$\gamma_9(i+t)$	$\gamma_7(i+t)$	$\gamma_5(i+t)$	γ_3 (i+t)	γ_2 (i+t)	$\gamma_1(i+t)$	$C_j(i+t)$	R^2		
DQj+9											
i+1	0.01	0.96						24.15	0.94	0.27	17.73
i+2	0.03	0.88						52.52	0.86	0.08	49.69
i+3	0.06	0.82						58.37	0.81	0.01	67.26
i+4	0.09	0.77						52.70	0.78	-0.05	76.26
i+5	0.11	0.73						47.64	0.76	-0.11	85.33
DQj+7											
i+1	0,02	0,02	0,88					-8,36	0,89	-0.01	7.70
i+2	0,04	0,02	0,78					-14,45	0,80	0.08	16.05
i+3	0,05	0,04	0,69					-22,26	0,74	0.02	22.58
i+4	0,06	0,06	0,62					-30,41	0,70	0.01	26.23
i+5	0,07	0,06	0,57					-41,55	0,68	0.01	27.00
DQj+5											
i+1	0,00	0,00	-0,03	1,00				30,12	0,98	0.47	10.78
i+2	0,02	-0,02	-0,06	0,98				64,09	0,94	0.26	39.95
i+3	0,04	-0,07	0,02	0,95				99,95	0,90	0.09	83.54
i+4	0,07	-0,11	0,09	0,91				128,02	0,86	0.00	125.75
i+5	0,09	-0,12	0,12	0,88				0,12	0,83	-0.05	169.07
DQj+3											
i+1	0,03	-0,01	-0,01	-0,01	0,96			-5,21	0,92	0.30	1.72
i+2	0,05	-0,02	-0,03	-0,01	0,89			-12,61	0,83	0.04	-14.60
i+3	0,08	-0,03	-0,07	-0,01	0,82			-21,13	0,76	-0.06	-34.24
i+4	0,10	-0,04	-0,09	-0,01	0,78			-24,87	0,70	-0.04	-43.82
i+5	0,11	-0,06	-0,11	0,00	0,74			-22,02	0,66	-0.03	-44.58
DQj+2											
i+1	0,02	0,03	0,02	0,00	0,00	0,95		-13,07	0,96	0.32	-4.50
i+2	0,04	0,08	0,05	0,01	0,01	0,88		-13,99	0,90	0.10	-5.01
i+3	0,06	0,11	0,08	0,03	0,03	0,81		-11,67	0,84	-0.07	-8.53
i+4	0,09	0,11	0,11	0,04	0,03	0,76		-8,20	0,80	-0.11	-13.91
i+5	0,12	0,11	0,15	0,04	0,01	0,71		-8,09	0,76	-0.08	-28.16
DQj+1											
i+1	0,03	0,02	0,01	0,02	0,04	0,01	0,94	-24,01	0,96	0.29	67.80
i+2	0,09	0,02	-0,01	0,04	0,04	0,07	0,85	-61,38	0,91	0.20	153.64
i+3	0,13	0,01	-0,01	0,06	0,04	0,10	0,77	-90,12	0,87	0.12	230.73
i+4	0,17	0,02	0,00	0,08	0,06	0,11	0,71	-111,39	0,83	0.09	299.52
i+5	0,21	0,03	0,04	0,10	0,09	0,10	0,66	-129,12	0,81	0.07	371.19

The flood flows of C, L, V, N, M, P and S are forecasted for the last 15 flood seasons s by using Eq.6.30 and Eq.6.33. These forecasted flows are compared with actual observed flows, and then flood forecast accuracy is estimated by using the persistence index PI.

The patterns of errors in these forecasts with different lead times are observed. Normally the flood forecast errors are positive in flood accession and negative in flood recession above certain critical discharge values. Then there are the remaining error after updating with which one has to live with. Therefore, random values from the error probability distribution must be selected and must be added to or subtracted from crisp forecast to generate flood forecast band.

6.6.3.4. Results and Discussion

The comparative plot of observed and forecasted discharges along with remaining errors is shown in Fig.6.17. This Fig. shows the results of 1 and 2 day flood forecasting results for C, L, V, N, M, P and S. It is evident from Fig.6.21 that the remaining errors are negligible at all gauges in 1 day lead time, however, the errors goes up to 10,000 m^3/sec in two days lead time.

The flood forecasting errors of P and S for fourth and fifth day lead time are presented in Fig-6.22. The error ranges from 10,000 to 20,000m^3/sec. The error time series as plotted on the bottom of Fig.6.22 shows that positive errors are always followed by positive errors and vice versa. Secondly, the flood forecast is lagging behind the observed discharge time series, especially in the rising limb of flood hydrograph. This means, this approach fails to predict the sudden rise in discharge in the rising limb. The high dependence on the past discharges in predicting future flows is the main reason of this failure.

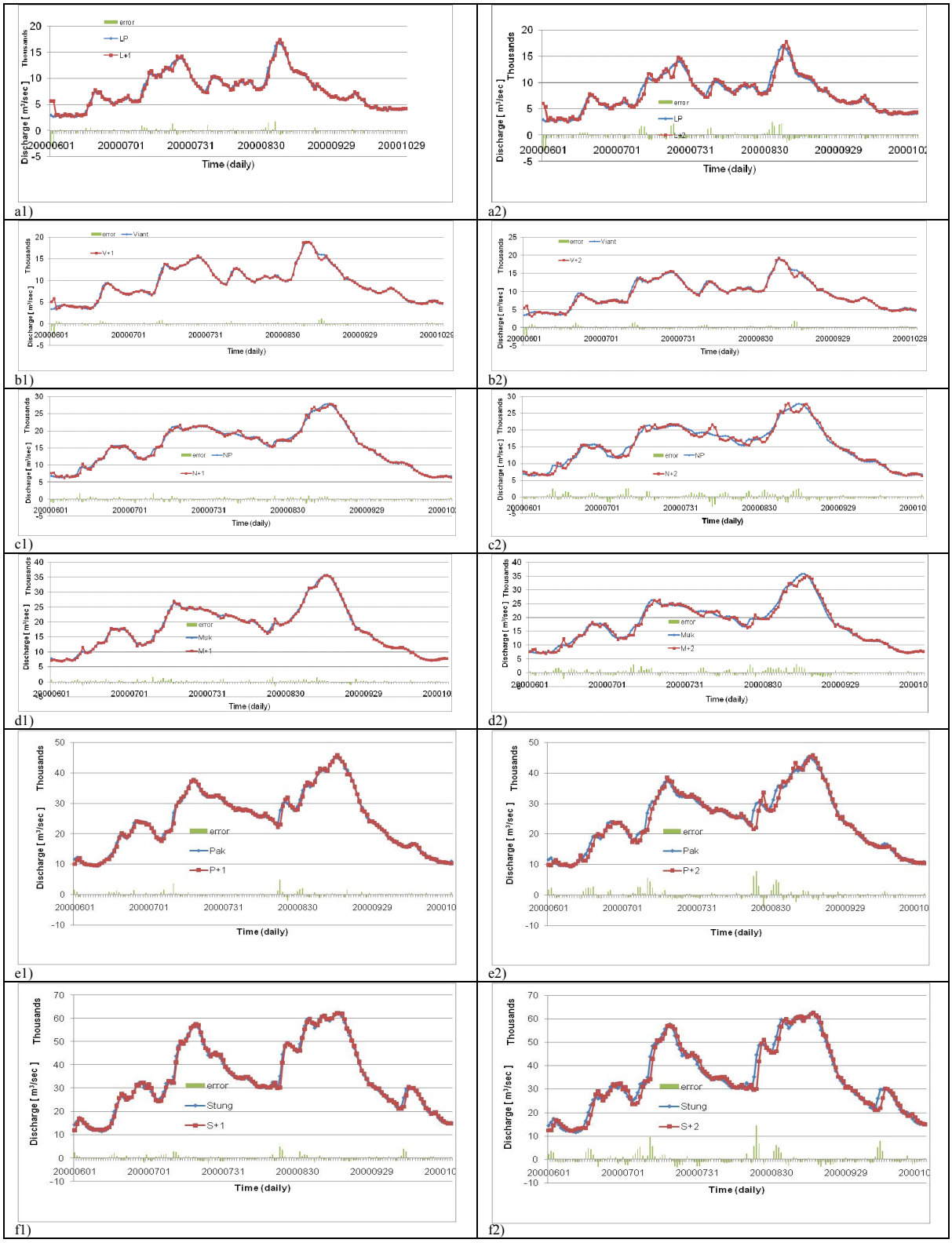


Fig. 6.21: Flood forecast by multi-regression for C, L, V, N, M, P and S from top to bottom respectively, right column with one day lead time, left column with two days lead time

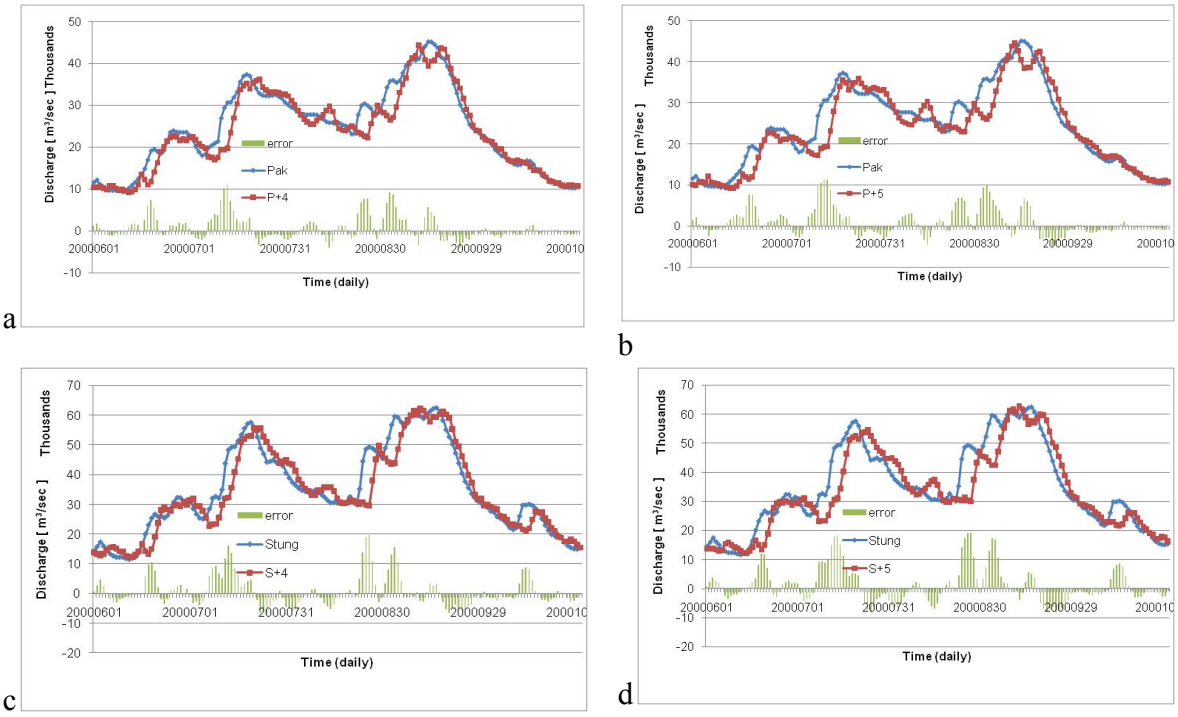


Fig. 6.22Flood forecast by multi-regression for P and S from top to bottom respectively, right column with four days lead time, left column with five days lead time

The probability distribution plots of the forecasting errors of C, L, V, N, M, P and S are shown in Fig.6.23. In each case, the spread of errors increase with the lead time. The forecasting errors of lead time greater than two days are biased toward positive x-axis, which means the flood forecasts underestimate the observed flow in the longer lead times.

The efficiency of flood forecast by multi-regression is gauged with the help of persistence index (PI) is shown in Fig.6.24. The efficiency of all forecasting gauges, except Nakhon Phanom, for first three days lead time varies from 0.5 to 0.7. However the efficiency decreases with increase in lead time from 4^{th} to 5^{th} day.

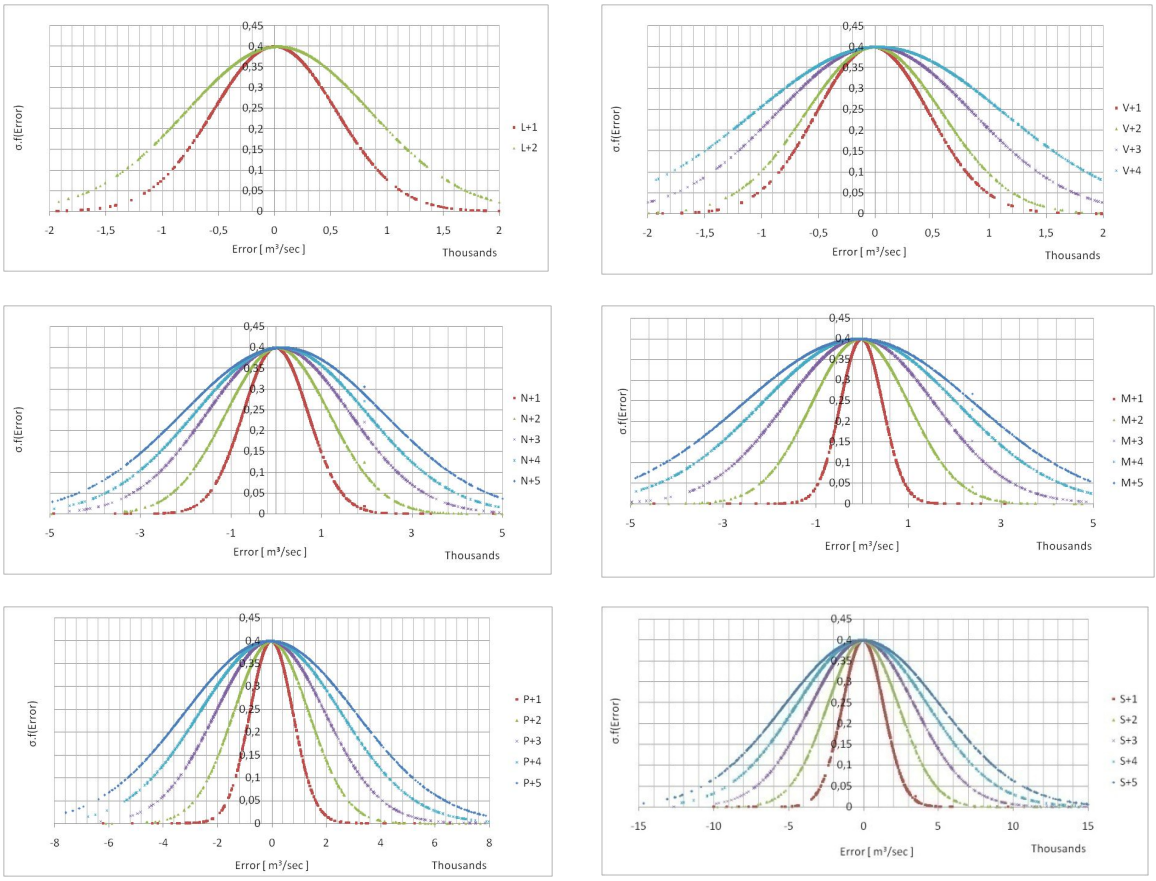


Fig. 6.23: Probability distribution of errors for 1 to 5 days forecast of C, L, V, V, M, P and S based on 1991 to 2000 flood season data

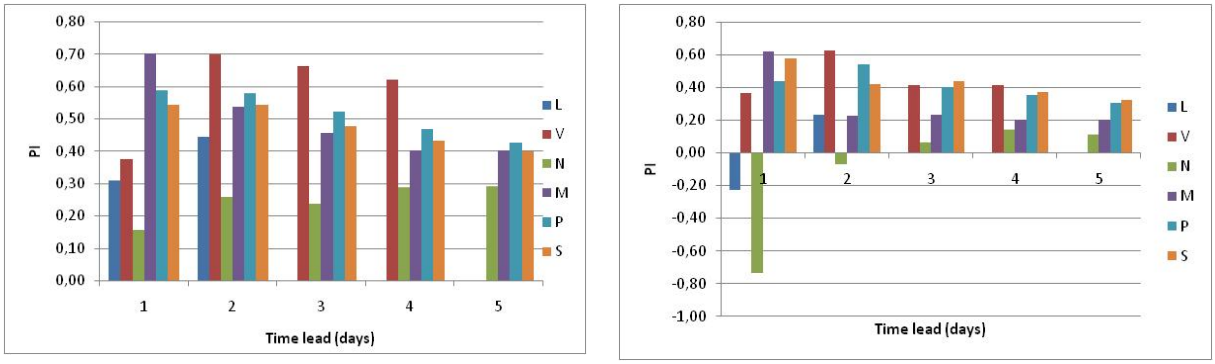


Fig. 6.24: Efficiency of 1 to 5 days flood forecast by PI for C, L, V, V, M, P and S based on left column; 1991 to 2000 and right column; 2000 to 2005 flood season data

The errors of flood forecast by multi-regression are further compared with the errors of no model forecast. In the case of no model forecast, the observed discharge of today is taken as the estimate of future discharges.

Tab. 6.5: Standard deviation of flood forecast (for 1991 to 2000 and 2000 to 2005)

Gauge	Without forecast (1991 to 2000)	Forecast based on multi regression (1991 to 2000)	Without forecast (2001 to 2005)	Forecast based on multi regression (2001 to 2005)
L+1	640	549	649	622
L+2	1101	821	1059	910
V+1	623	493	674	479
V+2	1081	599	1117	634
V+3	1461	860	1454	1005
V+4	1778	1096	1716	1251
N+1	712	716	818	724
N+2	1288	1145	1417	1222
N+3	1799	1579	1886	1700
N+4	2248	1896	2293	2033
N+5	2648	2219	2650	2406
M+1	826	474	987	469
M+2	1536	1098	1762	1209
M+3	2165	1613	2439	1833
M+4	2721	2104	3028	2413
M+5	3219	2483	3543	2848
P+1	1212	831	1225	758
P+2	2118	1411	2167	1366
P+3	2887	2053	2984	2108
P+4	3546	2613	3690	2780
P+5	4122	3137	4295	3390
S+1	2007	1397	1974	1242
S+2	3508	2515	3460	2310
S+3	4760	3540	4714	3300
S+4	5809	4489	5737	4231
S+5	6708	5290	6597	5057

The respective errors are compared in terms of standard deviations. The standard deviation of errors of multi-regression model is compared with the standard deviation of errors of no model

forecast. This comparison of standard deviation is given in Table-6.5, which indicate that standard deviations reduce by 50% in 1 day lead time from the no model forecast to the forecast by multi-regression. Again this reduction in standard deviation is valid for all forecasting gauges except for Nakhon Phanom. But this reduction in standard deviation of errors decreases with increase in lead time.

6.7. Summary

A no model forecast, i.e. assuming discharge of today will be the discharge of the future, serves as first bench mark for improvement by any forecasting model. Application of Type 0 Model, both by moving back in time and moving back in time and space, established a span of upcoming future discharges based on the empirical analysis of discharge time series without real time forecast: i.e. based on historical minimum, maximum, average, discharge differences. This analysis is further augmented by provision of the probability distributions of discharge differences of 1 to 5 days lead lead time. Consideration of seasonality in probability distribution and in the estimation of minimum, maximum, average, discharge differences further narrowed the future discharge span.

Type 1 Model is applied in two versions, one by moving back in time, and the second moving back in time and space. The results of flood forecasts by moving back in time and space simultaneously is found better than the forecasting results of moving back in time only. The multi-regression model is found to be the best approach for estimating lateral inflows. Forecasting results of the multi-regression model are further improved with the help of updating by forecasting error regression.

Forecasting efficiency for the second version of Model 1 ranges from 0.5 to 0.7 in terms of PI for first three days, however it decreases sharply above three days lead time. Again, the forecasting efficiency for Nakhon Phanom, i.e. 0.2 to 0.3 is much lower than the forecasting efficiency observed at other forecasting gauges.

7. Rainfall and Discharge Data Based Forecasting (Type 2 Model)

7.1. Type 2 Models

During application of the regression model on the Mekong River, it was observed that the use of persistence between subsequent discharges is sufficient to produce flood forecasts with acceptable efficiency, i.e. PI = 0.45 to 0.70 up to two days lead time. However, the efficiency of flood forecasts by this model reduces from PI = 0.4 to 0.3 with increase in lead time from the 3rd to the 5th day (Chapter-6). It is likely that rainfall forcings reduce the capability of the type 1 model to produce effective flood forecasts for lead times greater than two days. This loss of efficiency is due to the fact that the direct influence of rainfall is neglected. This can be tested by plotting the errors of the Type 1 model forecasts against daily rainfall hyetographs. The empirical evidence indicates that use of rainfall data becomes essential. If a significant correlation exists between rainfall and errors, then linear regression could be used directly to reduce the forecast errors by means of the rainfall time series. However, it is less likely to obtain good correlation between errors and the daily rainfall data because, the Type 1 model, already explained part of the discharges caused by rainfall forcing. Therefore it is difficult to justify to use the rainfall data again for explaining the unexplained portion of the errors. Consequently, instead of improving model 1 forecasts, it is more logical that the daily rainfall should be correlated directly with lateral runoff.

The distinctive hydrological feature of a rainfall – runoff relation in catchments is the time delay of the rainfall before it runs off. It is informative to obtain this time delay, expressed through a time lag T, before regressing rainfall occurrence $DP_{j,j+1}(i)$ against lateral discharges $DQ_{j,j+1}(i)$. Then the lateral discharge can be directly calculated from the rainfall by means of simple linear regression, provided each daily impulse of rainfall $DP_{j,j+1}(i)$ actually affects each daily lateral discharge $DQ_{j,j+1}(i)$ with time delay T. However, the runoff $DP_{j,j+1}(i)$from daily catchment rainfall is distributed in time over more than one day except in very small catchments of steep areas. Consequently, it is better to use a rainfall runoff model to distribute the rainfall over time to shape the discharge hydrograph. This approach results in the Type 2 model.

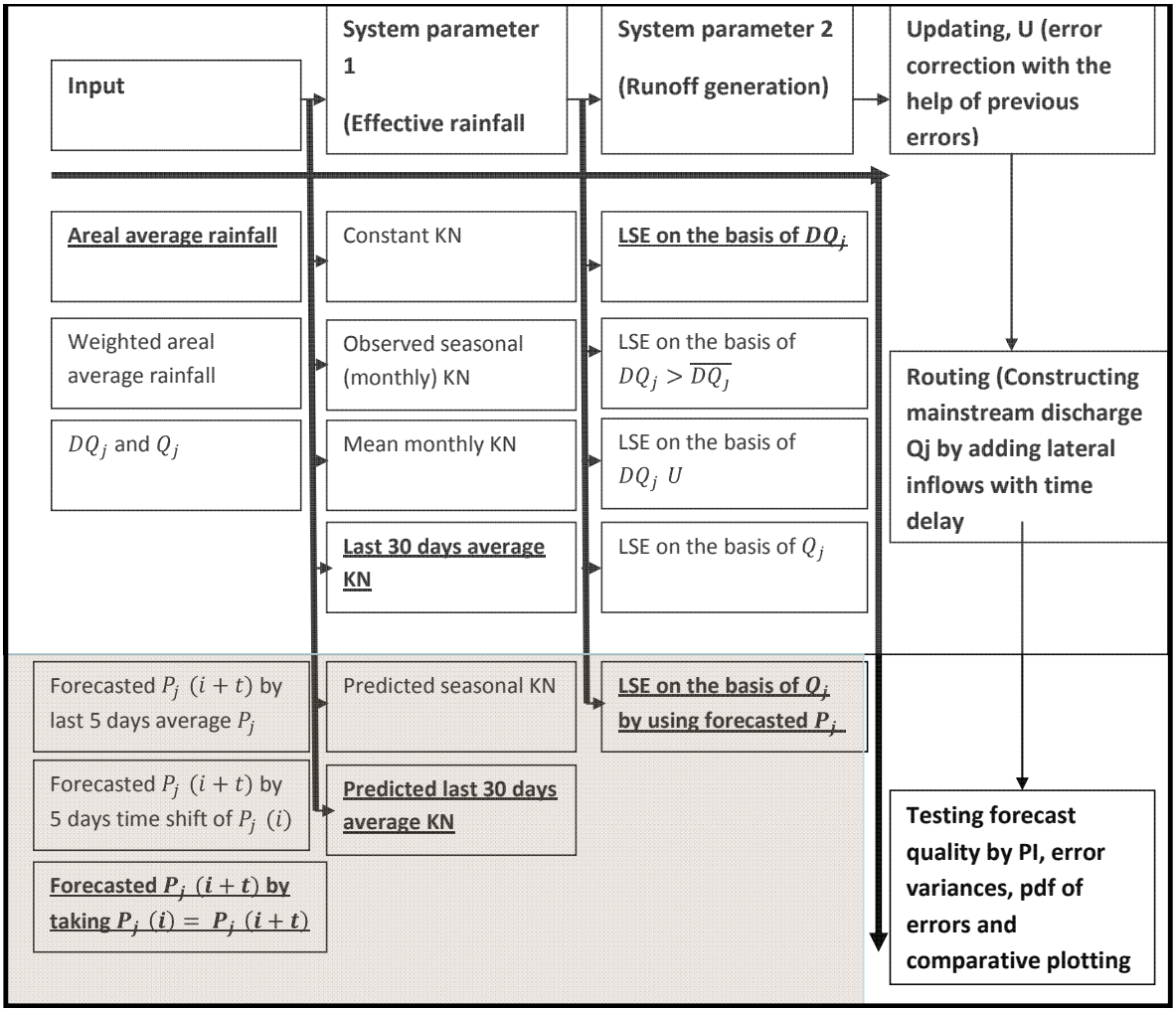


Fig. 7.1: Flow chart for the development of Type 2 Model

The work flow chart for the Type 2 Model for Mekong flood forecasting, is presented in Fig.7.1. Different options to prepare rainfall input, effective rainfall and runoff are listed in this Fig. The options preferred at the end are underlined. The non-shaded portion of Fig.7.1 pertain to the analysis mode of model development, whereas the shaded portions of Fig.7.1 pertain to the forecast mode, i.e to the use of the model with a forecasted rainfall input. This chapter explains the development and application of Type 2 Model by following the steps given in this Fig.

It was already discussed in chapter-3 that conceptual models serve well the purpose of flood forecasting in macro scale catchments. But prior to building the conceptual model structure, one has to answer three questions: what should be the spatial distribution of the sub-areas of the catchment of a large river in a modeling structure, both in the horizontal and vertical scale, how

to estimate the effective rainfall, and how to convert the effective rainfall into the runoff hydrograph.

7.1.1. Spatial distribution of a catchment

For the answer to question one for the horizontal scale, depending on the type of model used, the catchment is subdivided into a number of sub-basins, into grids, or into response units. However, the difficulty of measurement or estimation of rainfall and discharge in each distributed unit place limits on this division. The use of finer spatial resolution is of no additional benefit if the spatial resolution of the network of gauges for measuring observed data is coarser than that of the distributed sub-basins or grids. Secondly, even if it is assumed that high resolution rainfall and discharge data are available, this approach loses its validity because the efficiency is reduced when moving from micro to macro scale (Plate,2009) in balancing the equations of a flow from one scale to another. In principle, it would be possible to subdivide the areas of the sub catchments of the Mekong even further, into two or more units of macro scale size, provided that discharge gauges were available at some intermediate points in order to validate rainfall runoff models. However, no such gauges were available at intermediate points for this study, so that neither fully distributed modeling nor partial distribution into two or more sub-basins was possible. Consequently, all the sub-areas between subsequent discharge gauges on the main river are lumped into one catchment, and the discharge change $DQ_{j,j+1}(i)$ was calculated from the mainstream discharge gauges for these larger sub-catchments, as described in Chapter-6.

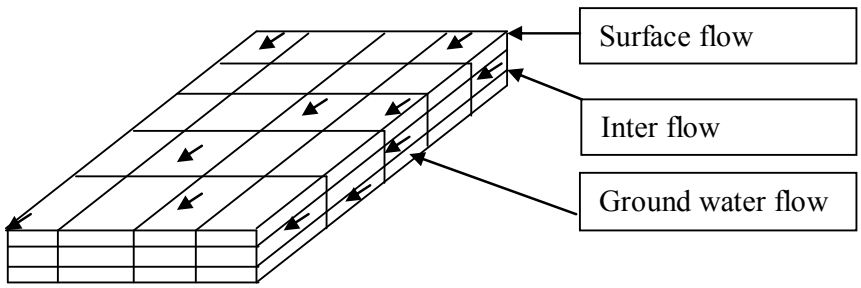


Fig. 7.2: Conceptualization of flow, spatial distribution

In answering the first question for the vertical scale one needs to consider the vertical structure of a catchment. For the conceptualization of flow components, one can divide the catchment into a number of horizontal layers piled up one on another from the top surface layer to deep down soil strata through which the water moves with different translation times. Among hydrologists, the most conventional division of water flow is the separation into surface (quick) and sub-surface

(delayed) flow. And the sub-surface flow can be broken down again into interflow that runs just beneath the surface layer with little time delay, and ground water flow.(Fig. 7.2). In unit hydrograph analysis the surface flow has been named as direct flow generated by rainfall from the present storm, and the delayed flow generated through draining of groundwater from the soil is called base flow. Accordingly, the base flow has to be separated from the total flow hydrograph in order to generate the direct flow discharges. An effective rainfall is used to generate direct flow, to which the pre-estimated base flow is added afterwards in order to obtain the total discharge. However, in conceptualization of flow in this study, base flow and surface runoff are lumped into one component, because in practice one cannot measure these components separately and one has to rely on physical models to determine each component. Furthermore, available models to sub-divide hydrographs into a number of different flow components are highly judgmental (Maidment, 1992). Therefore, in this study the observed flow is considered as generated by an effective rainfall obtained by multiplying the actual rainfall with an "adjustment factor" KN.

7.1.2. Effective Rainfall / Gain Estimation

In order to answer the second question, one needs to calculate the effective rainfall, from total areal average rainfall. In this regard, the importance of good measurement of rainfall cannot be denied (Hapuarachchi, 2007).The total areal average rainfall can be taken from the satellite rainfall estimates (SRE), or by taking the arithmetic or weighted average (for example, by Thiessen polygons) rainfall of the point gauges in the catchment, as was done for this study.

Once the total areal average rainfall has been calculated, the effective rainfall can be computed by different methods. According to the procedure of effective rainfall determination from the rainfall, rainfall-runoff models can be classified into two groups. For group-1 models the effective rainfall is calculated first and then it is used as input to a conceptual catchment model for runoff generation, whereas group-2 models (normally known as surface budget models) take the incident rainfall as input and calculate the infiltration and other losses as integral part of runoff estimation.

Effective rainfall is the component of the storm hyetograph which is not retained or infiltrated. For large scale catchments, where field measurements are not available, the soil conservation service (SCS) method, Horton equation or Green & Ampt method are frequently used to determine effective rainfall (Maidment, 1992) These methods are intended to balance the water volume, so that total runoff and total effective rainfall become identical. In the present study, a method similar to group-1 models is employed. An adjustment factor KN, which includes the runoff coefficient was used to compute the gain (the percentage of rainfall that is converted to

runoff) from the rainfall. For each month, an average is found by simply closing the mass balance between area-averaged rainfall and runoff, i.e. through the relation:

$$\sum_s Q_j\,(i) = \varphi \cdot KN_j(s) \cdot A_{0j} \cdot \sum_s P_j\,(i) \qquad (7.1)$$

where φ is a unit-conversion factor and $KN_j(s)$ is the adjustment factor for reach j and season s. Multiplying this coefficient KN with the total rainfall yields the total runoff volume, i.e. the coefficient KN is used to calculate rainfall excess available for runoff.. Initially a constant KN_j for the whole flood season was used. However, this assumption is not realistic enough. It was found that KN depends both on season and on reach. Studies on variability of catchment response to same rainfall input but of different catchment wetness conditions (Merz, 1997; Berthet et. al, 2009, Longobardi, et. al) have suggested the need for using variable or seasonal runoff coefficients.

The runoff coefficient is mostly related to the incident moisture condition and the infiltration capacity of a catchment. If the catchment is fully saturated, then the rainfall occurrences do not have to replenish the soil moisture deficit before turning into runoff. However, in cases of unsaturated and semi-saturated catchments, the soil moisture deficit has to be made up first. The saturation dynamics of different catchment is highly local, depending on the type of soil, its permeability, and the catchment's hydro-meteorological conditions. In the catchments of Mekong tributaries annual rainfall is distribution over a few flood months and no or very little rainfall occurs in other part of the year. Soil moisture increases from the onset of the flood season / rainy season and starts decreasing by the end of the flood season. Since the frequency of rainfall storms is high in the flood season with only few dry days in between the consecutive rainfall storms, the catchment remains in the saturation state for a long time, once it gains the saturation state. This behavior of the Mekong River is very typical of monsoon catchments, because in other rivers (not hit by typical monsoons) the span of dry days in between the subsequent rainfall storms may be high and hence each rainfall event first has to replenish the soil moisture to bring the catchment into the saturation stage. Consequently, in these catchments the antecedent rainfall occurrences must be considered in the computation of the runoff from the rainfall of each event. However, the behavior of the Mekong is different. Because it receives rainfall both from the south-west monsoon and typhoons, it shows a continuous increase in effective rainfall during the flood season as the catchment moves from unsaturated to saturated conditions. The catchment moves back to unsaturated condition once the flood season is over. This phenomenon is repeated every year and hence, a cyclic wave propagation like movement of monthly KN_j around the mean yearly KN_j (Fig.7.3) is produced. Therefore, in this study KN_j is computed empirically from observed rainfall runoff data. Four different methods, including assuming constant KN_j for whole flood season, have been tried.

7.1.2.1. Constant $KN_j(y)$

A constant value of KN for the whole season is obtained by calculating:

$$KN_j(y) = \frac{\sum_y Q_j(i)}{\varphi \cdot A_{0j} \cdot \sum_y P_j(i)} \tag{7.2}$$

$$\overline{KN_j(y)} = \sum_1^n \left[\frac{\sum_y Q_j(i)}{\varphi \cdot A_{0j} \cdot \sum_y P_j(i)} \right] \tag{7.3}$$

where y is an index for the days of the flood season of from May to October and $\overline{KN_j(y)}$ is the mean of $KN_j(y)$ of 1 to n annual records of (y days) flood seasons.

7.1.2.2. Mean monthly $KN_j(s)$

$$KN_j(s) = \frac{\sum_s Q_j(i)}{\varphi \cdot A_{0j} \cdot \sum_s P_j(i)} \tag{7.4}$$

where s = 1,2…6

$$\overline{KN_j(s)} = \frac{1}{n} \sum_1^n KN_j(s) \tag{7.5}$$

where $\overline{KN_j(s)}$ is the mean $KN_j(s)$ of 1 to n years for month s

7.1.2.3. KN_j Estimation by multi-regression

The assumption of mean monthly $\overline{KN_j(s)}$ occurrence in each flood season ignores the $KN_j(s)$ variability from dry to wet years and the effect of rainfall distribution among different months of

the flood season. Consequently, instead of using $\overline{KN_j(s)}$, the actual $KN_j(s)$ can be used in the analysis mode to provide the gain for rainfall runoff modeling. However, $KN_j(s)$ is not available in the forecasting mode because its computation requires the data of P and Q for whole month which is not available. Therefore, a method is developed to estimate $KN_j(s)$ by correlating it with the previous month $KN_j(s-1)$ with the help of linear regression:

$$\mathrm{KN_j(s+1) = \alpha_j(s) + Ç_j(s) \cdot KN_j(s) + \varepsilon_{cj}(s+1)} \qquad (7.6)$$

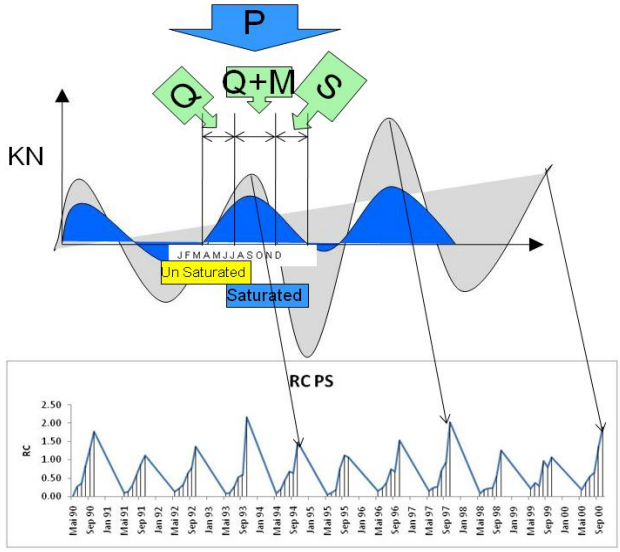


Fig. 7.3: Conceptualization of KN development with time, Top a) Variation of mean KN along flood season, bottom b) Variation of KN along the flood season for year 1990 to 2000

where coefficients α and Ç are empirical constants depending on the month and the area $A_{k,}$, and s denotes the month of the year. The error term $\varepsilon_{cj}(s)$ was further regressed on other factors, to reduce the error term. These other factors could be: current month's rainfall, $P_j(s)$, accumulative rainfall up to current month, $AP_j(s)$, current month discharge $Q_j(s)$, accumulative discharge up to current month $AQ_j(s)$, and KN_j of first month of flood season, i.e. $KN_j(in)$. This dependence of $KN_j(s)$ on number of different factors is exploited by using multiple linear regressions. This approach has the advantage, that the non-linearity of the dependency of KN_j on the season is covered by means of a piecewise linear relationship.

Let $\varepsilon_{cj}(s+1)$ be expressed through a linear regression equation

$$\varepsilon_{\mathrm{cj}}(\mathrm{s}+1)=\overline{\mathrm{KN}}\left\{\frac{\beta\cdot\overline{\mathrm{KN_{jC}}(\mathrm{s})}}{\overline{\mathrm{KN}}}\cdot\frac{\mathrm{KN_{jC}}(\mathrm{s})}{\overline{\mathrm{KN_{jC}}(\mathrm{s})}}+\frac{\gamma\cdot\overline{\mathrm{KN_{j}}(\mathrm{s})}}{\overline{\mathrm{KN}}}\cdot\frac{\mathrm{KN_{j}}(\mathrm{s})}{\overline{\mathrm{KN_{j}}(\mathrm{s})}}+\frac{\sigma\cdot\overline{\mathrm{P_{jc}}(\mathrm{s})}}{\overline{\mathrm{KN}}}\cdot\frac{\mathrm{P_{jc}}(\mathrm{s})}{\overline{\mathrm{P_{jc}}(\mathrm{s})}}+\frac{\tau\cdot\overline{\mathrm{p_{jc}}(\mathrm{s})}}{\overline{\mathrm{KN}}}\cdot\frac{\mathrm{P_{jc}}(\mathrm{s})}{\overline{\mathrm{P_{jc}}(\mathrm{s})}}+\frac{\upsilon\cdot\overline{\mathrm{Q_{jc}}(\mathrm{s})}}{\overline{\mathrm{KN}}}\cdot\frac{\mathrm{Q_{jc}}(\mathrm{s})}{\overline{\mathrm{Q_{jc}}(\mathrm{s})}}+\frac{\varphi\cdot\overline{\mathrm{Q_{jc}}(\mathrm{s})}}{\overline{\mathrm{KN}}}\cdot\frac{\mathrm{Q_{jc}}(\mathrm{s})}{\overline{\mathrm{Q_{jc}}(\mathrm{s})}}\right\} \quad (7.7)$$

with $\overline{KN}=$ average value of $\mathrm{KN_j(s+1)}$. Furthermore, the over bar denotes averages for every quantity for the location j and season s+1. Note that all quantities in the brackets have been made dimensionless by their respective averages, which are known from the original data analysis. Consequently, one can express Eq.7.8 by a more compact expression:

$$\varepsilon_{\mathrm{cj}}(\mathrm{s}+1)=\overline{\mathrm{KN}}\left\{\beta\cdot\frac{\mathrm{KN_{ic}}(\mathrm{s})}{\overline{\mathrm{KN_{ic}}}(\mathrm{s})}+\gamma\cdot\frac{\mathrm{KN_{init}}(\mathrm{s})}{\overline{\mathrm{KN_{init}}}(\mathrm{s})}+\sigma\cdot\frac{\mathrm{P_{i}}(\mathrm{s})}{\overline{\mathrm{P_{i}}}(\mathrm{s})}+\tau\cdot\frac{\mathrm{P_{ic}}(\mathrm{s})}{\overline{\mathrm{P_{ic}}}(\mathrm{s})}+\upsilon\cdot\frac{\mathrm{Q_{i}}(\mathrm{s})}{\overline{\mathrm{Q_{i}}}(\mathrm{s})}+\varphi\cdot\frac{\mathrm{Q_{ic}}(\mathrm{s})}{\overline{\mathrm{Q_{ic}}}(\mathrm{s})}\right\}+\varepsilon_{\mathrm{ccj}}(\mathrm{s}) \quad (7.8)$$

7.1.2.4. *KN_j Estimation by moving average regression*

There are two problems in using pre-estimated monthly approximates $KN_j(s)$: first, the variation of $KN_j(s)$ within the month is not accounted for, secondly, the potential of the most recent available data of KN_j, is not exploited in estimating current and future KN_j value. Therefore, the average of last n days Q_j to P_j ratio is used as approximation of current day $KN_j(d)$. In this way, each day (d) gets a separate value of KN_j i.e.:

$$KN_j(d)=\frac{\frac{1}{n}\sum_{k=1}^{n}Q_j(d-n+k)}{\varphi\cdot A_{0j}\cdot\frac{1}{n}\sum_{k=1}^{n}P_j(d-n+k)} \quad (7.9)$$

where n is the number of days over which average is taken that range from current day $(d-n+n)$ to day $(d-n+1)$ in the past.

In order to estimate KN_j of future days, when data of P and Q are not available, a first order Markov Chain model can be tested. However, instead of using a Markov chain on a fixed monthly period, a moving average window of n days is used in order to estimate future days $KN_j(d+t)$.

$$KN_j(d+t) = \frac{\frac{1}{n}\sum_{k=1}^{n} Q_j(d-n+k+t)}{\varphi \cdot A_{0j} \cdot \frac{1}{n}\sum_{k=1}^{n} P_j(d-n+k+t)} \tag{7.10}$$

where t = 1,2…5 denotes lead time from the current day d The relation between $KN_j(d) and\ KN_j(d+t)$ can be established by using the basic Markov chain method:

$$KN_j(d+t) = \theta_j(d).KN_j(d) \tag{7.11}$$

$$KN_j(d+t) - \overline{KN_j(d+t)} = \theta_j(d).[\,KN_j(d) - \overline{KN_j(d)}] \tag{7.12}$$

$$\theta_j(d) = \frac{\left[KN_j(d+) - \overline{KN_j(d+t)}\right]}{[\,KN_j(d) - \overline{KN_j(d)}]} \tag{7.13}$$

$$KN_j(d+t) = \overline{KN_j(d+t)} + \theta_j(d).\left[\,KN_j(d) - \overline{KN_j(d)}\right] \tag{7.14}$$

7.1.3. Conversion of effective rainfall into runoff

For answering the third question, the distribution of effective rainfall over time as runoff from a sub-catchment has to be considered. It can be assumed that $DQ_{j,j+1}(i)$ is only due to the input

from rainfall in tributaries. The lumping of each sub-basin into one area with uniform rainfall and uniform basin characteristics allows one to use the unit hydrograph approach. Then observed effective rainfall and $DQ_{j,j+1}(i)$ can be used in a simple unit hydrograph analysis. This conceptualization assumes a linear system response.

The ordinates of unit hydrographs can be obtained in a number of different ways as explained in chapter-3. One way is to estimate a unit hydrograph ordinate by fitting any distribution function, which starts at zero and has a long tail with an asymptote of zero, as for example the gamma distribution. A unit hydrograph u(t) in form of a gamma distribution (Nash Cascade) is:

$$u(t, n) = \frac{k}{\Gamma(n)} \cdot (k \cdot t)^{n-1} \cdot e^{-kt} \qquad (7.15)$$

k and n are parameters to be obtained from calibration, $\Gamma(n)$ is the gamma function of n, and t = $(i - 0.5)\Delta t$. The shape of the gamma distribution is controlled by its parameters, i.e. n and k. In the analysis mode, these parameters have to be determined empirically to generate flow values close to the observed flows. The parameterization by trying different n and k parameter combinations could be based on different methods, for example the methods of least squares, maximum likelihood, or method of moments.

Prior to trying different n and k values to determine least square errors between observed and generated flows by a trial and error approach, it is time efficient to establish a possible range of n and k values. The product (n-1)*k is the time to the maximum of the gamma distribution. Putting this time as equal to the time lag T between rainfall and runoff for maximum cross correlation, yields a probable range of values of n and k. Although cross correlation analyses for different annual flood seasons yield different time lags, the approximate range of T can be obtained by this analysis. Possible combinations of n and k can then be determined that match the equality:

$$(n - 1).k = T \qquad (7.16)$$

n = number of reservoirs

k = Storage constant

T = time lag

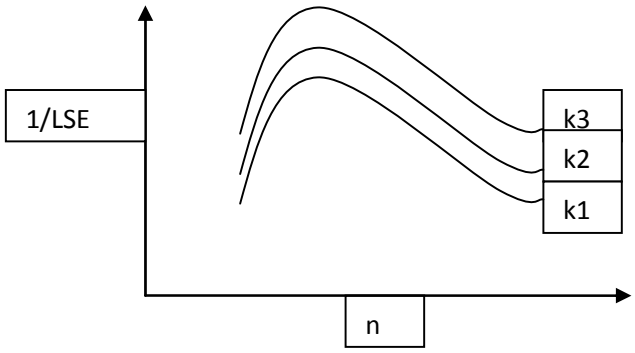


Fig7.4: Optimization of Gamma distribution functions; n and k

The n and k values are optimized by LSE. The manual way of optimization is to plot k as function of n for different k values, as shown in Fig.7.4. In each curve of this Fig. k is kept constant and n is varied along x-axis. On the y-axis is 1/LSE, where LSE is obtained by taking the squared differences of observed and calculated discharges. The optimum combination is a pair n and k which gives minimum LSE.

The fitness of observed to model discharges may be tested by using the Nash and Sutcliffe (1970) coefficient NSK. However, in this study the coefficient of determination RSQ is used as better measure to gauge the accuracy of the model. Reason for using RSQ in gauging the efficiency of the rainfall runoff model is because in forecasting the consideration of correlation between modeled to observed discharge is more important than one to one correlation of modeled to observed discharge as reflected in NSK. The RSQ indicates the potential of the model to produce discharges close to observed ones by providing the regression between modeled and observed discharges. The coefficient of determination shows the strength of linear relation between x and y as presented in Eq.7.17:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} \tag{7.17}$$

$$RSQ = r^2 \tag{7.18}$$

To sum up, LSE should be used to select optimum n and k, whereas RSQ should to used to gauge the efficiency of model which is based on these n and k parameters. Automatic optimization algorithms are available, and can be used instead of graphical plotting as shown in Fig.7.3, to find the best n and k parameters on the basis of LSE. The range of n and k is fixed

through Eq.7.18 in these optimization algorithms in order to obtain the optimum n and k within pre-selected range.

Two different types of rainfall input can be used for n and k parameterization, i.e. event based and continuous inputs. In the event based approach, parameterization is done separately for different storms. The event based approach gives different n and k values, one optimum for each storm event. Different n and k for different storms are caused by the nature of storms, i.e. by the differences in spatio-temporal distribution of rainfall and storm movement. In such a case, an average over all gamma distribution curves can be selected as optimum unit hydrograph for future flood events. The second approach is to take the rainfall hyetograph and the discharge time series of a whole season, which includes a number of different storms, and fit the gamma distribution with a single optimum pair of n and k. In this study, for parameterization of the rainfall runoff model development for forecasting, taking continuous data was preferred over event based data because it averages over all storms. Also, no base flow abstraction is needed.

Once the coefficient KN and the ordinates of unit hydrograph are estimated in the analysis mode, the standard unit hydrograph equation can be used to compute discharges from the rainfall data as:

$$DQ_j(i) = KN_j(s) \cdot A_{0j} \sum_{\omega=1}^{i} P_j(\omega) \cdot u_j(i-\omega) \qquad (7.19)$$

In this equation, ω is the time coordinate for the rainfall, $P_j(\omega)$ is the rainfall during time interval Δt at time $\omega \cdot \Delta t$, and KN basically is the runoff coefficient, but it also yields an empirical compensation for non-uniformity of the rainfall distribution. The rainfall data $P_j(i)$ of each basin with area A_{0j} and discharge increases $DQ_j(i)$ between adjacent stages are needed both in the forecast and in the analysis mode.

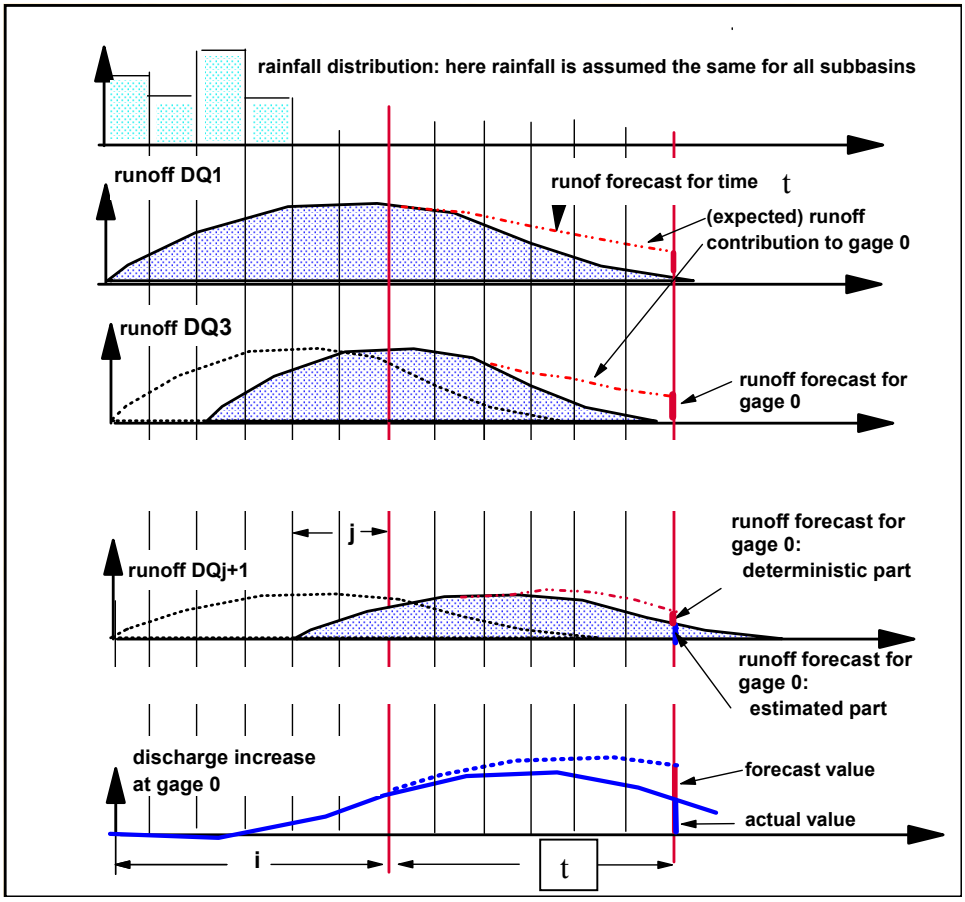


Fig. 7.5: Unit hydrograph approach to forecasting: schematic presentation

The application of a unit hydrograph model in the forecasting mode is schematically shown in Fig.7.5. For implementation of this schematic presentation, Eq.7.21 has to be combined with Eq. 7.22. Because part of the rainfall of today will be runoff in the future, the forecast DQF has two components DQ1 and DQ2:

$$DQF_j(i+t-j) = DQ1_j(i+t-j) + DQ2_j(i+t-j) \tag{7.20}$$

$$DQ1_j(i+t-j) = KN_j(s) \cdot A_{0j} \cdot \sum_{\omega=0}^{i} P_j(\omega) \cdot u_j(i-\omega) \tag{7.21}$$

All quantities for DQ1 for the forecast are known at time t except $KN_j(s)$. In order to calculate DQ1 exactly, KN has to be estimated in the forecast mode. Unlike analysis mode, $KN_j(s)$ cannot be found by simply closing the mass balance between area-averaged rainfall and runoff through Eq.7.1 because both the rainfall and runoff of the running month (s) are unknown for the whole month unless we are sitting at the end of month. Consequently, $KN_j(s)$ has to be estimated first by any one of the mentioned approaches, i.e. constant KN, average monthly KN, multi-linear regression, or moving average.

Suitability of these methods can be tested by checking the efficiency of the adapted model in generating the discharges. Therefore, further explanations of the applicability of these methods are given in the application section of this approach to the Mekong.

In determining DQ2, both the future rainfall and $KN_j(s)$ are not known and have to be forecasted in order to determine DQ2:

$$DQ2_j(i+t-j) = KN_j(s) \cdot A_{0j} \sum_{\omega=i}^{i+t-j} PF_j(\omega) \cdot u_j(i-\omega) \tag{7.22}$$

In the absence of a weather forecast model which produces future rainfall forecasts, one has to make assumptions of the future rainfall occurrences in the forecasting period (i+t). Very simple future rainfall assumptions could be:

$$PF(i+\omega) = P(\omega) \tag{7.23}$$

$$PF(i+\omega) = \frac{1}{k} \sum_{\omega=i}^{i-k} P_j(\omega) \tag{7.24}$$

Both approaches have been used in this study.

7.2. Application of the conceptual model to the Mekong River

7.2.1. Input Data

Areal average daily rainfall and the time series of discharge increases $DQ_j(i)$ between adjacent stages, as described in Chapter 6, are used as input to develop the model and to test it. The data time series is divided into two parts, i.e. 1990 to 2000 and 2001 to 2005 for simulation and validation respectively.

The areal averaged rainfall $P_j(i)$ was calculated from the recorded rainfalls of the selected gauging stations (Table.7.1) for each of the six sub-basins. The location of ground rainfall gauges used for average rainfall determination is shown in Fig.4.1. In order to estimate areas associated with each rainfall station, the method of Thiessen polygon was tried for determination of weighted average rainfall, but it was not found useful because of the poorly organized location of rainfall gauges. In most of the cases, the rainfall gauges were found to be in clusters at one location in each sub-basin, leaving only one rainfall gauge, to take more than 60 % of total weight in the weighted average rainfall, hence introducing a strong bias in the average rainfall determination by Thiessen polygons. Trial and error calculations showed that the best results were found by using the simple area-averaged rainfall for further analysis. Consequently, the arithmetic average of rainfall of all the gauges of each sub-basin j is used for areal rainfall $P_j(i)$ determination. SRE (satellite rainfall estimates) were not used because their quality was not validated by ground based measurements and became one of the reasons for low efficiency of URBS model forecast as mentioned by Malone in the report on sensitivity analysis of URBS Model application on Mekong (Malone, 2009). The area of each sub-basin as needed in Eq.7.1 is calculated with the help of GIS maps and given in Table.7.1.

Tab. 7.1: Rain gauges of each sub-basin j

Reach	Area (Sq. Km)	Rain Gages
CL	59000	Muong Namtha
		Phongsaly
		Muong Ngoy
		Oudomxay
		Chiang Kham
		Xieng Ngeun
LV	24650	Xieng Ngeun
		Sayaboury
VN	50016	Ban Nape
		Muong May
		Muong Kao
		Thabok
		Paksane
		Bung Khan
		Ban Nasone
		N. Phanom
		Signo
NM	11006	Mukdahan
		Signo
		Thakek
		N. Phanom
MP	37993	Mukdahan
		Savannakhet
		Seno
		Ban Kengkok
		Muong Tchepon
		Khongsedone
		Saravan
		Laongam
		Pakse
PS	66757	Pakse
		Mounlapamok
		Soukhouma
		Muang Champasack
		Attopeu
		Nonghine

The unit hydrograph approach based on the gamma distribution, requires in addition to rainfall data three parameters for each section between adjacent gauges: the coefficient KN and the Nash cascade parameters k and n. For a data based approach, these have to been determined empirically from the data for each sub-basin j separately.

7.2.2. Effective Rainfall / Gain Estimation

7.2.2.1. Gain by Constant KN

The adjustment factor KN as a percentage of total rainfall was obtained by using Eq.7.2 and 7.3 with the help of 10 years data from 1990 to 2000. Results are given in Table.7.2. They show the variability of KN_j in different sub-basins with a maximum, i.e. 1.39 in NM and a minimum i.e. 0.46, in CL. The variability in KN among different sub basins is partly caused by different geographical conditions and soil types. Another cause of this KN_j variability is the size of the area used for calculation in each sub-basin. The size of the area used in Eq.7.1 is not the actual size of the contributing area of sub-basin j, because it is only the portion of the area on the left bank of the contributing sub-basin. The reason of using only the area of left bank tributaries is because it was assumed that only the left bank tributaries' rainfall contributed significantly to the inflow. But in actual practice, there is the portion of inflow from the right bank tributaries as well. Ignoring the contribution of runoff from the right bank tributaries makes the KN_j value larger than if it were a runoff coefficient, especially in NM, where it exceeded one.

The constant values of KN_j determined in the analysis mode for the whole season are tabulated in table 7.2. They were initially used in Eq.7.19 to produce lateral inflows. However, the resultant comparative plot of observed versus produced discharges show over-estimation in the initial part of the flood season and underestimation late in the flood season. These results lead to the consideration of seasonality of KNj

Tab. 7.2: Constant KN coefficient for each sub-basin j

Reach	CL	LV	VN	NM	MP	PS
P	2484	891	4281	720	2137	4748
Q	1135	554	2382	998	1987	3648
KN	0.46	0.62	0.56	1.39	0.93	0.77

7.2.2.2. Results with Monthly KN

Typical empirically calculated KN_j by using Eq.7.4 for each sub-basin as functions of the month in flood season are shown in Figs.7.6a and 7.6b. It can be seen in Figs.7.6a and 7.6b that there is

a significant trend, with very small coefficients at the beginning of the season, which increase with progress in time until reaching a maximum at the end of the season.

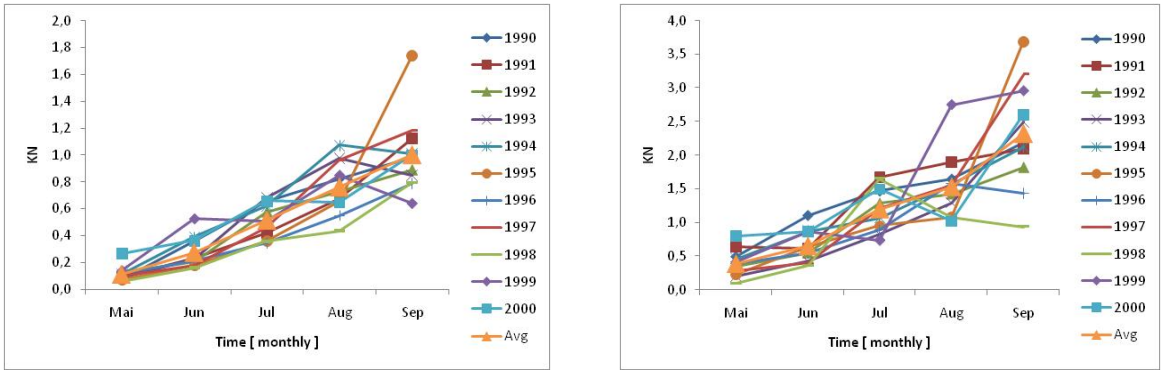


Fig. 7.6a: Typical adjustment factors for VN (left) and NM (right) for the years 1991 to 2000

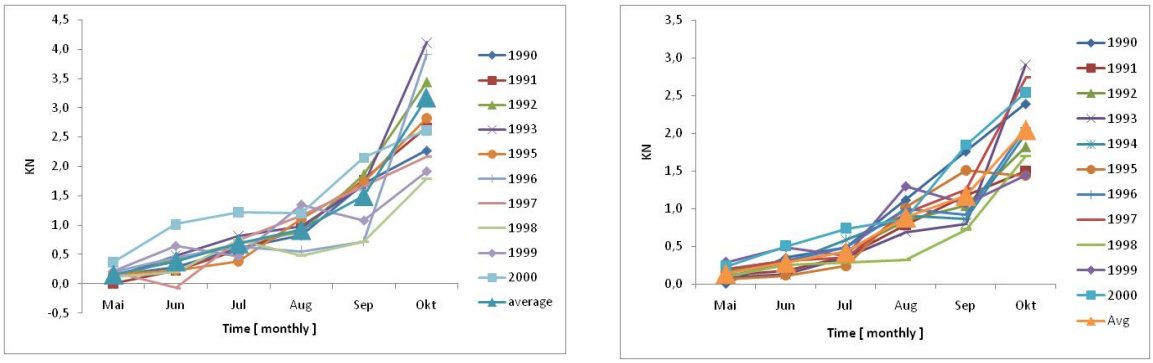


Fig. 7.6b: Typical adjustment factors for MP (left) and PS (right) for the years 1991 to 2000

Because the adjustment factor basically is a runoff coefficient (with a theoretical maximum of 1), it can be surmised that the trend reflects the special seasonal hydrologic conditions of the Mekong region. At the beginning of the season, before the advent of the monsoon, the soil is dried out. Further into the season it is saturated gradually so that KN_j increases and reaches a maximum at the end of the season. Note that the large increases at the end of the season, which are observed for some of the years, are due to the fact that low discharges are divided by small numbers of rainfall. Apart from these anomalies, the average trend is clearly significant.

The box plot of monthly KN illustrates further the behavior of KN_j variability within the flood season and also from one flood season to another as shown in Fig.7.7. In the analysis of data from the Mekong lateral tributaries, it was found that $KN_j(s)$ strongly depends on the season of the year. A surprisingly well fitting relation was found to exist between accumulative rainfall and

accumulative discharge for each reach between gauging stations – an indication of the well structured meteorological pattern caused by the regularity of the South West Monsoon (Fig.7.7).

Monthly KN plotting

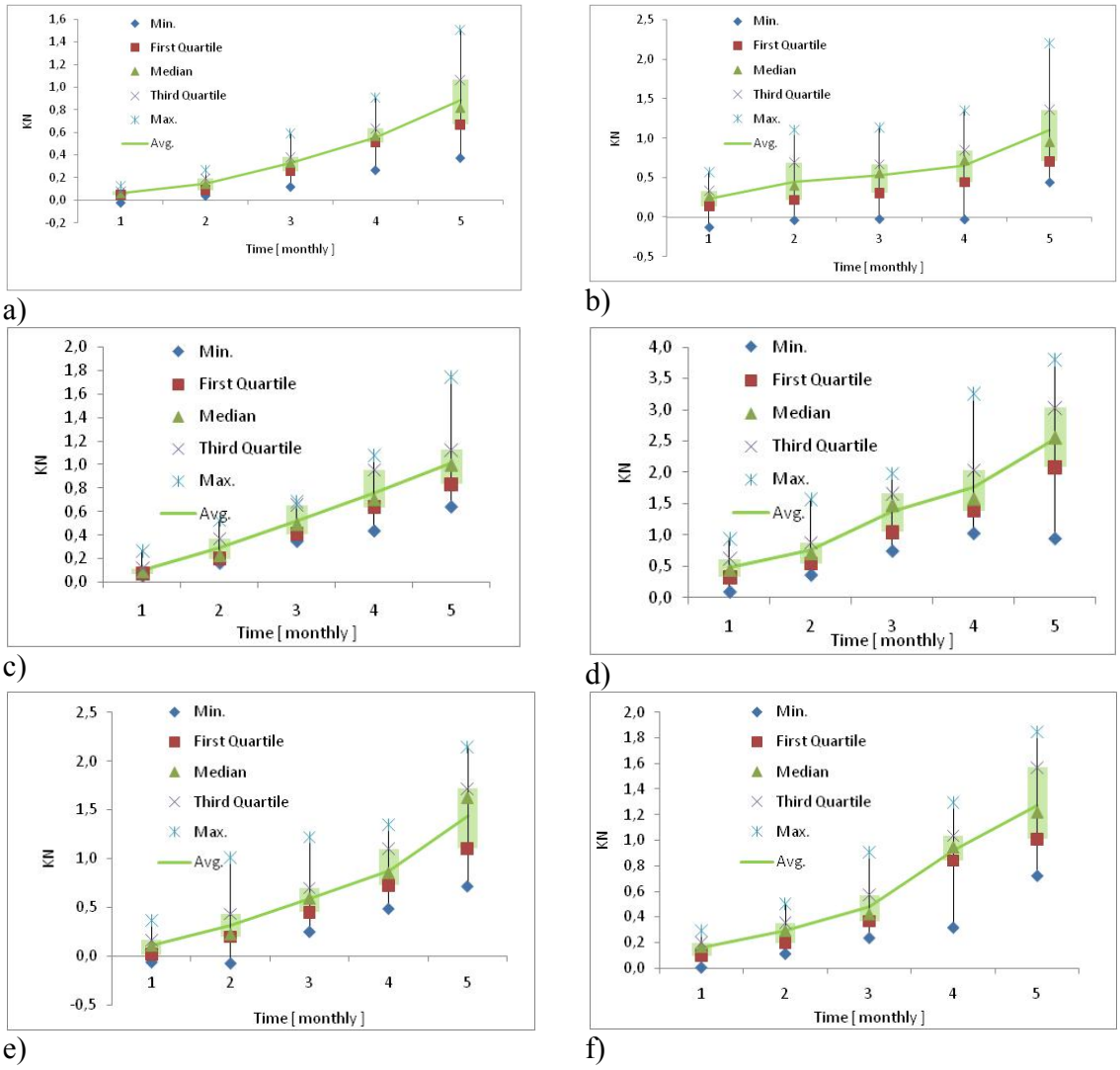
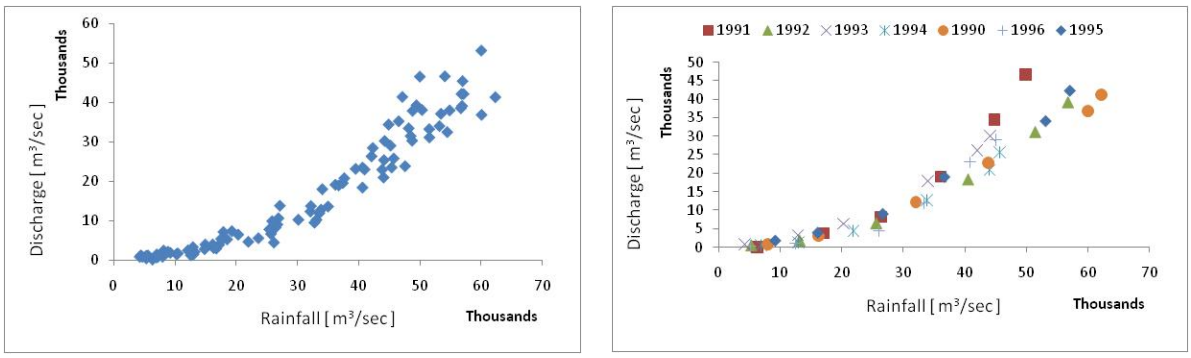


Fig. 7.7: KN box plot for each sub-basin, a) CL, b) LV, c) VN, d) NM, e) MP, f) PS

Monthly values of KN_j were noted to increase from first month of flood season, i.e. June to last flood month, i.e. October, as shown in Fig. 7.7. This steady rise of KN_j, observed in almost every

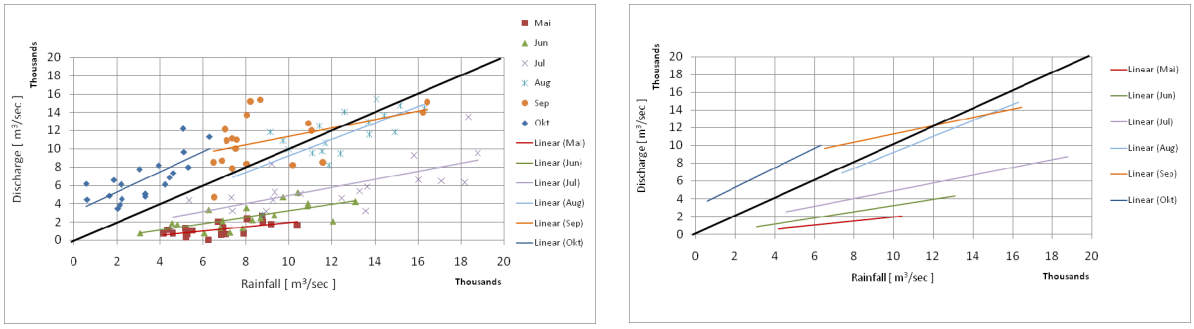
year, was a result of catchment wetness condition and soil permeability. First improvement in the assumption of constant KN_j was, therefore to use mean monthly$\overline{KN_j(s)}$ in the analysis mode.



a) b)

Fig. 7.8: Scatter plot of accumulative rainfall and accumulative runoff for PS a) collective scatter plot for flood season 1990 to 2005, b) scatter plot for flood season 1990 to 1995

The results of using mean monthly$\overline{KN_j(s)}$ to generate discharges removed the problem of over and underestimation of discharges in the early and late flood season. It is because the relation of rainfall to runoff changes along the flood season, as indicated in Fig.7.9, that the trend lines have different slopes in the rainfall runoff scatter plot of each month. An interesting observation is that the slopes of these trend lines increase from May up to October. The scatter plot of October represents the falling limb and consists to a large part of base flow, and therefore represents data points of old water flow and only very little rainfall input.



a) b)

Fig. 7.9: Scatter plot of rainfall and runoff for PS a) scatter plot for flood season 1990 to 2005, b) trend line representation of the scatter plot for flood season 1990 to 2005

7.2.2.3. Gain by estimated monthly $KN_j(s)$ by multi-regression

The variability $KN_j(s)$ from one flood season to another can be seen in the spread of box plot in Fig.7.6. In most of the cases, the spread is wide and thus reducing the efficiency of $\overline{KN_j(s)}$ to represent the actual$KN_j(s)$. In order to improve this situation, multi-regression is used to estimate the $KN_j(s)$ separately for each month and for each flood season, as discussed in section 7.2. The data of 1990 to 2000 are used to build a regression model and to determine the regression weights of Eq.7.6 to 7.9. This regression model is then validated on data for 2001 to 2005.
The regression parameters of Eq. 7.6 and Eq.7.8 are given in Table.7.3 to 7.6

Tab. 7.3: KNs+1 computation coefficient for VN

Month	α	β	γ	σ	τ	υ	Φ	Ç
Jun	1,11							0,11
				0,00		0,00		
Jul	0,70							0,25
		-2,81	2,73	-3,35	3,08	9,03	-8,63	
Aug	0,73							0,08
		0,16	-0,09	1,24	-1,30	-0,91	0,89	
Sep by jul	1,17							0,41
		-1,81	-1,31	0,85	-1,84	2,89	1,07	
Oct	-0,09							2,52
		-1,09	-2,31	9,97	-12,42	5,50	0,08	

Tab. 7.4: KNs+1 computation coefficient for NM

Month	α	β	γ	σ	τ	υ	φ	Ç
Jun	0,64							0,08
				0,00		0,00		
Jul	2,51							-0,03
		-0,02	2,45	-2,15	2,89	3,01	-6,20	
Aug	1,11							0,04
		-2,52	0,59	1,21	-1,28	-0,57	2,59	
Sep	1,06							0,07
		0,27	-2,04	0,06	-0,73	1,58	0,86	
Oct	3,64							-0,37
		16,60	1,64	4,07	-3,65	-3,96	-14,65	

Tab. 7.5: KNs+1 computation coefficient for MP reach

Month	α	β	γ	σ	τ	υ	φ	Ç
Jun	1,08							0,17
				0,00		0,00		
Jul	0,77							0,36
		2,76	-2,38	5,45	-5,75	-9,21	9,25	
Aug	0,32							0,75
		0,01	-0,14	2,23	-2,18	-1,22	1,32	
Sep	0,53							1,08
		-0,04	-0,66	1,15	-1,51	-0,96	2,00	
Oct	0,55							2,69
		-1,23	0,77	-0,77	1,70	5,83	-6,25	

Tab. 7.6: KNs+1 computation coefficient for PS reach

Month	α	β	γ	σ	τ	υ	φ	Ç
Jun	1,29							0,05
				0,00		0,00		
Jul	1,01							0,13
		-2,19	-0,28	-0,10	1,19	3,02	-1,76	
Aug	0,44							0,22
		-0,13	0,38	0,85	-1,18	-0,12	0,23	
Sep by jun	-0,03							0,93
		-2,03	-2,60	8,94	-8,51	-7,54	11,34	
Oct by jun	-4,93							3,01
		-67,94	49,01	-46,62	44,52	54,41	-35,08	

7.2.2.4. Gain by estimated monthly $KN_j(s)$ by moving average

The performance of the multi-regression model to predict $KN_j(s)$ was found to be not quite as good in the validation mode, as during analysis. Therefore the moving average KN_j over the last 30 days has been used as current day $KN_j(d)$ estimate by using Eq.7.13. The results of

simulation by Eq.7.15 show significant improvement in the rainfall runoff model performance in comparison to the other three KN_j estimation approaches - due to the obvious reason of using most recent information, as well as accounting for the variability of KN_j within the month.

In the analysis mode, the seasonal and daily effective rainfall is computed by each of the four methods explained in section-7.1.2.4. However, KN values obtained by using moving averages of last 30 days is preferred by virtue of its performance and therefore is adopted in the forecasting mode as standard. As an example the forecasting results of KN by four different approaches for reach RS in the year 2000 are shown in Fig.7.10. This Fig. clearly indicates that KN estimation by moving average is performing better than any of the other three approaches.

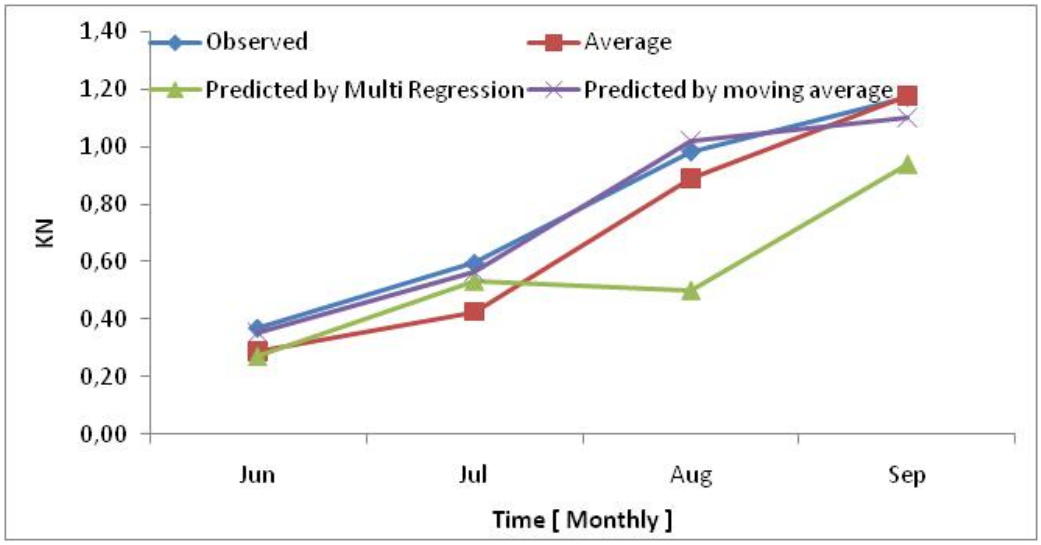


Fig. 7.10: KN estimation for reach PS of year 2000

However in the forecasting mode, the course of $KN_j(d+t)$ for the future t days has to be estimated according to Eq.7.14, where the parameters of Eq.7.14 are obtained from Eq.7.13. This approach would require separate parameters for each$(d+t)$, therefore a total of 153 parameter for 153 days d, which are however, actually not separate parameters but the ratio of $KN_j(d+t)$ difference from its mean $\overline{KN_j(d+t)}$ to $KN_j(d)$ difference from its mean $\overline{KN_j(d)}$.

7.2.3. Conversion of effective rainfall into runoff

In the analysis mode, Eq.7.17 is used on known input – output results from the situation to be studied. This analysis yields coefficients n and k of the gamma distribution.

7.2.3.1. Determination of the system parameters n and k

The parameters n and k describing the gamma distribution were obtained by trial and error optimization, yielding a stretching of the rainfall response over about 15 days with a peak after about 3 to 5 days, so that this approach fits the data fairly well for longer time forecasts.

Tab. 7.8: Estimation of n and k parameters for event based approach

Years	n	k	(n-1)*k
1990	3.4	2.1	5
1991	2.8	2.1	4
1992	7.1	1.8	11
1993	2.5	2.1	3
1995	2.4	2.1	3
1996	5.1	1.6	6
1997	4.0	1.3	4
1998	2.0	2.1	2
1999	14.3	0.5	7
2000	3.6	2.1	5
		Average	**5**

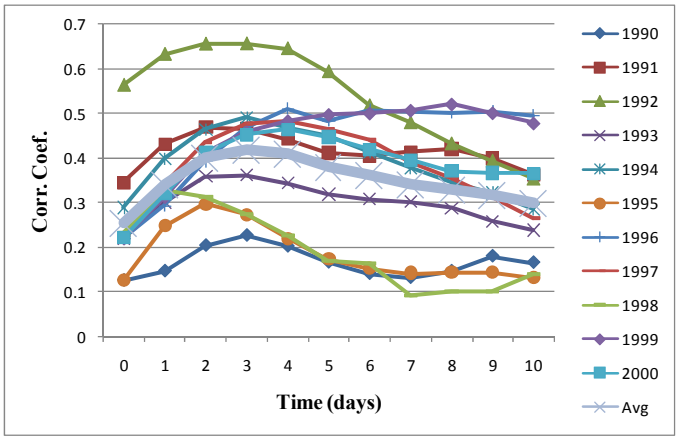


Fig. 7.11: Rainfall to runoff cross correlation coefficients for PS sub-basin

The cross correlation function of average rainfall $P_{j,}$ (i) and lateral discharges $DP_{j,j-1}$ (i) gave the time lag T between rainfall to runoff peak. As an example, the cross correlation function for the sub-catchment contributing to reach PS are presented in Fig. 7.11. It is clear from this Fig. that the maximum correlation between rainfalls and runoff occurs in the range of time lags between 3 to 4 days.

In the event based approach, runoff hydrographs for each of the historical storm events are fitted by means of a least square estimation of n and k for each storm. As an example, the n and k parameters for the best fit unit hydrograph from different storms are given in Table.7.8 for the sub-basin PS. One storm is selected per year. The value of k is fairly constant, i.e. 2.1 for most events, but the values of n show high variability.

In the season based approach, 11 years data of lateral rainfall and respective runoff (1990 to 2000) of each reach is taken together to be fitted with different n and k parameter combinations. The optimum n and k parameter combination is selected from these curves for each sub-basin j. But in the selection of n and k, the LSE optimization is applied in five different ways. In the first case, LSE optimization is applied by using original $P_{j,}$(i) and $DQ_{j,j-1}$ (i) data and accounting for all errors in the summation of LSE. In the second step, LSE are calculated for the observed and

modeled runoff $DQ_{j,j-I}$ (i) above certain thresholds. The threshold is fixed in order to obtain the n and k parameter, which are more efficient in high flows. In this case $\overline{DQ_{J,J-I}(i)}$ is taken as this threshold. In the third case, LSE are computed by taking the squared differences of observed and updated forms of modeled runoff from each n and k combination. Instead of taking the modeled runoff, the predicted errors are calculated and added to the modeled runoff to yield the updated runoff. LSE is obtained by taking squared difference of observed and updated modeled runoff. In the fourth case, all the sub-basins j are optimized simultaneously by applying the LSE function on the discharge difference (observed discharge $Q_j(i)$ minus discharge Q_jF (i) calculated from adding the modeled discharges $DQ_{j,j-I}$ (i) to upstream discharges with appropriate time lags. LSE of all the gauges $Q_j(i)$ are added together into a single summation, which is minimized by trial and error with the help of different n and k combinations of each sub-basin j. In the fifth case, the fourth step is repeated, but the forecasted rainfall of future days is used. In the third and fifth approach, an additional correction factor is used in order to remove the volumetric error of rainfall and runoff for each sub-basin j. This factor is changed with different combinations of n and k in LSE optimization. The optimized n, k and volumetric correction factor, FKN are presented in Table-7.9 for each of the five mentioned cases.

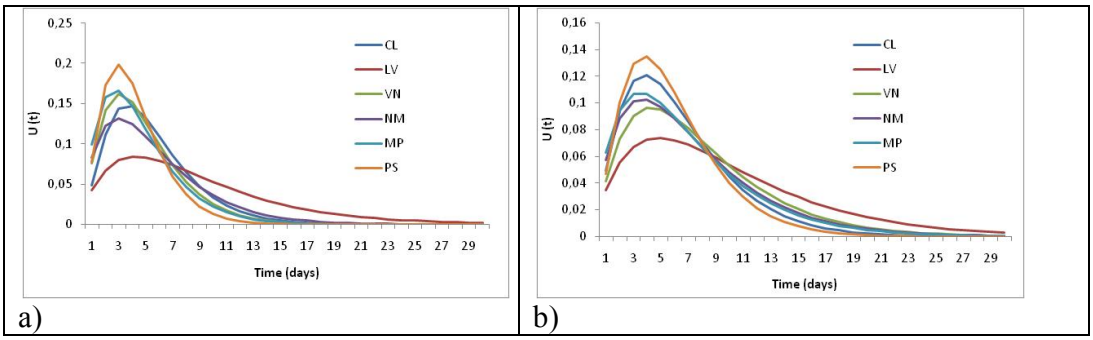


Fig. 7.12: a) Unit hydrograph ordinates as obtained by first case of n and k optimization of each sub-basin j, a) Unit hydrograph ordinates as obtained by the second case of n and k optimization for each sub-basin j

Once the KN and the optimum n and k values (as illustrated for Pakse-Stung Treng in Section 7.2.3.1 have been found for all reaches the ordinates of the runoff hydrographs can be determined, and for each reach lateral inflows can be calculated.

Tab7.9: Estimation of n and k parameters by 5 different cases

	n1	k1	FKN1
CL	2,49	2,55	1,00
LV	1,96	5,07	1,00
VN	2,23	3,48	1,00
NM	2,03	3,52	1,00
MP	2,03	3,38	1,00
PS	2,79	2,11	1,00
	n2	k2	FKN2
CL	2,97	1,83	1,00
LV	1,98	4,38	1,00
VN	2,71	1,79	1,00
NM	2,11	2,66	1,00
MP	2,44	1,86	1,00
PS	3,26	1,28	1,00
	n3	k3	FKN3
CL	2,16	3,72	1,01
LV	1,68	7,14	0,94
VN	2,51	2,67	0,75
NM	2,34	2,56	0,86
MP	2,66	1,81	0,80
PS	2,13	3,12	1,14
	n4	k4	FKN4
CL	8,83	0,34	0,42
LV	1,85	8,09	1,63
VN	2,76	2,64	0,91
NM	2,55	3,17	1,53
MP	4,80	0,69	0,65
PS	2,35	3,04	1,37
	n5	k5	FKN5
CL	1,43	12,32	1,28
LV	1,61	20,22	2,36
VN	1,57	8,00	0,71
NM	1,63	7,25	1,24
MP	0,99	16,38	1,09
PS	1,75	7,28	1,02

An additional step was to try in one of the reaches, i.e. Pakse-Stung Treng a double cascade with two n and k parameter sets to check possible improvement over a single cascade in the conversion of rainfall to discharge. However, this analysis yielded hardly any improvement

compared to single cascade based discharge estimation (Fig.7.13). Thus, a single Nash cascade is adopted for computation of runoff from the rainfall in each sub-basin j of Mekong

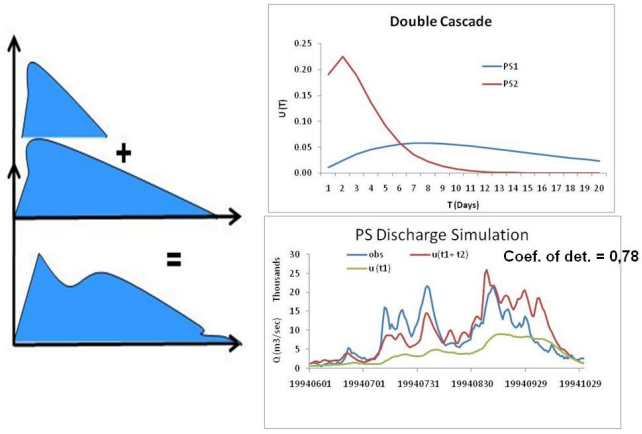


Fig. 77.13: a) Conceptual representation of double cascade, b) Unit hydrograph ordinates as dictated by optimized n and k parameters for PS, c) PS discharge simulation on the basis of double cascade

7.2.4. Rainfall runoff model usage for flood forecast

The estimated system functions can be used in the forecast mode in the rainfall runoff model for flood forecasts, provided that the rainfall input is available for the last 15 to 20 days and for the lead time of future forecasts.

In order to test the efficiency of this rainfall runoff model to produce flood forecast, it was used for three different kinds of assumed future rainfall inputs. In the first case it was assumed that both the future rainfall and the KN-values are known. This case gives the overall efficiency of the model: when the forecasts for future rainfall and KN- values are available as in the analysis mode. In the second case it was assumed that the future rainfall is known, but the KN-value is unknown, and therefore one of the KN forecasts is used. This case presents the loss of efficiency due to replacing the known KN with the predicted KN. As the model was tested on the historical period of 1990 to 2005, therefore the information on rainfall and KN-values was available for the forecasting period in a hind-cast mode. In the third case, both the predicted values of rainfall and of KN are used as input to the model. This case reflects the loss of model efficiency due to imperfect future KN and rainfall information.

7.2.4.1. Known rainfall and known KN assumption

$DQF_j(i+t-j)$ is calculated by using Eq.7.20 to Eq.7.22 and compared with the $DQ_j(i+t-j)$ obtained from the observed discharges. The performance of the rainfall runoff model in generating $DQF_j(i+t-j)$ in each sub-basin j is evaluated by RSQ and PI. Then the errors of these modeled $DQF_j(i+t-j)$ values are updated by using Eq.7.25. The regression parameters of Eq.7.25 are calculated by using 1990 to 2000 data. These parameters are given in Table.7.10.

$$\varepsilon_{fj0}(i+t) = \beta j(i+t).\varepsilon_{f0}(i+t-t) + \varphi j(i+t) + \varepsilon_{f1}(i+t) \qquad (7.25)$$

The updated values of $DQF_j(i+t-j)$ are then used in Eq.6.34 to produce flood forecasts for 1 to 5 days at L,V,N, M, P and S.

7.1.2.1.1. Result and Discussion

First case of n and k:

The n and k parameters obtained by applying LSE optimization in four different ways are used to produce $DQ_j(i+t)$ and $Q_j(i+t)$. The accuracy of $DQF_j(i+t)$ and $QF_j(i+t)$ varies from case to case, it is however improved by updating in each case. The accuracy, by using first case and second case of n and k optimization, of updated $DQ_j(i+t)$ in terms of RSQ for each sub-basin j is given in Fig.7.14. The improvement in $DQ_j(i+t-j)$ by updating can be seen, for example, in the case of PS in Fig.7.15. The comparison of observed and modeled PS discharges with and without updating shows significant improvements achieved by updating. However, these are results of updating with one day lead time. The efficiency of updated discharges decreases with increase in lead time.

Tab. 7.10: Regression parameters of updating equations for each sub-basin j with lead time 1 to 5 days

	Perfect P and KN with case 1 of n and k		Perfect P and fore. KN With case 1 of n and k		Fore. P and fore. KN With case 1 of n and k		Fore. P and fore. KN With case 5 of n and k	
Sub-basin	$\beta j(i+t)$	$\varphi j(i+t)$	$\beta j(i+t)$	$\varphi j(i+t)$	$\beta j(i+t)$	$\varphi j(i+t)$	$\beta j(i+t)$	$\varphi j(i+t)$
DQj+9								
i+1	0.92	41	0.92	37	0.92	40	0.93	27
i+2	0.80	98	0.81	85	0.78	106	0.84	64
i+3	0.70	156	0.72	135	0.58	203	0.74	103
i+4	0.61	203	0.63	175	0.42	287	0.64	143
i+5	0.53	244	0.56	210	0.30	344	0.56	177
DQj+7								
i+1	0.82	47	0.83	33	0.83	34	0.87	-16
i+2	0.69	78	0.72	55	0.70	59	0.78	-27
i+3	0.58	108	0.61	77	0.54	90	0.69	-38
i+4	0.49	131	0.52	93	0.39	120	0.59	-49
i+5	0.42	149	0.45	107	0.25	150	0.49	-61
DQj+5								
i+1	0.95	52	0.96	46	0.95	48	0.96	93
i+2	0.88	144	0.88	128	0.87	141	0.90	246
i+3	0.79	246	0.79	218	0.73	283	0.82	428
i+4	0.71	339	0.72	301	0.54	484	0.73	659
i+5	0.64	422	0.65	377	0.37	666	0.63	896
DQj+3								
i+1	0.89	51	0.90	38	0.90	38	0.91	10
i+2	0.75	118	0.77	88	0.74	99	0.79	23
i+3	0.64	170	0.67	128	0.49	196	0.61	42
i+4	0.57	203	0.60	155	0.28	276	0.43	61
i+5	0.51	229	0.54	178	0.15	325	0.30	74
DQj+2								
i+1	0.91	101	0.92	81	0.92	76	0.92	103
i+2	0.79	233	0.81	185	0.76	231	0.81	254
i+3	0.68	347	0.72	274	0.48	507	0.62	496
i+4	0.61	430	0.65	339	0.25	724	0.45	720
i+5	0.56	485	0.60	384	0.13	848	0.33	870
DQj+1								
i+1	0.92	139	0.92	116	0.92	119	0.93	121
i+2	0.81	323	0.82	274	0.76	360	0.83	295
i+3	0.71	504	0.71	432	0.51	736	0.71	516
i+4	0.61	670	0.61	580	0.31	1055	0.58	752
i+5	0.53	811	0.53	709	0.20	1231	0.47	949

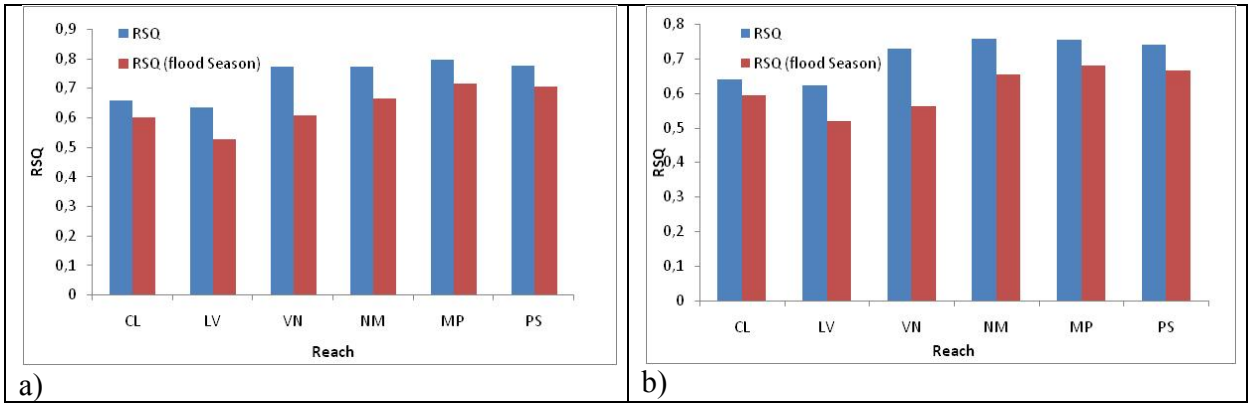


Fig. 7.14: Efficiency of lateral modelled discharges DQ_j (i+t) for each sub-basin j by RSQ, a) simulation period, b) validation period

The performance of lateral discharges $DQ_j(i + t - j)$, expressed in terms of the persistence index PI, is shown in Fig.7.16. With the exception of LV, the PI show almost equal efficiency in the simulation and validation mode. However, the efficiency increases with lead time and ranges from 0.2 to 0.5 from 1 to 5 day lead time. The reason why PI values increase from first to fifth day lies in the high discharge differences between current day and the fifth day, and this difference is large, whereas a part of this difference is explained by the forecast. This results in high PI values in comparison to the small discharge differences between current and future discharges for a one day forecast. Consequently, the forecasting model has to be very efficient to explain this small discharge difference for the one day interval. The same behavior can be seen in each sub-basin from CL to PS.

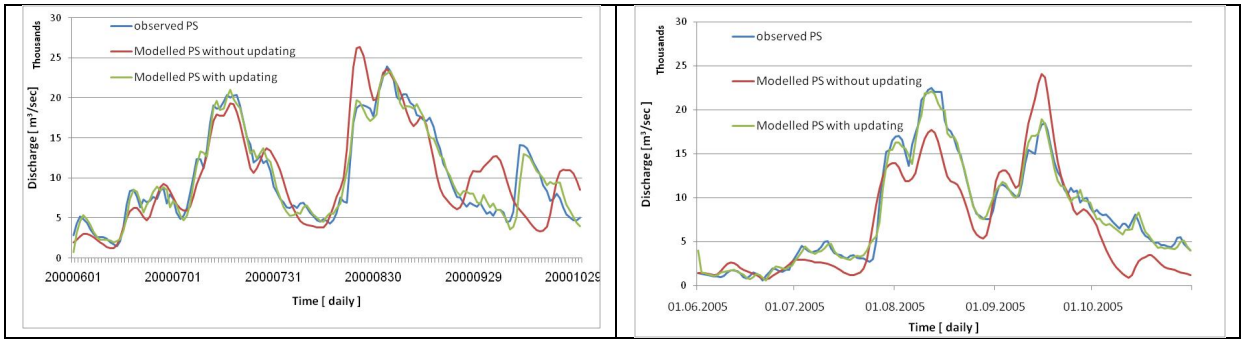


Fig. 7.15: Comparative plot of observed and lateral modeled discharges DQ_j (i+1) in simulation and validation mode with and without updating for sub-basin PS, a) flood selected from simulation period, b) flood selected from validation period

The possible reason of low PI in the case of LV is poor rainfall input because of the fact that this basin is represented by two gauges only which are not sufficient to cover the spatial diversity of rainfall.

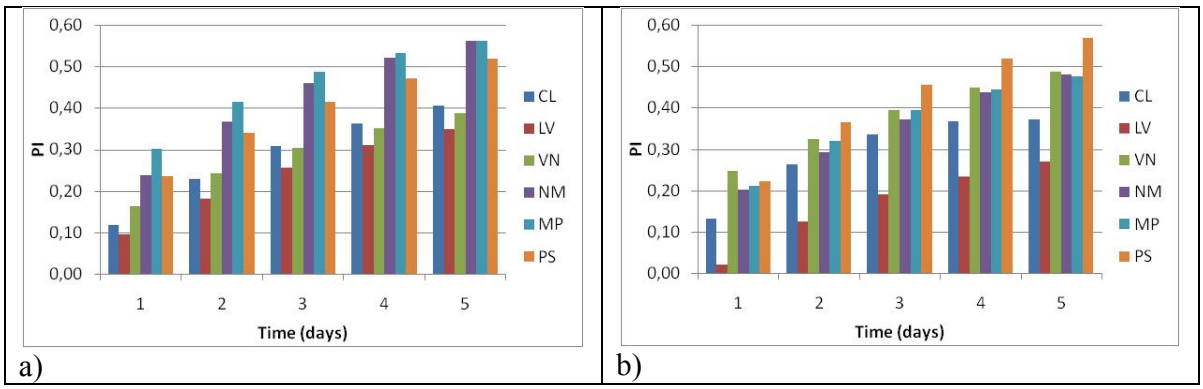


Fig. 7.16: Efficiency of lateral modeled discharges DQ_j (i+t) in terms of PI for each sub-basin j, a) for simulation period, b) for validation period

The forecasts of daily discharges in simulation and validation periods of 1 to 5 days lead time at L,V,N,M,P and S show similar results, i.e. that forecasting efficiency reduces with lead time, as presented in Fig.7.17 for the case of Stung Treng S. Comparisons of one and five day flood forecasts with the observed discharges in the selected high flood seasons show that in both the simulation period and the validation period the efficiency of forecasting model is more or less similar, which can be further inferred from the PI plots in Fig.7.18.

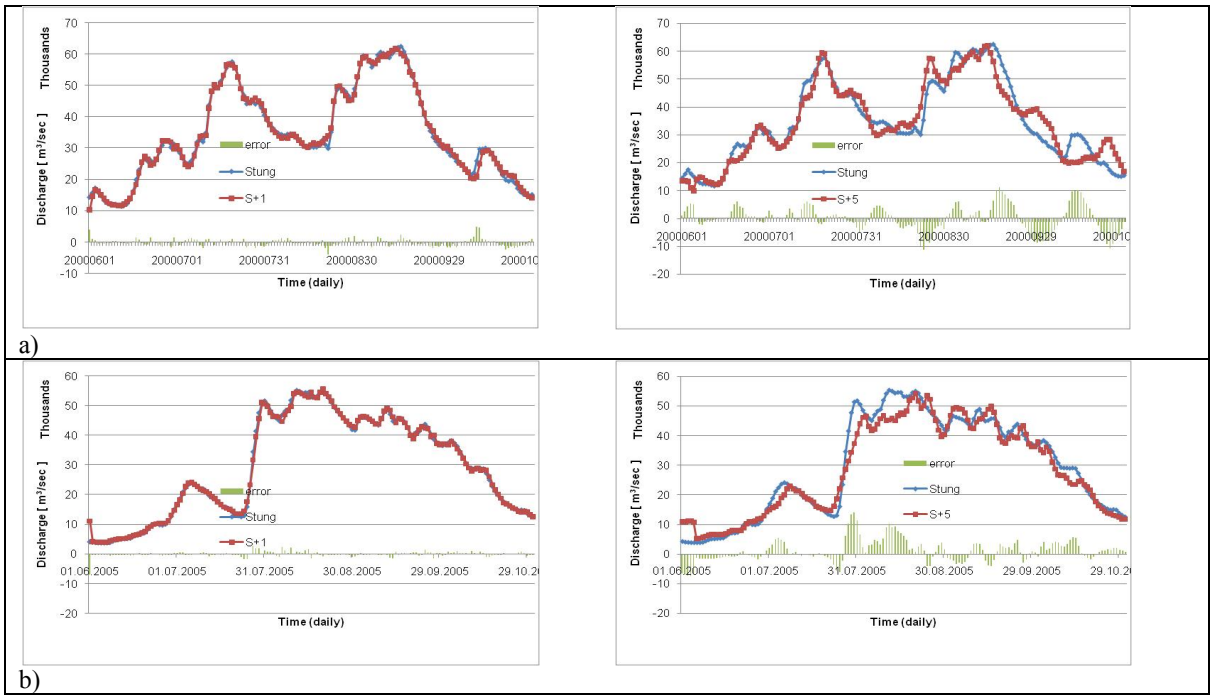


Fig. 7.17: Comparative plotting of 1 and 5 day forecast with observed discharges by first case at Stung Treng gauge S, a) selected flood year from the simulation period, b) selected flood year from the validation period

The value of PI ranges from 0.6 to 0.7 at the downstream gauges of M, P and S. For the further upstream gauge of L, PI is much smaller, i.e. 0.2 to 0.5. But the efficiency of second upstream gauge after L, i.e. V shows ranges from 0.3 to 0.7. Again the efficiency of model as per PI is slightly lower in the validation period than in the simulation period (Fig. 7.18)

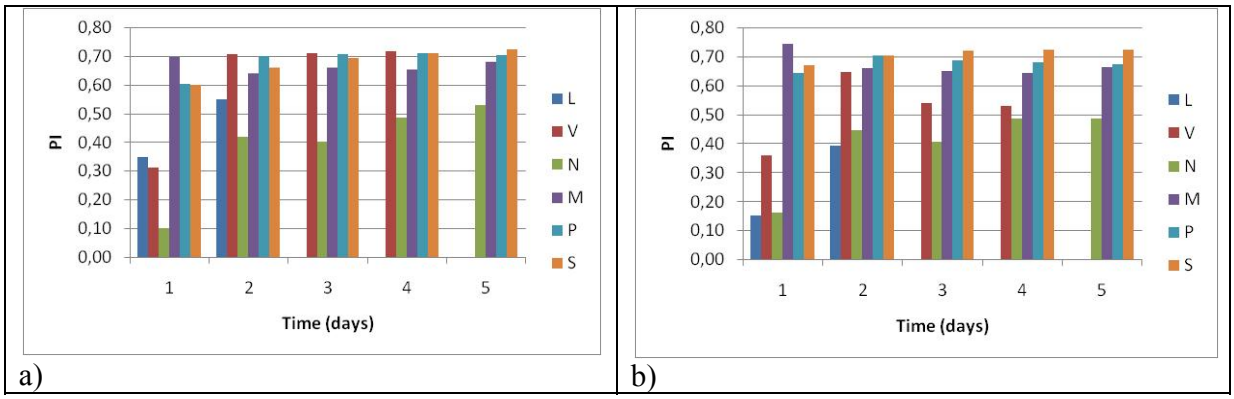


Fig. 7.18: Efficiency of forecasted discharges in first case for Q_j (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period

The spread of forecast errors is shown as normalized probability distribution in Fig.7.19. The error probability plot is necessary to show the error spread for the forecasting seasons in the simulation and the validation period in addition of observed and forecasted discharges comparative plotting and its efficiency in terms of PI. For example the error distribution for 1 to 5 days forecast of S of the simulation period is even a little smaller than of the validation period. Similar behavior has been observed in some cases of other gauges.

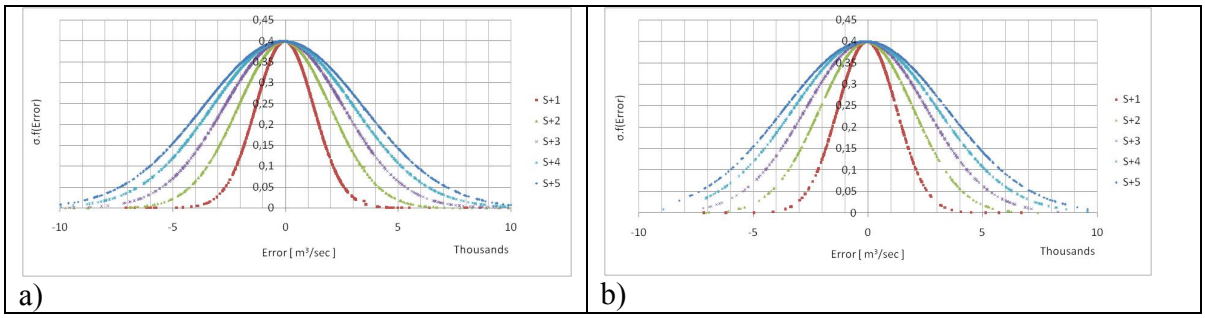


Fig. 7.19: Distribution of forecasted error of the first case at Stung Treng gauge S for 1 to 5 days flood season forecast, a) for simulation period, b) for validation period.

Second case of n and k:

The results of second to fourth case of n and k show similar characteristics of increase in efficiency of lateral discharges $DQ_j(i+t)$ with lead time and reduction in efficiency from simulation to validation period, as shown in Figs.7.22-7.25. After comparing the forecasting results of first case with second, third and fourth case, it can be stated that the first case is giving the best performance to produce $Q_j(i+t)$, as expressed by PI and error pdf. Although in the second case the n and k parameters were optimized by calculating LSE only on the basis of $DQ_j(i)$ values above average, yet the quality of the results in terms of PI are not better than for first case. A possible reason that no improvement was obtained, could be that PI is based on the discharges of the whole flood seasons and not only on high flows. And secondly, below average $DQ_j(i)$ – inflows from lateral sub-basin j may coincide with high flows at a mainstream gauge.

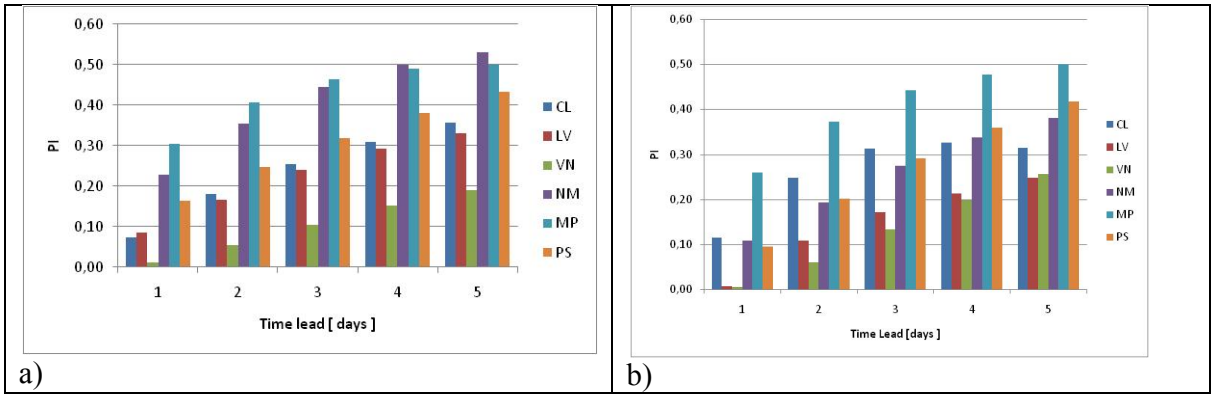


Fig. 7.20: Efficiency of lateral modeled discharges DQ_j (i+t) in terms of PI for each sub-basin j, a) for simulation period, b) for validation period (with second case of n and k)

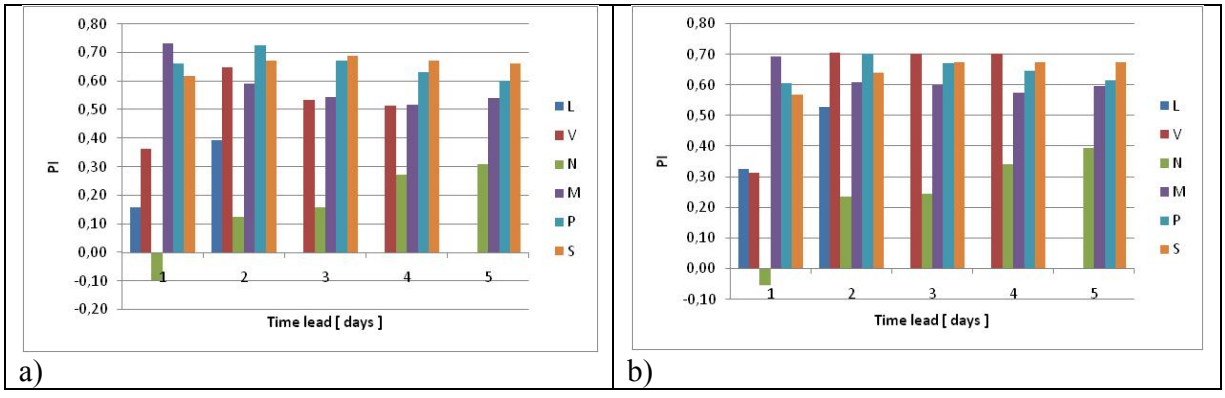


Fig. 7.21: Efficiency of forecasted discharges in second case for Q_j (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period (with second case of n and k)

Third case of n and k:

The third case did perform better than the first case in forecasting mainstream $DQ_j(i+t)$. In the optimization of n and k, the inclusion of the error correction routine is the reason of this improvement. But this case does not perform better than the first case in producing$Q_j(i+t)$.

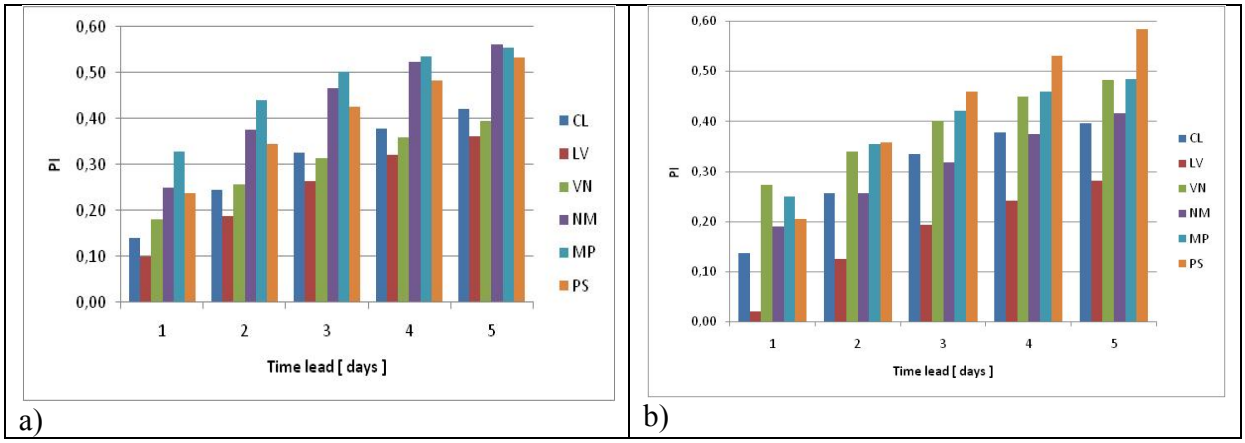


Fig. 7.22: Efficiency of lateral modeled discharges DQ_j (i+t) in terms of PI for each sub-basin j, a) for simulation period, b) for validation period (with third case of n and k)

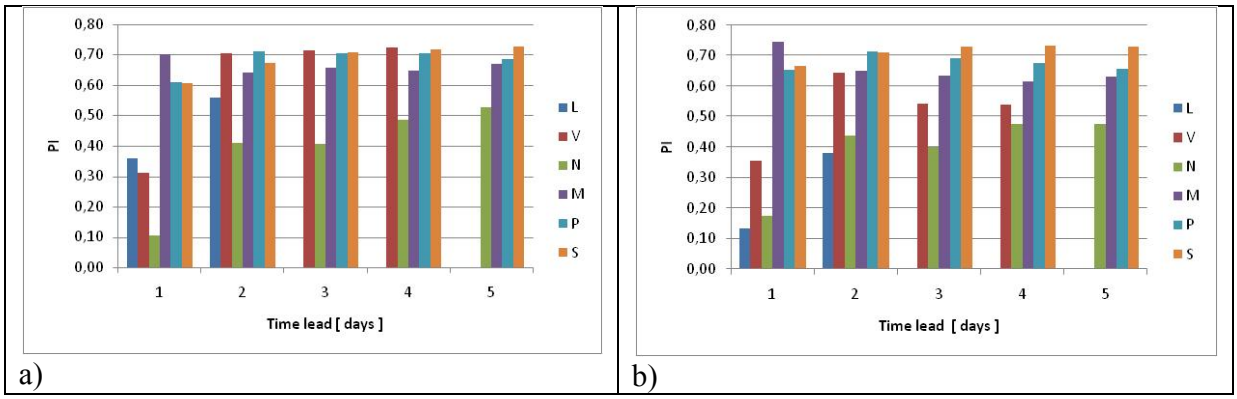


Fig. 7.23: Efficiency of forecasted discharges in second case for Q_j (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period (with third case of n and k)

Fourth case of n and k:

The negative PI has been noted in 1 day lead time forecast of CL in the fourth case of n and k. This is plausible, because in this case, the n and k parameter were optimized by concentrating on reducing the LSE of mainstream gauges $Q_j(i+t)$ instead of concentrating on lateral

inflows $DQ_j(i+t)$. However, this case does not bring much improvement in the results of mainstream$Q_j(i+t)$, except a small increase in the PI for P and S.

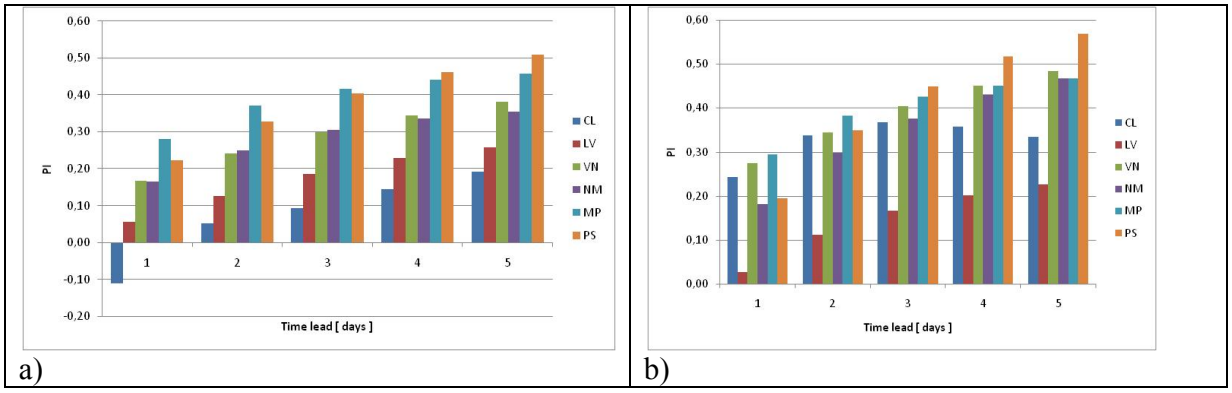


Fig. 7.24: Efficiency of lateral modeled discharges DQ_j (i+t) in terms of PI for each sub-basin j, a) for simulation period, b) for validation period (with fourth case of n and k)

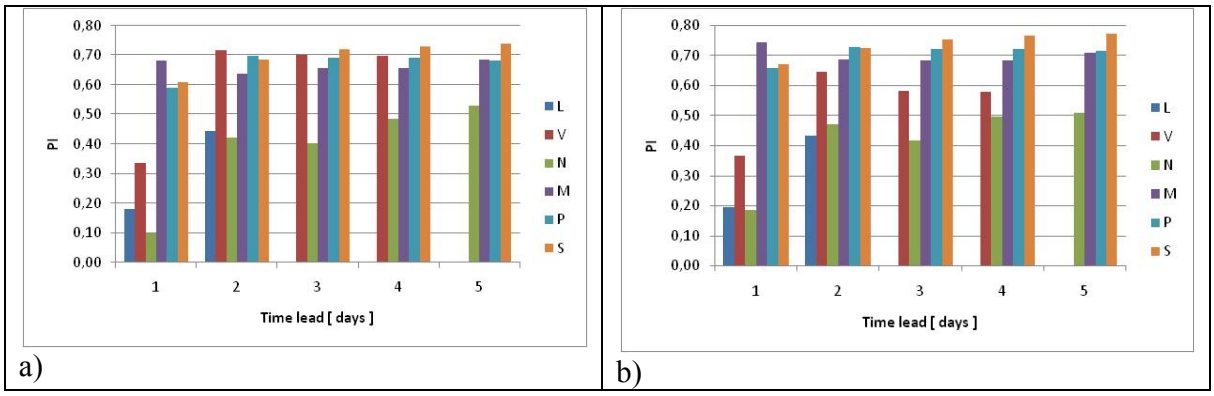


Fig. 7.25: Efficiency of forecasted discharges in second case for Q_j (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period (with fourth case of n and k)

7.2.4.2. Known rainfall and forecasted KN with first case of n and k

The comparative plots of observed and forecasted discharges at Stung Treng of 1 and 5 day lead time are shown in Fig.7.26. The comparative plotting of year 2000 represents the forecasting performance in the simulation period, and the comparative plotting of year 2005 shows the forecast quality in the validation period because year 2000 is a high flood year of simulation period and 2005 is also a year with high floods in the validation period. The daily discharges forecasts in the simulation and validation period of 1 to 5 days lead time at L,V,N,M,P and S by using the observed rainfall of forecasting period and predicting the value of

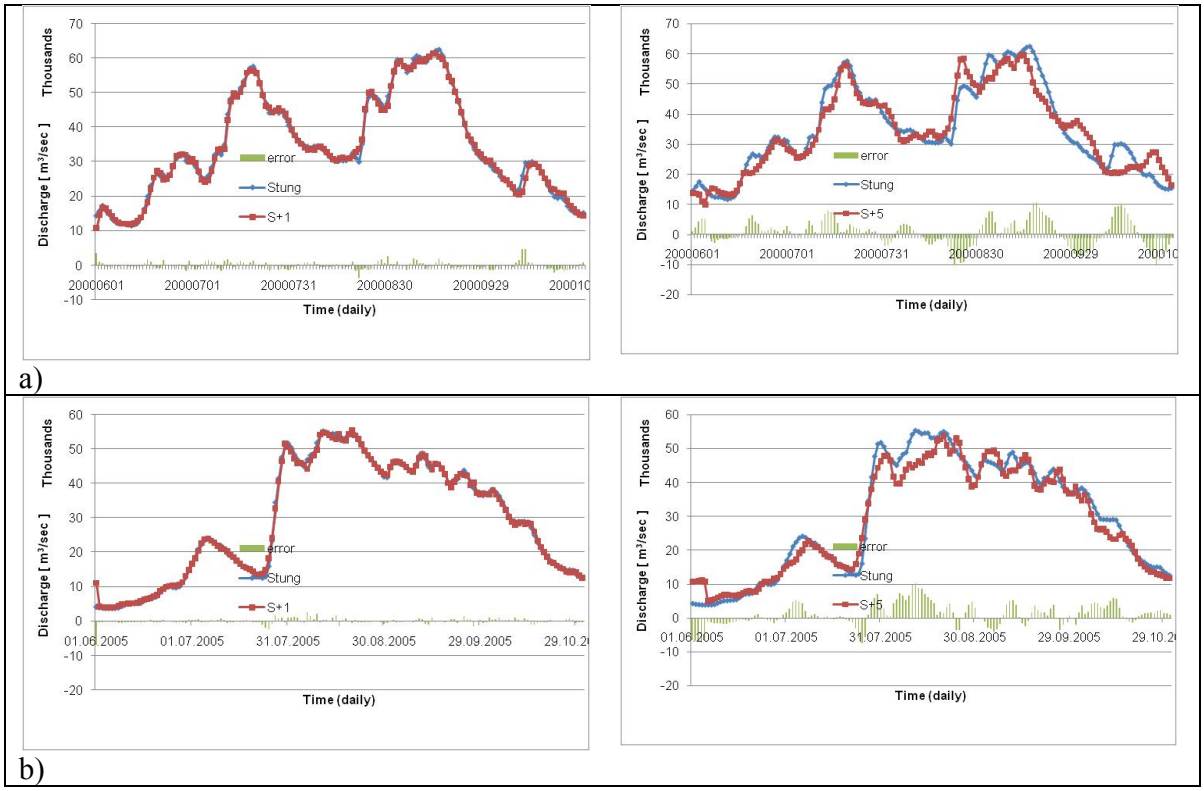


Fig. 7.26: Comparative plotting of 1 and 5 day forecast with observed discharges by first case at Stung Treng gauge S, a) selected flood year from the simulation period, b) selected flood year from the validation period

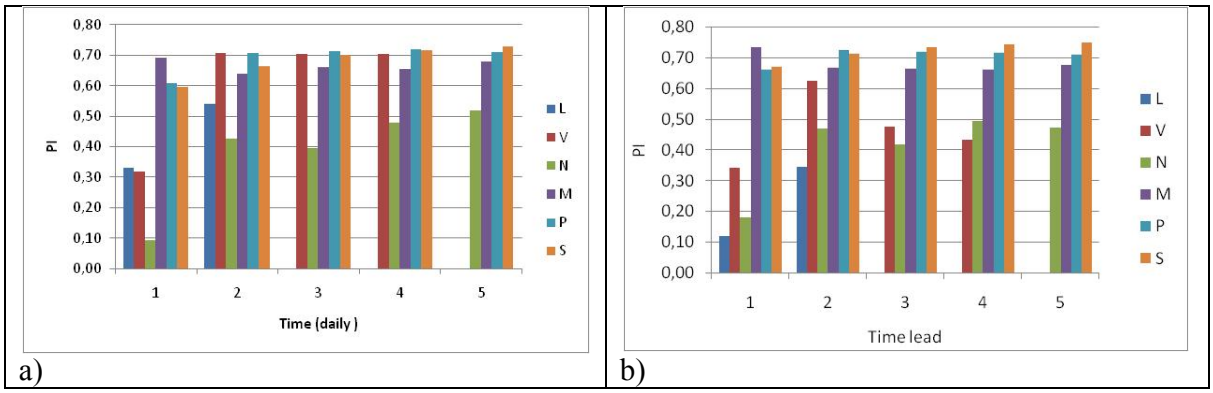


Fig. 7.27: Efficiency of forecasted discharges in second case for Q_j (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period

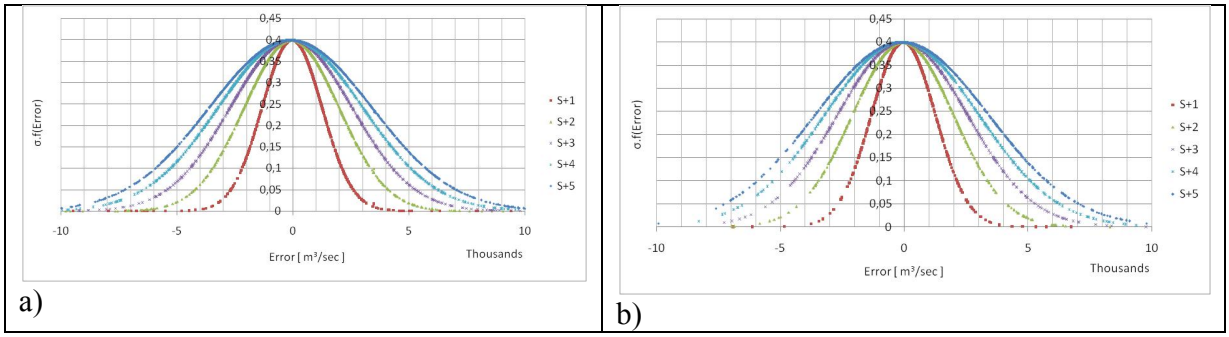


Fig. 7.28: Distribution of forecasted error of the second case at Stung Treng gauge S for 1 to 5 days flood season forecast, a) for simulation period, b) for validation period.

KN by Eq.7.14 shows that forecasting efficiency reduces up to 5% in some cases as presented in Fig.7.27. However, the efficiency of the last 3 downstream gauges, i.e. M, P and S in the simulation and validation period is similar, as can be seen in PI plotting of Fig.7.27. But the efficiency reduces from simulation period to validation period in the case of three upstream gauges, i.e. L, V, and N. It is difficult, however, to notice the difference in the error distribution of Stung Treng forecasting errors in Fig.7.28.

7.2.4.3. Forecasted rainfall and forecasted KN

With first case of n and k:

In the absence of good meteorological rainfall forecasts, three different methods of rainfall forecasting, i.e. last 5 days average rainfall as a future rainfall for forecasting period, 5 day forward shift of rainfall time series, and thirdly, taking today's rainfall as a rainfall forecast for next five days, have been tried to produce discharge forecast. However, only the results of the comparatively best method, i.e. taking today's rainfall as a rainfall forecast for next five days, are presented in Fig.7.29. The daily discharges forecast of 2 and 4 day lead time at Stung Treng S by using the forecasted rainfall and KN with the help of Eq.7.14 shows that this rainfall forecast is quite reasonable for 2 day lead times, but gets poor for the 4th day. The distinctive feature of 2nd and 4th day forecasting error is that forecasts are leading the actual observed discharges which means that both rise and fall of discharges in the forecast mode are occurring earlier than observed occurrences. Secondly, there are sudden and persistent rises and falls of the hydrograph which are due to repetitive inputs of either high or low rainfalls.

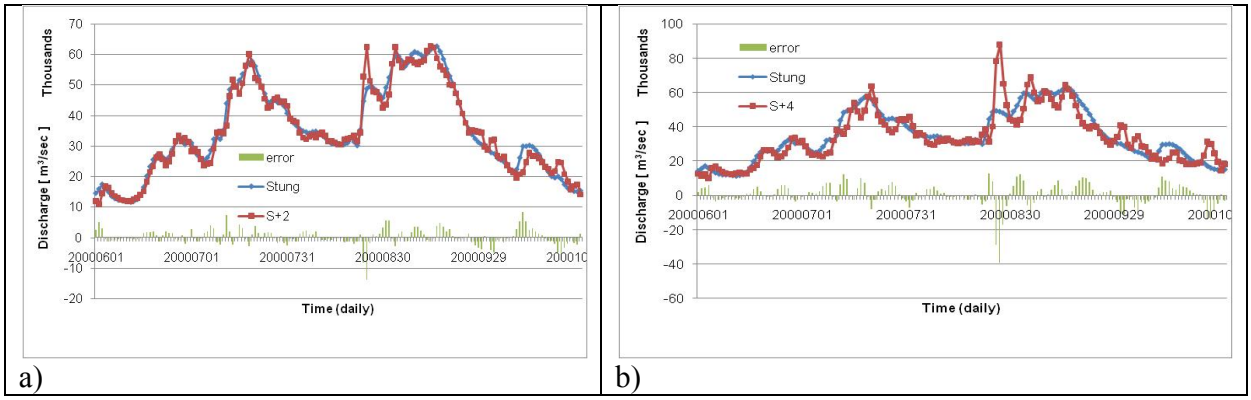


Fig. 7.29: Comparative plotting of 2 and 4 day forecast with observed discharges by third case at Stung Treng gauge S

The daily discharges forecast in the simulation and validation period of 1 to 5 days lead time at L,V,N,M,P and S by using the forecasted rainfall and KN by Eq. 7.16 and 7.24 respectively shows that forecasting efficiency reduces to PI = 0.3 in above fourth day forecast as presented in Fig.7.30.

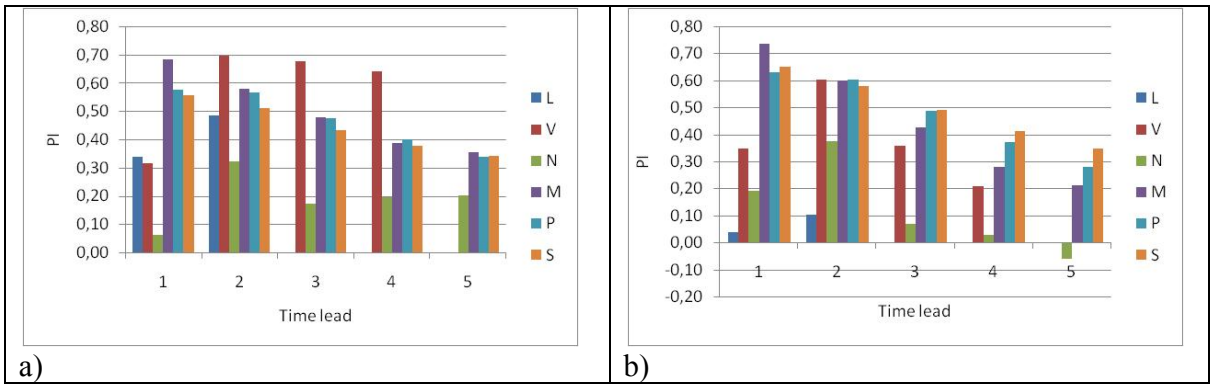


Fig. 7.30: Efficiency of forecasted discharges in third case for Qj (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period

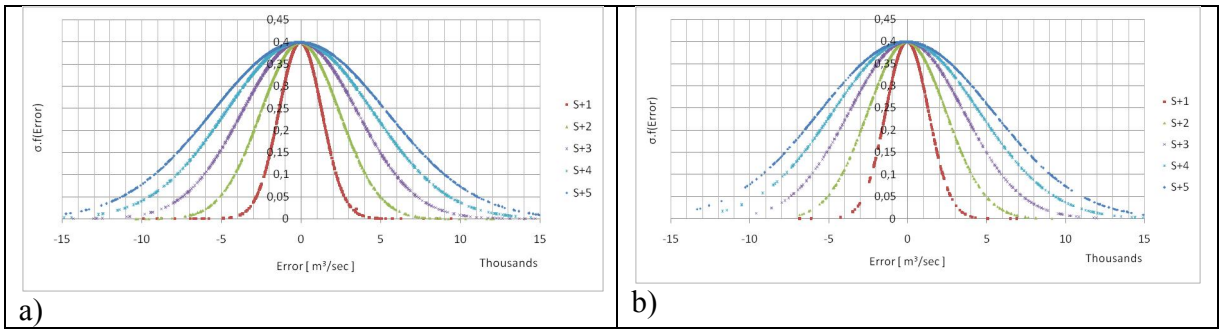


Fig. 7.31: Distribution of forecasted error of the third case at Stung Treng gauge S for 1 to 5 days flood season forecast, a) for simulation period, b) for validation period.

This difference is also noticeable in the spread of the forecasting errors distribution for Stung Treng, as shown in Fig. 7.31

With fifth case of n and k:

The fifth case of n and k was based on forecasted rainfall and therefore is expected to perform better in comparison to the first case of n and k. The efficiency of forecasting, as expressed by PI supports this argument. In comparison to PI for the first case (Fig.7.30), the PI for the fifth case (Fig.7.33) is high. Furthermore, comparative plotts of observed and forecasted discharges of S (Fig.7.32) are indicating less errors in comparison to errors produced by forecast by using first case of n and k (Fig.7.29).

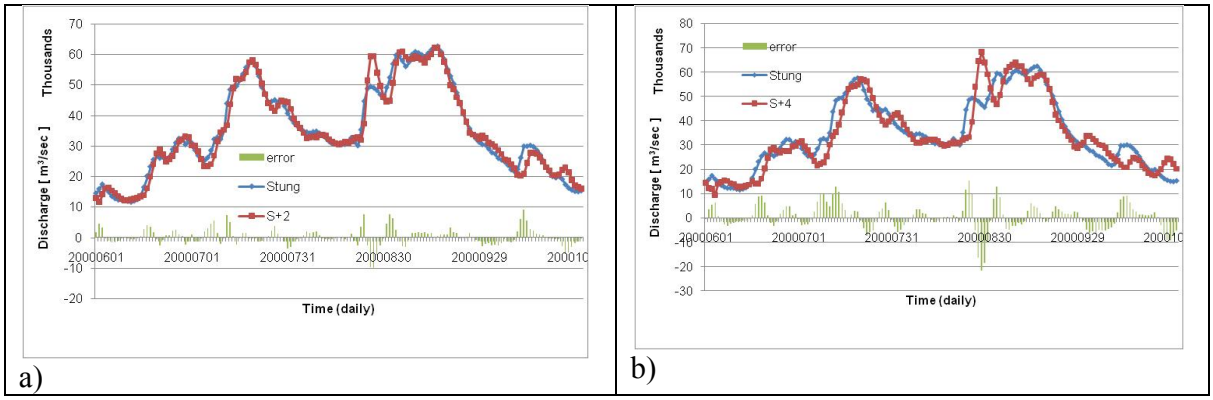


Fig. 7.32: Efficiency of forecasted discharges in third case for Q_j (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period

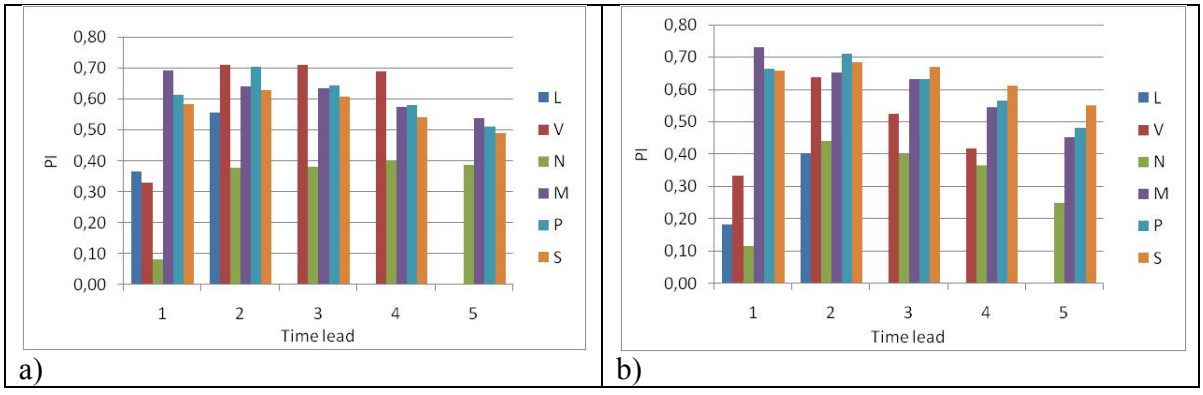


Fig. 7.33: Efficiency of forecasted discharges in third case for Q_j (i+t) in terms of PI at each gauge j, a) for simulation period, b) for validation period

7.3. Summary

Type 2 Model has been developed and used in this chapter in order to reduce the forecasting errors of Type 1 Model of 1 to 5 days forecast for Middle Mekong gauges, i.e. L, V, N, M, P and S. Instead of using any existing Type 2 Model, the forecasting routine is constructed from the available data base of 31 rainfall and 7 discharge gauges. The areas of the 6 reaches between 7 discharge gauges are considered as the lateral sub-basins j which contribute towards the main Mekong flow with different time lags. A rainfall runoff relation has been established for each sub-basin j by using lumped rainfalls as input into the rainfall - runoff model. Areal average rainfall is obtained both by Thiessen polygon and arithmetic average, however, arithmetic average is finally selected to be used for further calculations.

The effective rainfall available for runoff has been calculated by using an empirically derived "adjustment factor" KN, which is similar to a runoff coefficient. The monthly value of the coefficient KN is the ratio of the volumes of 30 days runoff to the volume of 30 days rainfall. KN is observed to increase uniformly along the flood season. In order to do forecasting, the KN - value is also to be estimated for future forecasting days along with the past and current day KN. Consequently apart from using the assumption of a constant KN, KN is computed by three other methods, i.e. as mean monthly KN, forecasted monthly KN and moving average KN. Owing to its performance, the method of moving average KN is finally selected to compute effective rainfall.

A Nash cascade with two parameters, i.e. n and k, is used to translate effective rainfall into runoff. The Nash cascade represents the unit hydrograph as a gamma distribution, whose shape is controlled by parameters n and k. The method of least squares is used for n and k determination with additional efficiency check by RSQ. Five different variants of n and k are produced by minimizing squared errors of: lateral discharges $DQ_jF(i+t)$, above average lateral discharges $DQ_jF(i+t) > \overline{DQ_jF(i+t)}$, lateral discharges after updating, discharges of mainstream $Q_jF(i+t)$ and discharges of mainstream by using forecasted rainfall. However, owing to its performance, the n and k optimization method based on LSE of lateral discharges $DQ_jF(i+t)$ is finally selected to compute runoff.

The method of split sampling is used for development and testing of the rainfall runoff model. Simulated lateral inflows obtained from the unit hydrograph model from each sub-basin j is produced using 1990 to 2000 rainfall data, and compared for validation against 2001 to 2005 data.

Tab. 7.11: Standard deviation of error by various cases of flood forecast (for simulation period)

		(first case of n and k)					(fifth case of n and k)
Gauge	Without forecast	Observed P and calculated KN	Observed P and forecasted KN	Observed P and forecasted KN (last 5 days avg.)	Observed P and forecasted KN (last 5 days time shift)	Observed P and forecasted KN (constant rainfall, P(i) = P(i+t))	Observed P and forecasted KN (constant rainfall , P(i) = P(i+t))
L+1	640	512	522	522	520	525	508
L+2	1101	730	746	802	789	809	731
V+1	623	510	509	509	509	512	503
V+2	1081	580	578	591	590	594	574
V+3	1461	781	790	843	827	834	780
V+4	1778	939	959	1123	1063	1122	981
N+1	712	672	672	675	683	682	677
N+2	1288	980	972	1025	1056	1034	1010
N+3	1799	1380	1389	1666	1628	1581	1408
N+4	2248	1602	1615	2165	2008	2064	1729
N+5	2648	1803	1826	2600	2359	2531	2065
M+1	826	465	462	462	469	465	461
M+2	1536	918	929	958	1008	951	928
M+3	2165	1256	1272	1537	1578	1450	1315
M+4	2721	1592	1609	2195	2145	2070	1784
M+5	3219	1811	1830	2752	2603	2645	2193
P+1	1212	781	771	767	801	753	763
P+2	2118	1173	1155	1357	1414	1255	1163
P+3	2887	1590	1556	2140	2115	2020	1737
P+4	3546	1944	1894	2897	2773	2800	2316
P+5	4122	2303	2236	3596	3376	3575	2895
S+1	2007	1285	1291	1316	1353	1307	1309
S+2	3508	2025	2051	2376	2476	2288	2147
S+3	4760	2591	2624	3648	3609	3543	2992
S+4	5809	3084	3113	4827	4616	4841	3953
S+5	6708	3486	3506	5751	5479	5941	4813

The modeled discharges of the river model for forecast time lags of 1 to 5 days are updated by adding the latest available information on lateral inflows, which were obtained for 1 to 5 days

lead time forecasts. The lateral inflows are checked for efficiency by PI, and subsequently used for mainstream discharge forecast. Forecasts are produced by three different rainfall input variants. In the first case, in order to test model efficiency, the observed historical future rainfall and KN computed from observed historical data is used as available information to produce forecasts. In the second case, the combination of observed future rainfall and forecasted future KN are used. And finally in the third case, forecasted future rainfalls and forecasted future KN are used. The quality of the model output in each case is assessed by comparative plotting of observed and forecasted discharges, by comparing index PI and by the probability density distributions of the remaining errors. In general, the efficiency index of flood forecasts at each gauge has been observed to be reduced from first to third case from PI 0.7 to 0.5 in 1 to 3 day lead time and from 0.7 to less than 0.4 in 4 and 5 day lead time.

The main reason behind the large efficiency loss for large lead times is the poor rainfall forecast, for which instead of using actual rainfall, the information based on the last available 5 days of rainfall is used to estimate future rainfalls. In the first case n and k values, which are optimized according to lateral inflows are used in conjunction with three different cases of rainfall forecasts. The fifth case of n and k optimization is designed to consider the case of unavailability of future rainfall, therefore, it is based on forecasted rainfall using today´s rainfall only. It led to some improvement, in general, efficiencies were obtained of PI = 0.7 for 1 to 3 day lead time and PI = 0.6 to 0.5 in 4 and 5 day lead time respectively.

The standard deviations of errors from different variants of the Type 2 Model are compared with the errors of no model forecast in Tables 7.12 and 7.13 for simulation and validation periods, respectively. These tables summarize the model's performance in comparison to no forecast at all (i.e. assuming that today´s value is the same as for the value at lead time t in the future) by presenting reduction in standard deviation of errors. In general, the standard deviation reduces by 50% from no forecast to forecast by perfect input of rainfall with first case of n and k optimization. However, the reduction in standard deviation decreases by replacing the perfect rainfall with forecasted rainfall. Still, comparatively less standard deviation of errors (one half in 1 to 3 day lead time and two third in 4 to 5 day lead time) is obtained by using the fifth case of n and k parameters (specially optimized case of n and k by using forecasted rainfall time series)

Tab7.12: Standard deviation of error by various cases of flood forecast (for validation period)

		(first case of n and k)					(fifth case of n and k)
Gauge	Without forecast	Observed P and calculated KN	Observed P and forecasted KN	Observed P and forecasted KN (last 5 days avg.)	Observed P and forecasted KN (last 5 days time shift)	Observed P and forecasted KN (constant rainfall propagation)	Observed P and forecasted KN (constant rainfall propagation)
L+1	649	567	574	574	599	560	552
L+2	1059	802	827	962	968	891	790
V+1	674	471	480	479	478	485	480
V+2	1117	607	626	649	647	642	610
V+3	1454	923	981	1122	1087	1049	928
V+4	1716	1117	1219	1595	1443	1518	1226
N+1	818	647	641	628	637	651	666
N+2	1417	949	932	965	1014	996	959
N+3	1886	1369	1355	1755	1713	1619	1376
N+4	2293	1574	1566	2398	2168	2241	1753
N+5	2650	1836	1867	3020	2648	2922	2229
M+1	987	432	446	445	446	457	452
M+2	1762	919	924	973	1016	987	954
M+3	2439	1300	1294	1687	1702	1584	1366
M+4	3028	1652	1634	2530	2385	2379	1905
M+5	3543	1881	1863	3260	2932	3160	2448
P+1	1225	669	668	664	700	643	666
P+2	2167	1087	1080	1239	1304	1127	1114
P+3	2984	1530	1504	2035	2037	1907	1734
P+4	3690	1914	1866	2869	2785	2725	2329
P+5	4295	2273	2198	3698	3474	3612	2959
S+1	1974	1111	1100	1111	1136	1116	1123
S+2	3460	1839	1811	2037	2198	1966	1906
S+3	4714	2424	2377	3204	3294	3027	2663
S+4	5737	2910	2819	4344	4289	4196	3498
S+5	6597	3327	3190	5339	5178	5318	4312

8. Mixed Model Forecasting

8.1. Mixed Model: Type-1 Model and Type-2 Model combined

It is confirmed by the coefficients in table 8.1 that model 1 performs better (or just as good as model 2) for short lead times, but that model 2 is much better at long lead times. The Problems occur when a rainfall forecast also has to be made. In this case the weighted averages of model 1 and 2 are useful. The benefit of weighted averages of Type-1 and Type 2 model were also found useful in the light of following observations.

The results of Type-1 and Type-2 models, presented in chapter 6 and 7 respectively, shows that the forecasts by Type-1 Model underestimate the observed discharges of the rising part of flood hydrograph in contrast to forecast by Type-2 Model (with forecasted rainfall input) which overestimates the observed discharges, and both models behave vice versa in a falling limb. It means that the forecast by Tpye-1 model is lagging behind the observed discharges and forecast by Type-2 model is leading the observed discharges in the rising and falling parts of the hydrograph. This tendency of under and over-estimation of Type-1 model in rising and falling limb respectively is because of the fact that forecast is dependent on the gradient of past discharges. The high dependence on past gradient results in under-estimation of initial rises of discharges in the rising limb of flood hydrograph and over-estimation of discharges in the initial decreases of discharges in the falling limb. However, this departure of future discharges from the past trend is explained well by Type-2 Model, because it is based on the information of past rainfall and future rainfall prediction scenarios instead of past discharge gradients. The forecast of future discharges by Type-2 models depends on accuracy of rainfall input and modeling accuracy. If the rainfall input is accurate, then the nature of Type-2 model is responsible for earlier rise and fall of future discharges in forecasting mode than actual rise or fall of observed discharges. The Type-2 model is based on unit hydrograph which is optimized in such a way that the rainfall is distributed in time with an early peak (according to minimum time delay observed between rainfall occurrence and runoff peak). This distribution of rainfall with early peak is responsible for overestimation of discharge in the rising limb and vice versa in the falling limb.
Another distinctive feature of Type-1 model is the steady rise and fall of forecasted discharges in contrast to abrupt and sudden rise and fall of forecasted discharges by Type-2 Model. The reason of steadiness by Type-1 Model is because future discharges are constructed from the previous discharge gradient information, which is steady due to the dynamics of large river. This steadiness of the gradient of past discharges is responsible for the steadiness of future discharges obtained from theType-1 model. On the other hand, the Type-2 model uses the information of rainfall which may change abruptly in time, therefore, the forecasted discharges based on rainfall information also show abrupt changes, in spite of being distribution over time by the unit hydrograph.
These two complementary features of flood forecasts from Type-1 and Type-2 Model suggest to use the two models in combination. Consequently, in this chapter the flood forecast by Type-1

and type-2 model is averaged and compared with the observed discharges. However, instead of taking simple averages, weights are assigned to the forecast of each model. These weights are determined by the standard procedure of multi-linear regression. The lagging and smooth change of discharge in the forecast in Type-1 Model has been cancelled in part by leading and abrupt change of discharge in prediction by Type-2 Model.

8.2. Mixed Model application on Mekong

The mixed model as linear combination of the individual model forecasts are expressed through Eq. 8.1, with regression weights α and β given in Table8.1, which are used to produce mixed flood forecasts:

$$\mathrm{QF_j(i+t)} = \alpha \mathrm{QF1_j(i+t)} + \beta \mathrm{QF2_j(i+t)} + \varepsilon_{f3j}(i+t) \qquad (8.1)$$

Tab. 8.1: Optimized parameters for the updating of flood forecast

	Observed P and KN		Observed P and for KN		Fore P and Fore KN	
Gauge	α(i + t)	β(i + t)	α(i + t)	β(i + t)	α(i + t)	β(i + t)
Qj+9 (L)						
i+1	0.391	0.606	0.397	0.600	0.404	0.593
i+2	0.205	0.793	0.186	0.811	0.391	0.606
i+3						
i+4						
i+5						
Qj+7 (V)						
i+1	0.659	0.337	0.645	0.351	0.650	0.346
i+2	0.369	0.624	0.387	0.605	0.481	0.512
i+3	0.233	0.757	0.296	0.694	0.444	0.546
i+4	0.158	0.832	0.217	0.772	0.493	0.497
i+5						
Qj+5 (N)						
i+1	0.589	0.409	0.594	0.404	0.603	0.394
i+2	0.335	0.663	0.321	0.677	0.402	0.596
i+3	0.221	0.771	0.225	0.766	0.464	0.529
i+4	0.183	0.808	0.195	0.795	0.548	0.444
i+5	0.149	0.841	0.165	0.823	0.586	0.404
Qj+3 (M)						
i+1	0.517	0.482	0.637	0.362	0.612	0.387
i+2	0.181	0.817	0.221	0.777	0.314	0.684
i+3	0.046	0.952	0.112	0.885	0.369	0.628
i+4	0.040	0.954	0.110	0.882	0.453	0.540
i+5	0.039	0.954	0.113	0.879	0.521	0.471
Qj+2 (P)						
i+1	0.419	0.581	0.432	0.568	0.475	0.525
i+2	0.049	0.952	0.252	0.748	0.521	0.479
i+3	0.095	0.906	0.244	0.756	0.580	0.419
i+4	0.091	0.910	0.225	0.775	0.607	0.391
i+5	0.098	0.900	0.220	0.777	0.632	0.364
Qj+1 (S)						
i+1	0.153	0.848	0.215	0.785	0.322	0.678
i+2	0.068	0.935	0.149	0.853	0.420	0.581
i+3	-0.040	1.045	0.090	0.914	0.520	0.482
i+4	-0.018	1.025	0.107	0.898	0.583	0.419
i+5	-0.011	1.019	0.114	0.893	0.609	0.392

The three different cases of flood forecasts by Type-2 Model have been discussed in 7.4.3.1, 7.4.1.2 and 7.4.3.3. To generate regression weights these three flood forecast cases are optimized separately with Type-1 Model. Therefore three set of weights are shown in Table-8.1. The normal trend of these weights shows that Type-1 Model has comparatively more weight (represented by α) in flood forecasts of shorter lead times, i.e.1 to 2 days, however its weight decreases with lead times longer than two days. The weights of Type-2 Model (represented by β) are smaller for shorter lead times, i.e.1 to 2 days, but increase with lead time. This decrease and increase of α and β with the lead time is logical because the Type-2 model becomes more meaningful in explaining the future discharges with larger lead times because of the effect of the rainfall input in comparison to Type-1 Model without rainfall input.

8.2.1. Mix forecast: Case-1 (Observed Rainfall and calculated KN)

The Case-1 of Type-2 Model as given in section 7.4.3.1 is combined with Type-1 Model by Eq.8.1. The comparative plotting of observed and forecasted flow by Mix model for Stung Treng with 1 to 5 day lead time is shown in Figs.8.1a, b, c, d, e. This shows that the first flood peak is forecasted accurately by this approach, but the second flood peak is over-estimated in the rising part. The forecasted discharges are rising before the rise of observed discharges even for the 5 days flood forecast. This leading behavior of flood forecasts by the Mix Model can be useful in Case-3 of Type-2 Model mixing with Type-1 model, because rainfall forecasts of this case is based on neighboring past rainfall events. These past rainfall events are shifted to the future and can reduce the leading of future discharges by Case-1.

Figs.-8.2 a and b present the distribution of errors of 1 to 5 day flood forecasts for station Stung Treng in the simulation and validation period. There is no significant difference in the overall spread of error in the simulation and validation period.

This finding of similar performance of Mix Model in the simulation and validation period is further supported by PI as given in Figs-8.3a and b. The PI of flood forecasts at L, V, N, M, P and S show that except at L and N the flood forecasts of all other gages are quite acceptable, with PI above 0.6.

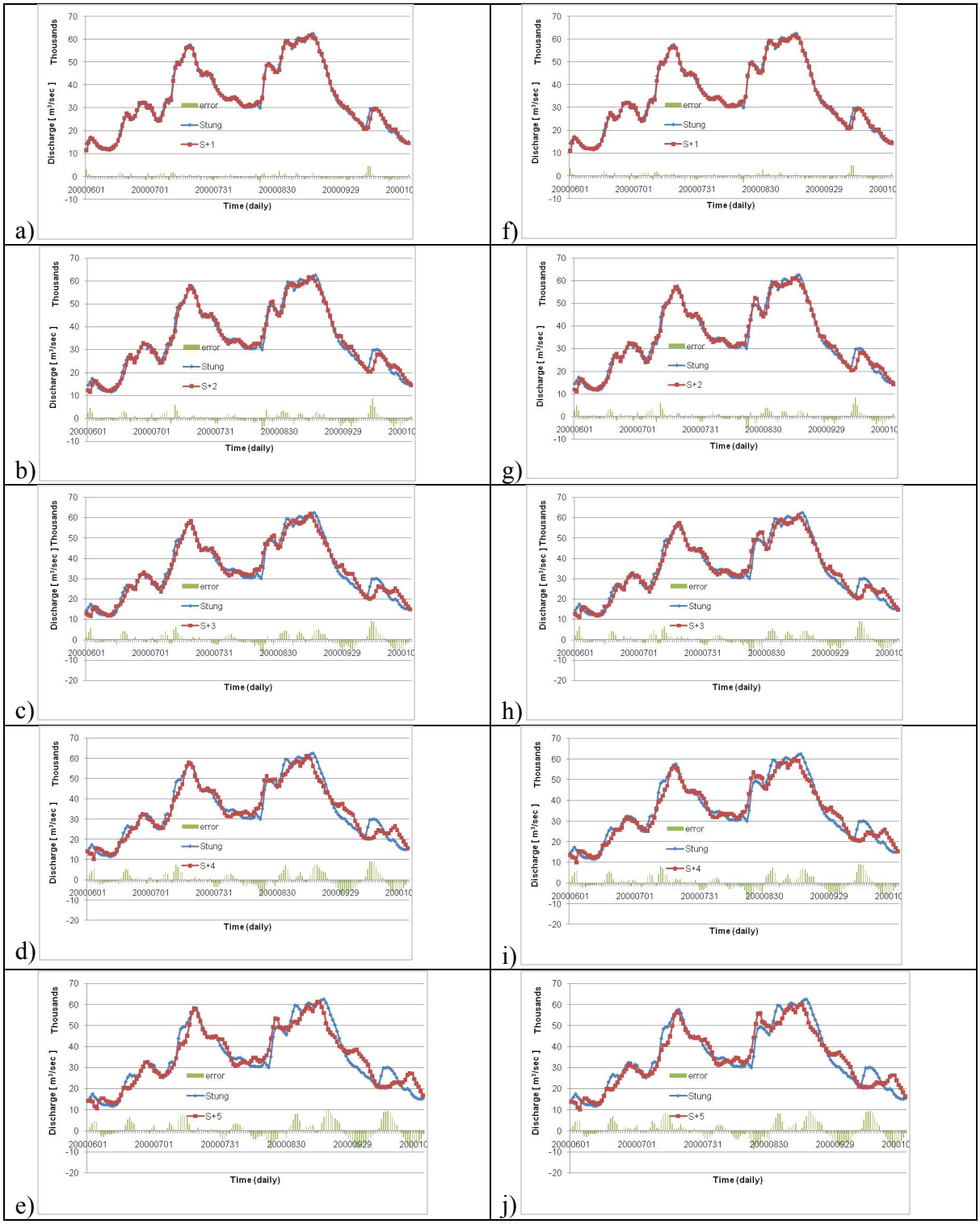


Fig. 8.1: Comparative plotting of observed and forecasted flows of Stung Treng, left column: 1 to 5 day forecast (a,b,c,d,e) with the case of observed rainfall with calculated KN, right column: 1 to 5 day forecast (f,g,h,i,j) with the case of observed rainfall with forecasted rainfall

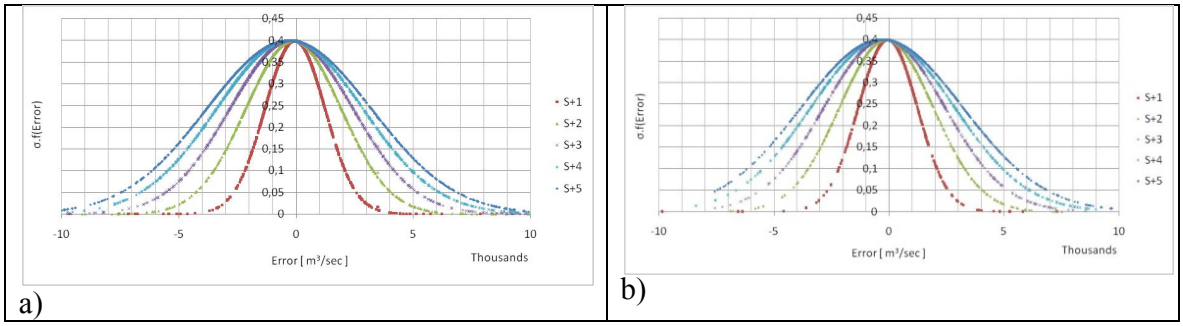


Fig8.2: Pdf of errors of the forecasted flows of Stung Treng for 1 to 5 day forecast with the case of observed rainfall with calculated KN, (a) simulation period, (b) validation period

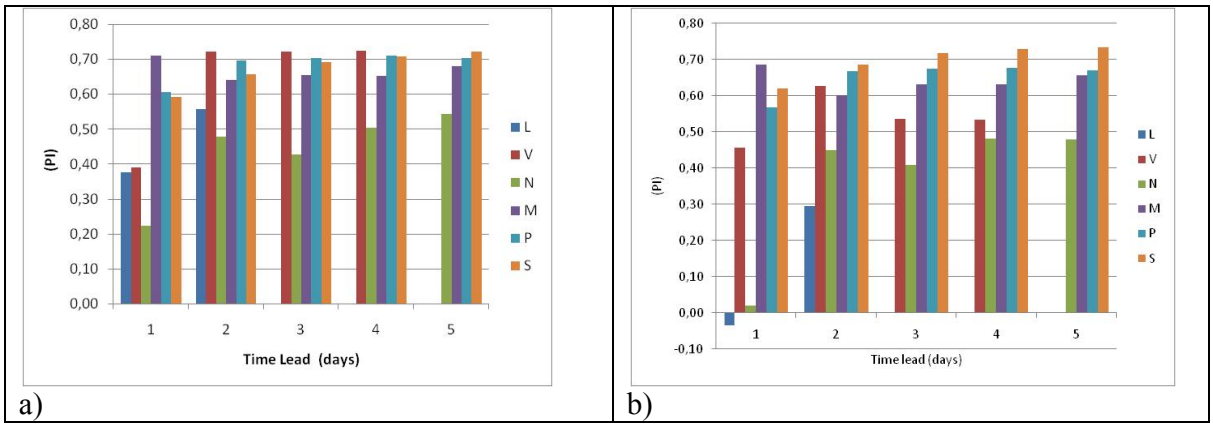


Fig. 8.3: Quality of flood forecast with the case of observed rainfall with calculated KN in terms of PI, left column: simulation period, right column: validation period

8.2.2. Mix forecast: Case-2 (Observed Rainfall and forecasted KN)

The Case-2 of Type-2 Model as given in section 7.4.3.2 is combined with Type-1 Model according to Eq.8.1. The comparative plotting of observed and forecasted flow by Mix model for Stung Treng with 1 to 5 day lead time is shown in Figs.8.1f, g, h, i, j. It shows results similar to those of the combination of Case-1 of Type-2 Model with Type-1 Model. In this case also the first flood peak is forecasted accurately, whereas the second flood peak is over-estimated in the rising limb.

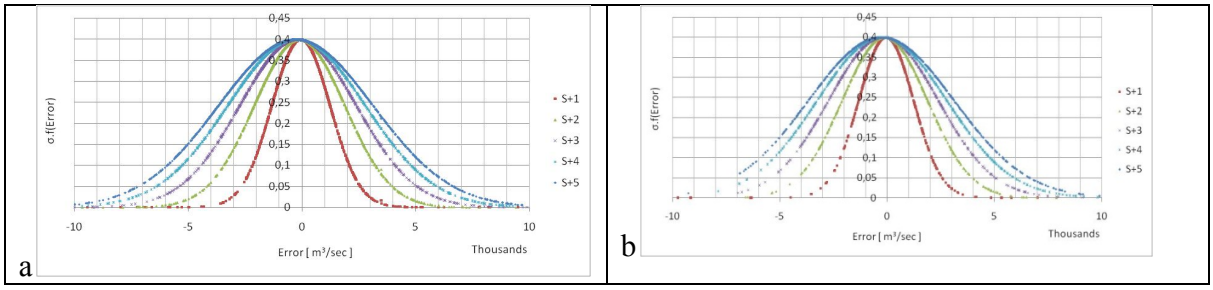


Fig8.4: Pdf of errors of the forecasted flows of Stung Treng for 1 to 5 day forecast with the case of observed rainfall with forecasted KN, (a) simulation period, (b) validation period

The distribution of errors of 1 to 5 days flood forecast of Stung Treng in the simulation and validation mode shows almost identical spread (Figs.8.4a and b).

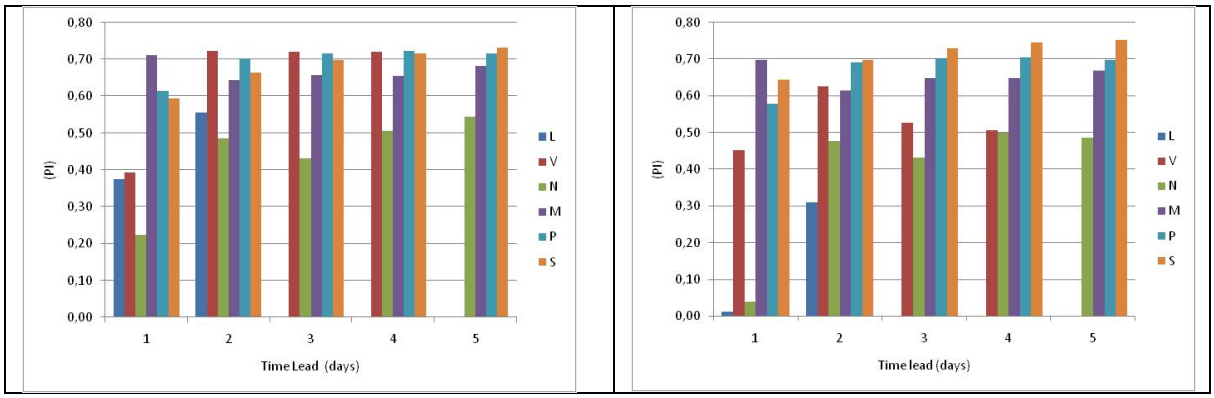


Fig. 8.5: Quality of flood forecast in terms of PI, left column: simulation period, right column: validation period

The PI of 1 to 5 day flood forecast by Mix-Model of L, V, N, M, P and S shows that except at L and N the flood forecast of all other gages is fairly good, i.e. PI above 0.6.

8.2.3. Mix forecast: Case-3 (Forecasted Rainfall and forecasted KN)

The first two cases of Type-2 Model combination with Type-1 Model in Mix Modeling yield a forecasting efficiency of acceptable level, i.e. above 0.6. However, the forecasting efficiency reduces in the third case of Type-2 Model combination with Type-1 Model.

Method-1: Last 5 day average rainfall

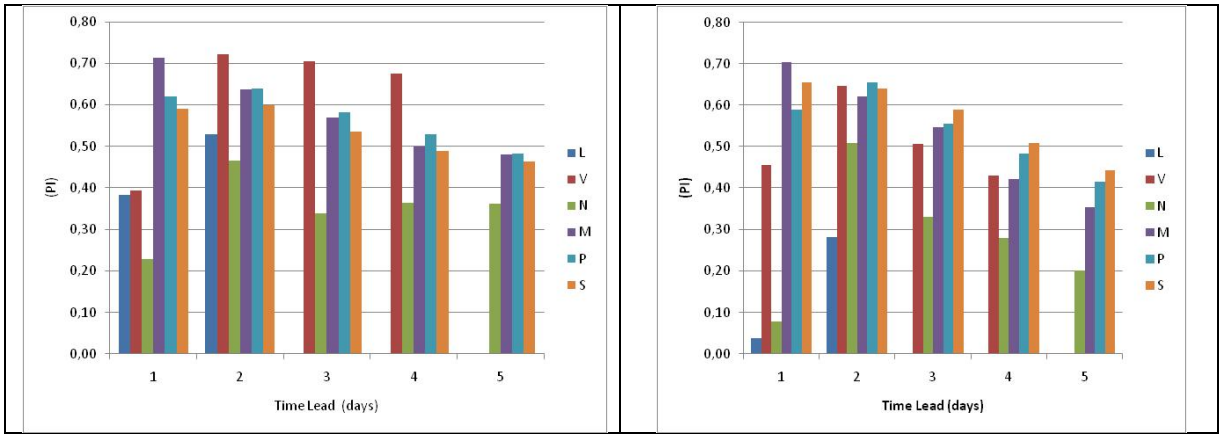


Fig. 8.6: Quality of flood forecast in terms of PI for first case of forecasted rainfall, left column: simulation period, right column: validation period

The reduction of flood forecast accuracy by Mix Model is a function of the accuracy of rainfall forecasts. Three different methods are used to generate future rainfall forecasts as given in Chapter-7. The results of combination of 3 different methods of Case-3 of Type-2 Model with Type-1 Model are given in Figs.8.7, 8.8 and 8.9.

Method-2: Time shift of 5 days

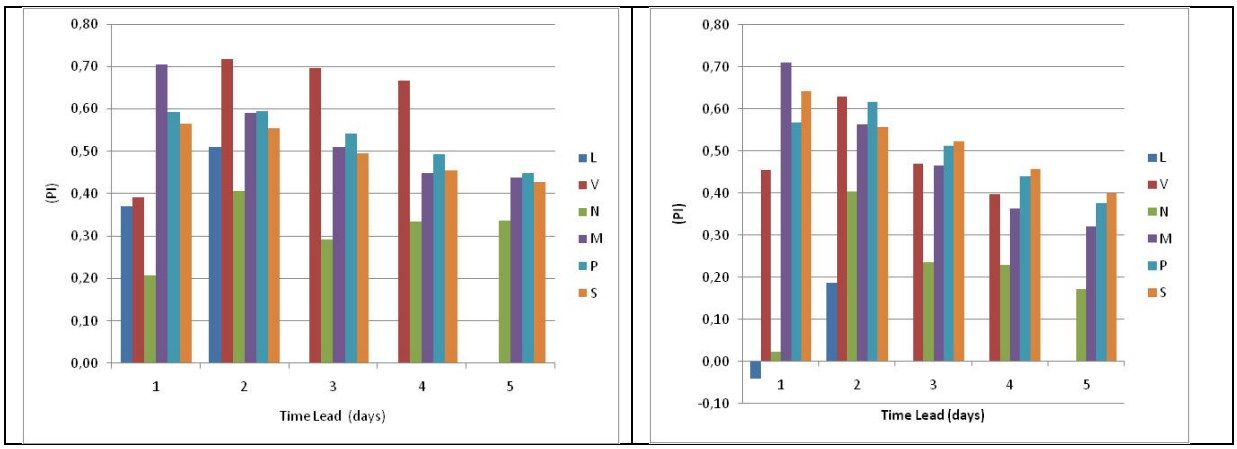


Fig. 8.7: Quality of flood forecast in terms of PI for second case of forecasted rainfall, left column: simulation period, right column: validation period

The Mix-Model results of Method-1 and 2 are shown in Fig-8.6 and 8.7. Mix-Model with rainfall forecast of Method-2 is performing better than Mix-Model with rainfall forecast of Method-1. The best forecasting results are obtained in the combination of Method-3 of Case-3 of Type-2 Model with Type-1 Model. The PI of 1 to 5 days flood forecast of V, M, P and S ranges from 0.5 to 0.7 (Fig.8-8.). However, in each of these three cases of Mix-Model results, the flood forecast

performance in the validation period is less accurate than forecasting performance in the simulation period.

Method-3: Using time index i rainfall for i+1 to i+5 rainfall forecast

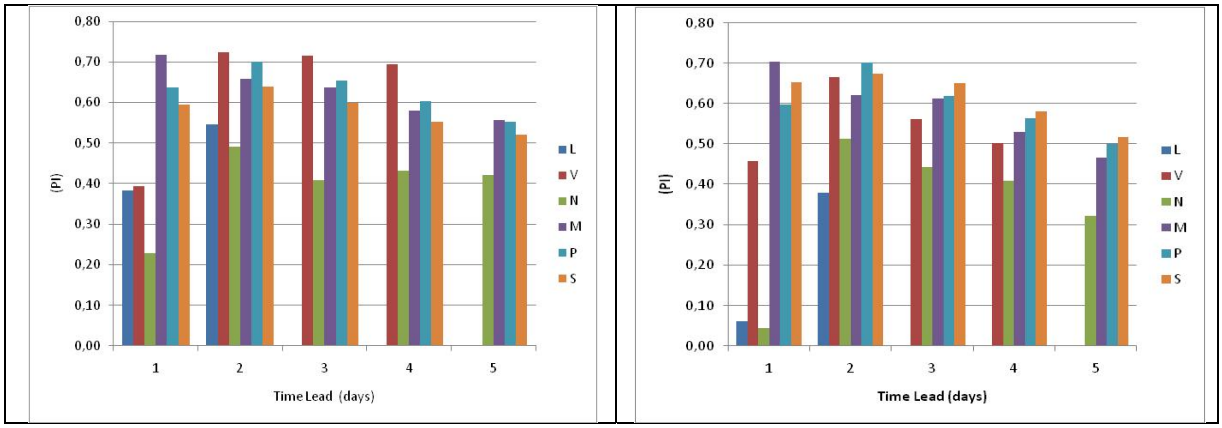


Fig. 8.8: Quality of flood forecast in terms of PI for third case of forecasted rainfall, left column: simulation period, right column: validation period

8.2.4. Results of Flood Forecast by Mix-Model (Method-3 of Case-3 of Type-2 Model in combination with Type-1 Model)

The accuracy of flood forecast reduces with the lead time. Similarly, the errors increase with the forecasting lead time at each forecasting gage. Therefore, a comparative plotting of flood forecasts with the observed flows is discussed separately for 1 to 5 days lead time.

8.2.4.1. Flood forecast of L, V, N, M, P and S with one day lead time

The flood forecasting results of 1 day lead time for L, V, N, M, P and S by Mix-Model are in Figs.8.9 to 8.14. Each figure contains two columns, the left column shows the comparative plotting of observed and forecasted discharges of year 2000, and the right column shows the comparative plotting of discharges for year 2005. The year 2000 is year with the highest floods in the simulation period, and year 2005 is year with the highest floods in the validation period. These figures show that one day forecast is close to perfect in most of the cases. The errors remain near zero, although in some rare cases the error goes up to 2000 m^3/sec.

.

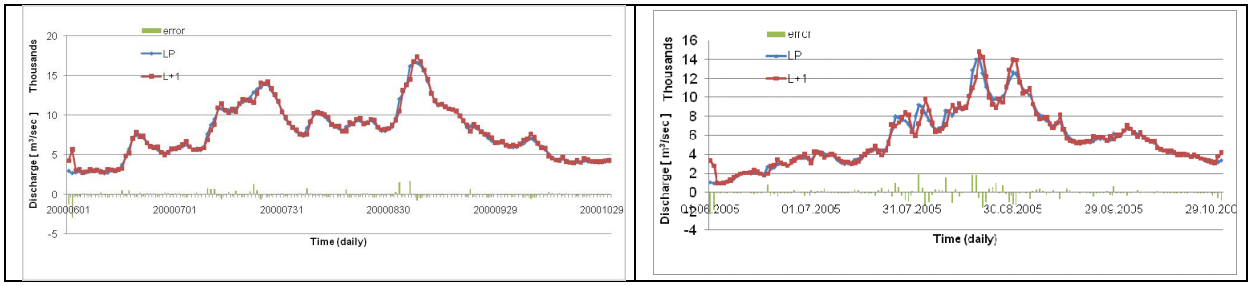


Fig. 8.9: Comparative plotting of observed and 1 day flood forecast of Luang Prabang (L), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

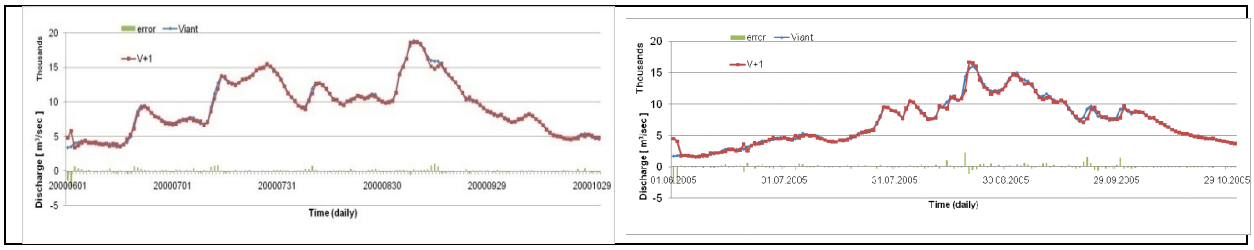


Fig. 8.10: Comparative plotting of observed and 1 day flood forecast of Vientiane (V), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

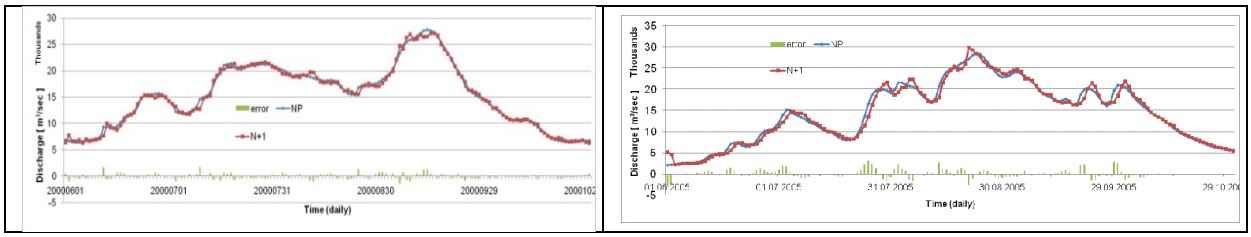


Fig. 8.11: Comparative plotting of observed and 1 day flood forecast of N. Phanom (N), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

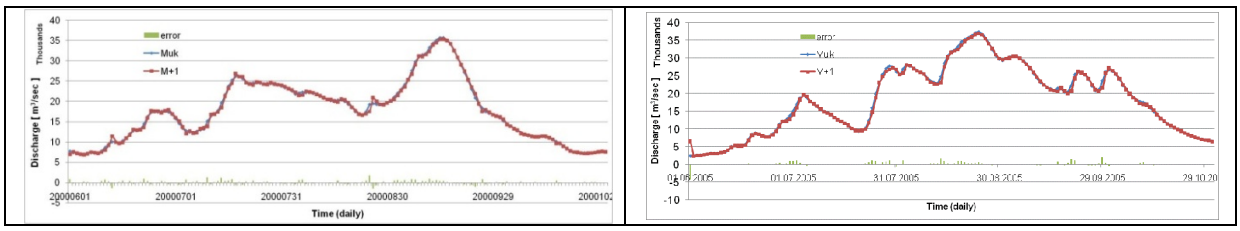


Fig. 8.12: Comparative plotting of observed and 1 day flood forecast of Mukdahan (M), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

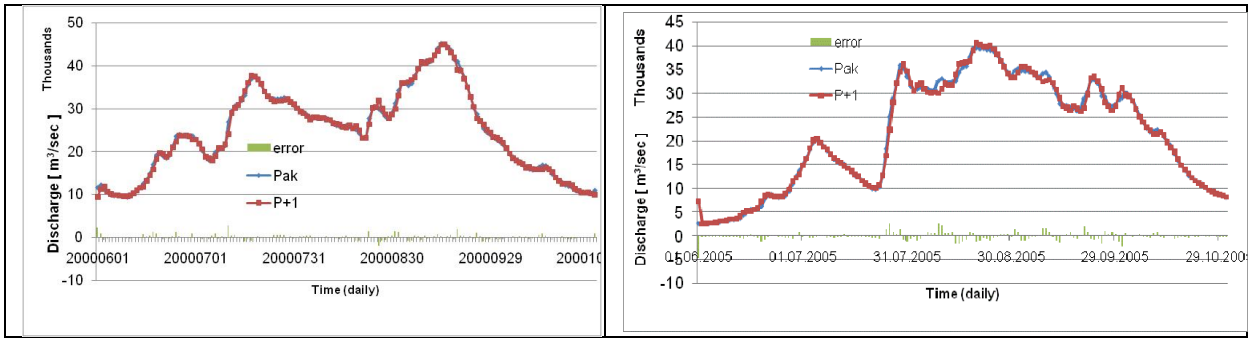


Fig. 8.13: Comparative plotting of observed and 1 day flood forecast of Pakse (P), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

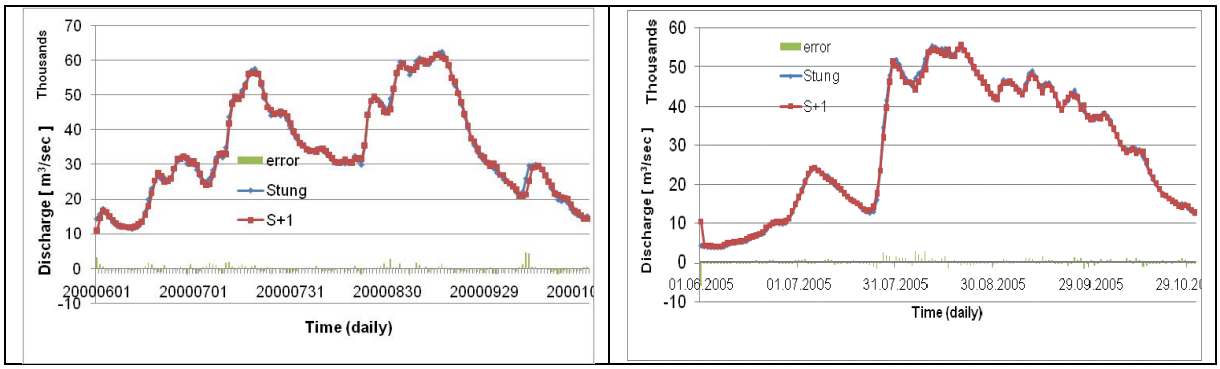


Fig. 8.14: Comparative plotting of observed and 1 day flood forecast of Stung Treng (S), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

8.2.4.2. Flood forecast of L, V, N, M, P and S with two days lead time

Fig-8.15 to 8.20 shows the flood forecasting results of L, V, N, M, P and S for two days lead time. The errors of two days flood forecast go up to 4000 m^3/sec, and they are comparatively larger than 1 day flood forecasts. But in most of cases, the forecasting error still remains close to zero. In general, the errors are observed in the cases of sudden rise of discharges -which seemingly are the result of rainfall occurrences after the forecasting day

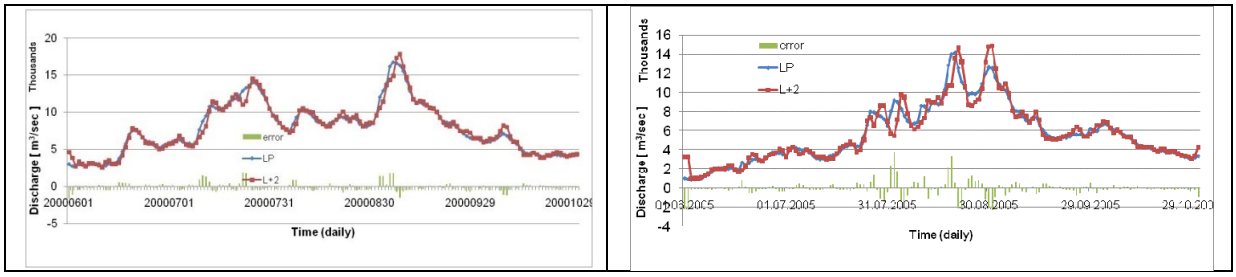


Fig. 8.15: Comparative plotting of observed and 2 day lead time flood forecast of Luang Prabang (L), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

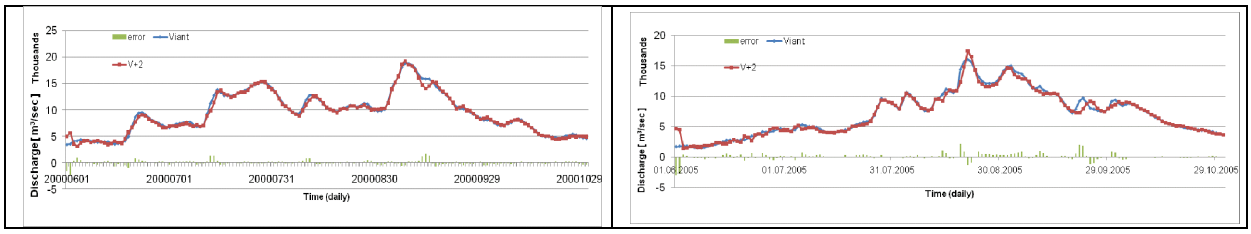


Fig. 8.16: Comparative plotting of observed and 2 day lead time flood forecast of Vientiane (V), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

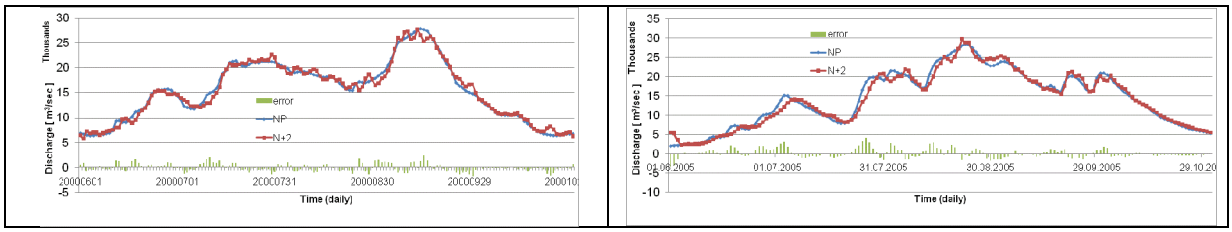


Fig. 8.17: Comparative plotting of observed and 2 day lead time flood forecast of N. Phanom (N), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

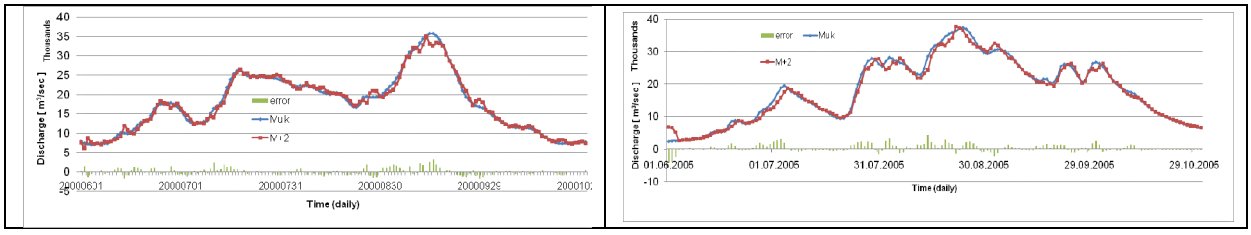


Fig. 8.18: Comparative plotting of observed and 2 day lead time flood forecast of Mukdahan (M), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

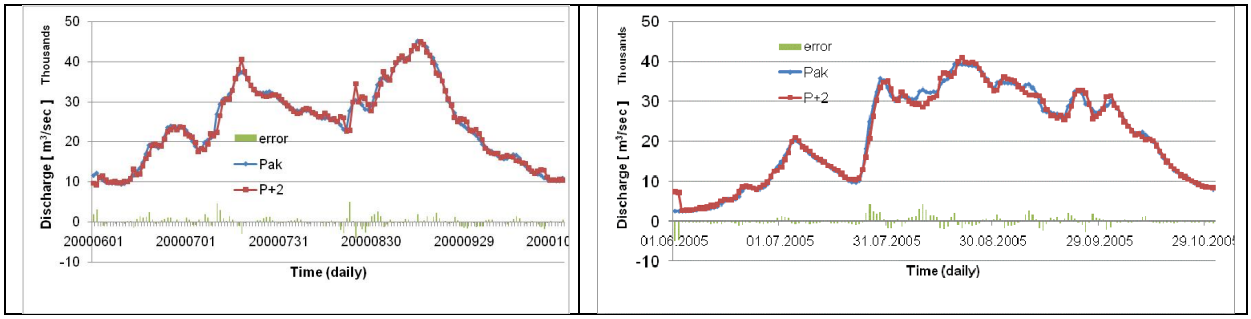


Fig. 8.19: Comparative plotting of observed and 2 day lead time flood forecast of Pakse (P), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

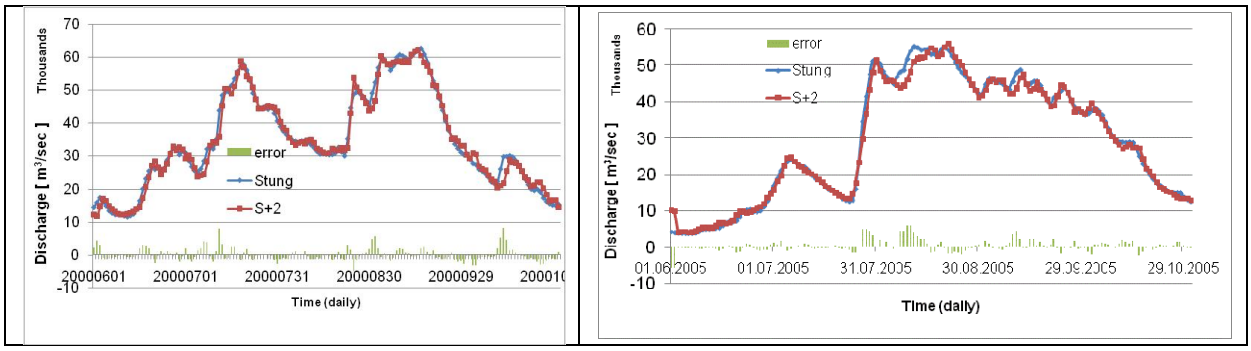


Fig. 8.20: Comparative plotting of observed and 2 day lead time flood forecast of Stung Treng (S), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

8.2.4.3. Flood forecast of V, N, M, P and S with three days lead time

Fig-8.21 to 8.25 shows the flood forecasting results of V, N, M, P and S for three days lead time. The forecasting errors of V, N and M remain under 5000 m^3/sec, however, more of the errors exceed 5000 m^3/ sec although they remain below 10,000 m^3/sec in the case of P and S. Positive errors are observed in the rising limb of the flood hydrographs; and most of the errors are negative in the falling limb. This implies that for three day lead times, forecasts by Mix-Model lag behind the actual observed discharges. This lag is apparently due to large lateral inflows in short times. For example, in the case of the second flood peak at S, the discharge increased from 12000 to 40,000 m^3/sec in just 5 days. However, the Mix-Model was able to forecast a large portion of this discharge and the errors remain under 8000 m^3/sec.

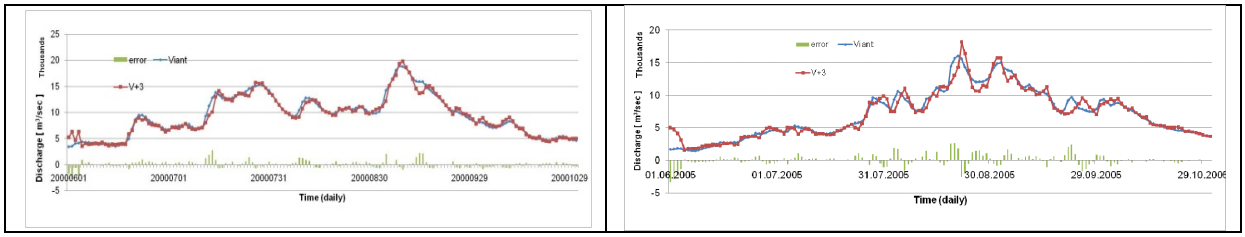


Fig. 8.21: Comparative plotting of observed and 3 day lead time flood forecast of Vientiane (V), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

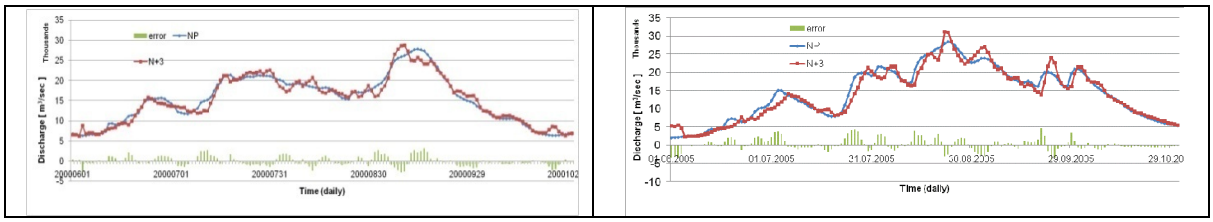


Fig. 8.22: Comparative plotting of observed and 3 day lead time flood forecast of N. Phanom (N), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

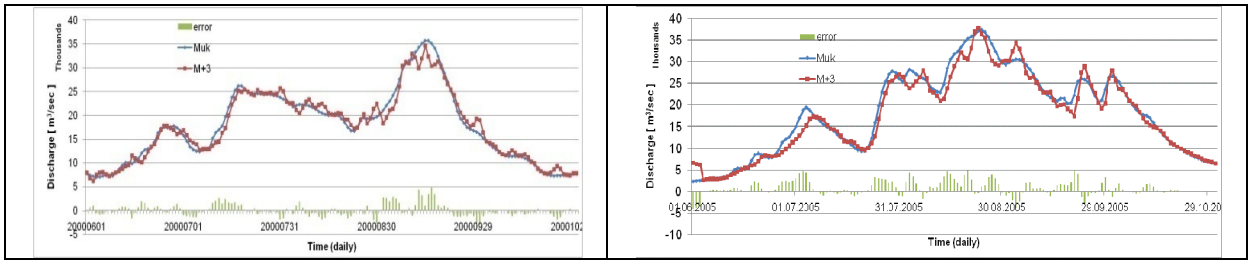


Fig8.23: Comparative plotting of observed and 3 day lead time flood forecast of Mukdahan (M), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

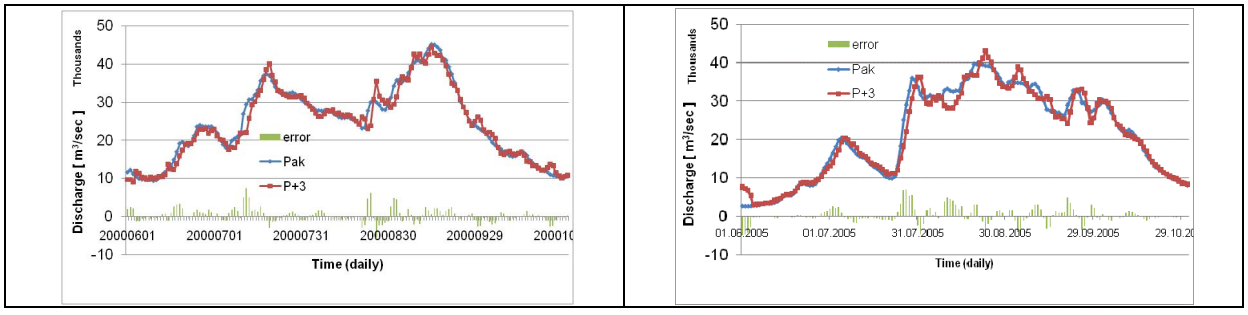


Fig. 8.24: Comparative plotting of observed and 3 day lead time flood forecast of Pakse (P), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

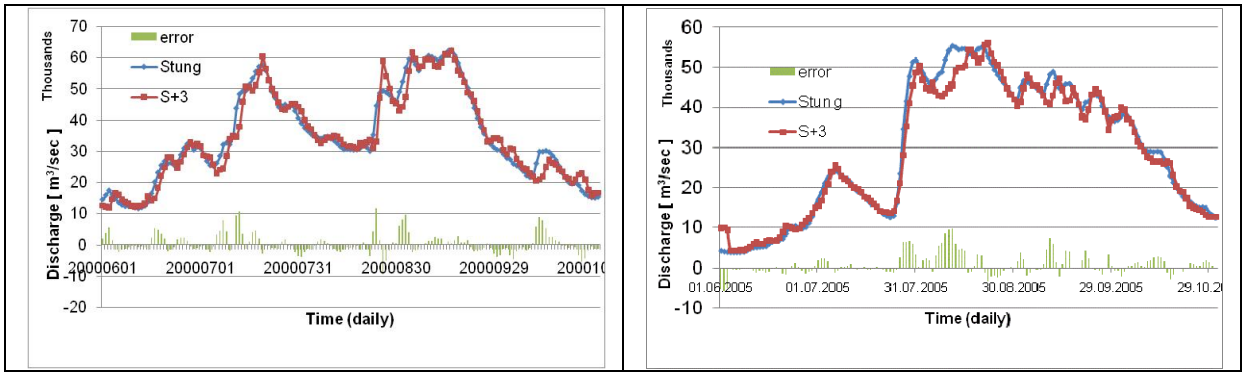


Fig. 8.25: Comparative plotting of observed and 3 day lead time flood forecast of Stung Treng (S), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

8.2.4.4. Flood forecast of V, N, M, P and S with four days lead time

Figs.8.26 to 8.30 shows the flood forecasting results of V, N, M, P and S for four days lead time. The forecasting errors of V, N and M remain under 5000 m^3/sec but in the case of P and S the errors exceed 5000 m^3/ sec. However, in almost all cases they remain below 10,000 m^3/sec.

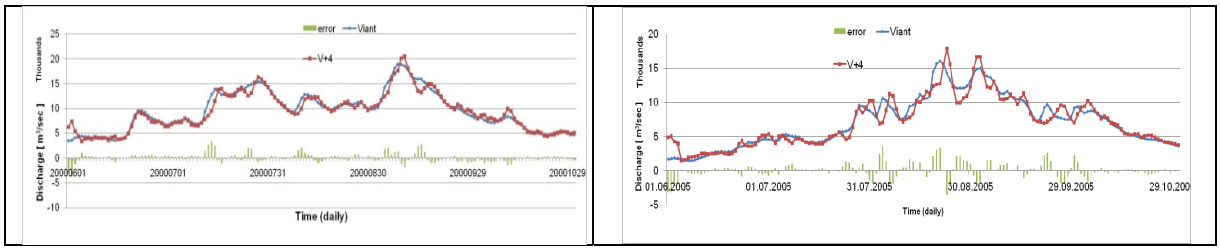


Fig. 8.26: Comparative plotting of observed and 4 day lead time flood forecast of Vientiane (V), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

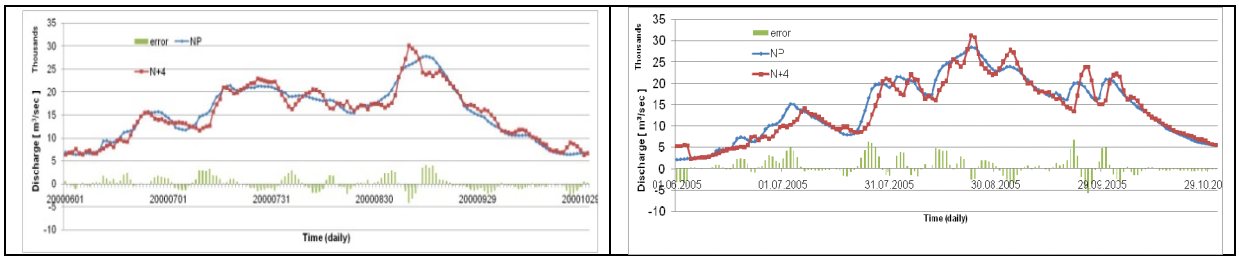


Fig. 8.27: Comparative plotting of observed and 4 day lead time flood forecast of N. Phanom (N), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

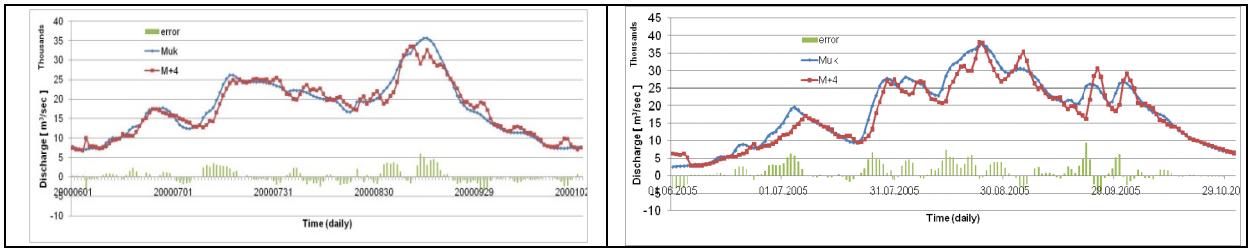


Fig. 8.28: Comparative plotting of observed and 4 day lead time flood forecast of Mukdahan (M), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

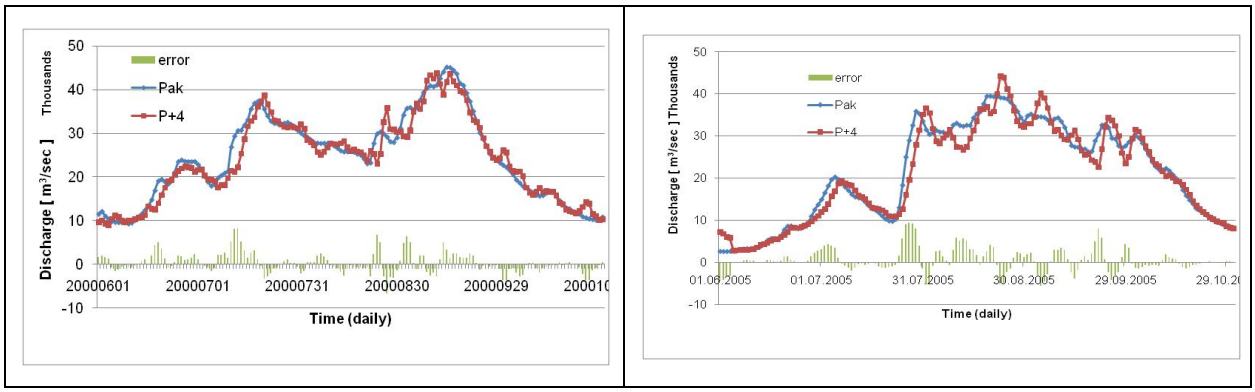


Fig. 8.29: Comparative plotting of observed and 4 day lead time flood forecast of Pakse (P), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

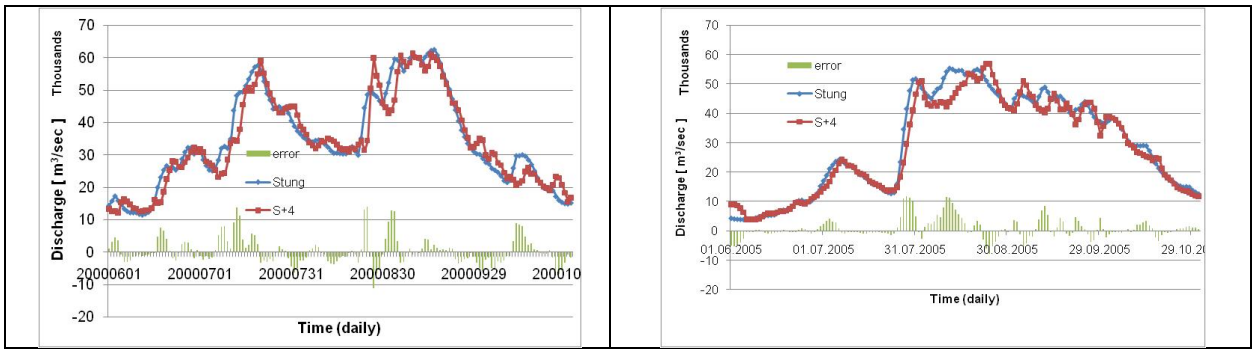


Fig. 8.30: Comparative plotting of observed and 4 day lead time flood forecast of Stung Treng (S), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period.

8.2.4.5. Flood forecast of N, M, P and S with five days lead time

Figs.8.31 to 8.34 show the flood forecasting results of N, M, P and S for five days lead time. In general, the forecasting errors of N, M and P range from 500 to 10,000 m^3/sec, but the

forecasting error of S increased up to 18,000 m^3/ sec in a total discharge of 50,000 m^3/sec. This high error in 5 days lead time strongly affects the average performance of Mix-Model.

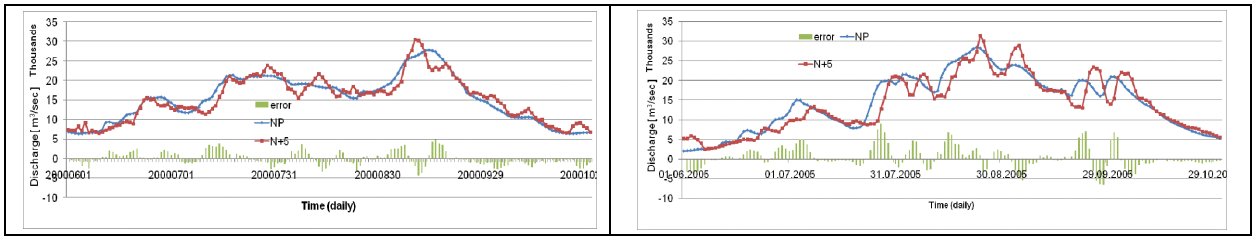


Fig. 8.31: Comparative plotting of observed and 5 day lead time flood forecast of N. Phanom (N), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

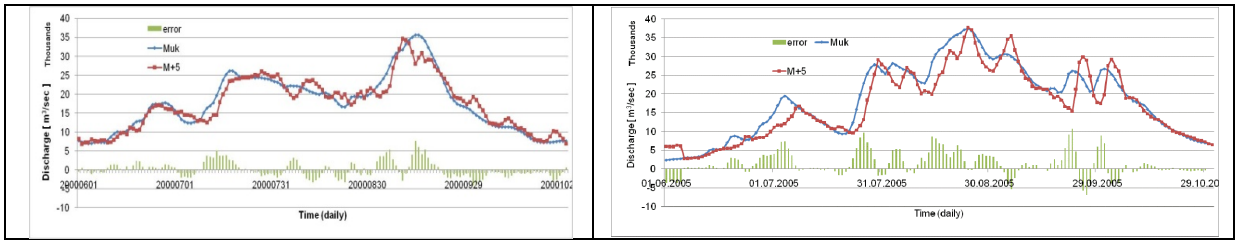


Fig. 8.32: Comparative plotting of observed and 5 day lead time flood forecast of Mukdahan (M), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

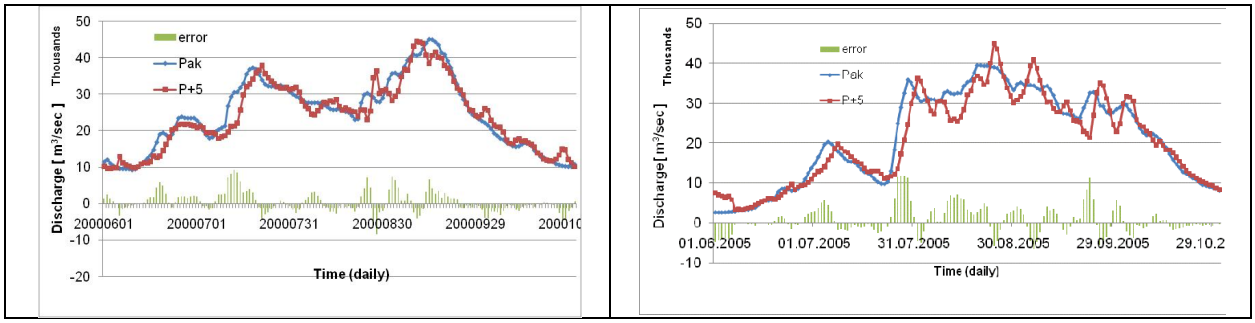


Fig. 8.33: Comparative plotting of observed and 5 day lead time flood forecast of Pakse (P), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

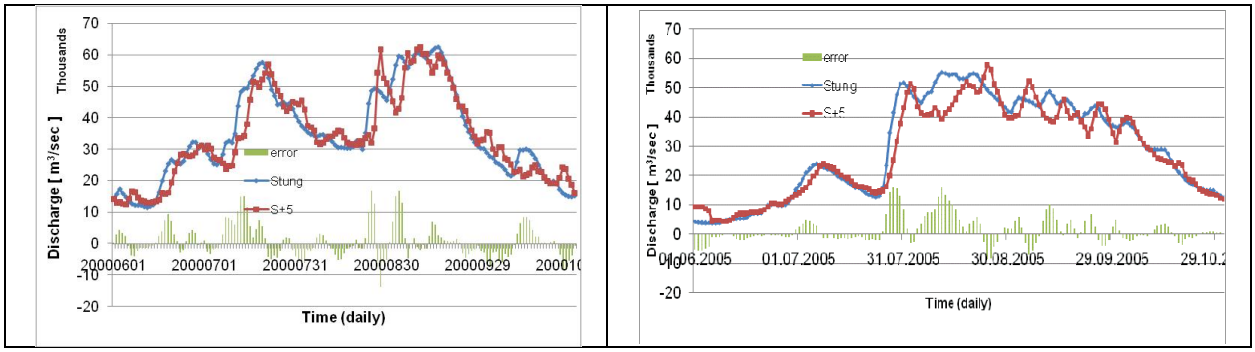


Fig. 8.34: Comparative plotting of observed and 5 day lead time flood forecast of Stung Treng (S), left column: for selected flood season (year 2000) from simulation period, right column: for selected flood season (year 2005) from validation period

8.2.4.6. Distribution of errors for the Flood forecast of 1 to 5 days lead time

Fig. 8.35 shows the probability density distributions of forecasting errors of L, V, N, M, P and S for 1 to 5 day flood forecasts in the simulation and validation periods. The error spread of the validation period is about the same as the error spread of simulation period at each forecasting gage. The error spread ranges from - 6000 to + 6000 m^3/sec for L, V, N and M. However this spread increases to 7000 m^3/sec in the case of P. In the case of S most of the errors remain within +/- 15000 m^3/sec.

8.2.5. Results of Flood Forecast by Mix-Model in terms of Water Levels (Method-3 of Case-3 of Type-2 Model with Type-1 Model)

The forecasting results of 1 to 5 days flood forecast for year 2000 as presented in Fig.8.9-35 are repeated in Fig.8.36-8.61 after converting observed and forecasted discharges into water levels. The New MRC ratings of year 2010 are used to convert discharges because the rating curves for year 2000 were not available to the writer. The reason of plotting observed and forecasted discharges in terms of water level was because the same discharge errors can be different in terms of water level at different stages. Normally, in the case of trapezoidal channels, errors are larger in terms of water depth at low stages for the same discharge, because of narrow channel with low discharge carrying capacity at low levels, which become large at high stages because of high discharge carrying capacity of the flood cross section.

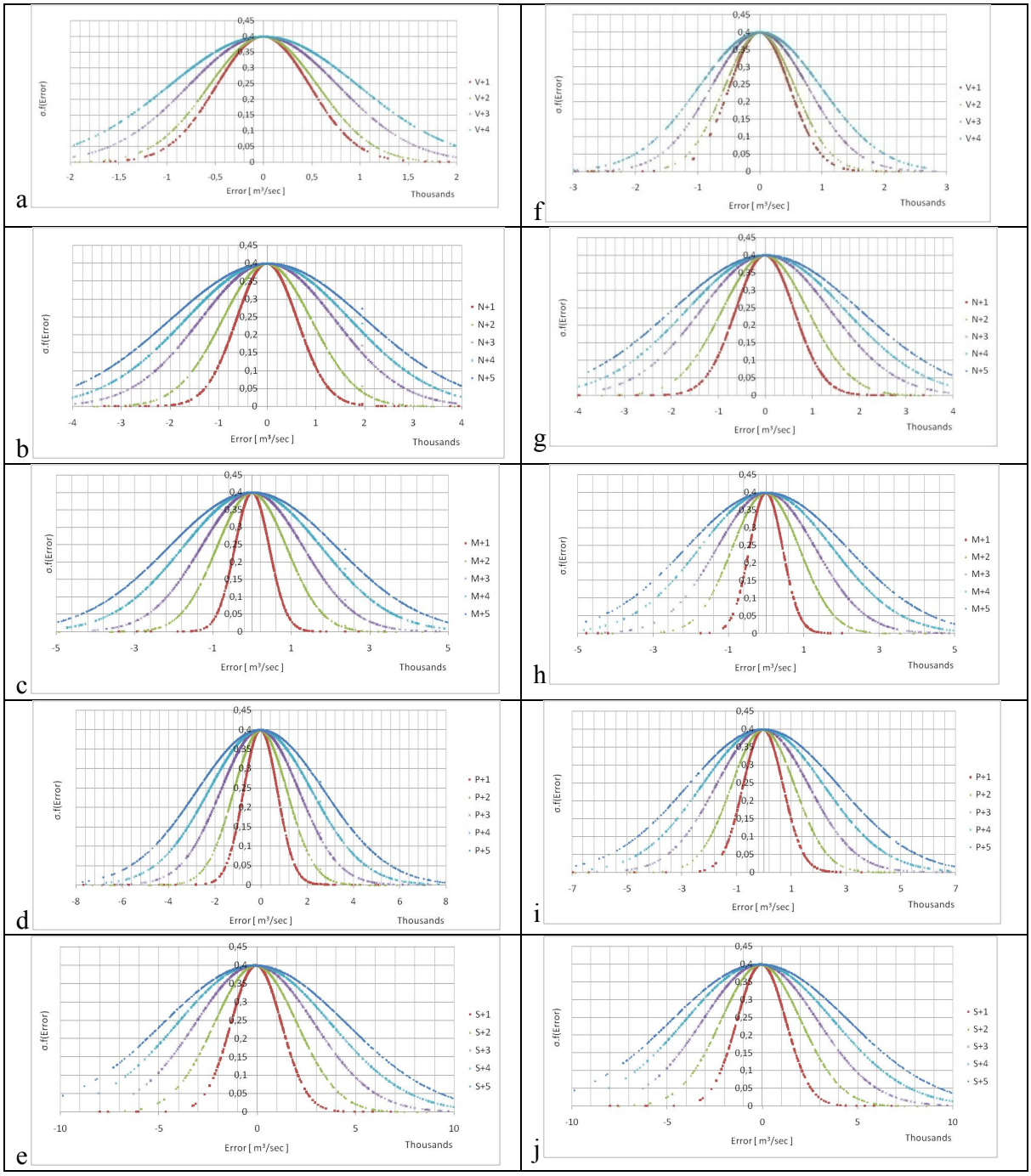


Fig. 8.35: Probability distribution of errors of 1 to 5 day flood forecast, left column: forecasting errors of L, V, N, M, P and S (a,b,c,d,e) in the simulation period, right column: forecasting errors of, V, N, M, P and S (f,g,h,i,j) in the validation period

Due to the fact that flood discharges correspond to high stages large discharge errors can be small in terms of water level. In order to analyse the behavior of discharge errors in terms of water depth a comparative plotting of observed and forecasted water levels is essential.

Further, flood forecasts are more meaningful in terms of water level to determine the critical and overflow level of the channel, whereas discharges are important to estimate the flood magnitude.

8.2.5.1. Flood forecast of L, V, N, M, P and S with one day lead time

The flood forecasting results of 1 day lead time for L, V, N, M, P and S by Mix-Model in terms of water level are presented in Fig-8.36 to 8.41. These figures show that one day forecasts are close to perfect because in most of the cases the error remains near zero, however in some few cases the error goes up to 30 to 50 cm.

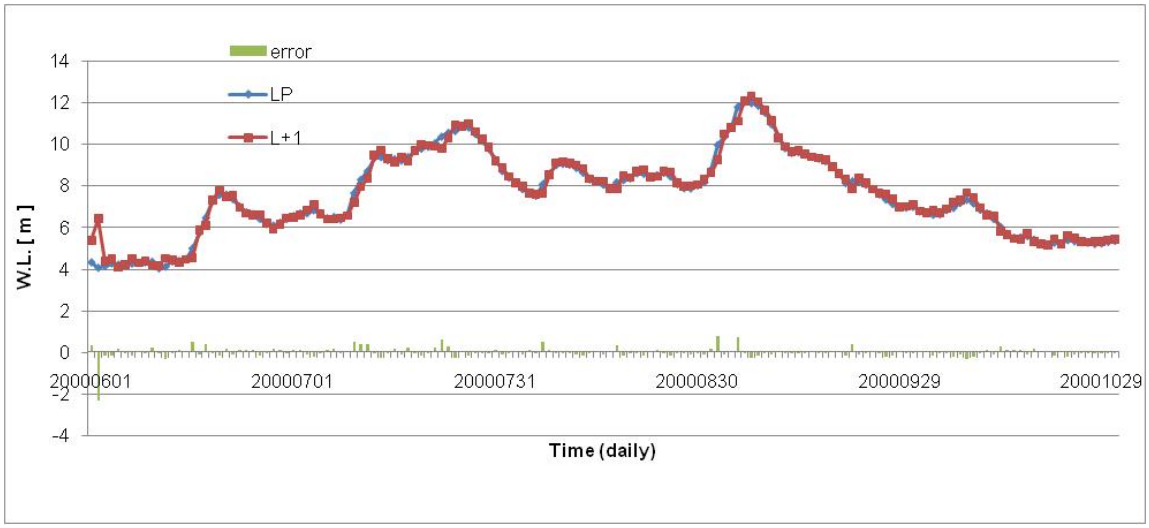


Fig. 8.36: Comparative plotting of observed and 1 day flood forecast (in terms of water level) of Luang Prabang (L)

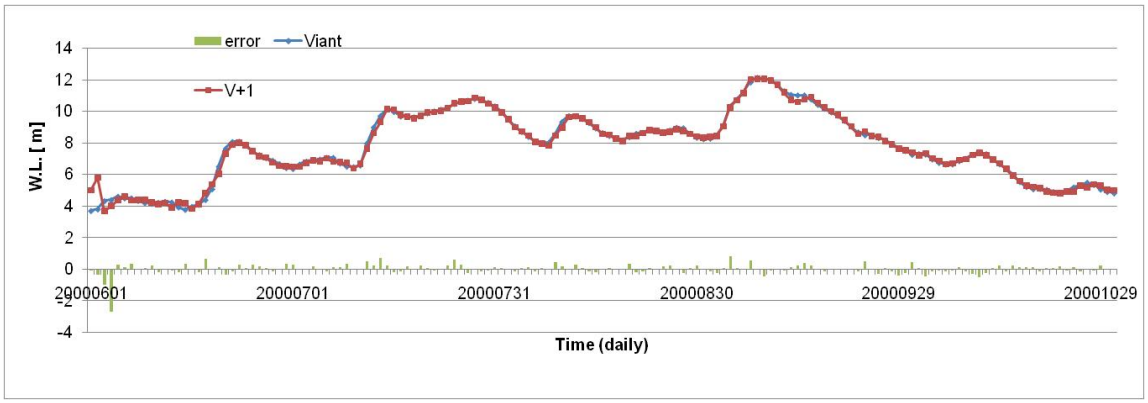


Fig. 8.37: Comparative plotting of observed and 1 day flood forecast (in terms of water level) of Vientiane (V)

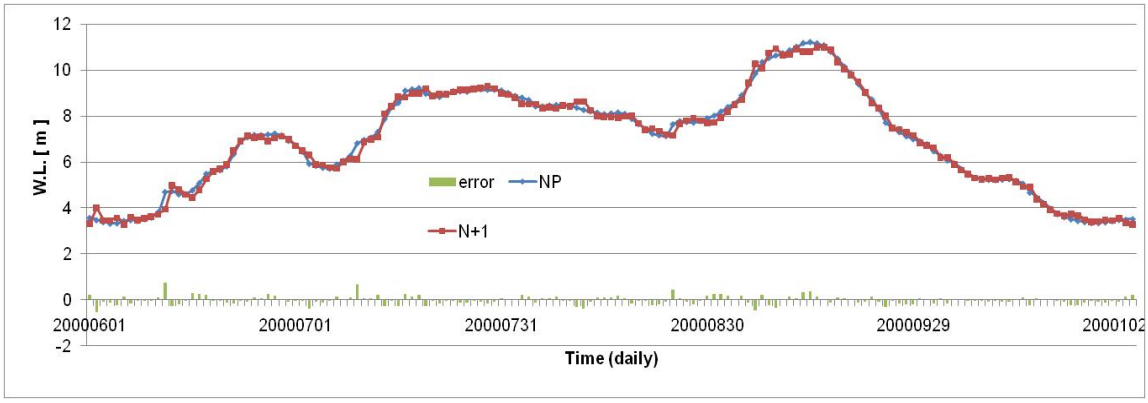


Fig. 8.38: Comparative plotting of observed and 1 day flood forecast (in terms of water level) of N. Phanom (N)

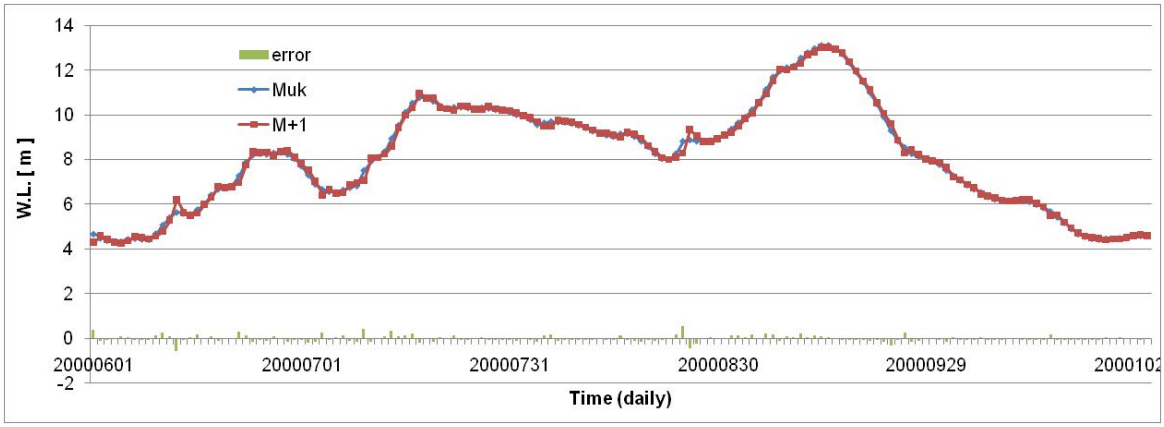


Fig. 8.39: Comparative plotting of observed and 1 day flood forecast (in terms of water level) of Mukdahan (M)

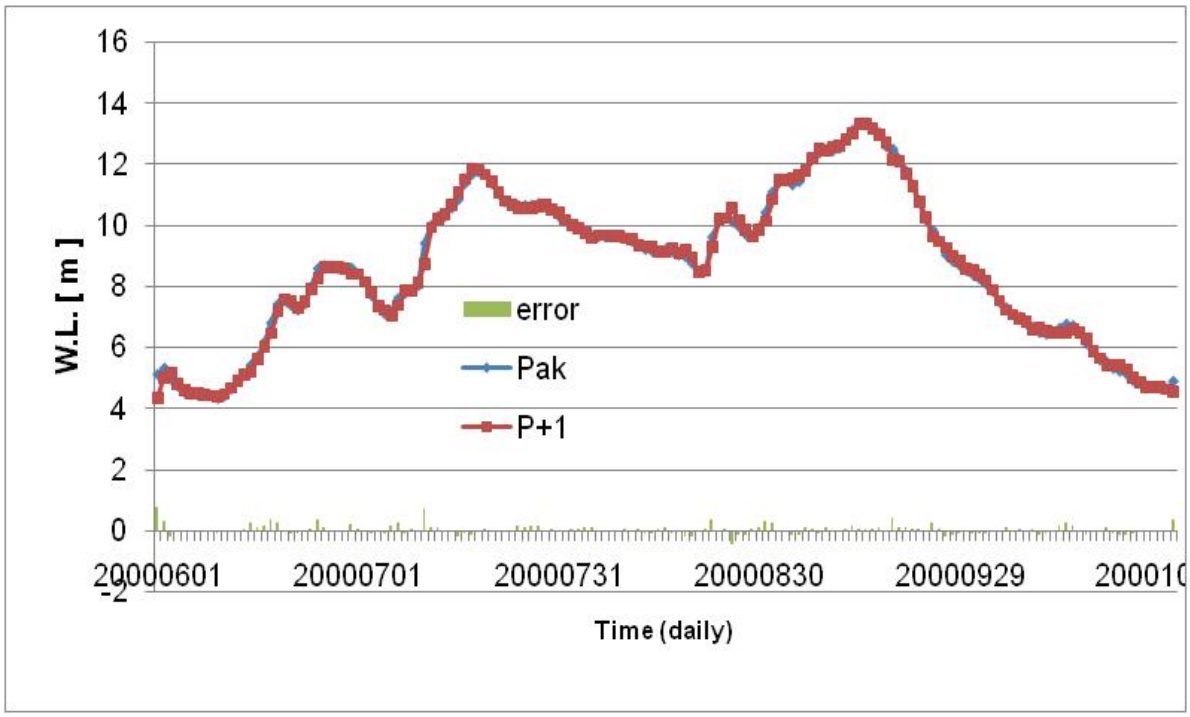


Fig. 8.40: Comparative plotting of observed and 1 day flood forecast (in terms of water level) of Pakse (P)

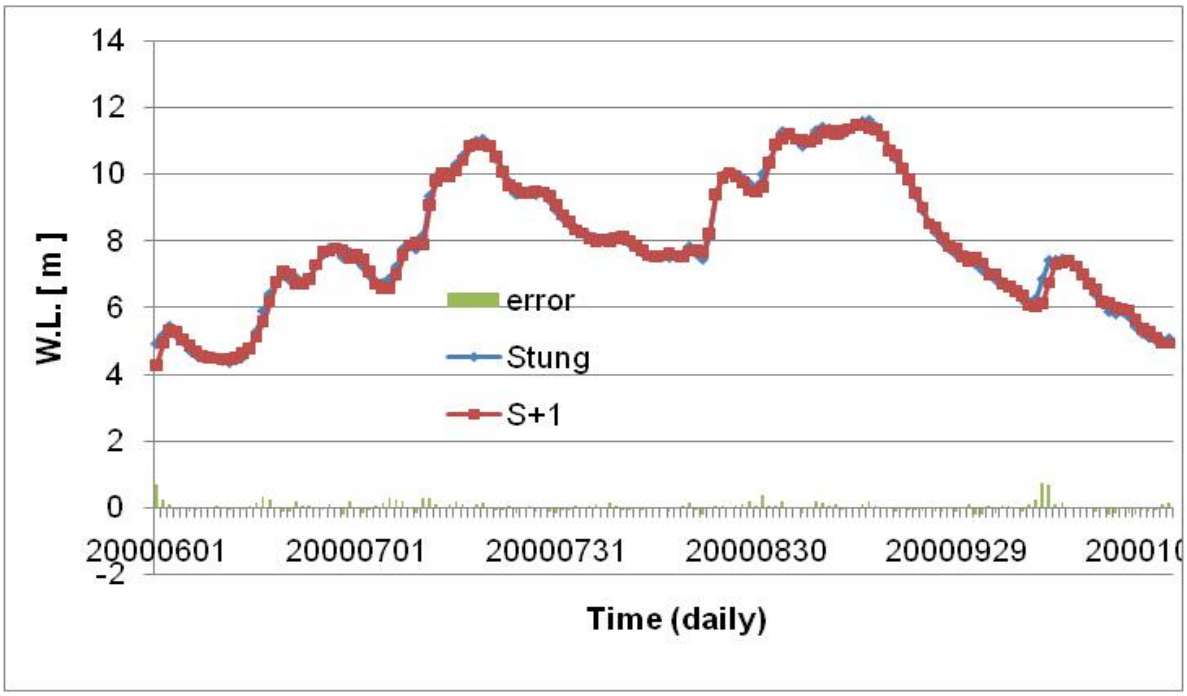


Fig. 8.41: Comparative plotting of observed and 1 day flood forecast (in terms of water level) of Stung Treng (S)

8.2.5.2. Flood forecast of L, V, N, M, P and S with two days lead time

Fig-8.42 to 8.47 shows the flood forecasting results of L, V, N, M, P and S for two days lead time in terms of water level. The errors of two days flood forecasts are also only a few cm in most of the cases, however, in some cases the error shoots up to 90 cm.

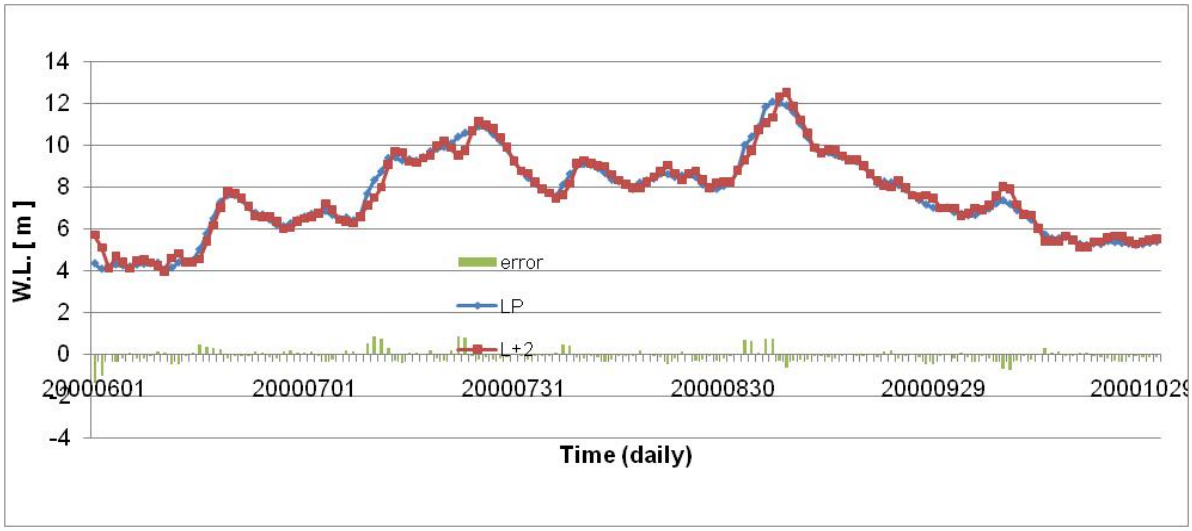


Fig. 8.42: Comparative plotting of observed and 2 day lead time flood forecast (in terms of water level) of Luang Prabang (L

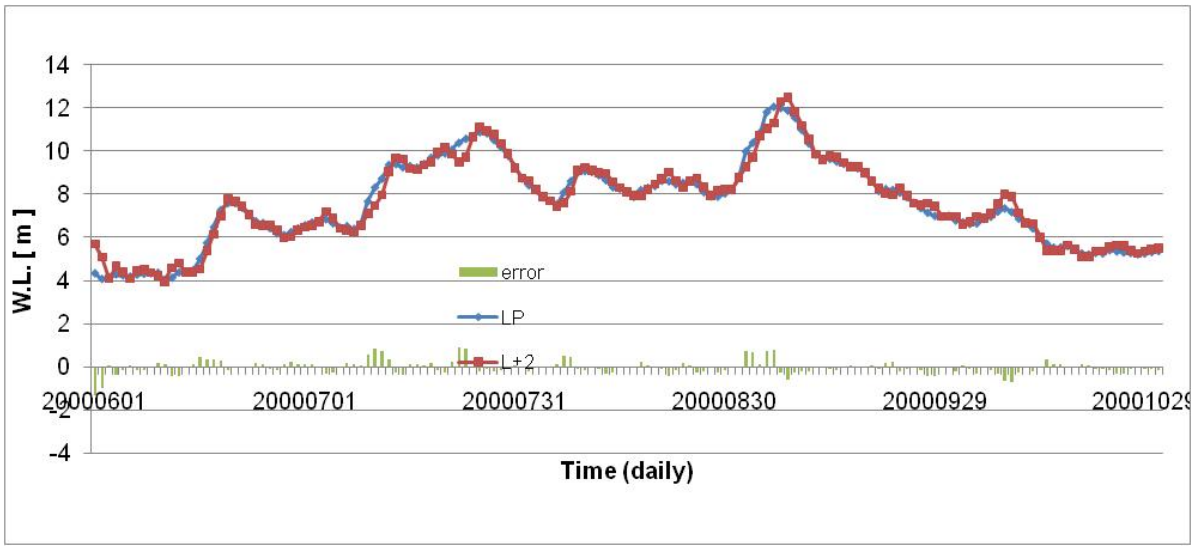


Fig. 0-18.43: Comparative plotting of observed and 2 day lead time flood forecast (in terms of water level) of Vientiane (V)

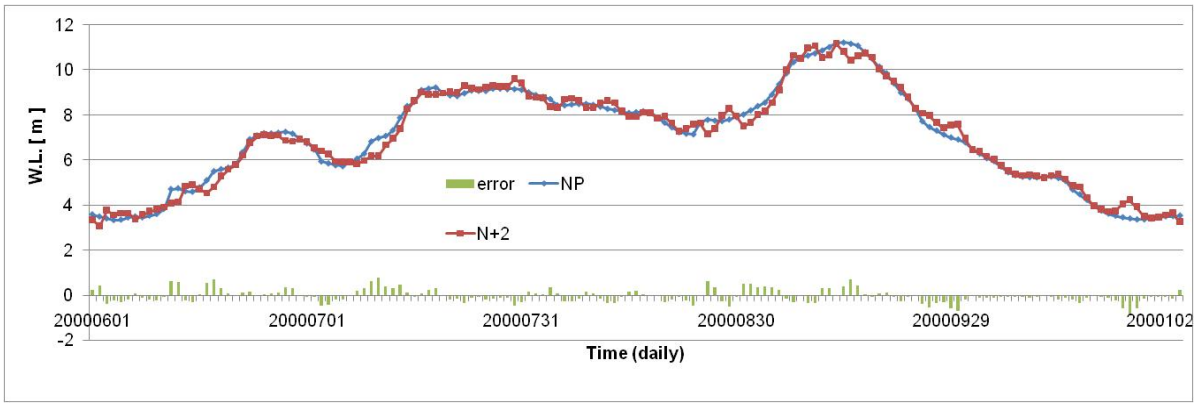


Fig8.44: Comparative plotting of observed and 2 day lead time flood forecast (in terms of water level) of N. Phanom (N)

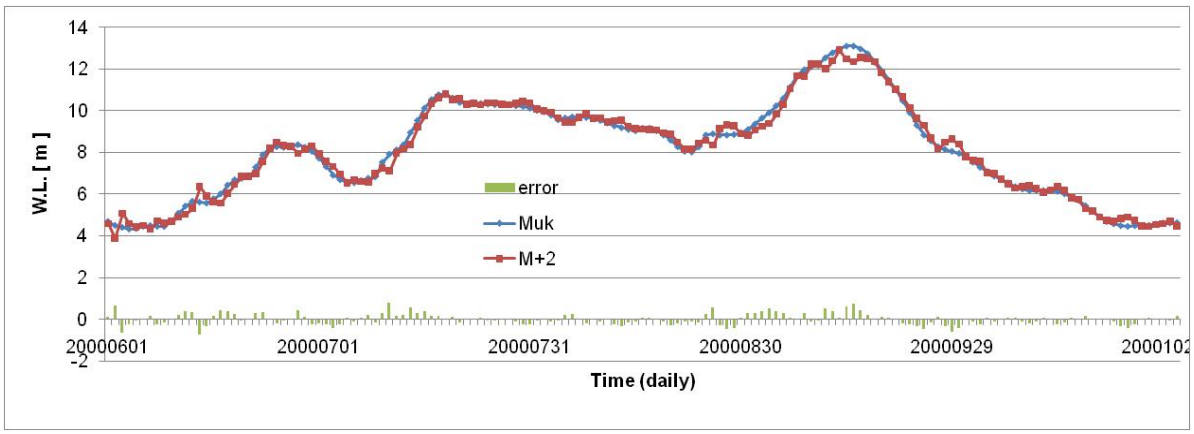


Fig. 8.45: Comparative plotting of observed and 2 day lead time flood forecast (in terms of water level) of Mukdahan (M)

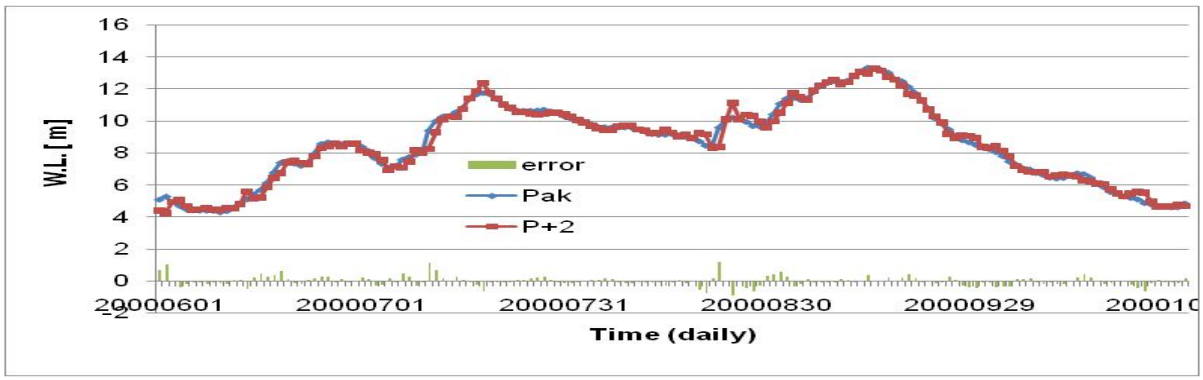


Fig. 8.46: Comparative plotting of observed and 2 day lead time flood forecast (in terms of water level) of Pakse (P)

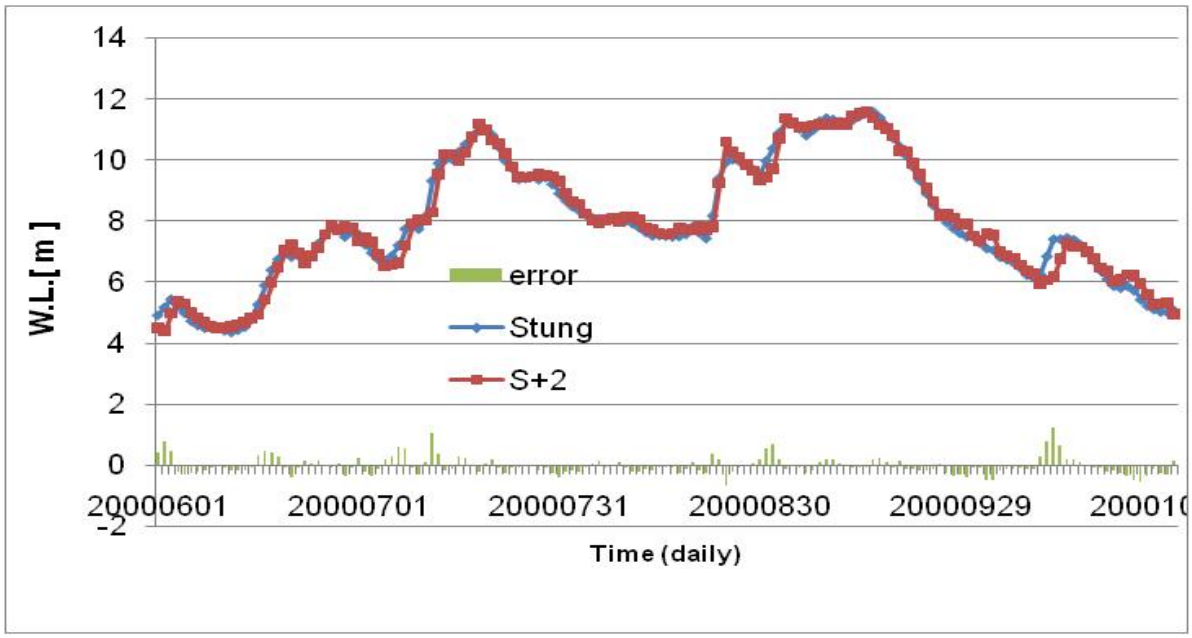


Fig. 8.47: Comparative plotting of observed and 2 day lead time flood forecast (in terms of water level) of Stung Treng (S)

8.2.5.3. Flood forecast of V, N, M, P and S with three days lead time

Figs.8.48 to 8.52 show flood forecasting results of V, N, M, P and S in terms of water level for three days lead time. In general, the forecasting errors of V, N and M remain under 1 m but errors may exceed 1 m in some rare cases at P and S. High errors are generally observed in the rising part of a hydrograph and not at the higher water levels corresponding to discharge peaks, which is useful because the flood forecast becomes more important at high stages.

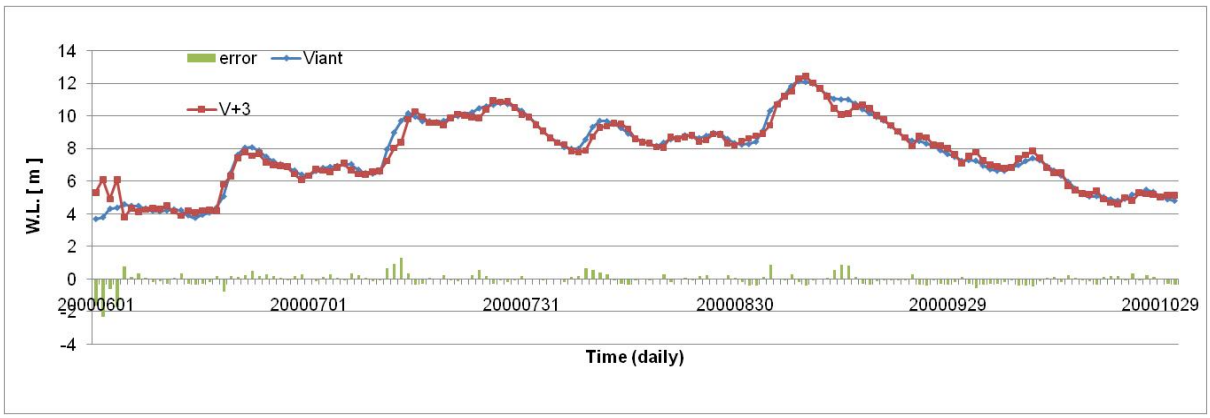


Fig. 8.48: Comparative plotting of observed and 3 day lead time flood forecast (in terms of water level) of Vientiane (V)

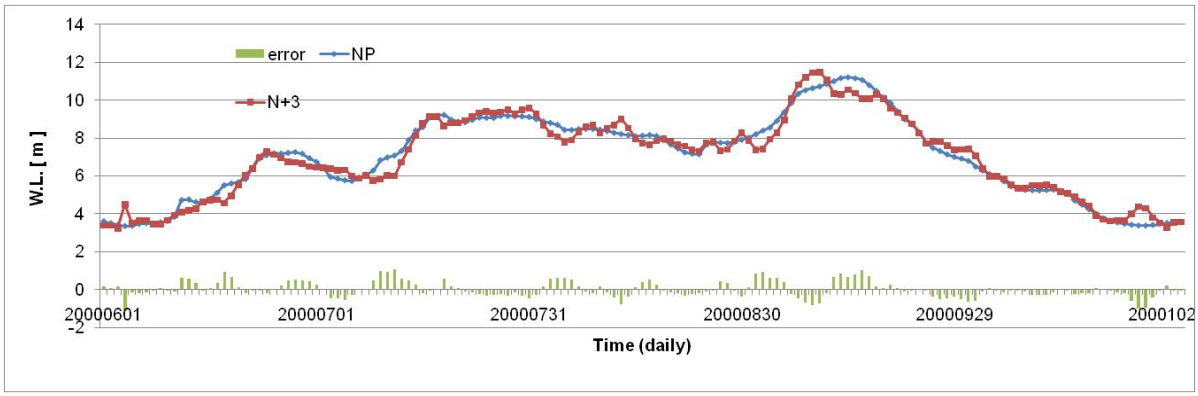


Fig. 8.49: Comparative plotting of observed and 3 day lead time flood forecast (in terms of water level) of N. Phanom (N)

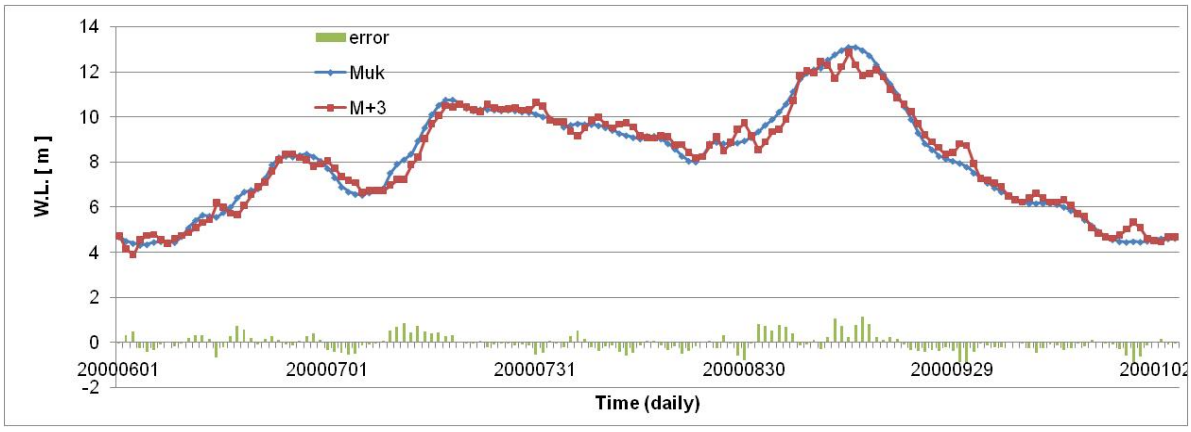


Fig. 8.50: Comparative plotting of observed and 3 day lead time flood forecast of Mukdahan (M)

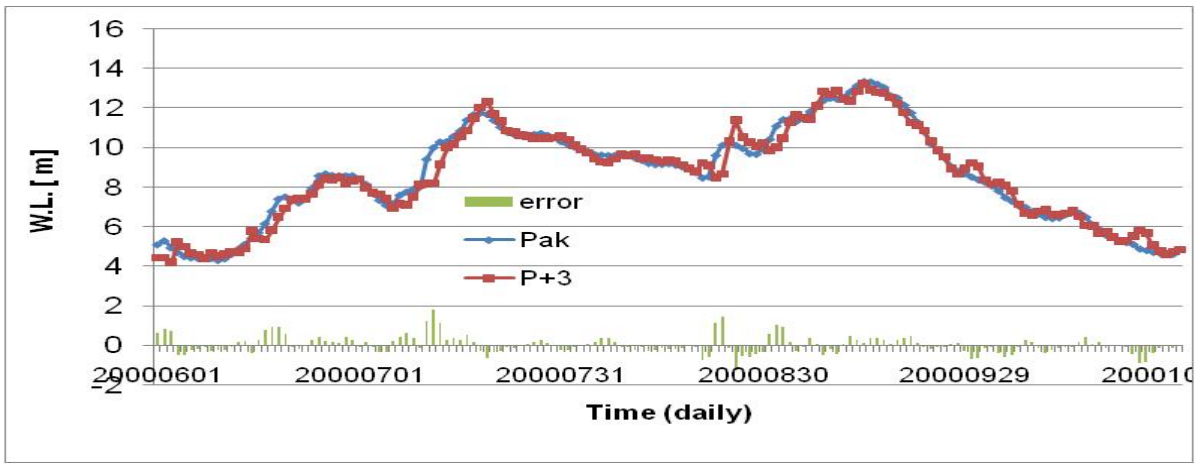


Fig. 8.51: Comparative plotting of observed and 3 day lead time flood forecast (in terms of water level) of Pakse (P)

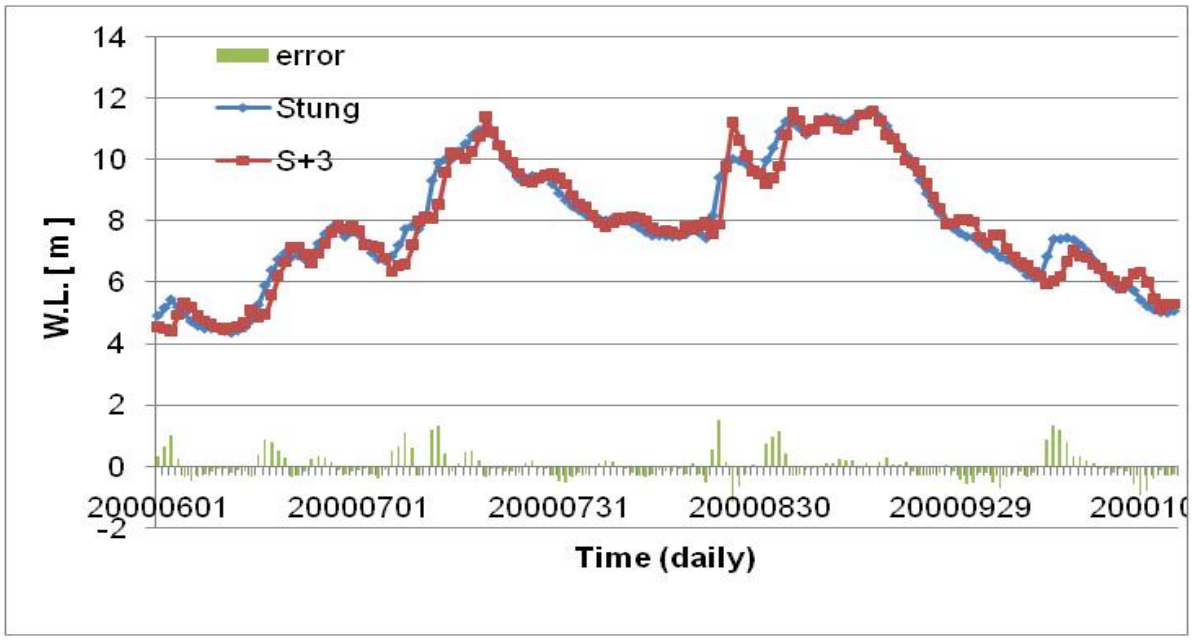


Fig. 8.52: Comparative plotting of observed and 3 day lead time flood forecast (in terms of water level) of Stung Treng (S)

8.2.5.4. Flood forecast of V, N, M, P and S with four days lead time

Fig-8.53 to 8.57 shows the flood forecasting results of V, N, M, P and S for four days lead time in terms of water level.. Forecasting errors of V, N and M remain between 0 to 150 cm but in the case of P and S the errors can exceed 150 cm and may go up to as high as 180 cm. Again the high errors are not exactly at the flood peak, but rather in the rising part of the flood hydrograph.

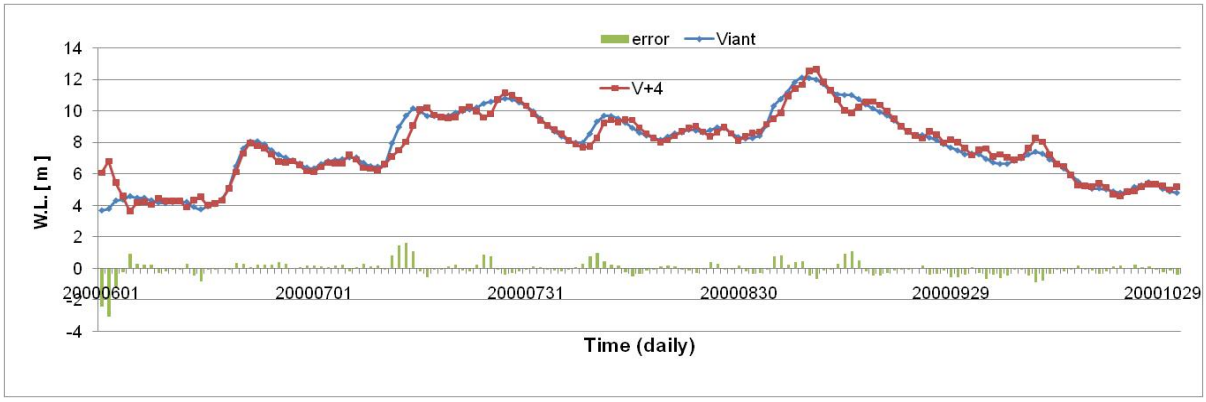


Fig. 8.53: Comparative plotting of observed and 4 day lead time flood forecast (in terms of water level) of Vientiane (V)

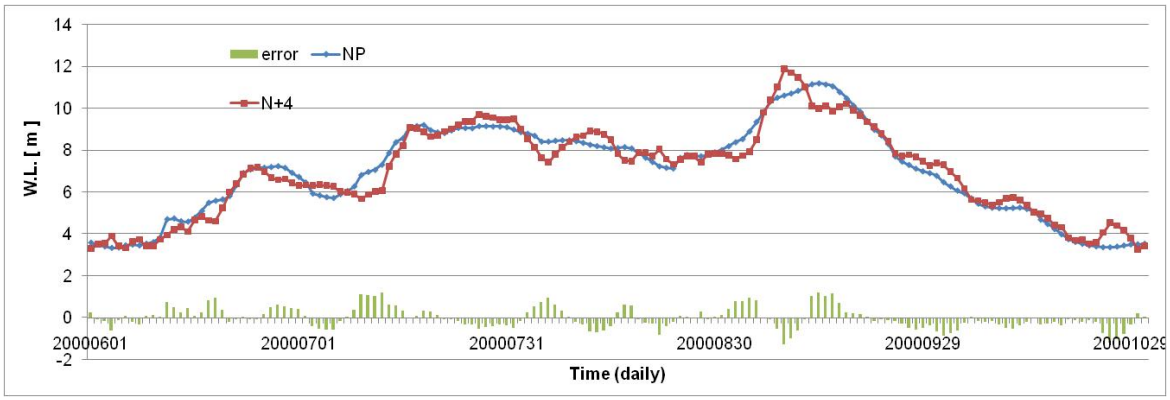


Fig. 8.54: Comparative plotting of observed and 4 day lead time flood forecast (in terms of water level) of N. Phanom (N)

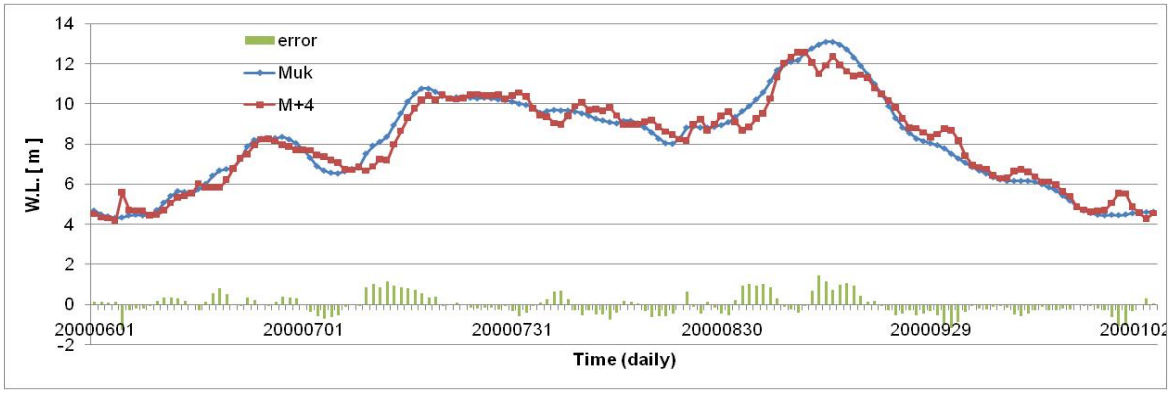


Fig. 8.55: Comparative plotting of observed and 4 day lead time flood forecast (in terms of water level) of Mukdahan (M)

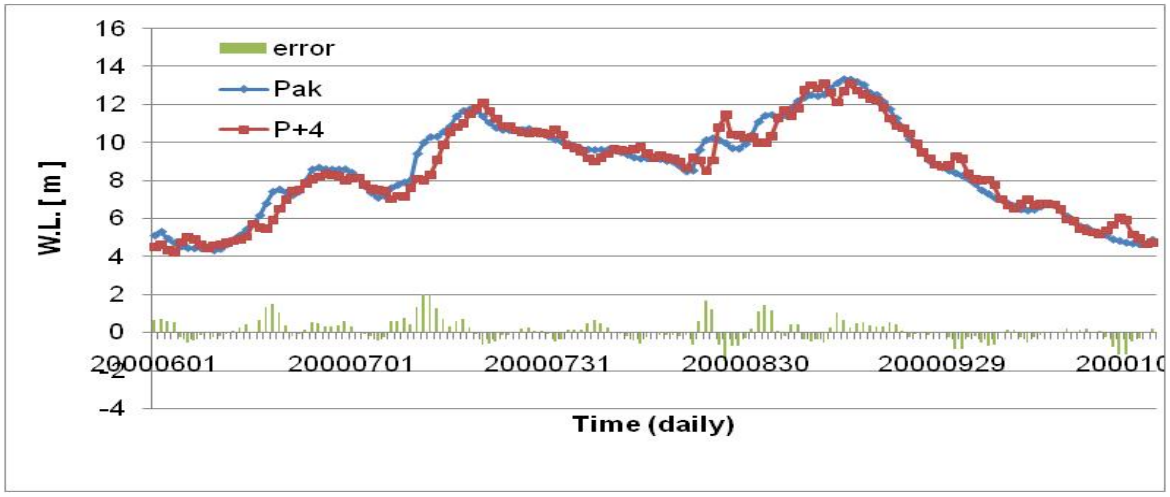


Fig. 8.56: Comparative plotting of observed and 4 day lead time flood forecast (in terms of water level) of Pakse (P)

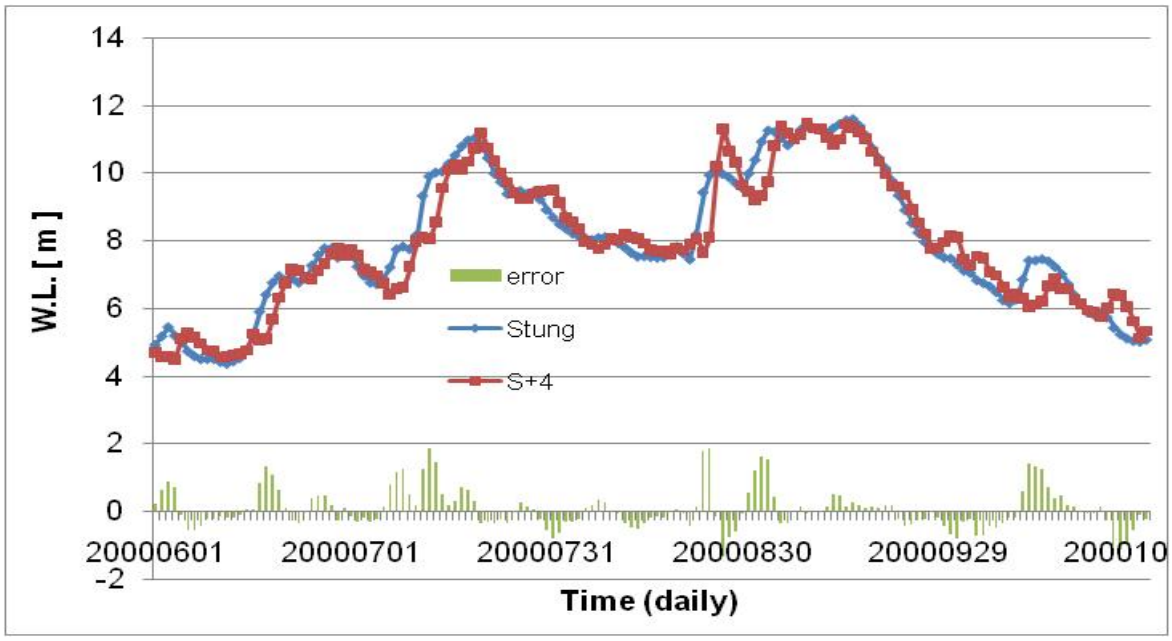


Fig. 8.57: Comparative plotting of observed and 4 day lead time flood forecast (in terms of water level) of Stung Treng (S)

8.2.5.5. Flood forecast of N, M, P and S with five days lead time

Figs. 858 to 8.61 shows the flood forecasting results of N, M, P and S for five days lead time in terms of water level. Forecasting errors of N, M, P and S range from 0 to 2 m. The typical behavior in 5 days forecast is that positive errors are followed by positive errors, and negative errors by negative errors. The possible reason of this continuous under or overestimation could be the rainfall input for 5 days, because instead of using actual rainfall, a repetition of the current day rainfall has been used as an estimate for future 5 days rainfall in the forecast by means of the Type-2 Model.

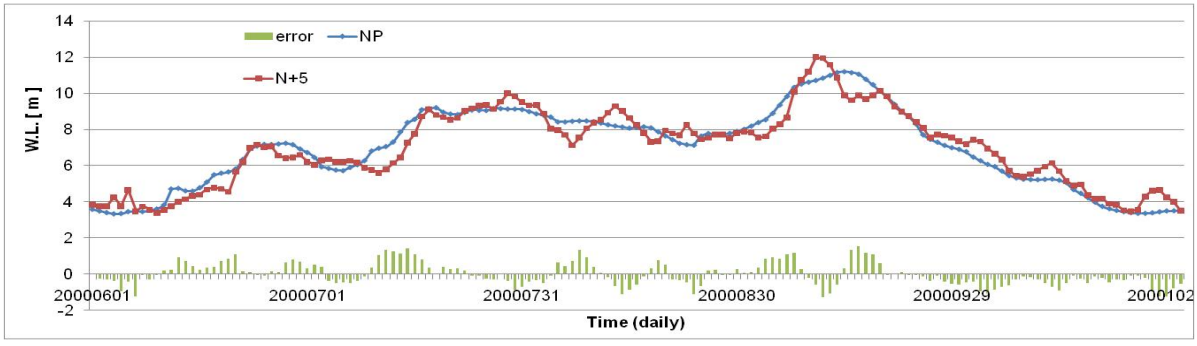


Fig. 8.58: Comparative plotting of observed and 5 day lead time flood forecast (in terms of water level) of N. Phanom (N)

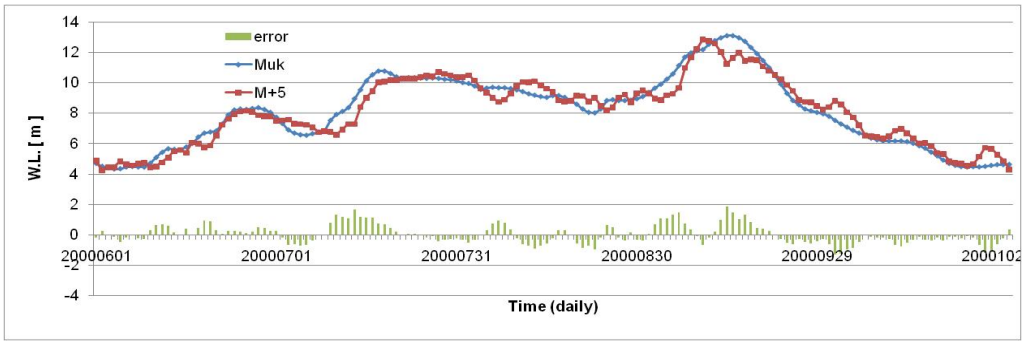


Fig. 8.59: Comparative plotting of observed and 5 day lead time flood forecast (in terms of water level) of Mukdahan (M)

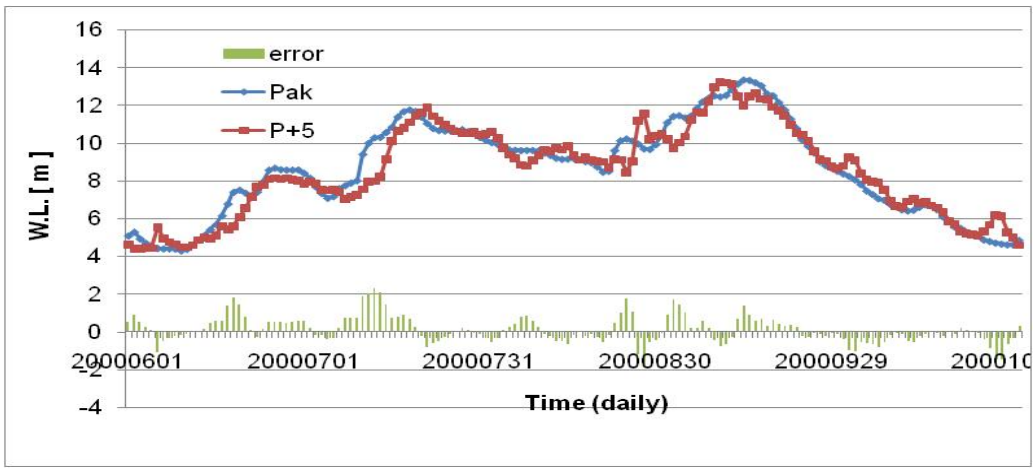


Fig. 8.60: Comparative plotting of observed and 5 day lead time flood forecast (in terms of water level) of Pakse (P)

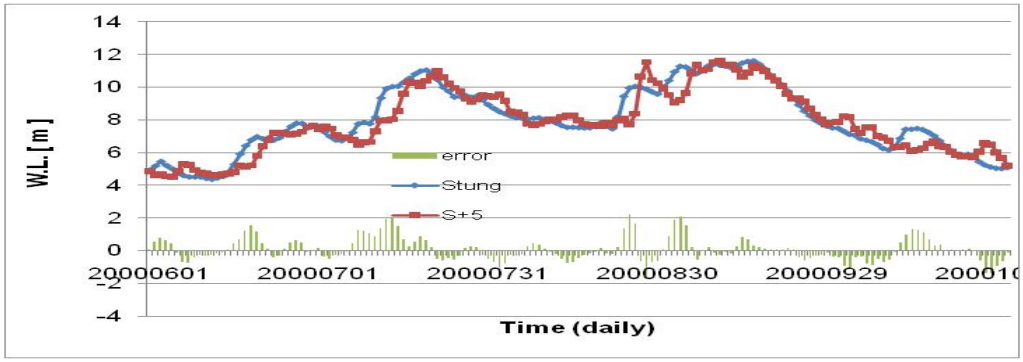


Fig. 8.61: Comparative plotting of observed and 5 day lead time flood forecast (in terms of water level) of Stung Treng (S)

It is apparent from Figs.8.36-61 that the errors do not remain constant during the flood season. Apart from comparatively more errors in the rising limb and comparatively less errors directly

under the flood peak, there is no relation between the stage and the errors. Therefore, it is useful to compute the mean absolute errors as shown in Fig.8.62. The mean absolute errors in 1 day forecast remain close to 1 cm for the downstream gauges, i.e. N, M, P and S but the errors are high in terms of head at L and V, i.e. close to 10 cm. The errors in 2 days lead time forecasts remains under 10 cm and in 4 days the errors ranges in between 38 to 45 cm. The maximum error, of the order of 40 to 52 cm can be seen for 5 days lead time forecasts. But interestingly, in 4 and 5 days forecasts, there are fewer errors in the upstream gauges in comparison to the downstream gauges.

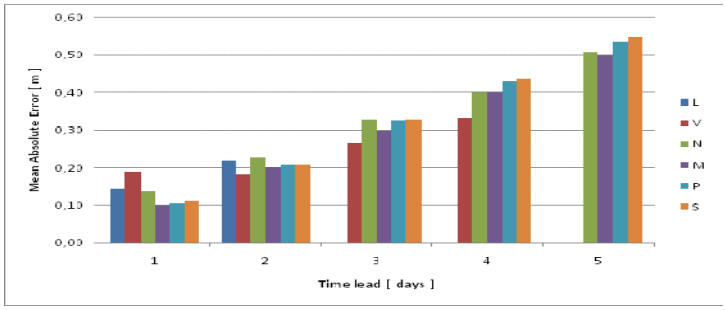


Fig.8.62: Mean absolute errors of 1 to 5 days flood forecast for L, V, N, M, P and S for year 2000 flood

8.3. Conclusion and Discussion

The 1 to 5 days flood forecast of C, L, V, N, M, P and S has been improved by taking the weighted average of the forecasts of Type-1 and Type-2 Models. The reason of this improvement was the complementary nature of the errors of flood forecast produced by Type-1 and Type-2 Model. The improvement in flood forecast can be seen by comparing the Tables-6.5, 7.12 and 7.13 with Tables-8.2 and 8.3. The standard deviations of errors from different variants of the Mix Model are compared with the errors of no model forecast in Tables 8.2 and 8.3 for simulation and validation periods, respectively. These Tables summarize the model's performance in comparison to no forecast at all (i.e. assuming that today´s value is the same as for the value at lead time t in the future) by presenting standard deviations of errors. The first two cases represent the error statistics of Mix-Model, where the forecasting errors of Type-2 Model are produced by using observed rainfall with calculated KN and observed rainfall with forecasted KN respectively. The last three cases present the forecasting error of Type-2 Model with three different methods of making rainfall forecasts, as discussed in Chapter-7. The comparison of first two cases of Mix-Model forecasts with no model forecasts indicate that the standard deviation of errors is reduced by 50% from no Model forecast to first two cases of Mix-Modeling forecasts. This demonstrates that the standard deviation of errors can be reduced to half for almost all gages, if the actual future rainfall is available. It also indicates the limit of Mix-model flood forecasting efficiency in the cases of perfect rainfall input

Tab. 8.2: Standard deviations (in m³/s) of various cases of flood forecast (for simulation period)

Gauge	Without forecast	Observed P and calculated KN	Observed P and forecasted KN	Observed P and forecasted KN (last 5 days avg.)	Observed P and forecasted KN (last 5 days time shift)	Observed P and forecasted KN (constant rainfall propagation)
L+1	640	505	505	502	507	502
L+2	1101	729	732	754	768	740
V+1	623	485	484	484	485	484
V+2	1081	568	566	569	574	567
V+3	1461	770	773	794	804	779
V+4	1778	931	940	1015	1027	984
N+1	712	630	629	628	636	627
N+2	1288	931	924	943	993	921
N+3	1799	1360	1358	1462	1514	1383
N+4	2248	1581	1579	1793	1833	1695
N+5	2648	1787	1787	2113	2153	2011
M+1	826	451	451	449	456	447
M+2	1536	923	920	928	985	901
M+3	2165	1272	1267	1420	1513	1307
M+4	2721	1604	1598	1925	2018	1765
M+5	3219	1820	1813	2319	2408	2145
P+1	1212	762	757	751	778	735
P+2	2118	1167	1153	1273	1349	1159
P+3	2887	1569	1540	1865	1952	1697
P+4	3546	1906	1867	2430	2524	2232
P+5	4122	2240	2197	2960	3057	2754
S+1	2007	1284	1282	1289	1325	1280
S+2	3508	2048	2035	2221	2338	2107
S+3	4760	2627	2611	3242	3377	3009
S+4	5809	3121	3083	4143	4281	3875
S+5	6708	3514	3462	4903	5062	4637

The last three columns of Table-8.2 show the reduction of standard deviation of errors from no Model to the cases of using imperfect rainfall forecast in Model-2. The best results are obtained by using 3rd method of rainfall forecast, i.e. using current day rainfall $p_j(i)$ as an approximation for future day's rainfall $P_j(i+t)$ in Type-2 Model. In this case, the reduction of standard deviation from no model forecast is as high as 50% for one to three day forecasting lead time, but it is reduced to 30% in the fifth day. This reduction in the efficiency of the forecasting model with lead time is due to the obvious reason of imperfect rainfall forecasts. The efficiency of Mix-Model can very likely be much improved for four and five day lead times by better rainfall forecasts.

The comparison of Table-8.2 and Table-8.3 shows that there is no significant difference between standard deviations of errors in the simulation and the validation period. It means that the Mix-Model is performing with similar efficiency in the simulation and validation mode.

Tab. 8.3: Standard deviation of various cases of flood forecast (for validation period)

Gauge	Without forecast	Observed P and calculated KN	Observed P and forecasted KN	Forecasted P (last 5 days avg.) and forecasted KN	Forecasted P (last 5 days time shift) and forecasted KN	Forecasted P (constant rainfall propagation) and forecasted KN
L+1	649	640	628	620	644	613
L+2	1059	863	858	886	941	822
V+1	674	489	490	488	489	488
V+2	1117	663	666	650	666	634
V+3	1454	973	986	1013	1049	953
V+4	1716	1153	1192	1288	1323	1205
N+1	818	809	801	784	808	799
N+2	1417	1050	1024	990	1090	986
N+3	1886	1442	1414	1534	1639	1400
N+4	2293	1640	1612	1929	1997	1749
N+5	2650	1898	1889	2347	2394	2164
M+1	987	547	536	529	523	530
M+2	1762	1097	1073	1056	1133	1055
M+3	2439	1453	1412	1561	1709	1438
M+4	3028	1783	1738	2176	2311	1952
M+5	3543	2010	1965	2690	2786	2431
P+1	1225	807	795	786	805	776
P+2	2167	1248	1201	1268	1339	1181
P+3	2984	1689	1618	1971	2067	1824
P+4	3690	2071	1975	2604	2724	2390
P+5	4295	2426	2314	3213	3332	2961
S+1	1974	1216	1179	1158	1181	1161
S+2	3460	1938	1904	2058	2280	1961
S+3	4714	2507	2443	2981	3209	2751
S+4	5737	2984	2879	3940	4143	3647
S+5	6597	3394	3256	4787	4985	4465

Similar to the results of Type-1 Model as presented in Table 7.12 and 7.13, in general, the standard deviation reduces by 50% for perfect input of rainfall with first case of n and k optimization. It means, there is not much difference between the results of Type-1 and Mix-Model when a perfect input of rainfall is taken as rainfall forecast. But in the cases where forecasted rainfall is used as an input, the standard deviation is comparatively smaller in the results by Mix-Models than by Type 1 Model. This standard deviation is comparable to that of

the fifth case (i.e. using n and k parameters with Type 1 Model). But since this case is physically less plausible, it is preferred to use the Mix-Model also in the absence of better rainfall forecast.

9. Conclusions and Recommendations

In the list of river basins exceeding 500,000 km^2 the Mekong has one of the largest peak discharges per unit area, (O' Connor, 2004). SE Monsoons and NW typhoons cause heavy rainfalls in the Lower Mekong. This rainfall in Northern Lao-PDR and on the Anman mountains along the border of Lao-PDR to Vietnam generate runoff into the main stem of the Mekong via lateral tributaries. The lateral inflows are superimposed on the Yunnan (Upper Mekong) discharge component of the Mekong. The Yunnan component contributes less than 20% of the total Mekong flow and its proportion in summer discharges is even less. The lateral inflows from left bank tributaries of the Mekong are the main reason of flooding.

The estimated flood damage to benefit ratio of 1 to 100 promote the idea of "living with the floods" in Lower Mekong Basin. However, a major effort is required to reduce the intangible costs in order to optimize social benefits of Mekong floods. Above all, losses of human lives must be avoided. The population density along the Mekong normally ranges from 30 to 50 persons per km^2 as per population density map of year 2000, of which many live in flood prone areas on the Mekong. In order too prepare people for beneficial floods, and to help protect them against harmful floods a good early warning of upcoming flood is necessary to save human lives. Flood forecasts of good quality are a decisive factor for a successful early warning. Flooding of once in every two years on the average along Middle Mekong requires effective forecasting.

In order to produce flood forecasts with 1 to 5 days lead time, at 6 gauges, i.e. L, V, N, M, P and S located in Lao-PDR and Cambodia along the Middle Mekong, a data based flood forecasting model was developed. In the analysis mode of the study, the model was developed directly from available data instead of fitting some imported model, i.e. a data based model was generated, in contrast to the use of imported models, such as models SSARR and URBS, which have already been tried by MRC, with limited success. In particular the simple linear storage model SSARR has resulted in poor performance in recent high floods of year 2000 and 2005, mainly because the parameters of this model were calibrated long ago and have never since been updated successfully.

The semi distributed, non-linear and conceptual URBS model is a runoff-routing networked model of sub-catchments based on centroid inflows. The rainfall excess is determined by separating each sub-basin into pervious and impervious areas. In pervious areas the loss is computed by a continuous loss model (rainfall is lost on all parts of the catchment), proportionate runoff model (only part of the catchment contributes to runoff) and the Manley Philips infiltration loss model (Rainfall is lost on all parts of catchment up to end infiltration). But since data on local soil moisture conditions are seldom available – (it is not known to the writer how the URBS model is calibrated, especially when it was found (indirectly from catchment gain)

that soil wetness changes from early to late flood months. Unlike SSARR, in application of URBS to the Mekong, the whole Mekong basin is divided into hundreds of sub-basins. The application constraints of distributed models in flood forecasting for large rivers has been discussed in detail in Chapter-3, where it is indicated that the over-parameterization of this model does not help to get acceptable results

Both models SSARR and URBS are based on the principle of bottom to top modeling which was opposed by Young (2002) to be used in flood forecast. In both models, no additional sub-routine was given in order to describe the uncertainty with each forecasted value. Normally, every forecast is subject to uncertainty. If no measure of the uncertainty is communicated to the user, he is to believe that the forecast is true. Both the models need Satellite Rainfall Estimates (SRE) data, which up to today has not yet been validated by ground based measured data. This problem was noted by Malone (2009) who stated that the quality of URBS forecast is a function of input, and that the poor quality of URBS forecasts, as seen in flood season 2008, could be attributed to poor input of rainfall by SRE. Further, both models measure the quality of forecast by average absolute error which is not a good measure of the uncertainty of the results and thus. Not well suited to judge the performance of a forecasting model.

In order to set an objective standard to check the performance of the newly developed model, a performance index PI was selected as objective measure. Unlike the well known Nash Sutcliffe (NS), which gauges the performance of data generation model by dividing the variance of discharge differences of modeled QF(i+t) minus observed Q(i+t) through the variance of discharge differences of observed Q(i+t) minus mean $\overline{Q(i)}$, PI measures the model performance by dividing the variance of the discharge differences of modeled QF(i+t) minus observed Q(i+t) through the variance of the discharge differences of observed Q(i+t) minus present value $Q(i)$, i.e. by measuring the performance against the assumption of no changes in discharge over the lead time. NS was found to be a suitable objective function for a design model where the best available estimate of future discharge is the mean discharge, but not for flood forecasting where the best estimate of future discharges is the last observed discharge. In addition to PI, the probability distribution of errors was also determined to check forecast bias and the error probability distribution on the basis of a detailed error analysis.

The Mekong is a long River, and it takes 9 days for the flood wave to propagate from C to S. In the travel of flood wave from C to S, the flow is augmented by lateral tributaries. Therefore, at each gauge along Mekong, there is one portion of discharge which is coming from upstream, and another portion which is added by the lateral tributaries. With the help of the mass balance equation the known gauged discharges from upstream was routed downstream. The unknown lateral inflows were forecasted with the help of two data based models, i.e. Metric (Type 1 Model) and Hybrid Metric Conceptual (Type 2 Model). Unlike a complex physical model, the data based models were structured exclusively from the available data base of 7 discharge and 31 rainfall gauges. Instead of concentrating on replicating the physical accurateness of a catchment

and channel physics, optimisation of the functional accuracy of model in forecasting future flows was preferred. But prior to the application of Type 1 and Type 2 Models, the potential of an empirical probability based Type 0 Model was tested.

Type 0 Model was based on taking average discharge differences at each day for all years of observation from current day to future forecast day and computing probability distribution of discharge differences from current day to future forecast day. The addition of average positive and negative discharge differences to current lateral discharge formed upper and lower limits of future discharges. Similarly, adding the 5 and 95% occurrence probability discharge to current discharge gave the other upper and lower limits. The problem was that in these approaches the span was too narrow and many actual future discharges exceeded the lower and upper limits. In a second approach the full range of available data was used, however, in this case the span was too wide and forecasts on the basis of this approach were not useful.

Type 1 Model was developed by means of discharges only, using daily discharge data from 1960 to 1990 for analysis, and data from 1990 to 2005 for validation. In order to forecast future flows at each gauge along Mekong, the lateral inflows were forecasted first. The lateral inflows were computed on the basis of linear regression of forecasted value with time delayed upstream discharges. Multi- regression was preferred to simple regression because simple auto regression of lateral inflows suggested that good correlation between subsequent inflows decreased rapidly with increase in time steps from one to five days – which then was compensated by adding contributions from further upstream lateral inflows by means of simple linear regression. In this manner, by regressing the lateral inflows against lateral inflows of subject reach and all available upstream reaches with 1 to 5 days time steps forecasts for lateral inflows with 1 to 5 days lead time were generated.

The generation of lateral inflows enabled to do discharge forecasts for 1 to 5 days lead time at 6 gauges along Mekong in the flood season, i.e. June-October. In general, the efficiency of the forecasts of lateral inflows, as expressed by PI, ranged from 0.5 to 0.7 for the first two days, but decreased sharply with more than three days lead time. The forecasting efficiency for Nakhon Phanom was found to be even lower, i.e. 0.2 to 0.3 in comparison to other forecasting gauges.

Further application of the Type 1 Model on the Mekong River to the forecasts of river discharges in the Mekong, showed that the use of persistence between subsequent discharges was sufficient to produce flood forecasts with acceptable efficiency, i.e. PI = 0.5 to 0.70 up to two days lead time. However, the efficiency of flood forecasts by this model reduced from PI = 0.4 to 0.3 with increase in lead time from the 3rd to the 5th day. It was assumed, and found true later, that rainfall forcings reduced the capability of multi-regressions to produce effective flood forecasts for lead times greater than two days.

Consequently, rainfall information was used in developing the Type 2 Model, in order to compensate the shortcomings (failure to produce good forecast above two days lead time) of the

Type 1 Model. Some type of rainfall runoff model was required in order to convert the rainfall hyetograph into the runoff hydrograph. On the basis of extensive literature review of the requirements of large rivers flood forecast in the context of Mekong, a Hybrid Metric Conceptual approach has been selected for the generation of runoff – resulting in the Type 2 Model. Separate rainfall runoff models were constructed, one for each of the 6 sub-basins. Daily rainfall data of 33 rainfall and 7 discharge gauges were used. The data of 1990 to 2000 were used for simulation and 2005 to 2005 for validation by using the standard split sampling approach. After analyzing three different options, i.e. SRE, areal average, weighted areal average, to obtain areal rainfall, a simple areal average was selected, because SRE by NOAA was found to be not validated by ground based reality, and Thiessen polygon weighted areal averages were giving poor results because of uneven distribution of rainfall gauges - in some cases one of the many gauges was getting more than 60% of the weight because of its typical location.

The areal average rainfall was converted into effective rainfall with the help of an empirically introduced adjustment coefficient KN. KN is defined as the ratio between seasonal, i.e. cumulative monthly runoff to monthly cumulative rainfall. The typical behavior of runoff response against the rainfall during the course of the flood season suggested the use of variable runoff to rainfall ratio. The analysis of runoff to rainfall ratio has indicated seasonality, i.e. KN increases from first to last month of flood season. Instead of using mean monthly KN for each flood season, KN was forecasted with the help of multi linear regression and moving average method. On the basis of comparative performance, the forecast of KN by moving average was preferred at the end. This seasonal KN was used along with a 2 parameter Nash cascade in order to convert effective rainfall into runoff. The n and k parameters of the Nash cascade were selected on the basis of LSE (least squares error). LSE was applied on four different sets of output data, i.e. lateral runoff, lateral runoff above certain threshold, lateral runoff after updating, and mainstream discharges. In the fifth case, the input data of observed rainfall was changed into forecasted rainfall and LSE optimization was applied on mainstream discharges. The n and k values optimized by LSE application to lateral runoff was preferred because of its performance in forecasting mainstream discharges. The optimized n and k values, obtained on the basis of LSE of lateral runoff after updating, performed well for forecasting lateral discharges. But as the goal was to forecast mainstream discharges the first case of n and k optimization was preferred.

The lateral inflows computed by the best variant of KN and n and k parameter combination were updated on the basis of error corrections by using the correlation between subsequent errors. The forecasting efficiency of a Type 2 Model with perfect rainfall data (assuming future rainfall is known exactly) for 1 to 5 day forecast at mainstream gauges, as expressed by PI, ranged from 0.6 to 0.7, except at Nakhon Phanom where it remained of the order of 0.1 to 0.5. Forecasting efficiency reduced when the perfect input data was replaced with a forecasted rainfall for future days and thus ranged from 0.5 to 0.6 for the first 3 days lead time, and 0.2 to 0.4 for 4 and 5 days lead time. However, by using the fifth case of n and k parameters, the forecasting efficiency improved, as expressed by PI, which ranged from 0.6 to 0.7 for the first 4 days lead time, and

from 0.5 to 0.6 in 5 days lead time. An exception was the forecasting efficiency at Nakhon Phanom, which was lower i.e. ranging from 0.1 to 0.4. Although, the fifth case of n and k improve the efficiency of Type 2 Model to forecast mainstream discharges even with poor rainfall input (taking rainfall of today as future rainfall for next five days), it is not recommended for application because the typical shape of unit hydrograph for this case did not seem to be realistic. Therefore, it was necessary to find an alternate approach for producing better forecast with more plausible first case of n and k, even with poor rainfall forecasts.

The typical compensating pattern of forecasting errors in Type 1 Model and Type 2 Model suggested the use of average of these two as a forecast. The optimal weighted average was determined by the standard regression technique. The forecasting efficiency of mixed modeling, as expressed by PI, ranges from 0.6 to 0.7 in first 4 days lead time, and is also reduced to range from 0.4 to 0.5 in 5 days lead time.

It has been found that in a data scarce catchments like that of the Mekong, a Metric Type 1 Model can be used for up to 2 days lead time flood forecast and Hybrid Metric Conceptual Type 2 Model can be used for 3 to 5 days forecasts. But the efficiency of Type 2 Model in longer lead times was restricted by rainfall forecast quality. The use of Mix model has been preferred at the end because it combined the information of discharge persistence extrapolation of Type 1 Model with information of rainfall forcings into runoff as generated by Type 2 Model.

Since both Type 1 and Type 2 Models were based on empirical analysis of discharge and rainfall time series in the analysis mode, it was assumed that this empirical relation will hold also good in the future. But these functional dependencies of data are the result of system physics of the basin, whose foot prints are captured by the data – and work only if there are no major anthropogenic changes in the basin. On the other hand, global climate change effects only alter the input of the data based model if the spatial pattern of rainfall changes in Mekong sub-basins. However, change in the temporal pattern of rainfall is less likely to change the performance of a Type 2 Model.

The quality of Mixed Model for Pakse, as expressed by PI, is 0.55 to 0.7 for the flood periods of 1990 to 2005 against 0.3 to 0.5 of SSARR for flood season of 2005. It is not possible to compare the results of data based model with URBS results directly because the writer didn't have access to recent discharge data (2005 onwards), and URBS was not operational in the previous years (1990 to 2005). However, weekly mean absolute errors of 20 cm to 280 cm have been noted for the year 2009, which in comparison to mean absolute errors (5 cm to 50 cm) of mixed model for 2000 flood season are much higher.

Apart from the issue of performance, the use of the data based model is data efficient, works efficiently and is simple to use in comparison to SSARR and URBS. Further, it needs the data of only 33 rainfall gauges and 7 discharge gauges in comparison to extensive data requirement of

SSARR and URBS. Further, the additional analysis of errors allows the data based modeling approach to communicate uncertainty band along with each forecast.

There is, however, a need for better rainfall forecast, because for longer lead times, i.e. for more than 4days, the efficiency of both models is impaired by rainfall events that occurred after the day of issuance of flood forecast. The information of typhoon tracks can be used in order to locate storm eye and may yield estimates of rainfall from these storms. A study of the possible relation of typhoon tracks with rainfall occurrences in different sub-basins of Mekong is recommended. The study should put emphasis on establishing some quantitative, or usable qualitative relation between these two, so that information on typhoons could be used in forecasting rainfall. The forecast by NOAA could also be used provided the SRE are validated by ground based measurements but until these are available the use of NOAA rainfall forecasts is not likely to be free from significant errors.

Performance of Type 2 model can be improved by adding more rainfall gauges, especially in the LV sub-basin, where there are only two gauges at the moment. The installation of more rainfall gauges and at more representative locations is likely to improve the performance of Type 2 Model, however, very likely only after revising the n and k parameters.

Apart from the quality of rainfall forecasts, the quality of the performance of the forecasting model is also affected by the quality of the rating curves. An erroneous conversion of water levels to discharges and vice versa is a likely cause of additional errors. The issue of ratings seems to be a problem, leading to poor performance of both Type 1 and Type 2 Models in particular at Nakhon Phanom which can be improved by better ratings.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J. (1986a): An introduction to the European Hydrologic System-Systeme Hydrologique Europeen, SHE, 1: History and philosophy of a physically-based, distributed modeling system. J. Hydrol.,Vol. 87, 45–59.

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J. (1986b): An introduction to the European Hydrologic System-Systeme Hydrologique Europeen, SHE, 2: Structure of a physically-based, distributed modeling system. J. Hydrol., Vol.87, 61–77

Apirumanekul C. (2006): Flood forecasting in the Mekong River Basin: an improvement plan for the flood forecasting system In *International Conference on Mekong Research for the People of the Mekong, 18-21.10.2006*. Chiang Rei, Thailand.

Acreman, M. C. (1985): Predicting the mean annual flood from basin characteristics. Hydrol. Sci. J. Vol.30, 37-49.

Beck, M. (1987): Water quality modeling : A review of analysis of uncertainty. Water Resourc. Res., Vol.23 8, 1393-1442.

Boyd, M. J. (1987): Head discharge relationship for culverts. *Proc. 21st Congress International Association for hydraulic Research,* Vol 6. Keynote addresses. 118-122

Bathurst, J. C., Wicks, J. M., and O'Connell, P. E. (1995): Chapter 16: The SHE/SHESED basin scale water flow and sediment transport modeling system. *Computer models of watershed hydrology*, V. P. Singh, ed., Water Resources Publications, Littleton, Colo. 563–594.

Bronstert, and., Plate, E.J. (1997): Modeling of runoff generation and soil moisture dynamics for hill slopes and micro-catchments. Journal of Hydrology Vol.198(1-4), 177 –195.

Beck, M. B., Kleissen, F. M. & Wheater, H. S. (1990): Identifying flow paths in models of surface water acidification. Rev. Geophys. Vol.28(2): 207–230.

Berthet, L. , Andreassian, V. Perrin and C. Javelle P. (2009): How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. Hydrology And Earth System Sciences Vol.13: 819-831.

Beven, k. (2001): Model validation: Perspective in hydrological sciences. ng. J. Appl. Meteorol.,Vol.11, 1203,1972

Beven, k. (2002): Toward a coherent philosophy for modeling the environment. Proc. R. Soc. Lond. A: 458, 1-20

Beven, k. (2000): Uniqueness of place and process representations in hydrological modeling Hydrology and Earth System Sciences, Vol.4(2), 203-213

Beven, K.J. and Freer, J. (2001): Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. J. Hydrol., Vol.249, 11-29.

Beven, K.J. (1993): Prophecy, reality and uncertainty in distributed hydrological modeling. Adv. Water Resour., Vol.16, 41-51.

Beven, K.J. (1996): A discussion of distributed modeling, Chapter 13A, In: Distributed Hydrological Modelling, J-C. Refsgaard and M.B. Abbot (Eds.) Kluwer, Dordrecht, 255-278.

Beven, K.J. (1997): Distributed modeling in hydrology: Application of the Topmodel concepts. Wiley, Chichester, 360

Beven, K.J. (2002b): Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modelling system, Hydrol. Process. 186–206.

Clark, M. Schoenbohm, L., Royden, L., Whipple, K., Burchfiel, B., Zhang, W. et al. (2004): Surface uplift, tectonics, and erosion of eastern Tibet from large scale drainage patterns. Tectonics 23 TC 1006, doi:10.1029/2002TC001402

Cunderlik, J.M. (2003):Hydrological model selection for the CFCAS project: Assessment of water resources risk and vulnerability to changing climatic conditions. Project Report 1, Canada

Chow, V. T., Maidment, D. R. and Mays, L. W. (1988): Applied Hydrology. McGraw-Hill.

Carrol; D.G. (2004): *URBS - Unified River Basin Simulator: A rainfall runoff routing model for flood forecasting & design, Version 4, April, 2004*

Carrol; D.G. (2007): *URBS - Unified River Basin Simulator: A rainfall runoff routing model for flood forecasting & design, Version 4.30, September, 2007*

Calver, A., and Wood, W. L. (1995): Chapter 17: The Institute of Hydrology distributed model. *Computer models of watershed hydrology*, V. P. Singh, ed., Water Resources Publications, Littleton, Colo.

Camarago, S. J., Robertson, A.W, Gaffney, S., Padhraic, Ghil, M. (2007): Cluster analysis of typhonon tracks. Part I: General properties. American Meteorological Society, Vol. 20, 3635-3653.

Delgado, J.M., Apel, H., Merz, B. (2010): Flood trends and variability in the Mekong river Hydrol. Earth Syst. Sci., Vol.14, 407-418.

Dooge, J.C.I. (1960): Linear theory of hydrologic systems, *ARS Tech. Bull. No. 1468* Washington

Dooge, J. C. I. (1973): Linear theory of hydrologic systems, *Tech. Bull. no. 1468,* Agricult. Res. Serv., US Dept. of Agricult. Washington DC, USA.

Eastham J., Mpelaskova F., Mainuddin M., Ticehurst C., Dyce P., Hodgson G., et al. (2008):. Mekong River Basin water resources assessment: Impact of climate changes. CSIRO Water for a Healthy Country National Research Flagship report. CSIRO, Canberra. Available from: http://www.clw.csiro.au/publications/waterforahealthycountry/2008/wfhc-Mekong-WaterResourcesAssessment.pdf

Freeze, R. A., and Harlan, R. L. (1969): Blueprint for a physically-based, digitally-simulated hydrologic response model." J. Hydrol., Vol.9, 237–258.

Georgakakos, K.P and Bras, R. L. (1984): Hydrologically useful station precipitation model. Water Resour. Res.,Vol.20, 1585-1610

Glahn, H.R. and Lowry, D.A.(1972): The use of model output statistics in objective weather forecasting. J. Appl. Meteorol., Vol.11, 1203,1972

Gupta A. and Liew S. (2006): The Mekong from satellite imagery: A quick look at a large river. Geomorphology Vol.85 (3–4), 259–274.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1998): Toward improved calibration of hydrologic models: Multiple and no commensurable measures of information. Water Resour. Res, Vol.34(4), 751–764.

Garrick, M. Cunnane C. and Nash, J. E. (1978): A criterion of efficiency for rainfall-runoff models. J. Hydrol. Vol.36, 375-381

Gilchrist(1974): Statistical forecasting-the state of the art. OMEGA, J. Mgmt. Sci.,Vol.2, 733-749

Haan, C. T. (1979): Statistical methods in hydrology. The Iowa State University Press, USA

Halcrow Group Limited. (2001): Working Paper No 1 - Hydrological Review In *Mekong River Commission - Water Utilisation Project Component A: Development of Basin Modelling Package and Knowledge Base (WUP-A)*.64.

Hapuarachchi, H.A.P., Takeuchi, K., Fukami, K., Inomata, H., and Zhou, M. (2007): Importance of Hapuarachchi, H.A.P., Takeuchi, K., Fukami, K., Inomata, H., and Zhou, M. (2007): Importance of rainfall measurements for the flood forecasting of the Mekong River Basin.

Hochschild, V. (1999): Parameterization of Hydrological Models: The Contribution of Remote Sensing to Water Resources Management. Proceedings of the MODSIM'99, *International Congress on Modeling and Simulation,* 06.-09.12.99, Hamilton, NZ

Jain, S. K. and Sudheer, K. P. (2008): Fitting of hydrological models: a close look at the nash-sutcliffe index. J. Hydrol. Engg. ,Vol.13 , 981-986.

Kanning, W. Pich, S. and Pengel, B.E. (2008): Flood forecasting accuracy for the Mekong River Basin In *6th Annual Mekong Flood Forum (AMFF-6)*.

Kachroo R. K. (1992): River flow forecasting. Part. A discussion of the principles. J. Hyd., Vol. 133, 1-15

Kitanidis, P.K. and Bras, R.L. (1980a): Real-Time Forecasting With A Conceptual Hydrologic Model.1. Analysis Of Uncertainty. Water Resour..Res. Vol.16,1025-1033.

Kitanidis, P.K. and Bras, R.L. (1980b): Real-Time Forecasting With A Conceptual Hydrologic Model.2. Applications And Results. *Water Resources Research* Vol.16,1034-1044.

Krzysztofowicz, R. and Herr, H. D. (2001): Hydrology uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. Journal of Hydrology, Vol.249, 46-69.

Krause, P. Boyle, D. P. and Bäse , F. (2005): Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences ,Vol.5 , 89-87.

Kadioglu, M. Şen, Z. (2001): Monthly precipitation-runoff polygons and mean runoff coefficients. Hydr. Sci. J.,Vol.46 , 3-11.

Kachroo, A.K. and G.C.Liang.(1992): River flow forecasting. Part 2. Algebraic development of linear modelling techniques. J. Hydrol., Vol.133, 17-40.

Kachroo, R. K., Sea, C. H. Jemenez, H. and Saxena, R.P. (1992): River flow forecasting.Part 3. Application of linear techniques in modelling rainfall-runoff transporations. J. Hydrol. , Vol.133, 41-97.

Laursenson, E.M. and Mein, R.G. (1990): RORB - *Version 4 Runoff Routing Program User Manual*, Monash University, Dept of Civil Engineering

Lettenmaeir, D.P and Wood, E.F. (1992): Hydrologic Forecasting. *Hand book of hydrology*, 26.1-26.30

Lardet P. and Obled C. (1993): Real-time forecasting using a stochastic rainfall generator. J Hydr. Vol.(162), 391-408.

Lindenmaier, F, Shahzad, M.K., Ihringer, J. and Plate, E.J. (2010) :Early Warning system improvement. Wisdom project report, for BMBF, 59pp Liu S., Lu P., Liu D. and Jin P. (2007): Pinpointing source of the Mekong and measuring its length through analysis of satellite imagery and field investigations. Geo-spatial Information Science Vol.10 (1) 80–87.

Liu, S., Lu P., Liu, D., and Jin, P. (2009): Pin pointing source and measuring the lengths of principal rivers of the world. Int. J. Digital earth Vol.2 (81), 80-87

Longobardi A, Villani P, Grayson RB andWestern A.W. (2003): On the relationship between runoff coefficient and catchment initial conditions. *In Proceedings of MODSIM 2003;* 867–872.

MRC. (2006): Annual Mekong Flood Report (2005). Mekong River Commission: Vientiane, Lao PDR;82.

MRC Annual Flood Report (2005): Mekong River Commission, Vientiane, Lao PDR. 82pp. ISSN 17283248

McMahon, G.M. & Muller, D.K. (1986): The application of peak flow parameter indifference curve technique with ungauged catchments, *Hydological & Water Symposium, Institution of Engineers* Australia, National Conference, Publication No. 82/3, 110-114

Mekong River Commission, (2004): Review of hydro-meteorological data base of the MRC data secretariat

MRC Annual Flood Report (2005): Mekong River Commission, Vientiane, Lao PDR. 82pp. ISSN 1728

Mekong River Commission, (2005): *Overview of the Hydrology of the Mekong Basin*. Vientiane, Lao PDR, MRC.

Mekong River Commission, (2010): *State of the Basin Report*, 2010. Vientiane, Lao PDR, Mekong River Commission

MRC. (2007): *Annual Mekong Flood Report 2006*. Mekong River Commission: Vientiane;76.

MRC. (2008): *Annual Mekong Flood Report 2007*. Mekong River Commission: Vientiane, Lao PDR;86.

Malone T. (2009): *Sensitivity analysis and evaluation of the MRC Mekong flood forecasting system. MRC report 67pp.*

Merz, B. and Plate, E. J. (1997): An analysis of the effects of spatial variability of soil and soil moisture on runoff. Water Resources Research , Vol.33, 2909-2922.

Malone, T. (2006). Australian Bureau of Meteorology, *Roadmap Mission for the Development of a Flood Forecasting System for the Lower Mekong River Basin- Technical Report*, Mekong River Commission, September 2006

Nash, J.E. (1958): The form of the instantaneous unit hydrograph, *General Assembly of Toronto*, *Compt. Rend.* Vol.3 IASH Publ. No. 42, Gentbrugge. pp. 114–118

Nash, J.E. (1960): A unit hydrograph study with particular reference to British catchments, Proc. Inst. Civ. Eng. Vol.**17,** 249–282

Nash, J.E. and Sutcliffe, J.V. (1970): River flow forecasting through conceptual models, Part I - A discussion of principles. J. Hydrol.Vol.10, 282-290.

Natale, L., Todini andE. (1976a): A stable estimator for linear models,1, theoretical development and Monte Carlo experiments. Water Resources Research Vol.12 (4), 664– 671.

Natale, L. and Todini, E. (1976b): A stable estimator for linear models, 2. Real world hydrologic applications. Water Resources Research Vol.12 (4), 672– 676.

O'Connor J.E. and Costa J.E. (2004). The world's largest floods, past and present – their causes and magnitudes. US Geological Survey Circular. Number 1254.

O'connell, P. E., Nash, J. E. and Farrell, J. P. (1970): River flow forecasting through conceptual models part II-The brosna catchment at ferbane. J. Hydrol., Vol.10 , 317-329.

Penman, H. L. (1961): Weather, plant and soil factors in hydrology. Weather, Vol.16, 207–219.

Pengel, B.E., Malone, T. AND Hartmann, M. (2008): Towards a new flood forecasting system for the Lower Mekong River Basin In 4th International Symposium on Flood Defence. Toronto, Ontario, Canada;71-71 - 71-79.

Pich S. (2006): *Practical User Guide - Flood Forecasting and River Monitoring Operation In unpublished report of the Mekong River Commission* - Regional Flood Management and Mitigation Center, Comission MR ed. RFMMC: Phnom Penh;24.

Ponce, V. M. (1989): Engineering Hydrology, Principles and Practices. Prentice-Hall, Englewood Cliffs, New Jersey.
http://ponce.sdsu.edu/330textbook_hydrology_chapters.html

Plate, E.J. (2007): Early warning and flood forecasting for large rivers with the lower Mekong as example. J. of Hydro-envir. Res. Vol.1, 80-94.

Plate, E.J. and Hewitt, M.M. (2002): Proceedings of the international expert meeting on early warning for the Mekong River, 27th February - 1st March 2002 In *MRC Conference Series*. Mekong River Commission: Phnom Penh;297.

Plate, E.J. and Insigsiengmay, T. (2005): Early warning systems for the Lower Mekong River. Water International Vol.30, 99-107.

Plate, E.J. and Lindenmaier, F. (2008): Quality Assessment of forecasts In *Annual Mekong Flood Report 2008*. Phnom Penh.

Plate, E.J. (2009): Classification of hydrological models for flood management. Hydrol. Earth Syst. Sci., Vol. 13, 1939-1951

RFMMC. (2004): Draft - Review of hydro-meteorological database of the MRC Secretariat In *Data Management Manual for hydro-meteorological archive database of the MRC Secretariat - Progress Report*. Mekong River Commission: Vientiane;58.

Rockwood D.M. (1972): User manual for COSSARR model: A digital computer program designed for small to medium scale computers for performing streamflow synthesis and reservoir regulation in conversational model. US Engineer Division, North Pacific.

Rodriguez-Iturbe, B., Carr, R. I., Garcia, R., Allen, J. E., Rabideau, D., Rubio, L., and McIntosh, R. M. (1978). Acute poststreptococcal glomerulonephritis in Maracaibo,I. Studies on circulating immune complexes and immuno-globulins: Evidence for a circulating immune complex disease. *Kidney International (in press)*.

Refsgaard, J. C., and Storm, B. (1995): Chapter 23: MIKE SHE. *Computer models of watershed hydrology*, V. P. Singh, ed., Water Resources Publications, Littleton, Colo.

Seyfried, M. S., and Wilcox, B. P. (1995): Scale and nature of spatial variability: Field examples having implications for hydrologic modeling. Water Resour. Res., Vol. 31, 173–184.

Shahzad, M.K., Lindenmaier, F., Ihringer, J., Plate, E.J. and Nestmann, F. (2009): Statistical ensemble flood forecasting for the Mekong River In *Annual Mekong Flood Report 2009*. Bankok;10.

Shahzad, M.K., Lindenmaier, F., Ihringer, J., Plate, E.J. and Nestmann, F. (2009): Statistical flood forecasting for the Mekong River In European Geo-Sciences Union Assembly, Vienna, Austria

Shahzad, M.K. and Plate, E.J. (2010): Data based flood forecasting for the Mekong River In *Annual Mekong Flood Report 2010*. Vientiane;10.

Shahzad, M.K. (2006): Prediction of long terms inflows on the basis of hydro-meteorological parameters by statistical approach. Master Thesis, CEWRE, U.E.T, Lahore, Pakistan

Shrestha, R.R. (2005): River flood prediction systems: toward complementary hydrodynamic, hydrological and data driven models with uncertainty analysis. PhD Thesis, UNiversity of Karlsruhe, Germany

Singh, Vijay P. (1992): Elementary Hydrology. Prentice-Hall.

Singh, V.P. and Woolhiser, D.A. (2002): Mathematical modeling of watershed hydrology. J. Hyd. Engg., Vol. 7, 270-292

Todini E. (2007): hydrological catchment modeling: past, present and future. Hydrol. Earth Syst. Sci., Vol. 11, 468-482

Tanaka H. 1999. Flood forecasting of the Mekong River in 1997 1999/14 FRP ed.: FAO, Bangkok;109-117.

Twidale C. (2004): River patterns and their meanings. Earth Science Reviews Vol.67 (3–4), 159–218.

Tandon S.K. and Sinha R. (2007): Geology of large river systems. In: A. Gupta, *Large rivers: geomorphology and management* (pp. 7–28). John Wiley & Sons Ltd, London.

Tokar, A.S. and Johnson, P.A. (1999): Rainfall-runoff modeling using artificial neural networks. J. Hydrol. Eng., Vol.4: 232-239.

Ta T.K., Nguyen V.L., Tateishi M., Kobayashi I., Satio Y. and Nakamura T. (2002a): Sediment facies and late Holocene progradation of the Mekong River Delta in Brente Province, southern Viet Nam: an example of the evolution from a tide-dominated to a tide- and wave dominated delta. Sedimentary Geology Vol.152, 313–325.

Todini, E. (1996): The ARNO rainfall-runoff model. J. Hydrol., Vol.175(1-4), 339-382.

US Army Corps of Engineers. (1987): *User Manual - SSARR Model: Streamflow Synthesis and Reservoir Regulation*.

Wood E, O'Connell P. (1985): Real-time forecasting. Hydro. Forecasting, Vol.15, 505–558.

Welch, G. and Bishop, G. (2006): An Introduction to the Kalam Filter

White, W.R. (2000): Water in River: Flooding, A contribution to the World Water Vision, International Association of Hydrualics Engineering and Research (IAHR) & Wallingford, U.K..

Young P.C.(2003): Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. Hydrological Processes Vol.17, 2195–2217.

Young P.C.(2002): Advances in real-time flood forecasting. Philosophical Transactions of the Royal Society London Vol.A 360: 1433–1450.

Internet References

http://www.fao.org/docrep/003/x0596e/X0596e01e.htm#P213_18188

http://www.sunwater.com.au

www.mrcmekong.org

List of Figures

List of Tables