# COMPARISON OF RAINFALL-RUNOFF MODELLING TECHNIQUES IN SMALL FORESTED CATCHMENTS

by

Lauren Donnelly

B.Sc. (honours), University of Alberta, 1990

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

in the Department of Geography



SIMON FRASER UNIVERSITY

December 1997

0-612-24122-X

## APPROVAL

Name: Lauren Marie Donnelly

Degree: Master of Science

Title of Thesis: Comparison Of Rainfall-Runoff Modelling Techniques In Small Forested Catchments

Examining Committee:

Chair: I. Hutchinson, Associate Professor

R.D. Moore, Associate Professor
Senior Supervisor

F. Gobas, Associate Professor
School of Resource and Environmental Management

M.C. Quick, Professor
Civil Engineering, University of British Columbia
External Examiner

Date Approved: December 5, 1997

## ABSTRACT

The key issue for operational users of hydrologic models is whether distributed models perform sufficiently better than lumped models to justify the increased time and effort required for their application. The research had two main objectives: (1) to compare a quasi-distributed model to two lumped models to determine if there is a benefit associated with the increased demand for catchment data and (2) to determine if the statistical significance of differences in model performance can be quantified.

The models tested were a lumped black-box model, a lumped conceptual model, and a quasi-distributed conceptual model which requires topographic analysis of a digital elevation model. The models were compared using the evaluation procedure proposed by V. Klemes, which has four levels: (1) a split-sample test, (2) a proxy-basin test for geographic transposability, (3) a differential split-sample test for climatic transposability, and (4) a proxy-basin differential split-sample test for both geographic and climatic transposability. Model performance was compared for two small forested catchments (19.8 and 38.3 ha) within the University of British Columbia Research Forest, approximately 50 km east of Vancouver. The model efficiency ($E_m$) defined by Nash was used to compare model performance for the entire storm hydrograph, the peak flows, and the time-to-peak values.

The statistical analysis, which included the Jackknife method and ANOVA testing, showed that at levels 1 and 2, using a significance level of 0.05, there are no statistical differences in model performance. This finding confirms there is no significant benefit in applying the quasi-distributed model and that the simpler lumped models would

provide acceptably similar simulations under those conditions. At level 3, the quasi-distributed model performed statistically significantly better than both lumped models in both catchments. The statistical analysis provides justification for using the quasi-distributed model when simulating runoff events larger than those used for calibration. The statistical analysis for level 4 indicated that the quasi-distributed model performed significantly better than the lumped models in one catchment but not the other, such that the quasi-distributed model is no worse than the other two models but may perform better in certain catchments.

This research has provided further information on the relative performance of quasi-distributed and lumped rainfall-runoff models, and demonstrated that the ANOVA design including the Jackknife method is a workable method and could be a valuable tool for assessing statistical significance of differences in model performance.

## ACKNOWLEDGEMENTS

First and foremost I would like to thank my senior supervisor, Dr. Dan Moore, for whose patience and guidance I am extremely grateful. Thank you to my committee members, Dr. Frank Gobas and Dr. Micheal C. Quick for their comments and suggestions. Thanks is also extended to Dr. Rob Balshaw who provided his expert assistance with the statistical analysis.

This milestone could not have been possible without the love, support, and sense of humour of my family. Thanks Brian, Tim, Kari, Maureen, Bill, Chris, Maria, Greg and those kids - Jae, Kristel, Mark, Devin, and Daniela, and especially Mom and Dad. Thanks guys, you all mean so much to me.

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

Many scientific disciplines use models to describe systems in simpler terms and to predict system response. In hydrology, rainfall-runoff models enable users the ability to forecast the runoff from a catchment from the amount of precipitation received by that catchment. During the last three decades rainfall-runoff models have become accepted and important tools in operational hydrology for estimating information required for water resources planning, design, and operation. Rainfall-runoff models are also important to researchers in gaining a better understanding of the processes involved within a hydrologic system.

In the 1960s and 1970s, rainfall-runoff models were relatively simple, spatially lumped and represented hydrological processes using algebraic equations incorporating empirical parameters, e.g., the UBC Watershed model (Quick and Pipes, 1972) and the HBV model (Bergstrom and Forsman, 1973). Increasing availability of computing power, coupled with a desire to simulate sediment and chemical transport pathways within a catchment, led to the development of more "physically based" and spatially distributed models, such as the Systeme Hydrologique Europeen (SHE) model (Abbot *et al.*, 1986, the Institute of Hydrology Distributed Model (IHDM) (Beven *et al.*, 1987), and TOPMODEL (Beven and Kirkby, 1979).

Despite the optimism associated with development of these newer, more complex models, there have been few studies that have tested whether more complex models actually perform better in operational applications than simpler models (Beven, 1989). Where comprehensive intercomparisons of hydrologic models have been conducted, such as the WMO (1975) study of conceptual runoff models and the WMO (1986) study of snowmelt-runoff models, the statistical significance of differences in model performance has not been satisfactorily addressed (Cavadias and Morin, 1985).

This thesis addresses two broad questions: (1) are more complex models superior to simpler models for operational applications, and (2) can the statistical significance of differences in model performance be established? The remainder of this chapter reviews the characteristics of hydrologic models and existing procedures for comparing models to provide the context for the specific research questions.

## 1.1 Characteristics of hydrologic models

### *1.1.1 Mathematical and spatial representation of processes*

As shown in Figure 1.1, rainfall-runoff model types range in terms of their spatial and mathematical representations. The simplest black-box models (e.g. regression equations) attempt no explicit representation of processes, while physically based models solve differential equations which represent field processes such as infiltration and overland flow hydraulics. Conceptual models represent processes in terms of algebraic equations which attempt to approximate the solutions to the governing differential equations.

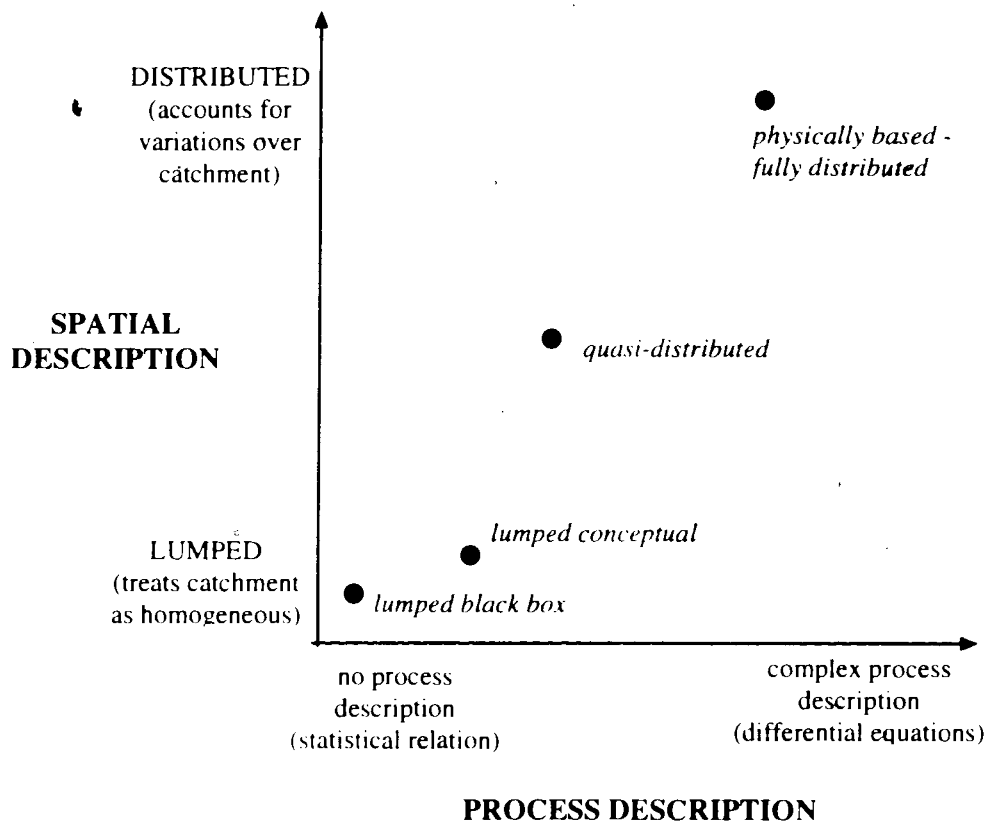



**Figure 1.1** Rainfall-runoff model types range in both spatial and process description.

In terms of spatial description, models are either lumped or distributed depending on whether the spatial distribution of hydrological parameters within the catchment is considered. A lumped model ignores spatial variability and treats the catchment as a homogeneous entity while a distributed model accounts for the spatial variability of the catchment.

An advantage of lumped models is their limited demand for spatial data. However, they need sufficiently long meteorological and hydrological records for their calibration, which may not always be available. These models make little or no use of information about topography, soil type and patterns and changes of vegetation. Therefore, use on another catchment, or predicting effects of land use change by changing parameter values, cannot be done with confidence (Beven *et al.*, 1984).

Distributed models, on the other hand, attempt to account for the spatial variability in the physical characteristics of a catchment. For example, since water flows downhill, representing the topography of a catchment using a distributed parameter should provide relevant information regarding the spatial variability of the flow (Ambroise *et al.*, 1996).

There are two types of spatially distributed models. A geometrically distributed model expresses variability in terms of actual points and their orientation within a catchment, whereas a probability-distributed model describes spatial variability statistically (Clarke, 1973). Probability-distributed models are also referred to as quasi-distributed.

In principle, distributed models should be superior to lumped models in terms of demanding less calibration of the parameters, when the model is applied to another catchment, since they are supposedly taking account of catchment characteristics. They do, however, require spatial data input such as a digital elevation model (DEM), while the lumped models do not. The key issue for operational users is whether distributed models perform sufficiently better than lumped models to justify the increased time and effort to acquire and process the spatial input data required for their application.

### *1.1.2 Operational versus research applications*

Any comparison and evaluation of rainfall-runoff models requires discussion of the type of application being assessed. Operational models are applied for purposes such as evaluating stormflow for delineating floodplain limits or designing reservoirs (e.g., Fedora and Beschta, 1989). Research models, on the other hand, are used where the purpose is to contribute to the understanding of hydrological processes such as infiltrability (Smith and Hebbert, 1979) or hillslope scale processes (Freeze, 1980).

Operational models usually tend to be empirical and spatially lumped with an emphasis on ease of application. Research models, on the other hand, being mainly concerned with the internal hydrological processes being modelled, are more complex, spatially distributed models. Fully distributed models, due to their extensive data demand, are unlikely to be used for operational purposes, at least not in the near future (Beven, 1989). The difference when evaluating the two different applications is that the complex process models must be compared against detailed field observations to assess performance for research purposes, whereas the operational models only attempt to simulate observed discharge values for forecasting purposes and are not usually overly concerned with the internal hydrological processes. The emphasis in this thesis is on evaluation of model performance for operational purposes.

### *1.1.3 Event-based versus continuous simulations*

Models can be used to simulate either long, continuous hydrographs or isolated events. Continuous simulations are used, for example, when the interannual variability of reservoir inflows is of interest (e.g. Bergstrom, 1979). Event-mode simulations are commonly used when interest is focused on catchment response to specific rainfall and/or snowmelt events, for example for computing design floods for structures (Hughes, 1984). This study used event-mode simulations to compare model predictions of storm hydrographs and peak flows in rainfall-driven catchments.

Continuous simulations commonly run with a coarser time resolution, typically one day for most snowmelt runoff models (WMO, 1986), while event-mode runs typically employ time steps of one hour or less. Another difference relates to specifying initial conditions such as reservoir storages and soil moisture conditions. This issue is not so critical in continuous simulations, especially for multi-year runs, where the effect of initial conditions typically becomes negligible after the first year of simulation. For event-mode runs, initial conditions can have a critical influence on modelled catchment response and must be controlled for in some way. For this study, initial soil moisture conditions were specified for the events. This issue is addressed in Section 2.5.

Use of event-mode simulations in this study allowed for a relatively large number of events to be included with a relatively high time resolution (1 hr), while minimizing the amount of meteorological and runoff data to be processed. It also avoided problems with simulating evapotranspiration and its effect on soil moisture in the periods between events.

## 1.2 Evaluation of model performance

### *1.2.1 Concept of operational validation*

The issues of model validation can be confusing for model developers and model users. Firstly, the terms verification and validation have been used interchangeably by some researchers but also have been distinctly defined by others. Rykiel (1996) defines model verification as a demonstration that the computer model is a correct implementation of the logical model. Validation has been defined as a demonstration that a model performs adequately for the intended application. As there is no set standard in the hydrological modelling definition of these terms, I will use the term validation to refer to the models' ability to simulate an independent data set other than from which it was calibrated although the usage may differ from that promoted by some hydrologic modellers.

Secondly, it is important to distinguish between operational validation, used in this study, and conceptual validity, which concerns a model's theoretical basis. An operationally validated model may work well for a specific use but may not be conceptually valid, i.e. a correct representation of the real system (Rykiel, 1996). For example, many simulation models are developed to meet practical management needs. These models are usually validated by comparing simulated to observed values to determine model performance. The ambiguity arises when inferences about the model's ability to reproduce reality are made from the validation results. A model's output may agree with observed data, but this correspondence does not guarantee that its internal structure is able to reproduce the actual processes operating in the real system. Any inferences made regarding the scientific basis of the model would be scientific hypothesis testing and not model validation.

Operational validation means that a model is acceptable for its intended purpose since it meets specified performance requirements (Rykiel, 1996). Validation does not require that the model applies to more than one condition unless that situation is part of the validation requirement. Good predictions do not have to be obtained only from a model that is entirely mechanistically correct and, conversely, invalidation does not imply that the scientific content of a model is wrong.

The models used in this study were operationally validated by simulating entire observed hydrographs of rainfall-runoff events for the purpose of model comparison. In this study, Klemes' (1986) evaluation scheme, introduced in the following section, and graphical and numerical criteria (Section 2.7.3) were used for the operational validation of the models.

### *1.2.2 Klemes' hierarchical approach to operational testing*

Klemes (1986) proposed a hierarchical method for the comparison of different types of hydrological models. The system, explained in detail in Section 2.7.2, is hierarchical since the modelling tasks are ordered according to their increasing complexity and demands on model capability.

Klemes' system includes tests for geographic and climatic transposability. Transposability refers to a model's ability to perform satisfactorily when applied to other catchments or climatic conditions for which it was not calibrated. Model transposability has long been recognized as the major aim and the most difficult aspect of hydrological simulation models (Klemes, 1986). In many countries, especially in the developing world, basic data for water assessment are sparse or in some cases almost non-existent. This lack of available data is one reason why it is important to develop realistic models that can be applied to ungauged catchments where a historical record of streamflow is not available. Despite this fact, relatively little effort has been expended on the testing of the transposability of existing model types in comparison to the number of published papers on hydrologic modelling.

The procedure recommended by Klemes consists of four levels of testing and aims to test (1) not only a model's ability to simulate current conditions in a given catchment (split-sample test), but also (2) its geographic transposability to other catchments within the same region (proxy-basin test). The proxy-basin test for transposability is crucial when dealing with the problem of rainfall-runoff modelling on ungauged basins. A model's transposability within a catchment (3) is also tested in how well it would reflect changes in climatic inputs or land use (differential split-sample test). The differential split-sample test can also be used to evaluate a model's ability to predict unusually large or rare events that may not be represented in the recorded runoff data. The highest level of testing involves evaluating model performance when testing for (4) geographic and climatic transposability simultaneously (proxy-basin differential split-sample test). Such universal transposability is the ultimate goal of hydrological modelling, a goal that may not be attained in decades to come (Klemes, 1986). However, models with this capability are in high demand and hydrologists are being encouraged to develop them despite the fact that so far even the much easier problem of simple geographical transposability within a region has not yet been satisfactorily resolved (e.g., Chiew and McMahon, 1994; Karnieli *et. al.*, 1994). It is important to implement a standard testing framework such as Klemes proposed so as to raise the level of operational credibility given to simulation models, to discourage exaggerated claims of model performance and to encourage research leading to better models.

### *1.2.3 The issue of statistical significance in model comparisons*

In addition to Klemes' framework, statistical analysis can be an important tool in contributing to the rigour and objectiveness in model comparison studies. In this thesis, I refer to statistical analysis as the evaluation of statistically significant differences between model performance values. Instead of merely being able to state that one model performed better than another at a certain level of testing, researchers need to be able to ascertain if one model performed statistically significantly better than another model or if the model performance values are not appreciably different. Statistical significance is an important issue since the difference between validation results of the models being compared may simply be a result of sampling variability.

The study by Cavadias and Morin (1985) was one of the few, if not only, that attempted to address the issue of statistical significance of runoff models. Cavadias and Morin used two different approaches to compute approximate confidence intervals for the validation results of the operational snowmelt runoff models compared in the international World Meteorological Organization intercomparison study (WMO, 1986).

The first approach was the use of a standard two-way analysis of variance and the second was the use of the Jackknife method.

The ANOVA method, using year and model as the two factors, provided unusable confidence intervals as the researchers found that the basic assumptions of the analysis of variance (homoscedasticity, independence and normality of the residuals) could not be met. Also, since the ANOVA was based on the pooled standard deviation, the difference in variances between models could not be accounted for.

The Jackknife statistic, however, was able to account for the different variabilities of the models since the method generates a distribution based on the data. Although Cavadias and Morin found that the Jackknife method could not provide non-symmetric confidence intervals and is valid only for identically and independently distributed random variables, the researchers determined that the Jackknife statistic allowed for a superior model comparison of validation data over the ANOVA method.

The Jackknife method, described in Section 2.7.3, was used in this study to calculate the differences in variability between the models' performance to allow for statistical analysis. Statistical analysis is a powerful final step in model comparison studies that gives meaning to model performance values, making the results conclusive.

## 1.3 Previous model comparison studies

This thesis focuses on a comparison of spatially lumped and quasi-distributed modelling techniques. Previous model comparison studies are reviewed in this section to provide a background to the research and to demonstrate the relative lack of model comparisons of lumped and distributed models found in the literature.

Numerous studies have compared models of the same type, including Singh's (1976) comparison of unit hydrograph models, the Burges and Lettenmaier (1977) study of time series models, the Weeks and Hebbert (1980) comparison of conceptual rainfall-runoff models, and the World Meteorological Organization (WMO) study comparing operational forecast models (Sittner, 1976). These studies do provide valuable information but since the crucial task in applied hydrology is often to determine the appropriate model for a given application, more and better comparative analyses of models of different types as opposed to a single model type are essential (Kundzewicz, 1986).

More recently, the relative merits of simple hydrologic models versus more complex hydrologic models have become a topic of debate (Beven and Binley, 1992; Grayson *et al.*, 1992; Beven, 1993). An assumption made by some model developers and users is that the more complex models are superior to black box or lumped models. Although several studies have assumed that there is an optimal level of model complexity (Van Genuchten, 1991; Jakeman and Hornberger, 1993), increasing model complexity is usually thought to increase model performance. Many studies comparing simple and complex models have shown, however, that simple models can perform as well as complex models (Naef, 1981; Wilcox *et al.*, 1990; Franchini and Pacciani, 1991; Chiew *et al.*, 1993).

These assumptions have also applied to the more specific comparison of lumped and distributed models. Many researchers feel that incorporating spatial variability of a basin into a model should promote confidence when simulating distributed output (Beven, 1989). The theoretical advantages of distributed models over lumped models, however, have not always proven to be valid in practice (Loague and Freeze, 1985).

Loague and Freeze (1985) compared two simple lumped models and a more complex distributed model using data from three small upland catchments. They found that the three model types used in their study all performed poorly but the simpler, less data intensive lumped models provided as good or better predictions than the more complex, distributed model. The models, however, were tested only at the lowest level of Klemes' tests (split-sample) and were not tested for transposability. Hughes and Beater (1989) found that simpler, lumped versions of the models they examined performed as well as the more complex, quasi-distributed versions. Like Loague and Freeze, Hughes and Beater did not test for transposability.

Distributed models, being less dependent on historical records and reflecting the catchment characteristics, should in principle perform better than lumped models when transposed to another catchment. Several studies have tested for transposability of quasi-distributed models but, unfortunately, did not compare the distributed model to another model during the study (e.g., Ambroise *et al.*, 1985; Chiew and McMahon ,1994; Karnieli *et al.*, 1994). Incidentally, the testing for transposability in these studies was considered to be unsuccessful by the authors.

Michaud and Sorooshian (1994) compared simulations from a complex distributed model, a simple distributed model, and a simple lumped model. When calibration was performed, the simple distributed model proved to be as accurate as the complex distributed model. The spatially lumped model performed very poorly. Although they did not use Klemes' test for transposability, they did compare the simulations produced without calibration. Without calibration, the complex distributed model was more accurate than the simple distributed model. This result is not to indicate that the more complex model may be used with greater confidence on an ungauged catchment, however. The authors concluded that the more complex models should not be abandoned despite their disappointing performance, but rather their potential deserves more research.

Refsgaard (1994, cited by Michaud and Sorooshian, 1994) used Klemes' system and subjected models to all levels of testing. He applied a lumped conceptual model, a distributed model of moderate complexity, and a distributed model of high complexity to three African watersheds. Refsgaard recommended that the lumped model, being the easiest to apply, be chosen over the other more complex models when calibration data are available. For ungauged watersheds, however, he recommended that distributed models be used ~~when~~ data on the physical characteristics of the watershed are obtainable.

Refsgaard and Knudsen (1996) compared three different models on three catchments using all levels of Klemes' testing scheme. They concluded that all models, when calibrated, performed equally well while the distributed models performed marginally better than the lumped model when there was no calibration. The results of these studies reinforce the need for operational models that reflect some sort of basin characteristics to allow for transposability.

With the results of these comparative studies, the development, application, and evaluation of complex distributed models continues to be an active area of research. The more complex models have their drawbacks such as overparameterization and data requirements, but their potential as research tools is encouraging, especially with the increasing availability of spatially distributed watershed data received from remote sensing (Beven, 1989, 1992; Beven and Binley, 1992; Jensen and Mantoglou, 1992; Grayson *et al.*, 1993). Although there appears to be agreement regarding the potential of the distributed types of models, there is no consensus as to whether they offer a significant improvement in performance when compared to the well proven lumped conceptual model type. Klemes' (1986) validation tests in combination with statistical analysis of the results are important to carry out in comparative studies to address the question of optimal complexity and transposability of rainfall-runoff models.

## 1.4 Research objectives

The research had two main objectives: (1) to compare a quasi-distributed model to two lumped models to determine if there is a benefit associated with the increased demand for catchment data and (2) to determine if the statistical significance of differences in model performance can be quantified.

The models used in this study were (1) a lumped black box rainfall-runoff model, (2) a lumped conceptual model, and (3) a quasi-distributed conceptual rainfall-runoff model, TOPMODEL. I compared model performance for the three different types of rainfall-runoff models by applying them to two small forested catchments within the University of British Columbia (UBC) Research Forest.

The working hypothesis was that the more conceptual, quasi-distributed model should perform better than the lumped models when transposed to another catchment and/or climatic conditions, since it accounts for the catchment topography and actual field processes. Also, the study explored the feasibility of carrying model comparisons through a further step of testing for statistically significant differences. This research contributes to the state-of-the-art of hydrologic modelling by providing further information on the relative performance of quasi-distributed and lumped rainfall-runoff models, specifically in forested catchments with shallow permeable soils. The research will also contribute to the progress of rainfall-runoff modelling in general by promoting a more rigorous method for model comparison.

### 1.5 Structure of the thesis

Chapter two presents the methods and data sources used in this research, including information on the study area, meteorological data, the selection of rainfall-runoff events, the generation of topographic information, and the model evaluation criteria. The models used are described in detail in Chapter three. The results of the model comparison are presented in Chapter four with the discussion of the results following in Chapter five. Chapter six will summarize the key findings and outline suggestions for further research.

# CHAPTER 2

# METHODS AND DATA SOURCES

This chapter describes the study area, streamflow and precipitation data, and the selection of rainfall-runoff events used in the study. Model evaluation and performance criteria such as the testing scheme and method of statistical analyses are presented.

## 2.1 Study area

The study focused on two catchments within the University of British Columbia (Malcolm Knapp) Research Forest, located approximately 50 km east of Vancouver, B.C. (Figure 2.1). The two catchments are the East-upper and the South watersheds as referred to by Feller (1988). The East-upper watershed will be herein referred to as East catchment. East catchment has an area of 38.3 ha with almost all (93%) of its area covered by mature forest. South catchment has an area of 19.8 ha with 100% of its area covered by mature forest. South catchment ranges in elevation from 175 m to 319 m while East catchment ranges from 280 to 447 m.

The climate of the forest is characterized by frequent cloudiness, wet, mild winters, cool and relatively dry summers, and a long frost free period (Klinka and Krajina, 1986). Precipitation, mainly produced by Pacific frontal systems, ranges from 2000 to 2500 mm per year of which over 70% falls between October and April (Utting, 1979). Less than 15% of the total precipitation occurs as snow because of the moderating effects of the Pacific and the low elevation (Utting, 1979).

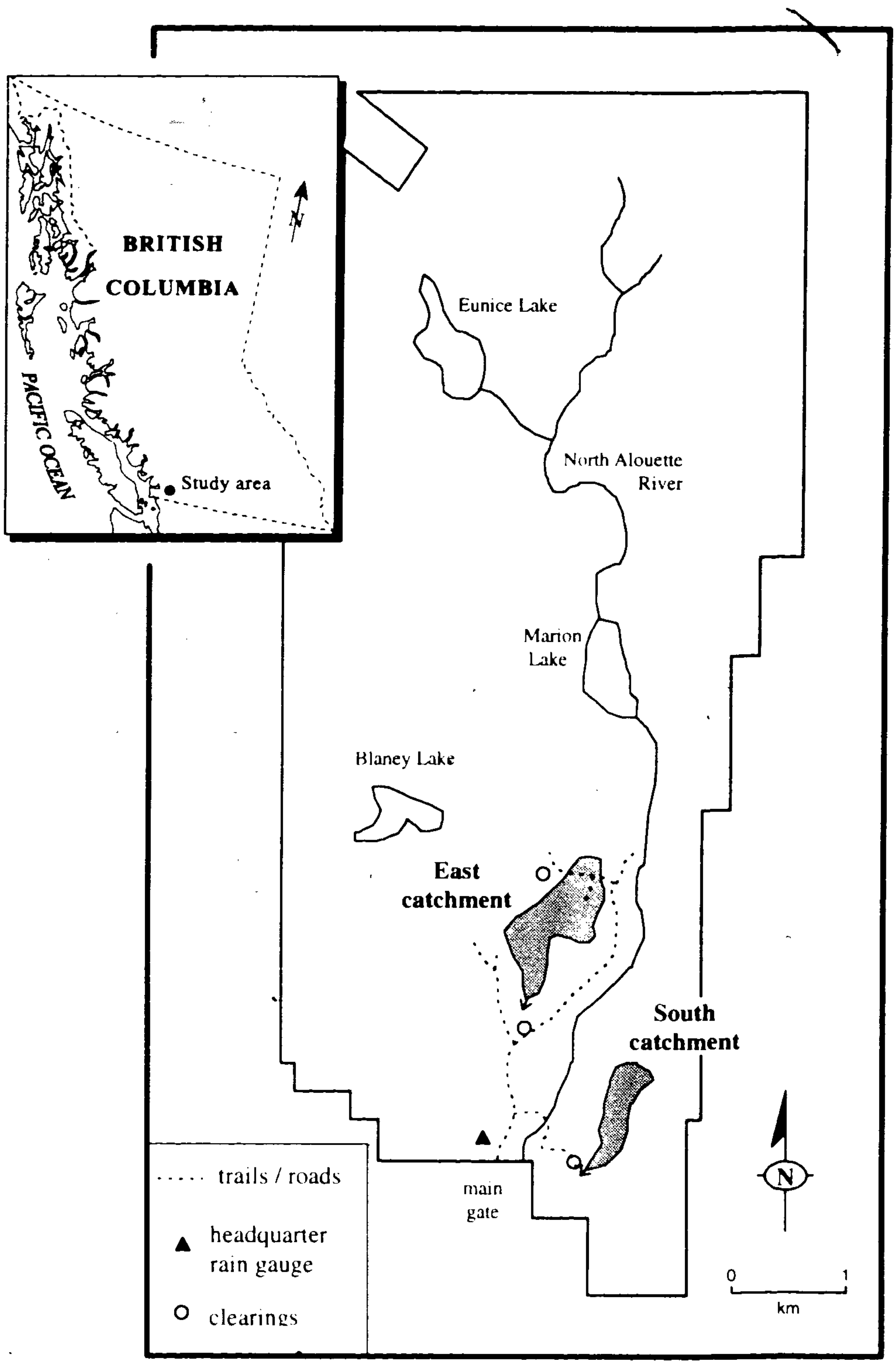



**Figure 2.1** Location of study catchments within the UBC Research Forest.

The Research Forest is typical of coastal forested mountainous terrain (Power, 1984). The dominant forest cover is western hemlock (*Tsuga heterophylla*) with smaller proportions of western red cedar (*Tsuja plicata*) and Douglas fir (*Pseudotsuga menziesii*). The stands are about 70 years old, having been logged and replanted in the 1920's, and have crown closures of 75-95 percent.

Klinka and Krajina (1986) summarized the geology at the Research Forest as predominantly quartz diorite, a finding supported by visible outcrops within the study catchments. The bedrock is overlain by thin and continuous deposits of glacial origin which are coarse-grained with average textural values of 57% sand, 41% silt, and 2% clay (Klinka and Krajina, 1986).

The soils within the Research Forest range from 0.5 to 2 m in depth and are highly permeable, underlain by a relatively impermeable compact till or bedrock at an average depth of 1 m (Utting, 1979). The dominant soil class is humic-ferric podzol and, texturally, the soils are coarse with Sandy Loam being typical (Utting, 1979). The soils in the Research Forest exhibit a strongly structured B horizon due to the presence of many roots, stones, and cemented aggregates (Tischer, 1986). Hydraulic conductivities are about $10^{-4}$ to $10^{-3}$ m $s^{-1}$ in the Research Forest soil, and about $10^{-7}$ to $10^{-6}$ m $s^{-1}$ in the underlying till (Utting, 1979; Cheng, 1988).

The Research Forest is relatively homogeneous in terms of climate, soil, and forest cover characteristics. The most significant difference between East and South catchments is the topography. This is important as the two catchments provide an excellent opportunity to compare the performance of quasi-distributed and spatially lumped models to determine if incorporating topographic information into a model provides better runoff estimates.

## 2.2 Streamflow data

Dr. M. Feller of the Department of Forest Science, UBC, has been measuring discharge in both South and East catchments since 1985 and has made the stage charts and corresponding stage-discharge equations available for this project. Stage charts for each event in each catchment were digitized at every change in slope to generate a file of stage-time data at irregular intervals. The stage-time values were converted into hourly discharge values by first interpolating the stage at six minute intervals, converting each stage value into a discharge value by using the stage-discharge relation for each catchment, and then integrating the ten, six-minute values into one average hourly discharge value. These hourly averages constitute the observed streamflow values used for evaluation of model performance.

## 2.3 Precipitation data

### *2.3.1 Data sources*

Historical climate data were obtained from the Atmospheric Environment Service (AES) for station 1103332 (Haney UBC RF Administration). The station, including a tipping bucket precipitation gauge, is located at the Research Forest headquarters within the UBC Research Forest at an elevation of 143 m, approximately 1.5 km from South catchment and 2 km from East catchment (Figure 2.1). The data include hourly and daily precipitation totals both for rain and snowfall. I compared the hourly precipitation data from the tipping bucket gauge for 24 hour periods to the daily totals to verify the correctness of the hourly values.

Gauges were placed in clearings outside of each of the two catchments for comparison to the headquarter data (Figure 2.1). Four gauges were placed in one clearing near South catchment and four gauges were placed in each of two clearings outside of East catchment, due to its larger size. The precipitation collected in the gauges was measured every two weeks from September, 1994 to June, 1995. These data were assumed to equal above-canopy precipitation. Data from Hetherington's (1976) Ph.D. thesis were also drawn upon to help define above-canopy precipitation at East catchment.

### 2.3.2 *Extrapolation of headquarter precipitation data*

The precipitation data recorded at the headquarter climate station were extrapolated to each catchment as above-canopy precipitation and used as input to a canopy storage (throughfall) model. I assumed uniform spatial distribution of rainfall within each catchment because of the relatively small sizes of the two catchments. It was assumed that above-canopy precipitation at each catchment ($P_c$) followed a simple proportional relation with precipitation at Research Forest headquarters ($P_{hq}$):

$$P_c = b\, P_{hq} \qquad (2.1)$$

where b is a constant of proportionality. Separate values of b were calculated for each catchment and each two week period, as follows:

$$b = P_c / P_{hq} \tag{2.2}$$

The values of b based on the bi-weekly intervals were then averaged for each catchment. It should be noted that the agreement between the bi-weekly values does not guarantee agreement at the hourly time scale.

At South catchment, a proportionality factor of 0.94 was found to apply (Figure 2.2). At East catchment, ratios were calculated for sites at the north and south ends and averaged. Due to problems with animals disturbing gauges at the north end of East catchment, data from Hetherington (1976) were used to estimate a ratio of 1.19. For the south end, a ratio of 1.13 was calculated (Figure 2.3). The average of the two ratios, 1.16, was therefore used in Equation 2.1 to estimate the above-canopy precipitation for East catchment from Headquarter data. There were fewer depth values collected at East catchment (Figure 2.3) than South catchment (Figure 2.2) because of the problem with animals chewing gauges at East catchment.

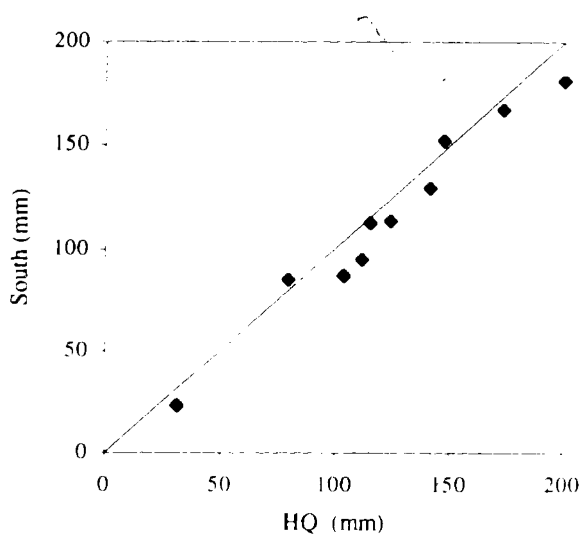



**Figure 2.2** Bi-weekly measurements of South catchment clearing depths versus Headquarter depths with 1:1 line.

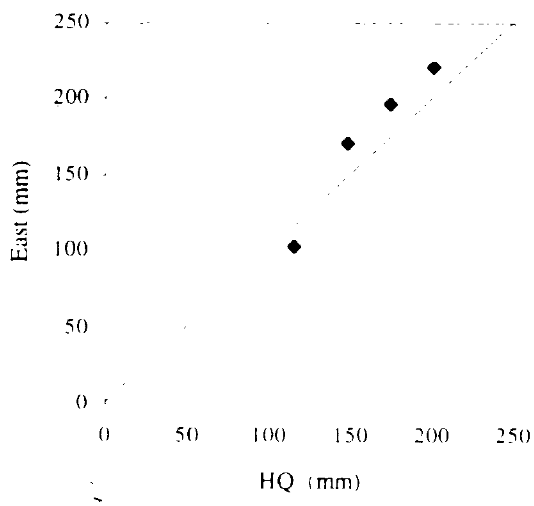



**Figure 2.3** Bi-weekly measurements of East (south end) catchment clearing depths versus Headquarter depths with 1:1 line.

## 2.4 Throughfall data

Throughfall is the portion of the above canopy precipitation that reaches the forest floor directly through gaps in the forest canopy and includes canopy drip. Quantitatively, it is the difference between precipitation and the combination of canopy interception loss and stemflow (Lee, 1980).

Throughfall was sampled every two weeks from September, 1994 to June, 1995, to calibrate a canopy-storage model independently of the rainfall-runoff models. Throughfall was sampled along one 100 m transect in South catchment and two 50 m transects in East catchment (Figure 2.4). Two transects in East catchment were used because of the larger size of the catchment and the ability to access East catchment from more than one location.

I randomly placed one gauge along each 10 m section of the transect and randomly relocated each gauge within the same 10 m section after each bi-weekly collection period to minimize sampling bias. If it was impossible to place a gauge at its randomly generated site because of a stump, rock, etc., then the gauge was placed 5 cm nearer the zero mark of the 10 m section.

The throughfall gauges consisted of funnels attached to 4 L polyethylene containers which were secured in the field by wooden stakes. The throughfall gauges had a funnel diameter of 106 mm yielding a gauge capacity of 450 mm of throughfall. Bi-weekly collection periods were adequate to guard against overflow.

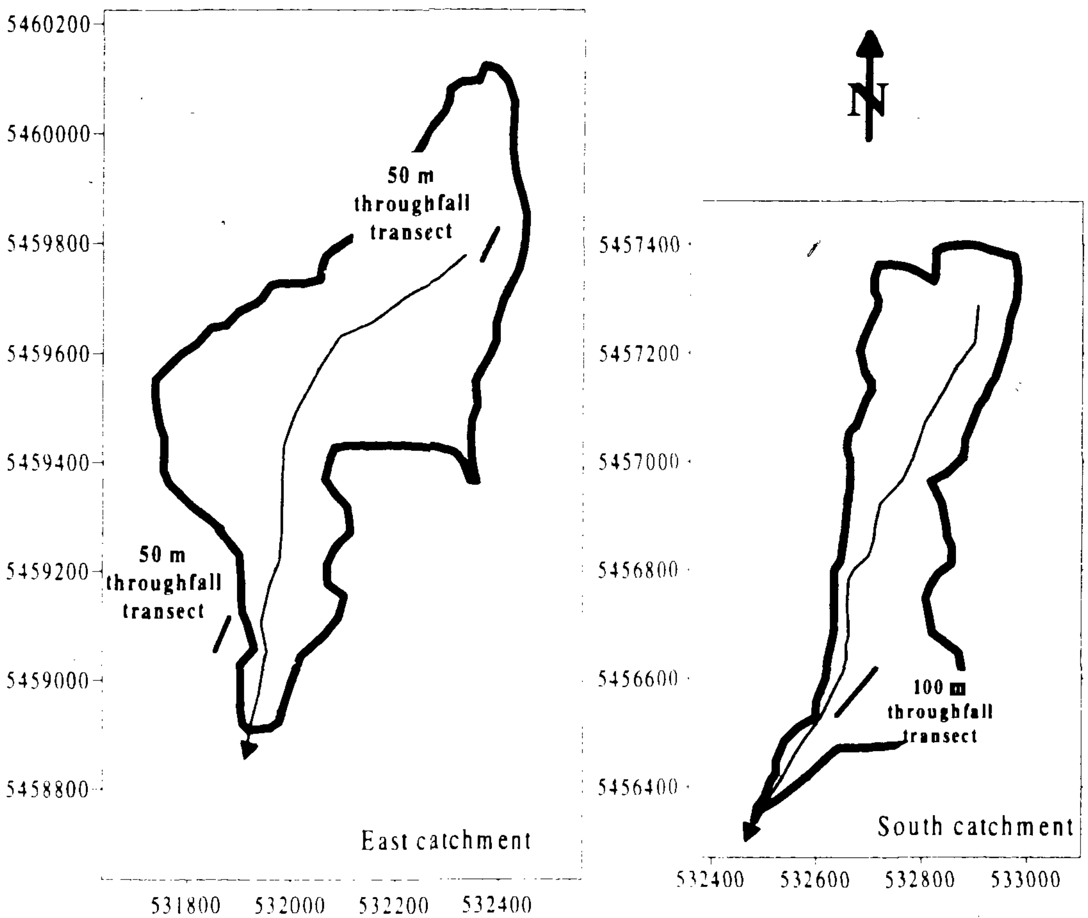



**Figure 2.4** Location of the throughfall sampling transects in East and South catchments. The UTM coordinates (m) are shown for scale with 10 m contour intervals.

For each catchment, I calculated the average precipitation for each collection period for the throughfall gauges under the canopy. Comparing the average throughfall values to the volumes in the clearings, the average interception loss was 23% with a range from 0% to 58%. The wide range of values is due to the varied precipitation events throughout the study period and the sampling variability. These values are comparable to a past interception calculation done at South catchment (Thompson, 1994) and with interception amounts from other studies (Rothacher, 1963; Patric, 1966; Loustau *et al.*, 1992).

## 2.5 Selection of rainfall-runoff events

Events were defined as periods of significant rainfall separated by at least six hours of rainfall intensities averaging less than 0.1 mm $h^{-1}$ (Harr, 1977; Pierson, 1980). The events were determined using stage and precipitation data that were available for both South and East catchments from 1985 to 1992. Each event started at least four (one hour) timesteps before precipitation began so that the first observed discharge value was still on a recession curve. I excluded events in which snowfall was recorded or the stage records indicated a freeze-up as the stage record for that event would not correctly represent the actual flow.

To avoid complications with specifying initial soil moisture conditions for each event, it was assumed for all model runs that there were no initial losses of throughfall inputs to soil moisture storage. An attempt was therefore made to include only events for which the catchment was initially 'wetted up.' Accounting for antecedent soil moisture in the events before the models are applied minimizes the possibility of confounding by having to optimize the soil parameters for each model. This procedure therefore results in a less ambiguous test of the transformation routines of the models.

In a first cut, I narrowed the list of rainfall-runoff events down to 63 for each catchment by including only events that occurred during the wet season, between the months of October and May, to minimize the effect of soil moisture losses to evapotranspiration. In a closer analysis of the remaining events, a hydrological response value was calculated which represents how 'wetted up' a catchment is. Following

Hewlett and Hibbert (1967), the hydrological response $HR$ for each event was calculated as

$$HR = Q_q / TF \qquad (2.3)$$

where $Q_q$ is the quickflow ($Ls^{-1}$) and $TF$ is the throughfall (mm). A constant separation slope of 0.0055 $L\,s^{-1}\,ha^{-1}\,hr^{-1}$ was used to divide the hydrograph of each event into quick and delayed flow (Figure 2.5).

I identified the events with significantly low hydrological response values ($HR < 0.30$) as not being sufficiently 'wetted up'. Figures 2.6 and 2.7 show that when the hydrological response is plotted against the initial discharge ($Q_{min}$), the event with an $HR$ less than 0.30 appear to be set off from the rest of the events, indicating the catchment in these events have been dried out sufficiently to require a large amount of the throughfall to infiltrate before quickflow begins.

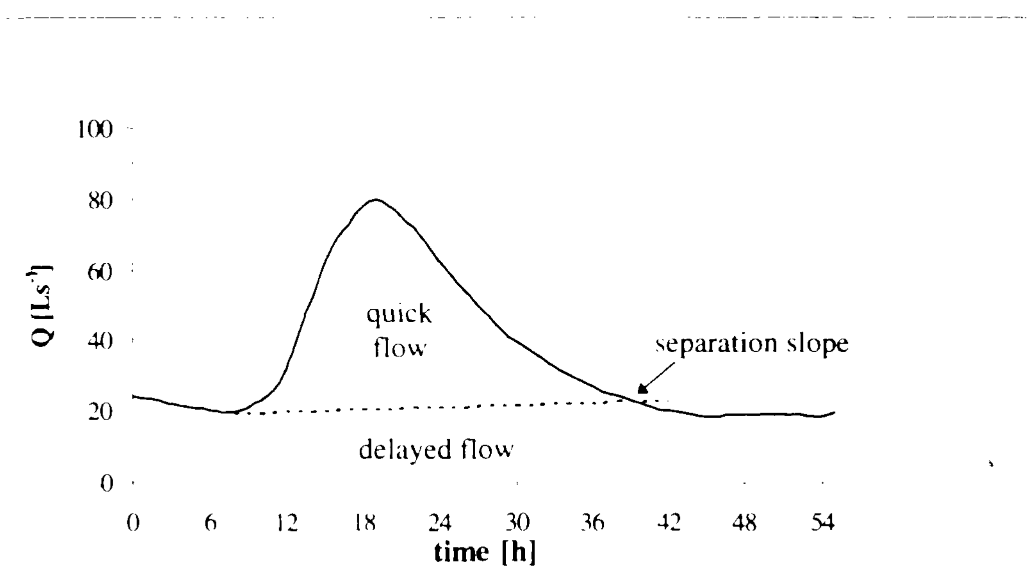



**Figure 2.5** Hydrograph representing separation of event 3, South catchment.

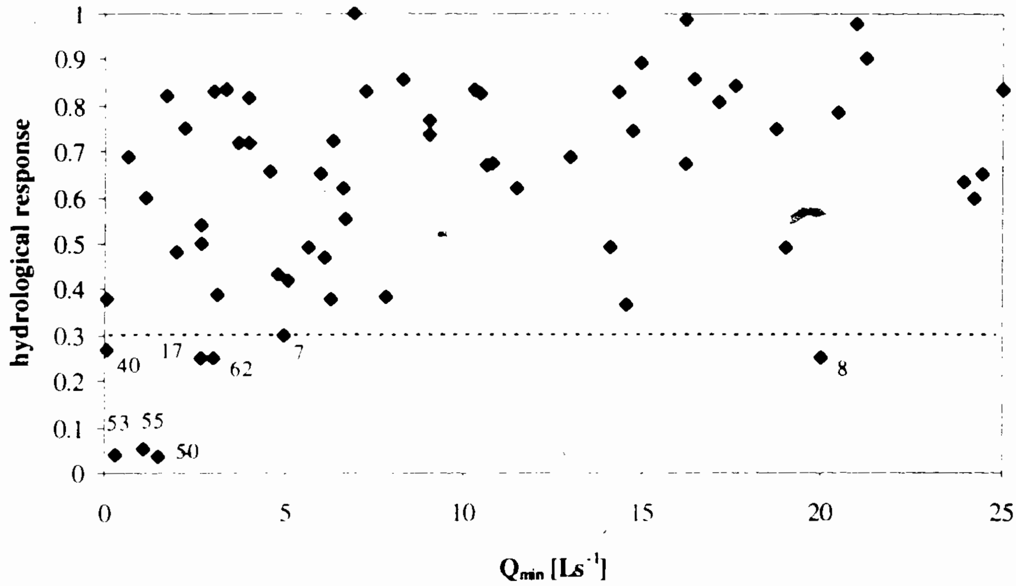



**Figure 2.6** Hydrological response of events in South catchment. All numbered events with an *HR* below 0.30 (below the dotted line) were deleted.

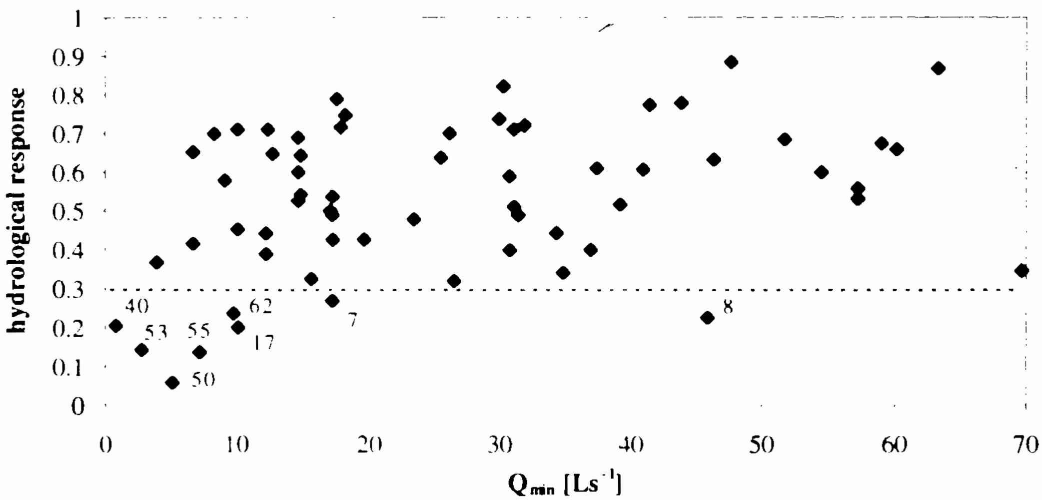



**Figure 2.7** Hydrological response of events in East catchment. All numbered events with an *HR* below 0.30 (below the dotted line) were deleted.

From the initial 63 events, eight events with an $HR < 0.30$ were eliminated in which substantial soil moisture losses appeared to have occurred. May events were also deleted to avoid events where there may have been notable evapotranspirative losses of soil moisture. As a result, 52 events remained for each catchment. The partitioning of these events into calibration and validation data sets is discussed in Section 2.7.2 with the description of the testing scheme. Event dates and data set breakdowns are shown in Appendix 1.

## 2.6 Topographic indices

The quasi-distributed rainfall-runoff model, TOPMODEL, represents catchment topography by means of the probability distribution of a topographic index, $\ln(a/\tan\beta)$, where $a$ is the area drained per unit contour and $\beta$ is the local slope angle. The $\ln(a/\tan\beta)$ attribute is an important component of many physically based geomorphic and hydrologic models, as it is assumed to characterize the spatial distribution of soil moisture, surface saturation, and runoff generation processes (Beven and Kirkby, 1979; Hornberger *et al.*, 1985; O'Loughlin, 1986; Moore *et al.*, 1988; Romanowicz *et al.*, 1993).

The topographic indices were derived from Digital Elevation Models (DEMs), which were generated for both South and East catchment by digitizing UBC Research Forest 1:5000 topographic maps. These maps have a 5 m contour interval, and were the most recent and of the largest scale available.

The DEMs have a 10 m grid size. Zhang and Montgomery (1994) applied the same quasi-distributed model used in this study, TOPMODEL, to forested catchments with moderate slopes. They showed that simulations using a 10 m grid size showed significant improvements over using a 30 m grid size but that no real benefit was gained with a finer resolution below 10 m. They argued that a 10 m grid size provided an optimal trade-off between topographic resolution and minimizing data storage requirements when modelling a moderately-steep forested catchment. Moore and Thompson (1996) found that topographic indices derived from an 8 m grid provided statistically significant predictions of soil saturation at a site adjacent to South catchment, while indices derived from a 16 m grid had little predictive power. Hence, the 10 m grid size is a reasonable choice for South and East catchments.

The topographic indices were generated by a computer program developed by Dr. R.D. Moore of Simon Fraser University. For each grid point within a catchment, the program calculates the upslope contributing area (A) draining through a point using a multiple flow direction algorithm described by Wolock and McCabe (1995), which is similar to that used by Quinn *et al.* (1991). The accumulated upslope area for any one cell is distributed amongst all of the eight neighbouring cells which are lower than it. The fraction of the area draining though each grid element to each downslope direction is proportional to the gradient of each downhill flow path, so that steeper gradients will naturally attract more of the accumulated area. The multiple flow path method is assumed to give a more realistic pattern of contributing area on the hillslope portion of the catchment than the single flow path method (Quinn *et al.*, 1991; Wolock and McCabe, 1995). Contour width ($w$) and slope angle ($\tan \beta$) are generated for each grid cell and, with the upslope contributing area (A), used to calculate the topographic index

$\ln(a/\tan\beta)$. The distribution of the topographic index for each catchment is shown in Section 3.5.3 with the description of the quasi-distributed model.

## 2.7 Model evaluation and performance criteria

### *2.7.1 Calibration and validation*

Calibration is the process whereby parameters of the model are adjusted to make the simulated output best match the observed data, based on evaluation of graphs and numerical indices. Validation involves running the calibrated model on an independent data set. Since data which contain larger hydrologic variability are more likely to result in more reliable parameter estimates, the calibration data sets for the first two levels of testing were chosen such that they cover the greatest variability of events. The validation data sets for the first two levels of testing were also chosen such that they also cover the variability of events to best test the model's ability to estimate a greater range of runoff events. The calibration and validation data sets for the third and fourth level of testing were determined by event volume as described in Section 2.7.2.

### *2.7.2 Klemes' (1986) hierarchical tests*

Klemes' hierarchical testing scheme was used to evaluate model performance. The system is hierarchical since the modelling tasks are ordered according to their increasing complexity.

*i) Split-sample test*

This is the most elementary test of a model and the most commonly used in the hydrologic literature. One calibration and one validation set is used for the split-sample test. Using the 52 events that remained after meeting the criteria of event selection, 26 events were available for each of the calibration and validation sets for the first and second level of testing. The parameters of the model are calibrated using the calibration set. The model is then tested by running the calibrated model using the input data of the validation set. The model passes the split-sample test if validation performance is adequate for the intended application. The split-sample test may provide useful information when the objective is to estimate and fill in missing streamflow data of events of a similar magnitude to those for which a model was calibrated.

*ii) Proxy-basin test*

This level of testing in Klemes' system evaluates geographic transposability. The calibration and validation data sets of events used for the proxy-basin test are the same sets used for the split-sample test of level 1. Instead of calibration and validation being carried out on one catchment, the model is calibrated on one catchment and validated on the other. Using the two gauged basins South and East, a model is calibrated on South catchment and validated on East Catchment, and also calibrated on East catchment and validated on South catchment.

*iii) Differential split-sample test*

For the differential split-sample test, the same approach as the split-sample test is followed but the calibration and validation data sets are divided by climatic conditions or event volume. For example, calibrating a model for a dry period and then running a test using data for a wet period provides a necessary, though not sufficient, test of whether the model is valid for predicting the effects of a change to a wetter climate. In this study, the calibration set comprises the small volume events and the validation set is the large volume events. The models are thus tested whether they can be reliably used to simulate events of a greater magnitude than those for which it was calibrated.

For the third and fourth level of comparison, in which the model is being verified on larger runoff volume events from which it was calibrated, the events needed to be split into large and small events. I ranked the events by runoff volume and split the events into two sets, small and large volume events, deleting the middle six events to emphasize the difference between the two sets (Appendix 1). The calibration set is composed of the small volume events and the validation set is composed of the large volume events since it is usually the change from small and average peak events to the rare large peak flows that are of concern.

*iv) Proxy-basin differential split-sample test*

This level of testing evaluates both the geographical and climatic transposability of a model and is the most stringent test. It was carried out by calibrating for the small events at South catchment and validating on the large events at East catchment, and conversely, calibrating for the small events at East and validating on the large events at South.

### *2.7.3 Model performance criteria*

*i) Numerical and graphical criteria*

The best basis for judging model performance is a comparison of model estimates with observations both quantitatively (numerically) and qualitatively (visually) (Willmott, 1984). Even if a model is based on the best available knowledge, comparison of the results with observations is the only way of establishing confidence in the simulation.

No single index of goodness-of-fit is suitable for describing how well a particular model performs. Therefore, I used a number of quantitative indices for model evaluation: (1) the Nash model efficiency value, (2) percent deviation, (3) mean and (4) standard deviation. These criteria, like any other criteria, are only estimates of model performance that are specific to the period modelled and dependent on the quality of the observed data (Bergstrom, 1991).

Among the most commonly used indices is the model efficiency coefficient $E_m$ of Nash and Sutcliffe (1970) (e.g. Beven *et al.*, 1984; Loague and Freeze, 1985; Gan and Burges, 1990; Chiew and McMahon, 1994). The model efficiency expresses the fraction of the measured streamflow variance that is accounted for by the model:

$$E_m = 1 - \left[ \frac{\sum_{i=1}^{n} (Q_{sim} - Q_{obs})^2}{\sum_{i=1}^{n} (Q_{obs} - \overline{Q}_{obs})^2} \right] \qquad (2.4)$$

where $Q_{sim}$ and $Q_{obs}$ are the average hourly simulated and observed discharges respectively at each hourly timestep $t$, and $\overline{Q}_{obs}$ is the average observed discharge over all timesteps. The model efficiency has a maximum value of 1.0 which represents a perfect fit between simulated and observed values. A model efficiency of zero indicates that the mean of the observed data is as efficient a predictor as the model, while a negative efficiency indicates that the model is a worse predictor than the observed mean.

The model efficiency has been shown to be the best objective function for reflecting the overall fit of a hydrograph (Servat and Dezetter, 1991). It is appropriate to use the model efficiency for event modelling since its strengths lie in evaluating the overall hydrograph shape and fit to peak flows while its weakness is assessing the fit to low flows. Therefore, the model efficiency value $E_m$ (Equation 2.4) was used to calibrate the models according to the overall fit of the hydrograph. The optimum parameter values for each calibration data set were determined by the maximum $E_m$ value attained.

The model efficiency $E_m$ was also calculated for event volumes, peak discharges, and time-to-peak values using the appropriate variables in Equation 2.4. In addition, percent deviations (WMO, 1986) were calculated for event volumes, peak discharges, and time-to-peak. For example, the following formula evaluates the percent deviation for peak discharge:

$$\text{deviation [\%]} = \sum_{t=1}^{n} \left[ \frac{Qpk_{obs} - Qpk_{sim}}{Qpk_{obs}} \right] * 100 \qquad (2.5)$$

where $Qpk_{obs}$ and $Qpk_{sim}$ are the observed and simulated peak discharges respectively at each hourly timestep $t$. The arithmetic mean and standard deviation were also calculated for the event volume, peak discharge, and time-to-peak. Performance of the models was assessed graphically using linear scale plots of simulated and observed hydrographs versus time, and scatterplots of observed and predicted values for peak discharge, runoff volumes, and time-to-peak.

*ii) The Jackknife procedure*

A problem with comparing model performance based on the numerical indices described above is that they depend on the selection of events used to calculate them. A different set of events would be associated with different performance indices.

A method is needed to account for this sampling variability and to determine statistically if one model is performing significantly better than another model under the conditions tested. For this purpose, the Jackknife procedure was used to generate sampling distributions of possible model performance values based on each data set of events. Variance estimates for these objective functions were then calculated and further statistical analysis performed to determine statistical significance.

The Jackknife procedure has become a general tool for estimating both the value of a statistic, its variance, and its confidence region in the field of ecology (Potvin and Roff, 1993) and other fields where the sampling distribution is difficult to derive or the distribution is likely to be highly skewed. It is considered a robust and multipurpose procedure since it is applicable and reliable under a wide variety of conditions (Miller, 1974; Neter *et al.*, 1982).

Specifically, the Jackknife procedure was used to generate estimates of a population mean of the root mean square error (RMSE), defined as:

$$\mathrm{RMSE} = \left[ n^{-1} \sum_{t=1}^{n} (Q_{\mathrm{sim}} - Q_{\mathrm{obs}})^2 \right]^{0.5} \tag{2.6}$$

where $Q_{\mathrm{sim}}$ and $Q_{\mathrm{obs}}$ are the simulated and observed discharges at each hourly timestep $t$. The $E_m$ and RMSE are related by the equation

$$E_m = 1 - \left[ \frac{n \cdot \mathrm{RMSE}^2}{\sum_{t=1}^{n} (Q_{\mathrm{obs}} - \overline{Q}_{\mathrm{obs}})^2} \right] \tag{2.7}$$

The Jackknife estimate is obtained by consideration of all possible subsets of the data in which one event has been eliminated from the original set. For example, to estimate J (the Jackknife estimate of the average root mean square error, RMSE) with $n$ events in a data set, an RMSE value, which I will call $J_i$, is calculated $n$ times, each time omitting one event. A pseudovalue ($S_i$) is calculated for each altered data set with its one deleted event as:

$$S_i = n\,\mathrm{RMSE} - (n-1)\,J_i \tag{2.8}$$

The Jackknife estimator J is simply the mean of the $n$ pseudovalues:

$$J = (1/n) \sum_{d=l}^{n} S_i \tag{2.9}$$

where $d$ refers to the number of datasets. Therefore, for each value of RMSE, a distribution of possible RMSE values for that data set are generated and statistical analysis can now be performed to determine if the RMSE estimates, and indirectly the Nash model efficiency coefficients, are significantly different.

The benefit of using the RMSE instead of the $E_m$ in the Jackknife procedure is that the RMSE does not have an upper limit, such that it makes the RMSE more amenable to statistical interpretation than the $E_m$. The concern is that the Jackknife procedure generates pseudo-$E_m$ numbers greater than unity, which are not valid for $E_m$ values. For example, a pseudo-$E_m$ value greater than one indicates a better than perfect fit and an RMSE less than zero, which is impossible. In addition, it was simple to transform the RMSE pseudo-values so that their distributions conformed to the assumptions of analysis of variance. As shown by Equation 2.7, the $E_m$ values are transformations of the RMSE values, so the results of analyzing RMSEs could be easily interpreted in terms of the model efficiency coefficients.

*iii) Analysis of Variance*

The Jackknifed model performance value for each model within each catchment was compared using two-way fixed effects analyses of variance (ANOVAs) at each of the four levels of testing to determine if model performance differences were statistically significant. The four ANOVAs used the fixed effects of model and catchment and included the model-catchment interaction. The ANOVAs are of a fixed-effects model type since the model and catchment factors have been specifically selected for analysis and have not been randomly selected. Therefore, no inferences can be drawn about any other levels of the two factors except the ones used in the study.

The RMSE pseudovalues calculated in this study are independent values and, after transformation by adding 1 and then taking the log of all pseudovalues, had common variance and met the assumptions of the ANOVA. The Tukey t method was used for the investigation of significant effects once they had been determined through the ANOVAs. The following chapter describes the models used in the study.

# CHAPTER 3

# DESCRIPTION OF THE MODELS

In this chapter, the characteristics of forested catchments are first discussed, with the following sections describing the structure, governing equations, and the calibration of parameters for the canopy storage model and each of the three rainfall-runoff models. The program code for each of the models can be found in Appendix 2.

## 3.1 Distinctive characteristics of forested catchments

### *3.1.1 Influence of forest canopy*

The hydrology of a forested catchment differs from that of a non-forested catchment. The canopy cover, depending on the density, intercepts much of the precipitation. In the UBC Research Forest, as in other temperate regions, about 30% of precipitation is intercepted and only 70% reaches the soil surface (Klinka and Krajina, 1986). This interception loss to a forest canopy greatly influences the amount of water reaching the soil surface and resulting in streamflow. It is crucial, then, to account for the amount of water actually reaching the forest floor in a canopy storage model before using the precipitation as input to the rainfall-runoff model.

It is also important to calibrate a canopy storage model as a separate component from the runoff models to avoid confounding. Confounding in the statistical sense refers to the influence of variables or factors which have not been properly taken into account. Since the objective of this study is primarily to test only the transformation routines of the rainfall-runoff models, it is important to separate out any interactions they may have with a canopy storage model. To minimize problems of interpreting model performance in the presence of model interactions, the canopy storage model has been calibrated independently of the streamflow data.

### *3.1.2 Forest soil infiltrability*

The high infiltrability of forest soils and subsequent subsurface flow dictates the hydrologic modelling approach required. With hydraulic conductivities typically about $10^{-4}$ to $10^{-3}$ $ms^{-1}$ in the Research Forest soil (Cheng, 1988) and the low rainfall intensities associated with the dominant frontal systems of the study area (Loukas and Quick, 1996), virtually all the water that reaches the ground surface infiltrates to become soil moisture and subsurface flow. The high infiltrability is a result of a permeable layer of humus on top of the mineral soil and tree roots and other vegetation that create macropores in the soil giving rapid vertical percolation. The high infiltration rates result in Hortonian overland flow (surface flow that has not infiltrated) being virtually non-existent. The subsurface flow is considered the most important component of the runoff of a forested catchment (Band *et al.*, 1993).

Saturation overland flow can also be a contributing factor to runoff due to the high infiltrability and shallow soils of the study catchments. Saturation overland flow occurs when subsurface flow is unable to discharge all the water that infiltrates into the soil. The water table rises to the ground surface and any throughfall falling on the saturated areas then runs off as overland flow. The source area for saturated overland flow is of variable size as it expands in area during a storm event, then shrinks as the soil drains (Bernier, 1985). Saturation overland flow usually occurs in areas adjacent to stream channels and other areas of the catchment where there is flow convergence. The quasi-distributed model used in this study accounts for saturation overland flow.

## 3.2 Canopy storage model

### *3.2.1 Structure and governing equations*

The amount of water reaching the soil surface, required as input to the rainfall-runoff models, was computed by solving the canopy water balance through the course of a storm. The canopy storage model was developed using some of the assumptions underlying Gash's (1979) analytical model of interception and has both a low demand for data and a simple but realistic approach to the interception process.

Figure 3.1 illustrates how during a storm, the canopy is wetted up and throughfall occurs when the canopy becomes saturated. The canopy storage model estimates this throughfall along with stemflow and free throughfall (precipitation through canopy gaps) to account for the total input reaching the soil surface.

For each ten minute time step, evaporation from the canopy ($E$) is calculated as

$$E_t = \overline{E}\,(S_{t-1} / S_c) \tag{3.1}$$

where $\overline{E}$ is the mean evaporation rate from a fully wetted canopy (mm h$^{-1}$), the canopy storage $S_{t-1}$ is the amount of water being held in the canopy at the beginning of the time

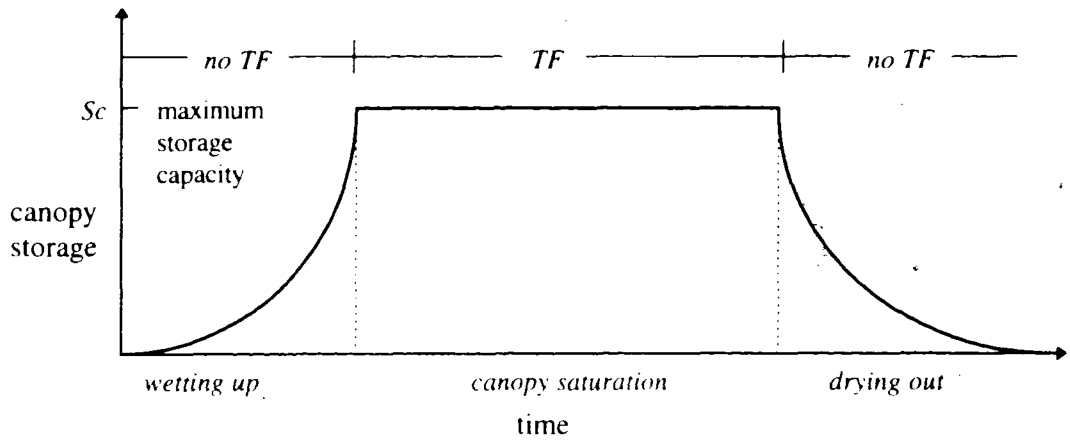



**Figure 3.1** Generalized variation of canopy saturation and throughfall (TF) over one storm event.

step (mm), and $S_c$ is the canopy storage capacity (mm). Following the assumptions of the Gash (1979) model, $\overline{E}$ is treated as a constant value for a given forest stand. The canopy storage capacity is defined as the amount of water left on the canopy in zero evaporation conditions when rainfall and throughfall have ceased (Gash and Morton, 1978).

As water evaporates and the canopy receives precipitation ($P_c$), the canopy storage is updated:

$$S_t = S_{t-1} + P_c - E_t \tag{3.2}$$

It is assumed that no canopy drip occurs unless the canopy storage capacity is full; hence, there is no drip during the wetting up phase.

Throughfall ($TF$) and stemflow ($SF$) occur when the computed canopy storage exceeds its storage capacity ($S \geq S_c$):

$$TF = S - S_c \tag{3.3}$$

$$SF = P_t(P_c) \tag{3.4}$$

where ($p_t$) is the proportion of the rainfall diverted to trunk and stemflow. Following the calculation of *TF* and *SF*, the canopy storage is set to $S_c$.

The throughfall and stemflow are integrated over six 10 minute time steps giving hourly values. For each hourly timestep, the total input to the soil surface used in the rainfall-runoff models (*I*) is computed as

$$I_t = TF_t(1-p) + P_{ct}(p) + SF_t \tag{3.5}$$

where $p$ is the free throughfall (gap) coefficient, which determines the amount of above canopy precipitation ($P_c$) falling directly to the forest floor through canopy gaps.

### *3.2.2 Calibration of parameters*

The canopy parameters are the free throughfall (gap) coefficient ($p$), the stemflow coefficient ($p_t$), and the canopy storage capacity ($S_c$). I calculated the gap coefficient ($p$) for each catchment by using crown closure codes from the UBC Research Forest map of forest cover. A weighted average of percent crown closure within each catchment was calculated by multiplying the amount of area covered by a closure class by the average percentage value of that closure class (Table 3.1). The gap coefficient was calculated as

$$p = 1 - \text{crown closure} \tag{3.6}$$

giving values of 0.159 and 0.203 for South and East catchment respectively. I assumed a value of two percent of total rainfall for the stemflow coefficient ($p_t$), based on past research done at the UBC Research Forest (Hutchinson and Roberts, 1981).

**Table 3.1** Distribution of crown closure for South and East catchments.

| | | percent of total catchment area in each crown closure class | |
|---|---|---|---|
| crown closure class (% closure) | average closure | South catchment | East catchment |
| class 0 (0 - 5 %) | 2.5% | | 4% |
| class 3 (26 - 35 %) | 30.5% | | 1% |
| class 5 (46 - 55 %) | 50.5% | | 2% |
| class 8 (76 - 85 %) | 80.5% | 59% | 56% |
| class 9 (86 - 95 %) | 90.5% | 41% | 37% |

The two remaining parameters, the canopy storage capacity ($S_c$) and the mean evaporation rate ( $\overline{E}$ ), were calibrated for each catchment by comparing measured and modelled throughfall and determining the maximum model efficiency values (described in Section 2.7.3). All combinations of a wide range of possible $S_c$ and $\overline{E}$ values were used in the search for optimal values. The initial ranges of values used were 0 to 10 mm for $S_c$ and 0 to 6 mmh$^{-1}$ for $\overline{E}$. These ranges more than cover the maximum values of 1.0 mm for $S_c$ and 0.46 mmh$^{-1}$ for $\overline{E}$ used by other researchers applying Gash-type throughfall models (Gash and Morton, 1978; Gash, 1979; Gash *et al.*, 1980; Pearce and Rowe, 1981; Hutjes *et al.*, 1990; Loustau *et al.*, 1992). The intervals used were 0.02 mm for $S_c$ and 0.001 mmh$^{-1}$ for $\overline{E}$ which provide parameter estimates which are precise enough for the purposes here.

Figures 3.2 and 3.3 show the response surfaces of model efficiency values for South and East catchments, with the bold lines representing a total volume error of zero percent. The Nash coefficients, shown by the grey contour lines, represent how well the throughfall model performed when comparing the observed and simulated throughfall for each time period for different parameter values. The zero percent error line indicates that the total overall volume of throughfall of observed values did not differ from the model simulated values as calculated by

$$\frac{\sum_{t=1}^{n}(TF_{obs} - TF_{sim})}{\sum_{t=1}^{n} TF_{obs}} \times 100 = 0\% \qquad (3.7)$$

where $TF_{obs}$ is the observed throughfall value and $TF_{sim}$ is the simulated throughfall value for each two week period, $t$. The optimum values for $S_c$ and $\overline{E}$ for each catchment were chosen by taking the values with the highest Nash coefficients that fell on the zero percent error line of total overall volume.

The resulting shapes of the response surfaces correspond with past research that concluded that canopy storage models are most sensitive to evaporation and least sensitive to canopy parameters (Gash, 1979; Loustau *et al.*, 1992). Evaporation from the saturated canopy and after rainfall has ceased are the major components of interception loss while evaporation during the wetting phase plays a minor role in interception loss.

The resulting throughfall model parameters are shown in Table 3.2 with the free throughfall and stemflow coefficients determined independently of model calibration, while the canopy storage capacity and mean evaporation rate were determined by the calibration of the canopy storage model.

**Table 3.2** Values of throughfall parameters for South and East catchments.

| parameter | South catchment | East catchment |
|---|---|---|
| free throughfall coefficient $p$ | 0.159 | 0.203 |
| stemflow coefficient $p_t$ | 0.02 | 0.02 |
| canopy storage capacity (mm) $S_c$ | 3.30 | 2.31 |
| mean evaporation rate (mm $h^{-1}$) $\overline{E}$ | 0.228 | 0.335 |

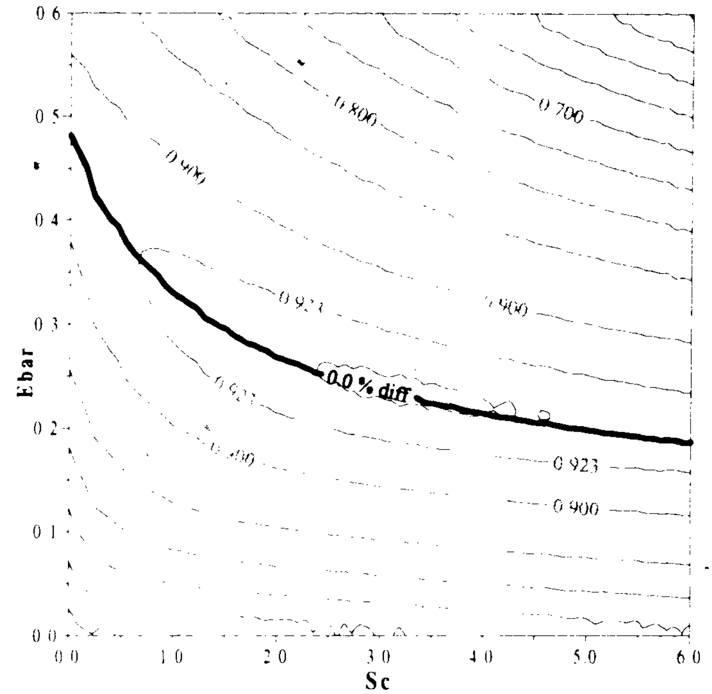



**Figure 3.2** Calibration response surface of the canopy storage model for South catchment. The contours represent values of the Nash coefficient, as defined by Eq. 2.4

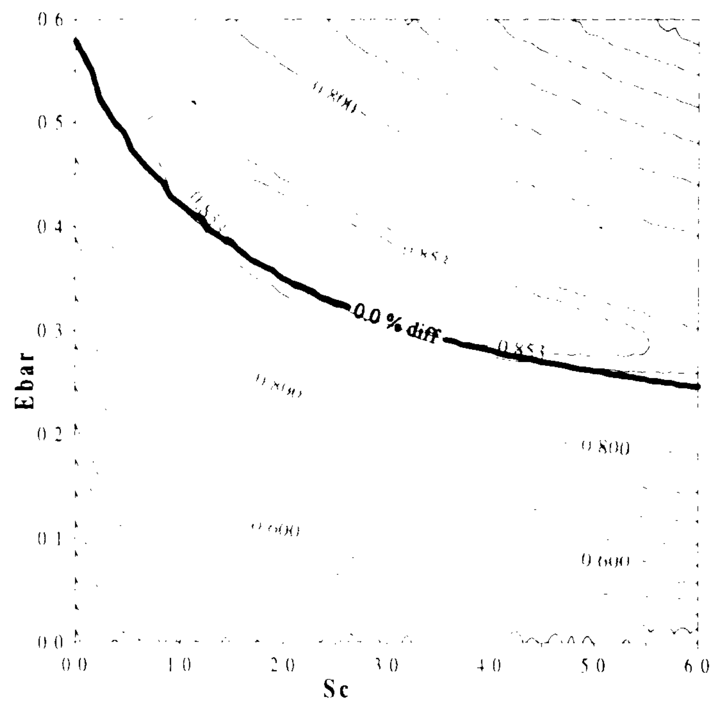



**Figure 3.3** Calibration response surface of the canopy storage model for East catchment. The contours represent values of the Nash coefficient, as defined by Eq. 2.4

## 3.3 Lumped black-box model

### *3.3.1 Structure and governing equations*

The simplest of the three rainfall-runoff models is the lumped black-box model which consists of two linear reservoirs in parallel (Figure 3.4). This model represents the catchment as being composed of one slow and one fast reservoir that simultaneously contribute to the outflow.

This dual reservoir model incorporates three parameters. The parameter $f_1$ determines a fraction of the total amount of precipitation input ($I$) reaching the soil

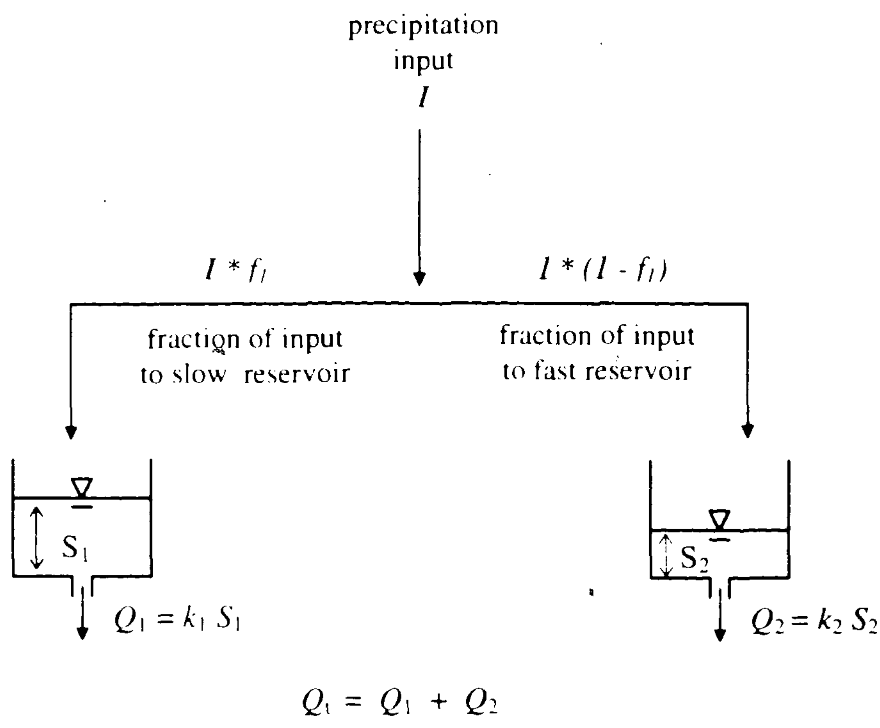



**Figure 3.4.** Schematic diagram of the black-box model, showing its parallel linear reservoir structure.

surface (as calculated by the canopy storage model) which enters the slow reservoir storage ($S_1$). The remaining fraction of the input ($1-f_1$) enters the fast reservoir storage ($S_2$). The remaining two parameters are the recession constants for each of the two reservoirs, $k_1$ and $k_2$ ($s^{-1}$), which control the rate of outflow from each reservoir. The changes in storage of the reservoirs are calculated at six-minute intervals as:

$$\Delta S_1 = (I \cdot f_1 \cdot A) - k_1 \cdot S_1 \tag{3.8}$$

$$\Delta S_2 = (I \cdot (1 - f_1) \cdot A) - k_2 \cdot S_2 \tag{3.9}$$

where $A$ is the area of the catchment ($m^2$). The discharge from each of the reservoirs at six minute intervals is calculated as:

$$Q_1 = k_1 \; S_1 \tag{3.10}$$

$$Q_2 = k_2 \; S_2 \tag{3.11}$$

The six minute discharge values are integrated to equal the total hourly discharge from the catchment:

$$Q_t = \left( \sum_{t=1}^{10} (Q_1 + Q_2) \right) / 10 \tag{3.12}$$

*3.3.2 Initial conditions*

Initial conditions were specified by assuming that, prior to the start of each event, all baseflow originates from the slow reservoir ($S_1$) and the fast reservoir ($S_2$) is empty. The storages are initialized as

$$S_{1\,(0)} = Q_{(0)} / k_1 \tag{3.13}$$

$$S_{2\,(0)} = 0 \tag{3.14}$$

where $S_{1(0)}$ and $S_{2(0)}$ represent the initial storages and $Q_{(0)}$ is the observed discharge at time t=0.

### 3.3.3 *Calibration of parameters*

Two sets of parameters for each catchment were necessary for the runoff models. One set was required for level 1 and 2 testing in which the calibration data set covered the range of events and the second set was required for level 3 and 4 testing in which the calibration set consisted of the small volume events. The method of calibration was the same for each calibration data set and is as follows.

The recession constant for the slow reservoir, $k_1$, was initially estimated as being equal to the slowest of the three recession coefficients that were first derived for the lumped conceptual model. The recession constants for the lumped conceptual model were calibrated, as described in Section 3.4.3, by fitting a function to eighteen recession curves. The function is

$$Q = Q_1 e^{-k_1 t} + Q_2 e^{-k_2 t} + Q_3 e^{-k_3 t} \tag{3.15}$$

where $Q$ is the discharge for timestep t; $Q_1$, $Q_2$, and $Q_3$ are the initial discharges from each of the three reservoirs of the lumped conceptual model, which vary from event to event; and $k_1$, $k_2$, and $k_3$ are the recession constants. The recession parameters were estimated by minimizing a loss function based on the squared differences between observed and predicted discharges as follows:

$$\varepsilon_d^2 = \frac{1}{n_r}\sum_{i=1}^{n_r}\left[\frac{1}{n_r}\sum_{j=1}^{n_i} d_{ij}^2\right] \qquad (3.16)$$

where $d_{ij}$ is the difference between predicted and observed discharges for the $j^{th}$ time in the $i^{th}$ recession segment; $n_i$ is number of observations in the $i^{th}$ segment (Moore, 1997).

Using the initial value of $k_1$, derived from the fitting of the recession curves for the calibration of the lumped conceptual model, the black-box model was run using iterative loops of all possible combinations of the two remaining parameters, $f_1$ and $k_2$. Model efficiency ($E_m$) values were generated for each combination of parameter values. The range of $f_1$ used was 0 to 1 with an interval of 0.002. A range of 0 to 0.5 $h^{-1}$ with an interval of 0.0002 $h^{-1}$ was used to calibrate $k_2$. The initial $k_1$ parameter was adjusted to maximize the best fit yielded by the $f_1$ and $k_2$ parameters. The $f_1$ and $k_2$ parameters were then adjusted to finally obtain the three optimum parameters for the lumped dual parallel reservoir model (Table 3.3). The response surfaces of model efficiency (Nash values) show how the model performed in each catchment using the various $f_1$ and $k_2$ parameters with the initial optimum $k_1$ value (Figures 3.5 and 3.6).

**Table 3.3** Parameters for South and East catchments for the dual reservoir model.

| | South catchment | | East Catchment | |
|---|---|---|---|---|
| parameters | level 1 | level 3 | level 1 | level 3 |
| $f_1$ | 0.010 | 0.018 | 0.160 | 0.260 |
| $k_1$ ($h^{-1}$) * | 0.0076 | 0.0096 | 0.0098 | 0.0082 |
| $k_2$ ($h^{-1}$) * | 0.0704 | 0.0360 | 0.0614 | 0.0378 |

* converted to units of $s^{-1}$ in model program

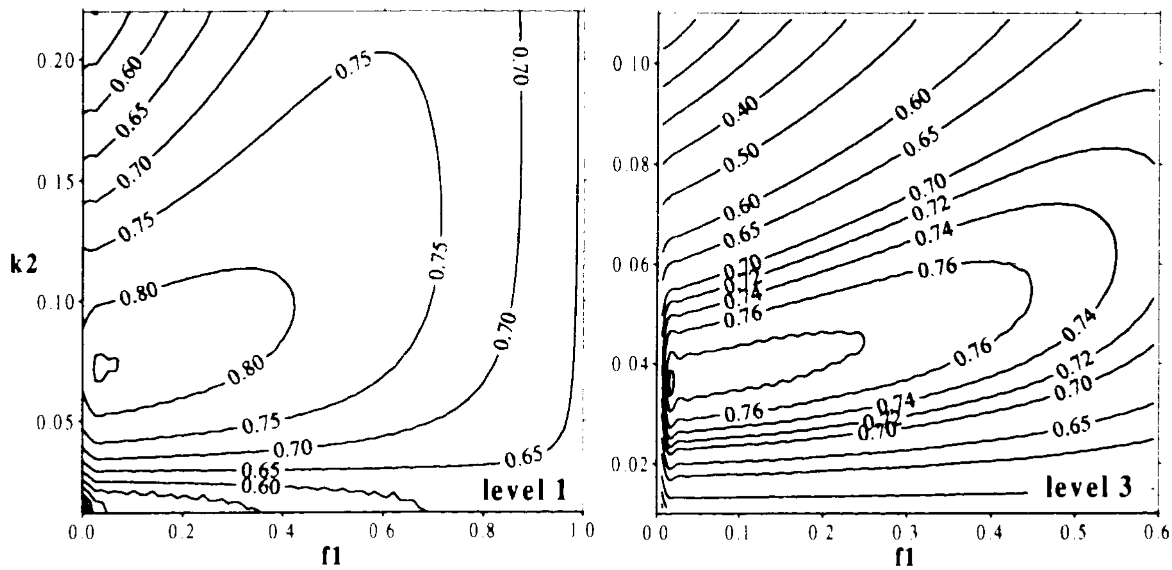



**Figure 3.5** Calibration response surfaces for the black-box lumped model for South catchment. The contours represent values of the Nash coefficient, as defined by Eq. 2.4.

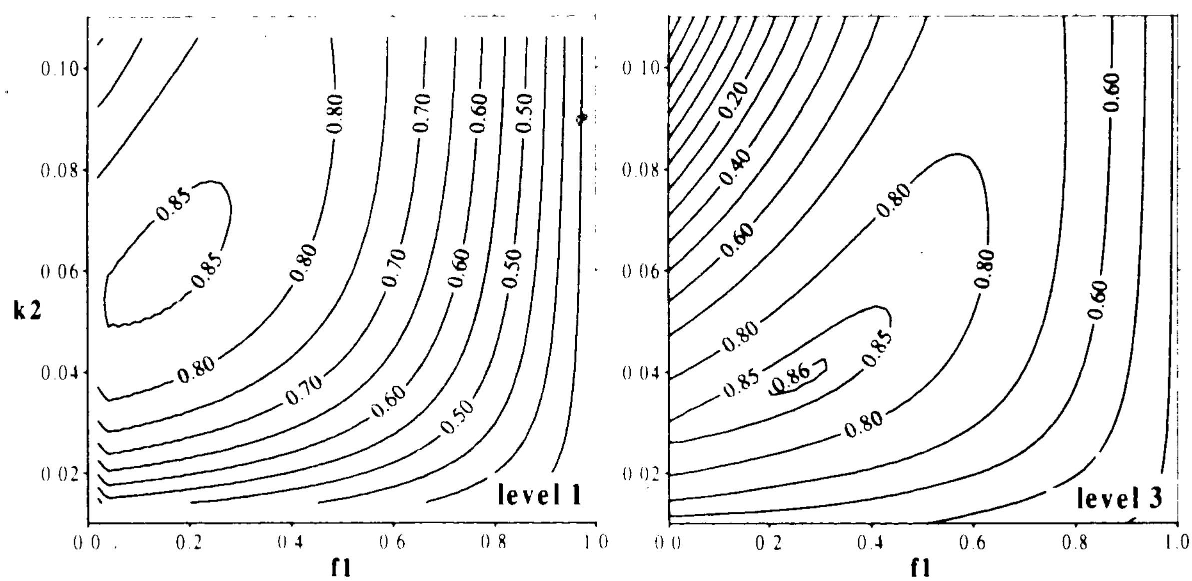



**Figure 3.6** Calibration response surfaces for the black-box lumped model for East catchment. The contours represent values of the Nash coefficient, as defined by Eq. 2.4.

## 3.4 Lumped conceptual model

### *3.4.1 Structure and governing equations*

The lumped conceptual model consists of three linear reservoirs in a serial configuration (Figure 3.7). Unlike the lumped black-box model, the lumped conceptual model is structured to represent the dominant hydrologic pathways in the study catchments. The three reservoirs represent (1) an upslope zone in which surface saturation never occurs, (2) a near-stream zone in which surface saturation does occur, at least transiently during storms, and (3) a surface storage comprising saturation overland flow and channel storage. The storages in the three reservoirs are denoted $S_1$, $S_2$, and $S_3$, respectively. Each reservoir has an associated recession constant $k_i$, where 'i' represents the reservoir number. The upslope reservoir is assumed to have the lowest recession constant and the overland flow and channel reservoir the highest.

Water drains from the upslope reservoir to the near-stream zone at a rate given by

$$Q_1 = k_1 S_1 \tag{3.17}$$

where $Q_1$ is the discharge ($Ls^{-1}$). Water discharges from the second, subsurface reservoir to the overland flow and channel storage at a rate given by

$$Q_2 = k_2 S_2 \tag{3.18}$$

Catchment outflow is assumed to equal the discharge from the third reservoir, computed as

$$Q_3 = k_3 S_3 \tag{3.19}$$

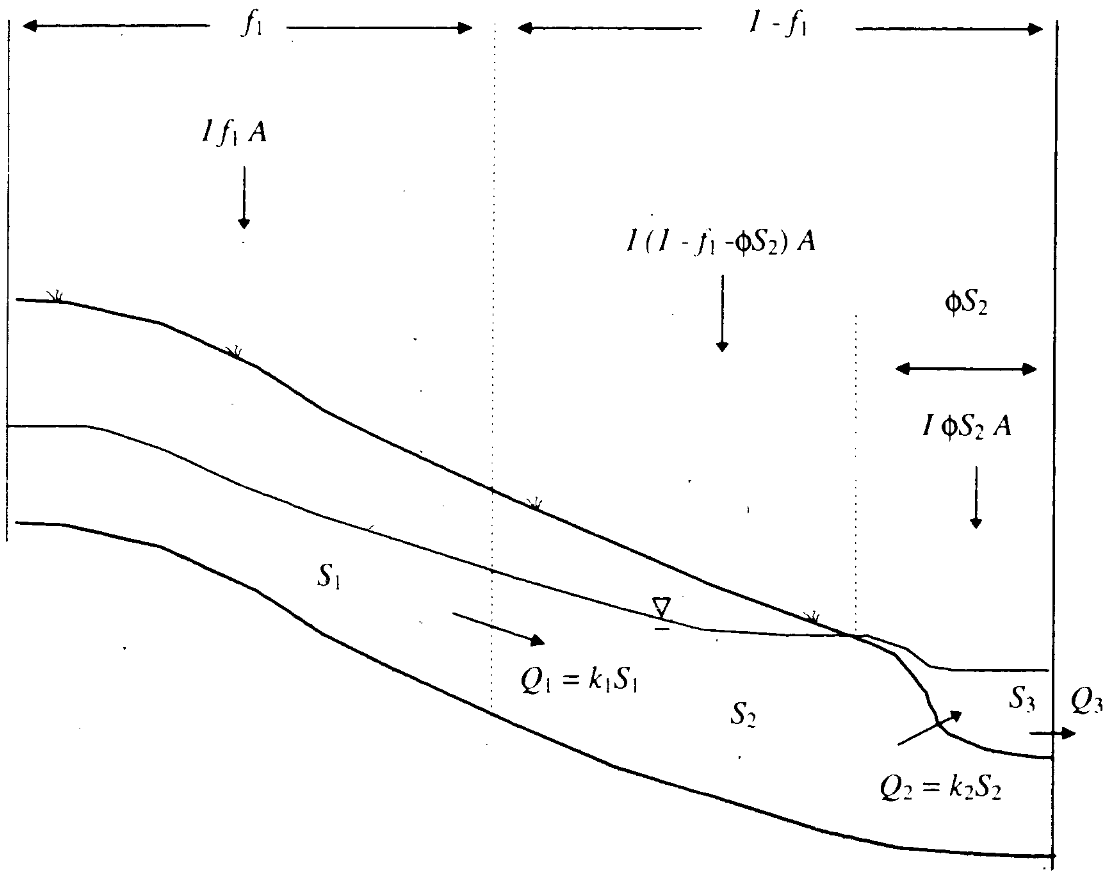



**Figure 3.7.** Schematic diagram of the lumped conceptual model, illustrating its triple serial reservoir structure.

The parameter $f_1$ represents the fraction of the catchment which functions as the upslope reservoir, while the quantity $1 - f_1$ represents the fraction of the catchment in the near-stream zone. The fraction of the catchment which is saturated to the ground surface is assumed to be proportional to the subsurface storage in that zone, and is computed as

$$f_{sat} = \phi S_2 \tag{3.20}$$

where $\phi$ is a parameter determined by calibration.

The model is based on a set of differential equations which represent the water balances of the three reservoirs.

$$\frac{dS_1}{dt} = I \cdot f_1 \cdot A - k_1 S_1 \tag{3.21}$$

$$\frac{dS_2}{dt} = I(1 - f_1 - \phi S_2)A + k_1 S_1 - k_2 S_2 \tag{3.22}$$

$$\frac{dS_3}{dt} = I \cdot \phi S_2 \cdot A + k_2 S_2 - k_3 S_3 \tag{3.23}$$

The differential equations were integrated at 6-minute intervals using a fourth-order Runge-Kutta scheme based on the algorithm described by Press *et al.* (1986). The 10 values of $Q_3$ for each hour were then averaged and used as the model output for the 1-hour time step for comparison with the observed discharge.

### *3.4.2 Initial conditions*

Initial conditions were specified by assuming that each rainfall event followed a sufficiently long baseflow period to assume near-steady-state such that

$$Q_1 = Q_2 = Q_3 \tag{3.24}$$

The three storages were therefore initialized at the beginning of each event as

$$S_{1(0)} = Q_0 / k_1 \tag{3.25}$$

$$S_{2(0)} = Q_0 / k_2 \tag{3.26}$$

$$S_{3(0)} = Q_0 / k_3 \tag{3.27}$$

where $Q_0$ is the initial observed discharge.

*3.4.3 Calibration of parameters*

Recession constants for the three reservoirs, $k_1$, $k_2$, and $k_3$, were determined by fitting a recession equation to eighteen event recession curves using the approach described by Moore (1997). The function is:

$$Q = Q_1\,e^{-k_1 t} + Q_2\,e^{-k_2 t} + Q_3\,e^{-k_3 t} \tag{3.28}$$

where $Q$ is the discharge for time t; $Q_1$, $Q_2$, and $Q_3$ are the initial discharges from each of the three reservoirs, which vary from event to event; and $k_1$, $k_2$, and $k_3$ are the corresponding recession coefficients, which are assumed to be constant for all events. The recession parameters were determined by minimizing a loss function (Equation 3.16) based on the squared differences between observed and predicted discharges described in Section 3.3.

Using the fitted recession values, $k_1$, $k_2$, and $k_3$, the model was run using iterative loops of combinations of the two remaining parameters, $f_1$ and $\phi$. Model efficiency ($E_m$) values were generated for all combinations of the parameters. The range of $f_1$ used was 0 to 1 with an interval of 0.002. A wide range of different magnitudes ($10^{-2}$ to $10^{-20}$) for the $\phi$ parameter were used. All five parameters were then adjusted to optimum values which gave the maximum model efficiency value $E_m$. The optimum parameter values are shown in Table 3.4 and the response surfaces representing the model efficiency value results for each set of calibration events are shown in Figures 3.8 and 3.9.

**Table 3.4** Parameters for South and East catchments for the triple reservoir model.

| parameters | South catchment level 1 | South catchment level 3 | East catchment level 1 | East catchment level 3 |
|---|---|---|---|---|
| $f_1$ | 0.002 | 0.176 | 0.152 | 0.340 |
| $\phi$ | 2.1e-07 | 2.8e-07 | 6.3e-08 | 8.0e-08 |
| $k_1$ ($h^{-1}$) * | 0.0119 | 0.0121 | 0.0074 | 0.0071 |
| $k_2$ ($h^{-1}$) * | 0.0578 | 0.0463 | 0.0541 | 0.0513 |
| $k_3$ ($h^{-1}$) * | 0.1231 | 0.0757 | 0.1256 | 0.0932 |

* converted to units of $s^{-1}$ in model program

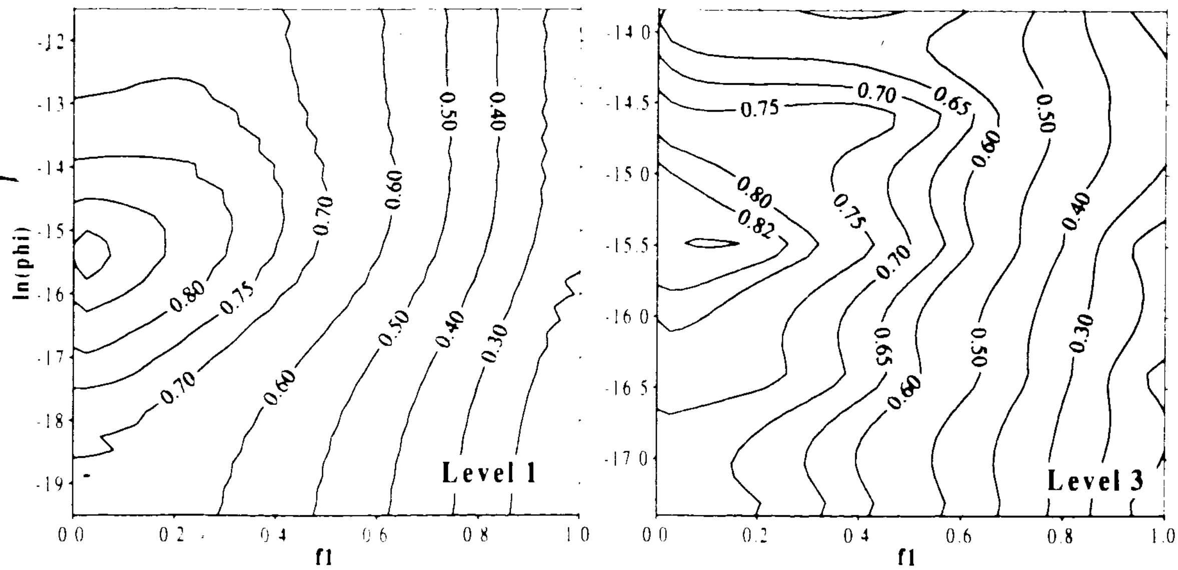



**Figure 3.8** Calibration response surfaces for the conceptual lumped model for South catchment. The contours represent values of the Nash coefficient, as defined by Eq. 2.4.

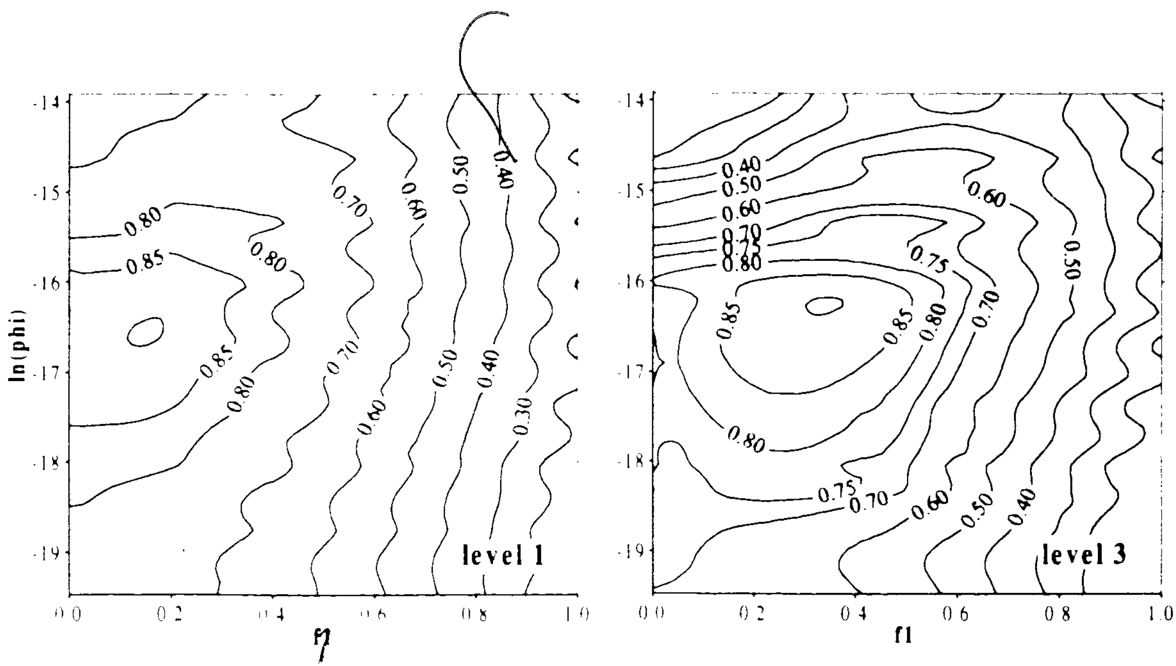



**Figure 3.9** Calibration response surfaces for the conceptual lumped model for East catchment. The contours represent values of the Nash coefficient, as defined by Eq. 2.4.

## 3.5 Quasi-distributed model

TOPMODEL (Beven and Kirkby, 1979) was selected to represent the quasi-distributed model type. TOPMODEL is not a definitive model structure but rather is a set of concepts that conceptually simulates distributed hydrologic response by using an index that represents the topography of the catchment. TOPMODEL uses the topographic index, $\ln(a/\tan\beta)$, to explicitly link topographic form to subsurface water flow and the production of surface runoff, where $a$ is the area drained per unit contour and $\beta$ is the local slope angle. The model is appropriate to use in this study since the shallow soil of the UBC Research Forest allows for the topography to significantly control the soil moisture distribution (Moore and Thompson, 1996).

The TOPMODEL concepts are gaining acceptance among researchers and have been applied to a variety of catchments and hydrological modeling problems (e.g. Beven *et al.*, 1984; Hornberger *et al.*, 1985; Quinn *et al.*, 1991; Durand *et al.*, 1992; Ambroise *et al.*, 1996). Figure 3.10 demonstrates the gain in popularity since the late 1980's. TOPMODEL has attracted attention because it is physically realistic and explicitly accounts for catchment topography, its mathematics are relatively straightforward and well documented, and it requires a relatively small number of parameters.

TOPMODEL is one of the few 'conceptual' models that accounts explicitly for the saturation excess overland flow mechanism and integrates the variable contributing area concept, both of which are essential to model the studied catchments accurately (Jordan, 1994; Iorgulescu and Jordan, 1994).

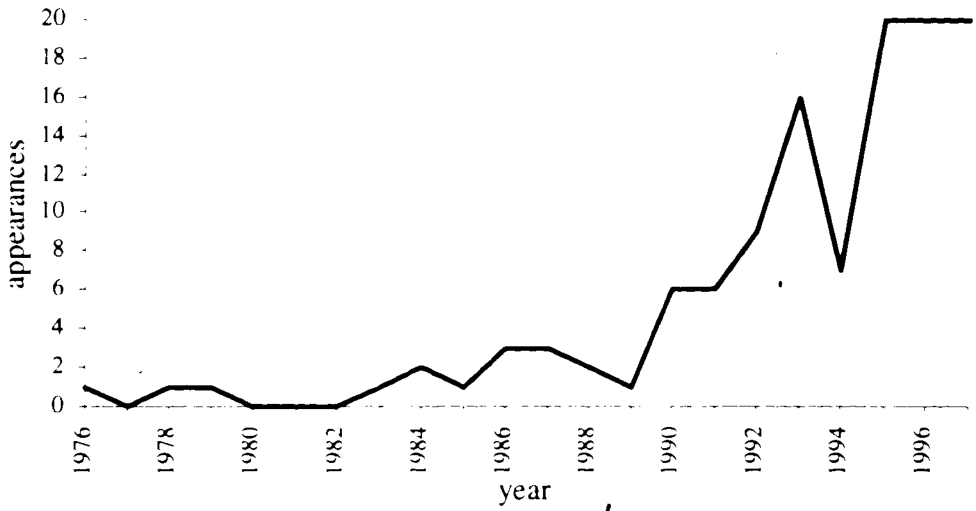



**Figure 3.10** Time series of frequencies of appearance of TOPMODEL in the published literature.

### *3.5.1 Structure and governing equations*

TOPMODEL is a quasi-distributed model which means that the hydrologic response is not accounted for at each specific point within the catchment, but rather is determined for the whole of the catchment by delineating different proportions of the catchment area which have the same topographic index (Figure 3.11). The use of the topographic index $\ln(a/\tan\beta)$, which can be considered an index of hydrologic similarity, is important to the simplicity of the TOPMODEL approach because it is not necessary to carry out calculations for every point, since every point with the same index value will have the same predicted response given the precipitation input (Ambroise *et al.*, 1996).

TOPMODEL assumes the soil water storage in a catchment can be represented by one linear and one non-linear storage for each $\ln(a/\tan\beta)$ increment. For each increment,

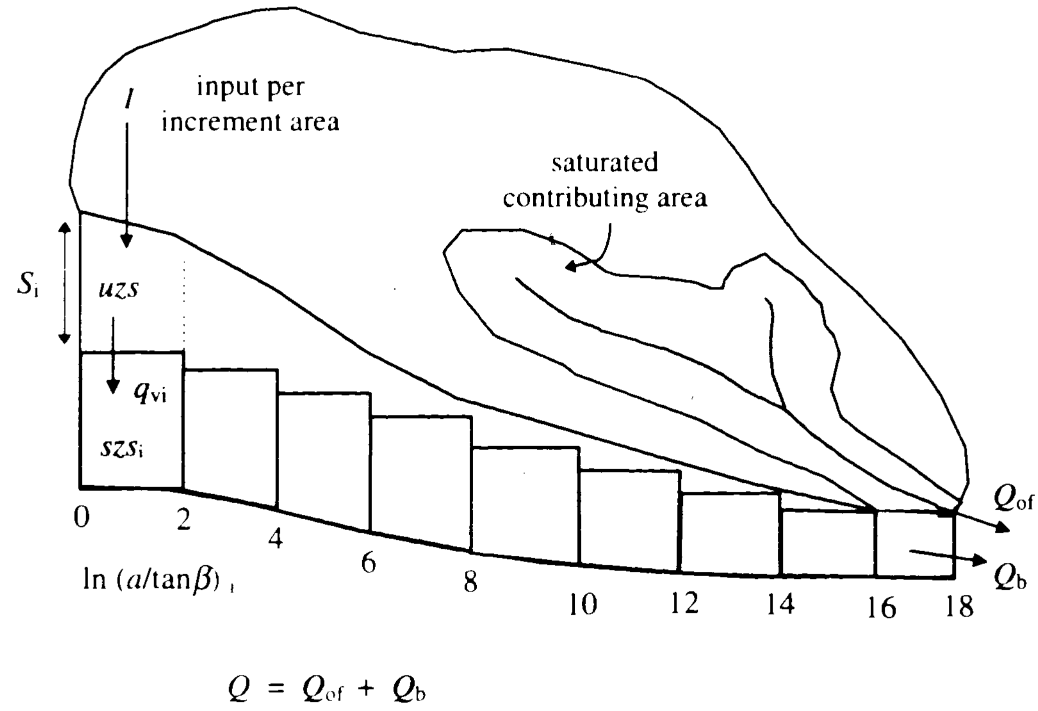



**Figure 3.11** Schematic diagram of quasi-distributed model TOPMODEL, modified from Hornberger *et al.* (1985).

water input first enters the unsaturated zone store ($uzs_i$). This water then flows vertically ($q_{vi}$) at a constant rate from the unsaturated zone store into the saturated zone store ($szs_i$):

$$q_{vi} = uzs_i \,/\, (td\ S_i) \qquad (3.29)$$

The rate of vertical drainage for each ln($a$/tan$\beta$) increment is a function of the unsaturated storage by the parameter $td$, the unsaturated zone time delay, and the soil moisture deficit, $S_i$, , which is equivalent to the quantity of water required to fill the unsaturated store.

The saturated zone acts as a nonlinear reservoir, with the water flowing laterally from the saturated zone as baseflow discharge, $Q_b$, determined by

$$Q_b = T_0 \tan\beta \exp(-\bar{S} / m) \qquad (3.30)$$

where $T_0$ is the lateral transmissivity, $\bar{S}$ is the mean catchment deficit, and $m$ is a recession parameter expressing the exponential decay rate of the saturated transmissivity with depth.

Saturation overland flow, $Q_{of}$, is generated when the saturation deficit for a certain increment becomes zero. The grid cells in the catchment which have those particular $\ln(a/\tan\beta)$ index values become saturated to the surface and make up the saturated contributing area. Areas of high values of $\ln(a/\tan\beta)$, i.e. areas of convergence or low slope angle, will become saturated first and as the catchment becomes wetter, the saturated contributing area will increase.

For each hourly time step, contributions of $Q_b$ and $Q_{of}$ from all the topographic increments are summed to give a total discharge for the catchment.

$$Q_t = Q_b + Q_{of} \qquad (3.31)$$

Two assumptions are made: (1) that there is an exponential relation between storage and discharge, and (2) that the direction of the local hydraulic gradient is parallel to the local ground slope (i.e. the water table is parallel to the surface). Since this research is operational, the validity of these assumptions of internal processes is not evaluated.

### 3.5.2 *Initial conditions*

Initial conditions were specified by assuming that each rainfall event followed a sufficiently long baseflow period such that there is no storage in the unsaturated zone.

$$uzs_{i(0)} = 0 \tag{3.32}$$

The saturated zone outflow parameter $szq$ is first initialized as:

$$szq = \exp\left(\ln(T_0) - \lambda\right) \tag{3.33}$$

where $\lambda$ is the areal integral of the $\ln(a/\tan\beta)$ index. The mean catchment deficit $\overline{S}$ is then initialized as a function of the $szq$:

$$\overline{S}_{(0)} = -m \;\; \ln\left(Q_0 / szq\right) \tag{3.34}$$

where $Q_0$ is the initial observed discharge. The mean catchment deficit dictates the local storage deficits for each topographic index increment which controls the total discharge from the saturated zone and the flow from the unsaturated zone once water input is received.

### 3.5.3 *Calibration of parameters*

Three parameters required calibration: $m$, $td$, and $T_0$. The optimum values of the parameters (Table 3.5) were determined by using iterative loops in the computer code which evaluated the best fit, by calculating the highest model efficiency value, from each possible combination of the three parameter values. The range used for $m$ was 0 to 50 mm with an interval of 0.1. The range used for $td$ was 0 to 1.0 $h^{-1}$ with an interval of 0.001. The range used for $T_0$ covered the magnitudes of $10^2$ to $10^{10}$ $mm^2h^{-1}$. The values chosen amply covered the ranges used by other researchers (Beven *et al.*, 1984; Durand *et al.*, 1992; Iorgulescu and Jordan, 1994; and others). Figures 3.12 and 3.13 show the response surfaces generated that have the highest model efficiency (Nash) values for the combination of the three parameters *td*, *m* and $T_0$ for each calibration set.

**Table 3.5** Parameters for South and East catchments for the distributed model.

| | South catchment | | East Catchment | |
|---|---|---|---|---|
| parameters | level 1 | level 3 | level 1 | level 3 |
| $m$ (mm) | 6.0 | 6.9 | 13.8 | 12.5 |
| $td$ ($h^{-1}$) | 0.130 | 0.112 | 0.036 | 0.031 |
| $T_0$ ($mm^2 h^{-1}$) | 1.1e+06 | 8.1e+04 | 4.9e+05 | 3.6e+05 |

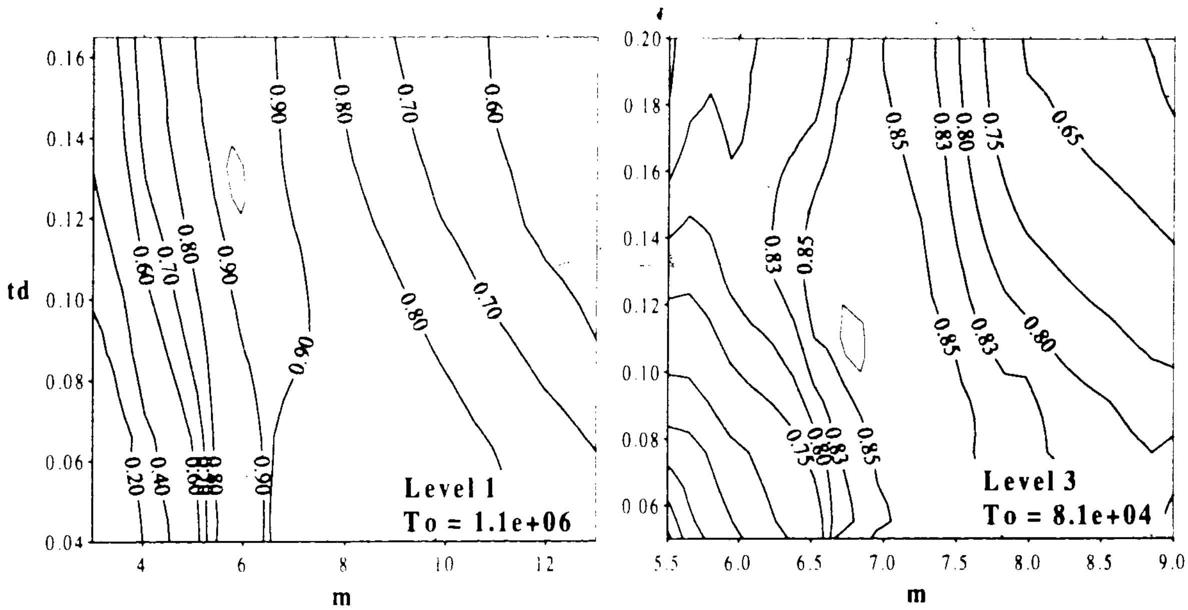



**Figure 3.12** Calibration response surfaces for the quasi-distributed model for South catchment having the highest Nash values for all combinations of the three parameters *m*, *td*, and *To*.

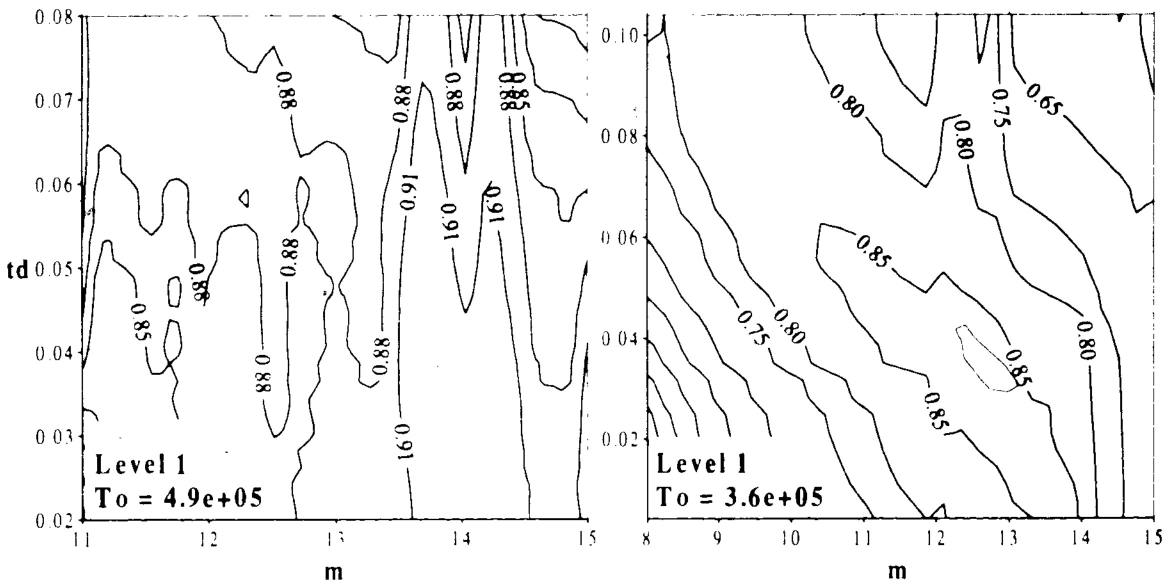



**Figure 3.13** Calibration response surfaces for the quasi-distributed model for East catchment having the highest Nash values for all combinations of the three parameters *m*, *td*, and *To*.

In addition to the calibrated parameters, the topographic index, ln($a$/tan$\beta$), and $\lambda$, the areal integral of the ln($a$/tan$\beta$) index, were generated from the digital elevation data. Figure 3.14 shows cumulated topographic index curves in which the percent area of catchment that would be saturated is calculated for each ln($a$/tan$\beta$) increment of 0.5. Figures 3.15 and 3.16 show the spatial distribution of the topographic index ln($a$/tan$\beta$) for South and East catchments. The higher the index value in a 10 m grid cell within a catchment, the wetter the cell and the more frequently that cell will be saturated during a storm event.

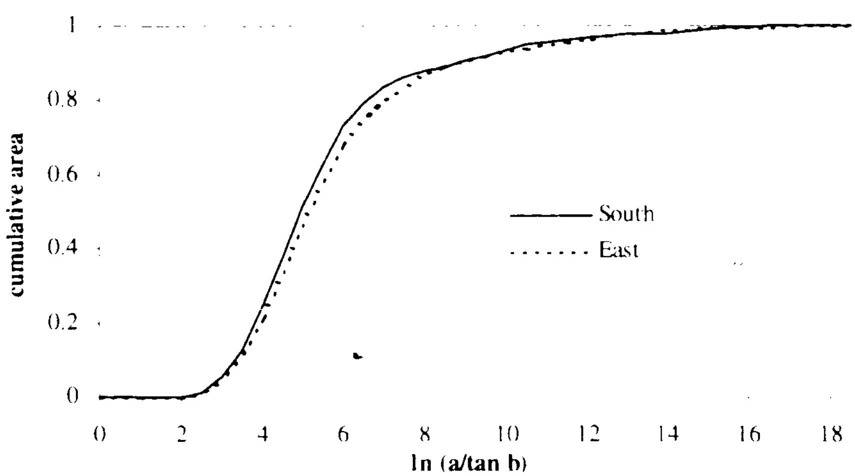



**Figure 3.14** Cumulative distribution of the ln($a$/tan$\beta$) index for South and East catchments.

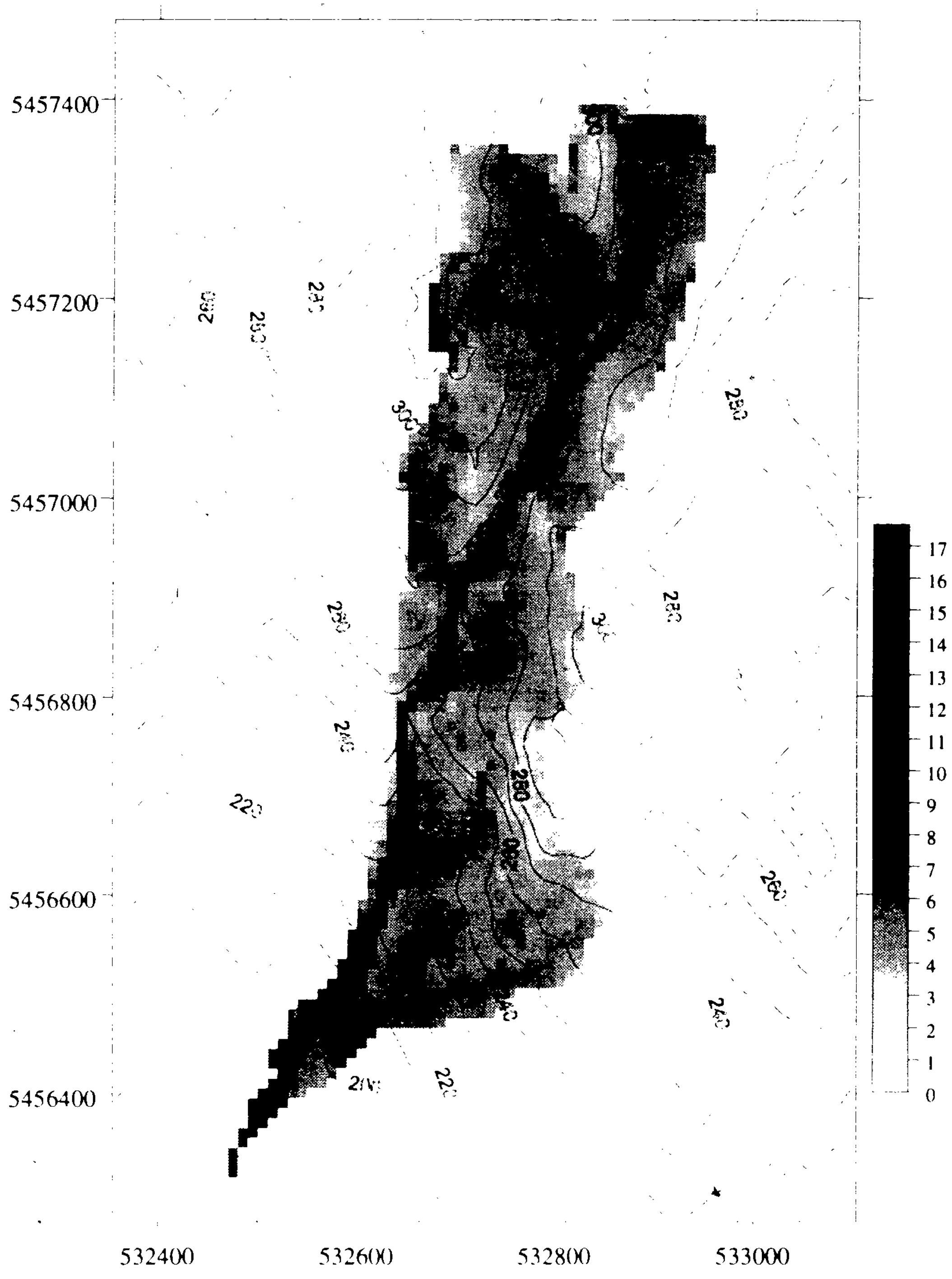



**Figure 3.15** Spatial distribution of the topographic index $\ln(a/\tan\beta)$ in South catchment using a 10 m grid size. Contour interval is 10 m with UTM coordinates (m) for reference.

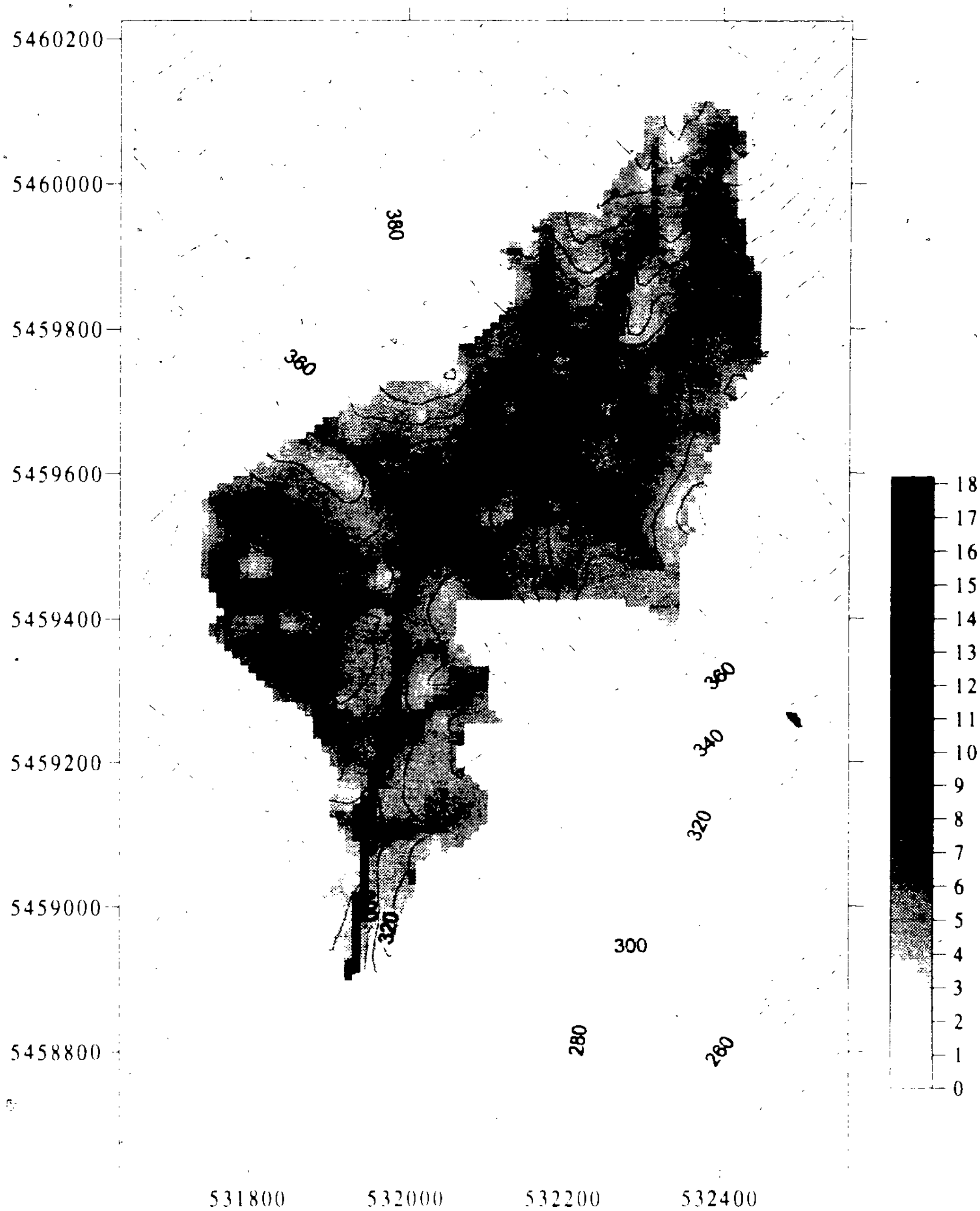



**Figure 3.16** Spatial distribution of the topographic index $\ln(a/\tan\beta)$ in East catchment using a 10 m grid size. Contour interval is 10 m with UTM coordinates (m) for reference.

# CHAPTER 4

# RESULTS

The results are presented in order from level one testing (split-sample) to level four testing (proxy-basin differential split-sample). Although performance statistics are presented for both calibration and validation runs, only a sample of hydrographs for validation runs are presented. One small volume and one large volume event for each catchment are represented in the hydrographs for graphical comparison. A complete set of scatterplots is found in Appendix 3.

The final section of this chapter presents a comparison of model performance. An important contribution is the introduction of the Jackknife technique, in combination with ANOVA, to provide a statistical assessment of the significance of differences in model performance.

## 4.1 Level 1 - Split-sample test

### *4.1.1 Level 1 - Split-sample test calibration results*

The split-sample level of testing tests the models' abilities to simulate events that are similar to those used for calibration within the same catchment. Table 4.1 shows the Nash coefficients for the calibration data set. The most important criterion for the model comparison is the $E_m(Q_r)$ value since it represents overall model performance and was the value used for the calibration of the model parameters. $E_m(Q_r)$ is the model efficiency value ($E_m$) of the average hourly discharge rate ($Q_r$) in which the simulated values are compared to the observed values at each hourly timestep for the entire rainfall-runoff event. The model efficiency values for event volume ($Q_{vol}$), peak discharge ($Q_{pk}$), and time-to-peak values ($t_{pk}$) are also shown. The percent deviations (WMO, 1975) for volume, peak discharge, and time-to-peak are also presented in the table as calculated by Equation 2.6.

When calibrated, all three models fit the observed values similarly in terms of the overall Nash value, $E_m(Q_r)$, although the quasi-distributed model yielded a slightly better fit for both catchments. The quasi-distributed model performed better than the lumped models in terms of simulating peak discharges ($Q_{pk}$), but worse in terms of simulating event volume ($Q_{vol}$). The quasi-distributed model best simulated the timing of the peak flow whereas the lumped models responded too quickly to the throughfall input.

Table 4.2 shows the test statistics for calibration of each of the models and allows for comparison to observed values for each of the indices. The standard deviation values reveal that all three models reproduced the variability in volume and time-to-peak but the lumped models did not simulate the variability in the peak discharges nearly as well as the quasi-distributed model.

**Table 4.1** Calibration goodness-of-fit indices for level 1 (split-sample) testing showing all model efficiency Em values and percent deviation values.

| | | $E_m$ | | | | % deviations | | |
|---|---|---|---|---|---|---|---|---|
| catchment | model | $Q_r$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| south | dual reservoir | 0.843 | 0.951 | 0.681 | 0.765 | +2.5 | -14.4 | -12.7 |
| | triple reservoir | 0.894 | 0.957 | 0.776 | 0.822 | +2.3 | -14.5 | -13.8 |
| | quasi-distributed | 0.931 | 0.899 | 0.934 | 0.975 | -16.6 | -6.6 | +0.1 |
| east | dual reservoir | 0.859 | 0.960 | 0.720 | 0.942 | +2.7 | -11.4 | -8.1 |
| | triple reservoir | 0.898 | 0.945 | 0.795 | 0.979 | +8.0 | -11.4 | -4.3 |
| | quasi-distributed | 0.930 | 0.940 | 0.872 | 0.905 | -11.7 | -10.4 | +6.0 |

**Table 4.2** Calibration test statistics for level 1 (split-sample) testing comparing the means and standard deviation of the means of the three models to the observed values. (n=26)

a. South catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 11.3 | 111.4 | 27.6 | 9.2 | 83.6 | 13.0 |
| dual reservoir | 11.6 | 95.4 | 24.1 | 7.7 | 42.6 | 13.9 |
| triple reservoir | 11.6 | 95.3 | 23.8 | 7.8 | 50.9 | 13.0 |
| quasi-distributed | 9.5 | 104.0 | 27.6 | 7.2 | 78.9 | 12.3 |

b. East catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 24.6 | 205.5 | 26.3 | 17.0 | 151.1 | 14.8 |
| dual reservoir | 25.2 | 182.1 | 24.2 | 15.7 | 82.0 | 13.9 |
| triple reservoir | 26.5 | 182.0 | 25.2 | 16.2 | 96.5 | 13.6 |
| quasi-distributed | 21.7 | 184.2 | 27.8 | 16.3 | 155.8 | 13.2 |

*4.1.2 Level 1 - Split-sample test validation results*

Table 4.3 shows the validation indices for all models at level 1. Comparable to the calibration runs, all three models performed similarly in terms of $E_m(Q_r)$ when validated, with the quasi-distributed model simulating the overall hydrograph better than the two lumped models in both catchments. Although the quasi-distributed model and the lumped models both underestimated the peak discharges, the quasi-distributed model was better at predicting the amount of peak flow. The lumped models were better at predicting overall event volumes, however. While the lumped models slightly overestimated event volumes, the quasi-distributed model considerably underestimated event volumes. The quasi-distributed model did not respond as quickly to the initial precipitation input at the beginning of an event as did the lumped models, thus overestimating the time-to-peak.

Table 4.4 shows the test statistics for validation of each of the models. Again, like the calibration run, the standard deviation values reveal that all three models simulated the variability in volume and time-to-peak but the lumped models did not reproduce the variability in the peak discharges as well as the quasi-distributed model.

The hydrographs (Figures 4.1 to 4.4) allow for a graphical comparison of simulated to observed hourly discharge rates for one small event and one large event (from the validation dataset) within each catchment. The quasi-distributed model provided a better fit of the entire hydrograph than the lumped models, especially for larger events. The quasi-distributed model simulated the peaks better but underestimated the rising and recession limbs, resulting in an underestimate of volume. The lumped models generally overestimated the rising limb, underestimated the peaks, and slightly overestimated the recession limb, resulting in overall event volumes being similar to observed but not simulating the observed peaks as well as the quasi-distributed model.

### *4.1.3 Level 1 - Split-sample test comparison summary*

Overall, the quasi-distributed model slightly outperformed the lumped models using the split-sample test, especially when simulating peak discharge and the variability of peak flows. The lumped models did slightly better when estimating overall event volumes.

**Table 4.3** Validation goodness-of-fit indices for level 1 (split-sample) testing showing all model efficiency Em values and percent deviation values.

| | | $E_m$ | | | | % deviations | | |
|---|---|---|---|---|---|---|---|---|
| catchment | model | $Q_r$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| south | dual reservoir | 0.797 | 0.914 | 0.589 | 0.833 | +3.7 | -15.9 | -14.3 |
| | triple reservoir | 0.861 | 0.916 | 0.708 | 0.864 | +3.6 | -15.9 | -12.5 |
| | quasi-distributed | 0.890 | 0.766 | 0.906 | 0.954 | -16.5 | -6.4 | +2.3 |
| east | dual reservoir | 0.811 | 0.916 | 0.559 | 0.865 | +2.4 | -13.7 | -9.9 |
| | triple reservoir | 0.869 | 0.888 | 0.683 | 0.898 | +7.9 | -13.6 | -3.9 |
| | quasi-distributed | 0.890 | 0.833 | 0.910 | 0.823 | -13.1 | -10.3 | +9.7 |

**Table 4.4** Validation test statistics for level 1 (split-sample) testing comparing the means and standard deviation of the means of the three models to the observed values. (n=26)

a. South catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 10.1 | 112.6 | 23.9 | 4.8 | 84.7 | 11.0 |
| dual reservoir | 10.5 | 94.7 | 20.5 | 4.1 | 35.9 | 11.5 |
| triple reservoir | 10.4 | 94.7 | 20.9 | 4.1 | 45.7 | 11.4 |
| quasi-distributed | 8.4 | 105.3 | 24.5 | 3.7 | 79.7 | 11.2 |

b. East catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 22.5 | 208.4 | 23.0 | 9.4 | 165.4 | 12.4 |
| dual reservoir | 22.7 | 179.8 | 20.7 | 8.4 | 66.4 | 11.5 |
| triple reservoir | 23.9 | 180.1 | 22.1 | 8.9 | 84.8 | 13.5 |
| quasi-distributed | 19.3 | 186.9 | 25.2 | 8.1 | 153.6 | 12.9 |

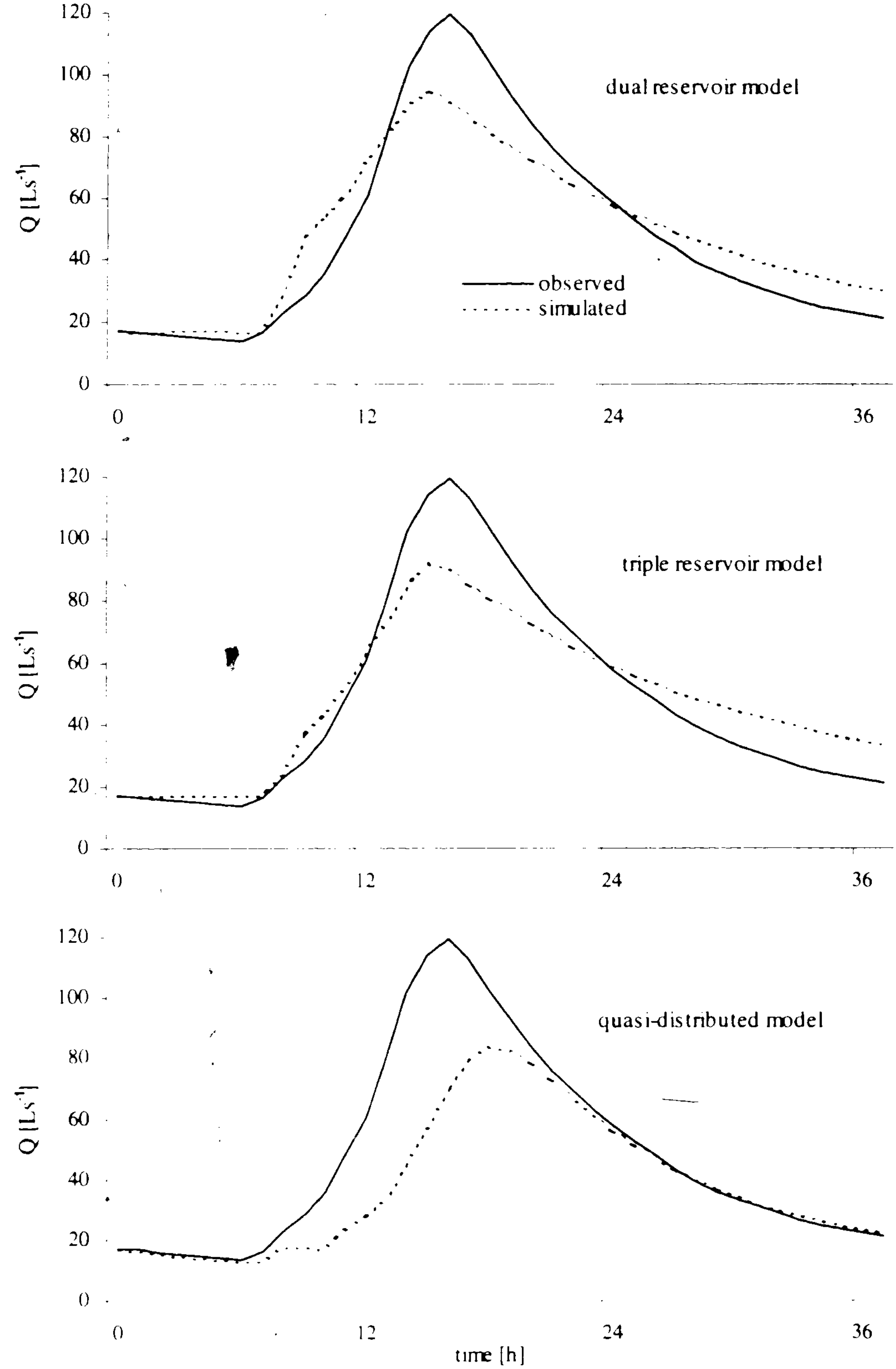



**Figure 4.1** Hydrographs of a small event (event 2), South catchment, level 1.

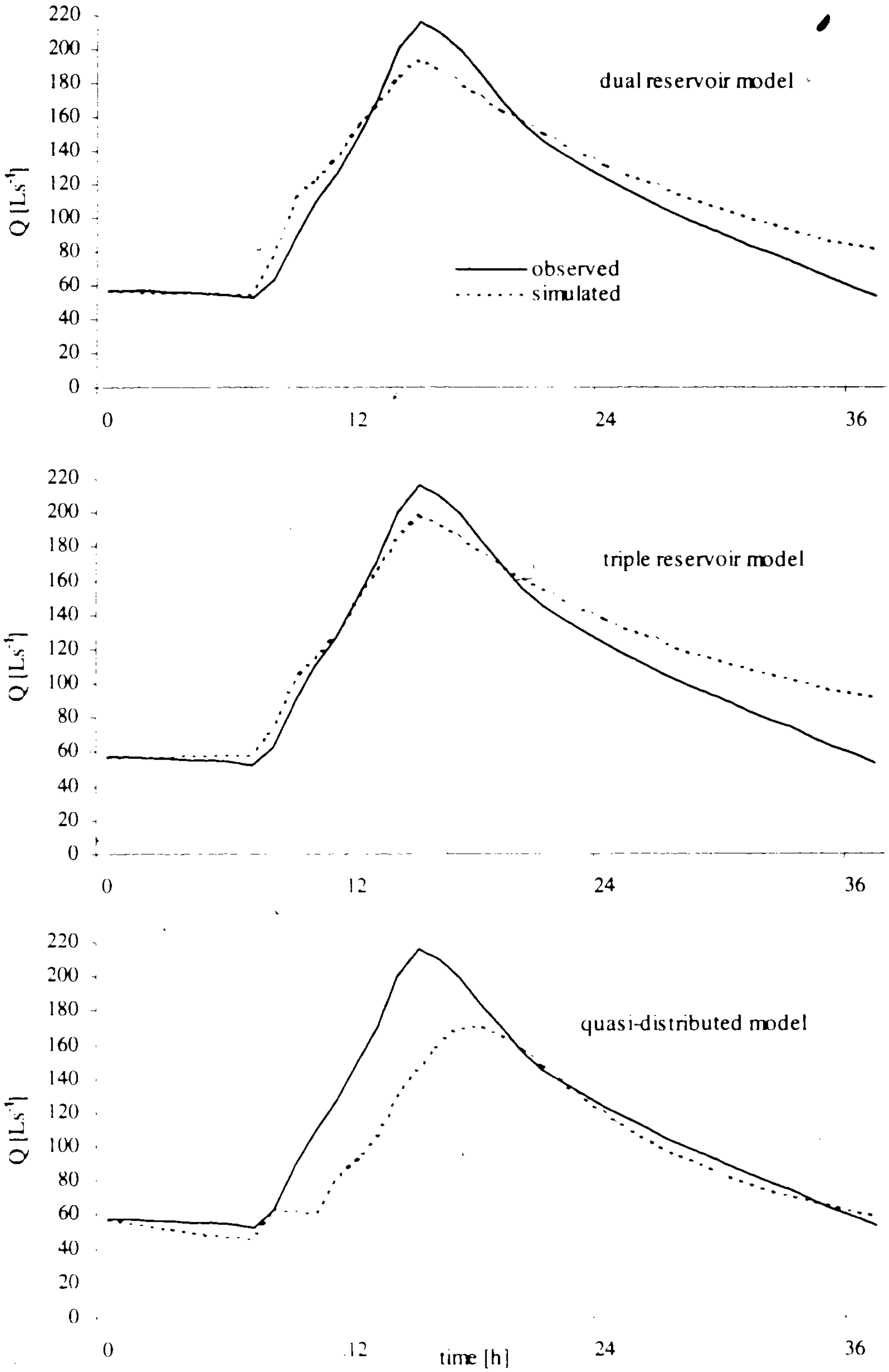



**Figure 4.2** Hydrographs of a small event (event 2), East catchment, level 1.

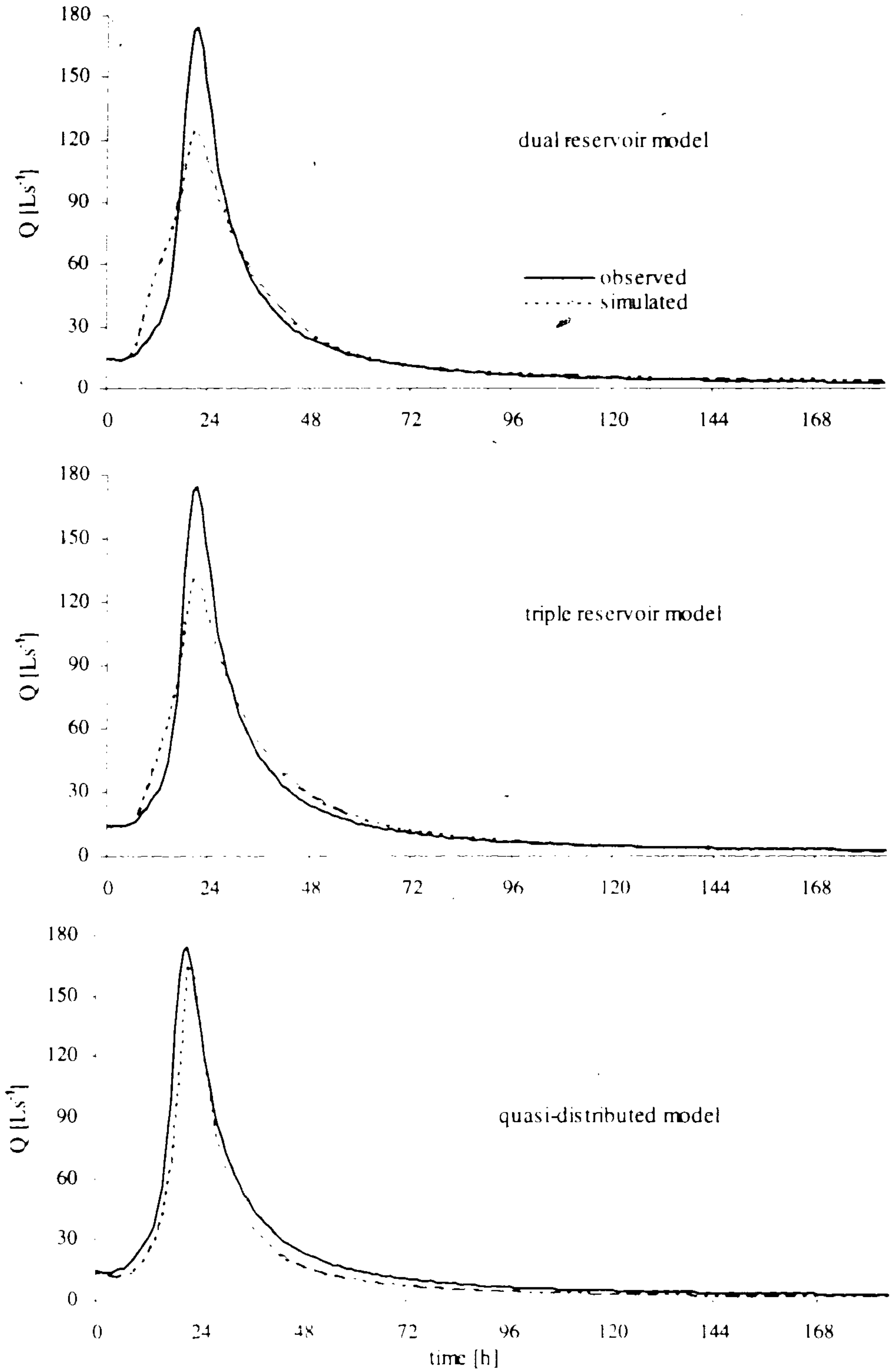



**Figure 4.3** Hydrographs of a large event (event 61), South catchment, level 1.

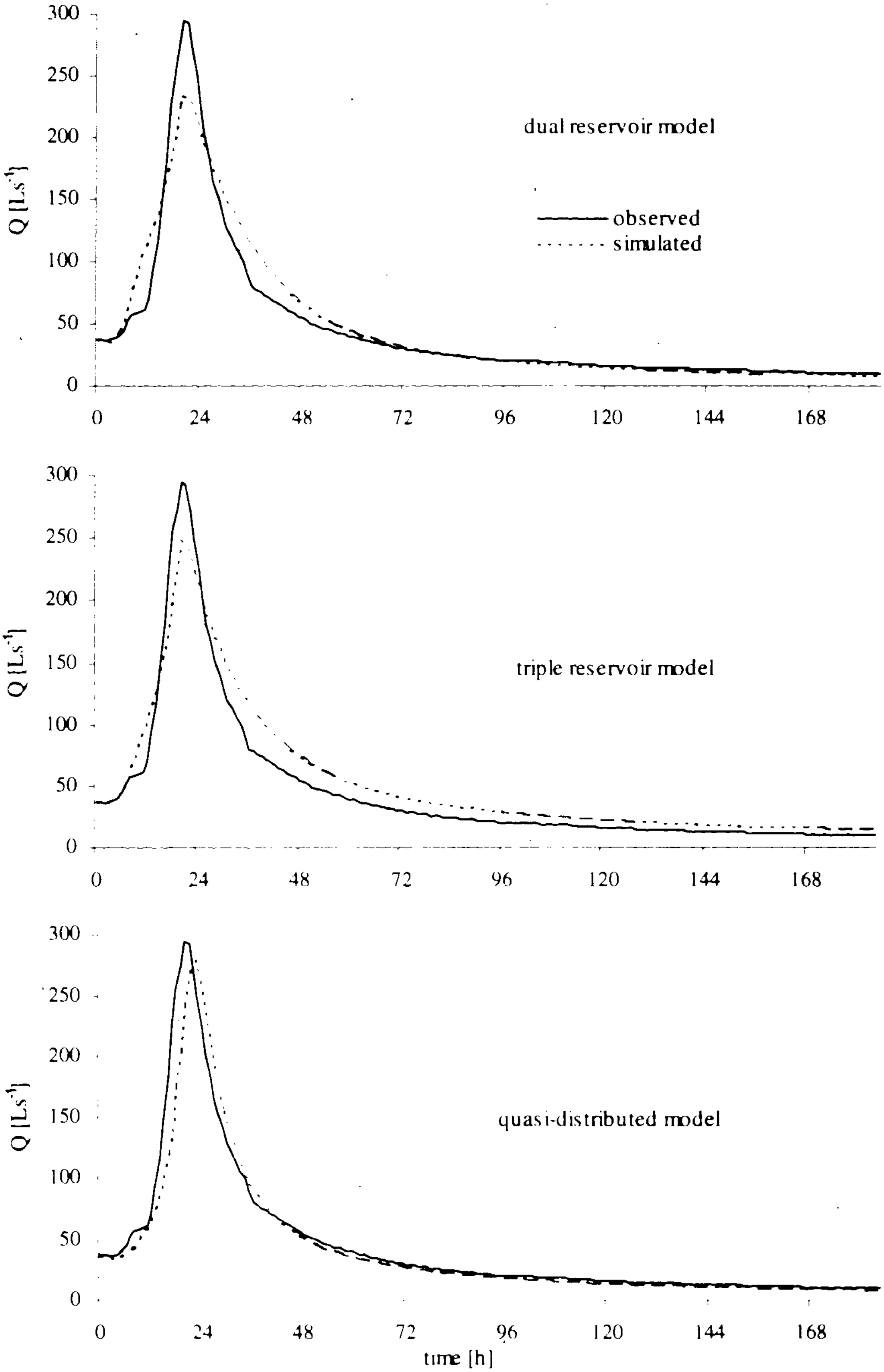



**Figure 4.4** Hydrographs of a large event (event 61), East catchment, level 1.

## 4.2 Level 2 - Proxy-basin test

### *4.2.1 Level 2 - Proxy-basin test validation results*

The second level of testing, the proxy-basin test, tests a model's ability to simulate runoff in a catchment for which it was not calibrated. The calibration statistics are the same as for level 1, and were presented in section 4.1.1. The models were geographically transposed and validated on the other catchment for level 2 testing.

As shown by the overall model efficiency values ($E_m(Q_r)$), in Table 4.5, the quasi-distributed model performed best when calibrated on South and validated on East catchment, but performed worse than both lumped models when transposed from East to South, even though it had the best calibration fits for both catchments. The lumped models performed comparably in both catchments. The quasi-distributed model performed poorer than both lumped models when simulating volume, particularly when transposed from East to South catchment. All three models did poorly in predicting peak discharge, especially going from East to South catchment. The models underestimated peak flows when validated in South and overestimated peak flows in East catchment.

Similar to performance values of the split-sample test, the lumped models were unable to reproduce the variability in peak discharges (Table 4.6), but were able to duplicate the variability in event volume and time-to-peak. The quasi-distributed model simulated the variability of peaks better than the lumped models when transposed to another catchment but overestimated variability in East catchment and underestimated variability in South catchment.

**Table 4.5** Validation goodness-of-fit indices for level 2 (proxy-basin) testing showing all model efficiency Em values and percent deviation values.

| | | $E_m$ | | | | % deviations | | |
|---|---|---|---|---|---|---|---|---|
| catchment* | model | $Q_r$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| East (c) | dual reservoir | 0.759 | 0.882 | 0.374 | 0.840 | -5.4 | -31.0 | -13.2 |
| South (v) | triple reservoir | 0.707 | 0.857 | 0.211 | 0.910 | -2.7 | -42.8 | +3.4 |
| | quasi-distributed | 0.652 | 0.588 | 0.403 | 0.766 | -24.7 | -45.4 | +13.6 |
| South (c) | dual reservoir | 0.777 | 0.828 | 0.673 | 0.853 | +12.3 | +4.8 | -11.4 |
| East (v) | triple reservoir | 0.734 | 0.793 | 0.712 | 0.847 | +14.1 | +26.2 | -11.9 |
| | quasi-distributed | 0.800 | 0.773 | 0.680 | 0.891 | -9.9 | +28.8 | -1.8 |

* (c) calibration catchment, (v) validation catchment

**Table 4.6** Validation run statistics for level 2 (proxy-basin) testing comparing the means and standard deviation of the means of the three models to the observed values. (n=26)

a. Calibrated on East catchment, validated on South catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| data | [$10^6$ L] | [$Ls^{-1}$] | [h] | [$10^6$ L] | [$Ls^{-1}$] | [h] |
| observed | 10.1 | 112.6 | 23.9 | 4.8 | 84.7 | 11.0 |
| dual reservoir | 9.5 | 77.7 | 20.8 | 3.7 | 29.3 | 11.5 |
| triple reservoir | 9.8 | 64.3 | 24.7 | 3.9 | 28.7 | 12.7 |
| quasi-distributed | 7.6 | 61.4 | 27.2 | 3.7 | 50.0 | 13.6 |

b. Calibrated on South catchment, validated on East catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| data | [$10^6$ L] | [$Ls^{-1}$] | [h] | [$10^6$ L] | [$Ls^{-1}$] | [h] |
| observed | 22.5 | 208.4 | 23.0 | 9.4 | 165.4 | 12.4 |
| dual reservoir | 24.9 | 218.3 | 20.3 | 9.2 | 81.4 | 11.6 |
| triple reservoir | 25.3 | 263.0 | 20.2 | 9.3 | 114.9 | 11.6 |
| quasi-distributed | 20.0 | 268.4 | 22.5 | 8.4 | 180.5 | 11.1 |

Figures 4.5 to 4.8 illustrate the contrasting simulations of the models when geographically transposed between catchments. The hydrographs of all three models show a flatter, less responsive simulation curve resulting in the underestimation of volumes and peak flows in South catchment. In East catchment, the simulation hydrographs of the lumped models are generally more peaked and overestimate the rising and falling limbs, leading to overestimation of volumes and peak flows. The simulated hydrographs of the quasi-distributed model in East catchment also overestimated the peak discharge but underestimated the rising and recession limb.

### *4.2.2 Level 2 - Proxy-basin test comparison summary*

When transposing from East to South catchment, all the models underestimated peak discharges and volumes, with the quasi-distributed model performing the poorest overall of the three models. Surprisingly, the dual reservoir model performed better than the other two, more complex, models. When transposing from South to East catchment, the quasi-distributed model performed better overall than the two lumped models, although it markedly overestimated peak discharge.

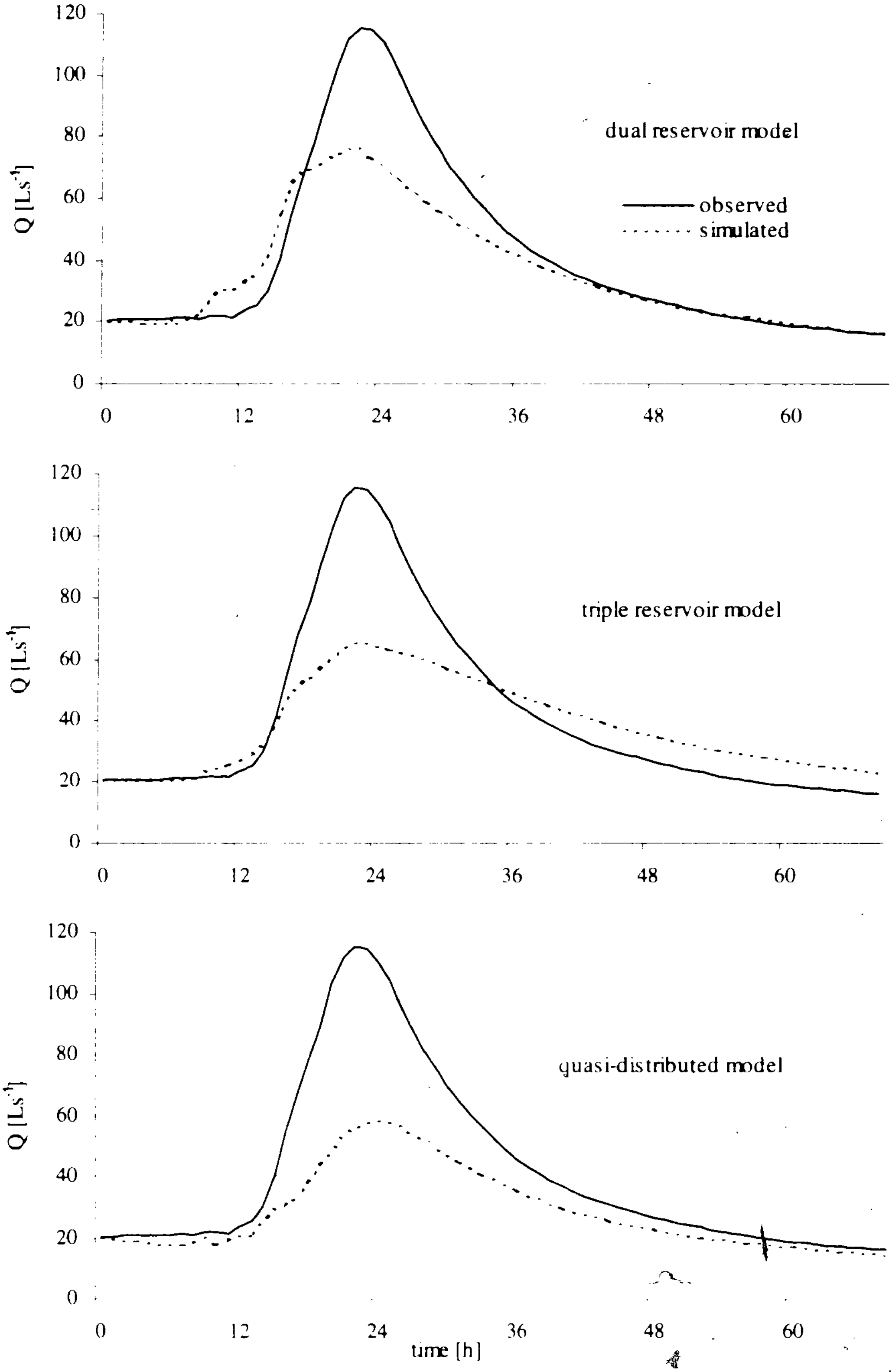



**Figure 4.5** Hydrographs of a small event (event 20), South catchment, level 2.

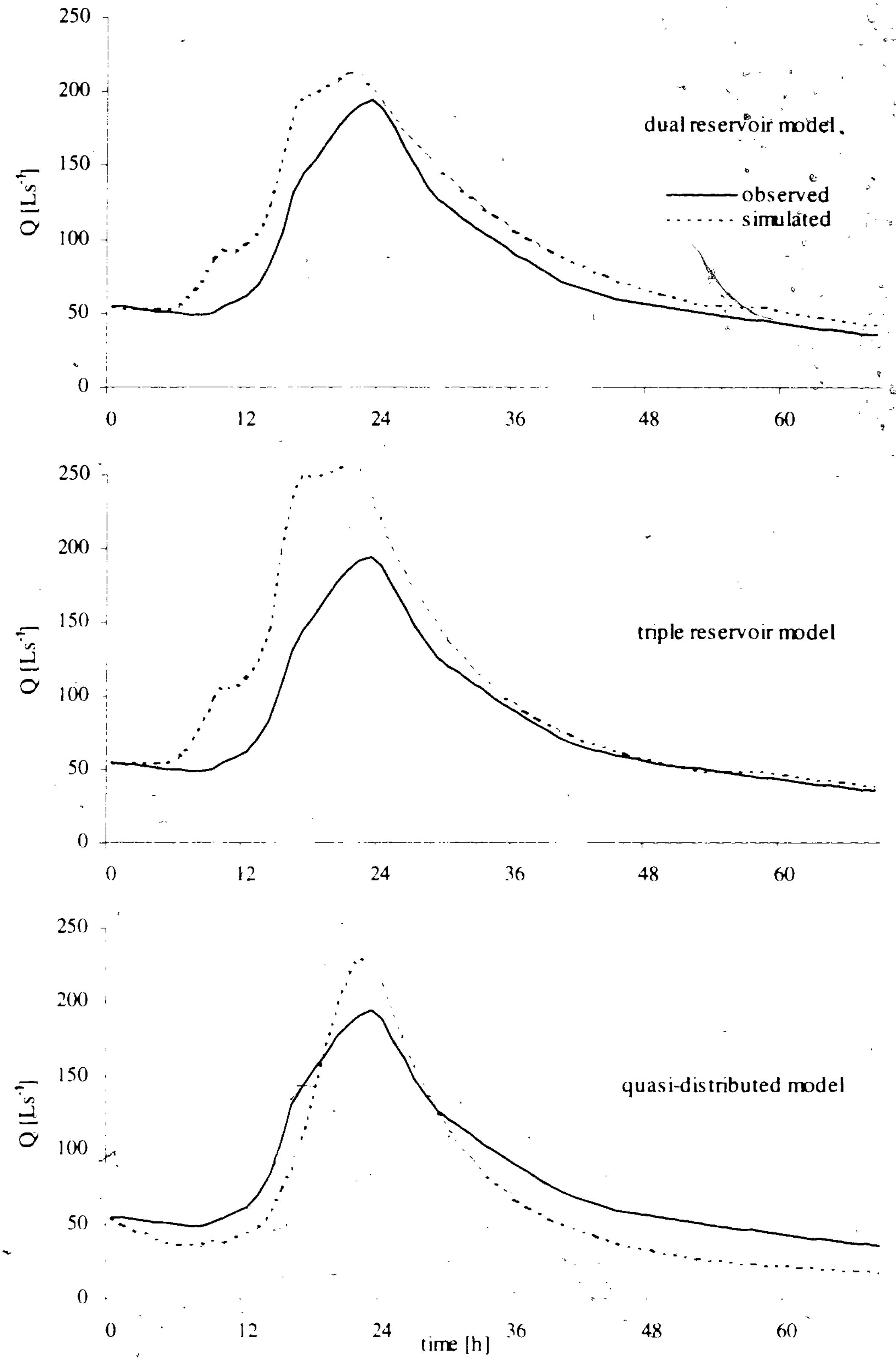



**Figure 4.6** Hydrographs of a small event (event 20), East catchment, level 2.

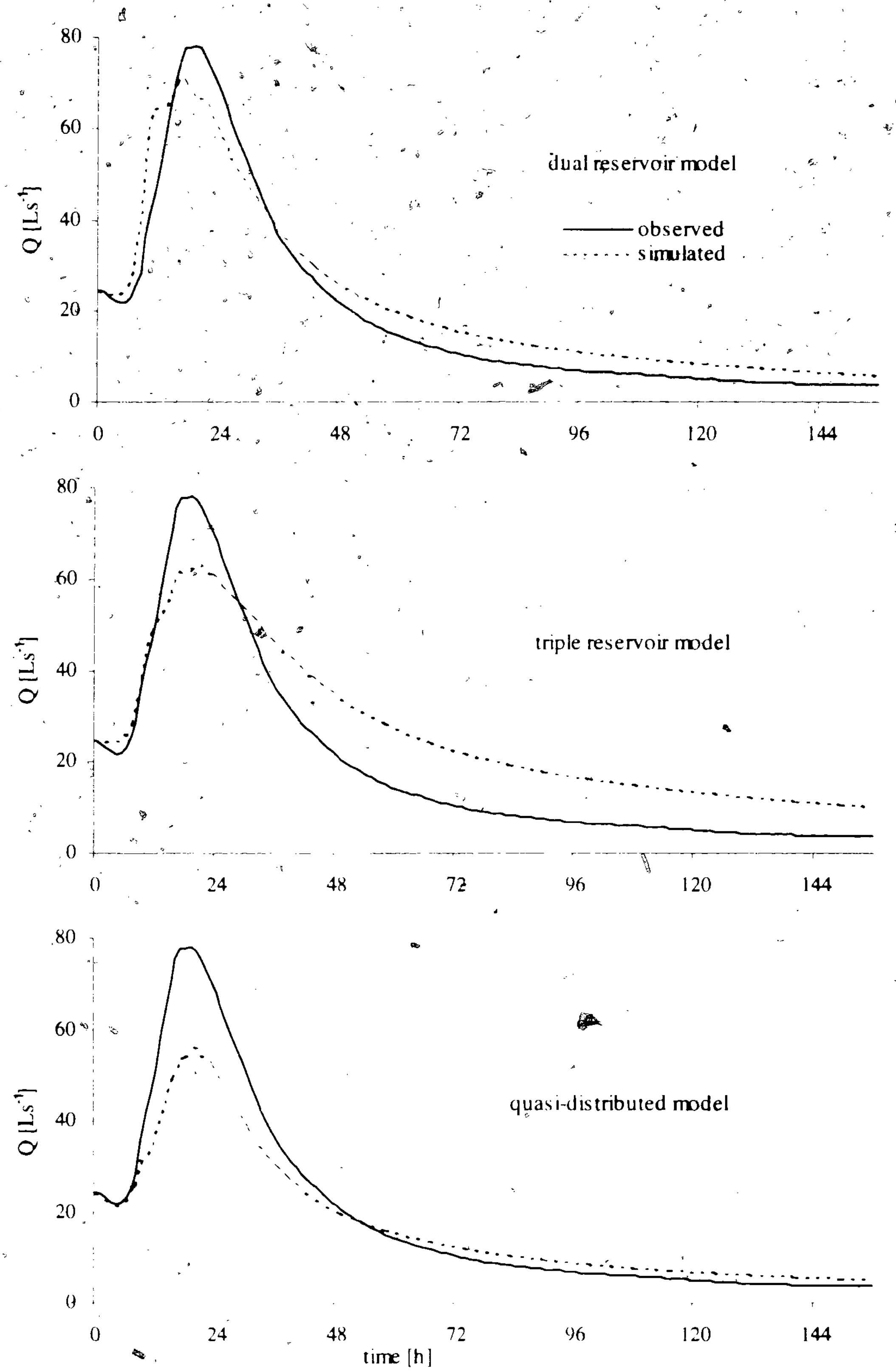



**Figure 4.7** Hydrographs of a large event (event 15), South catchment, level 2.

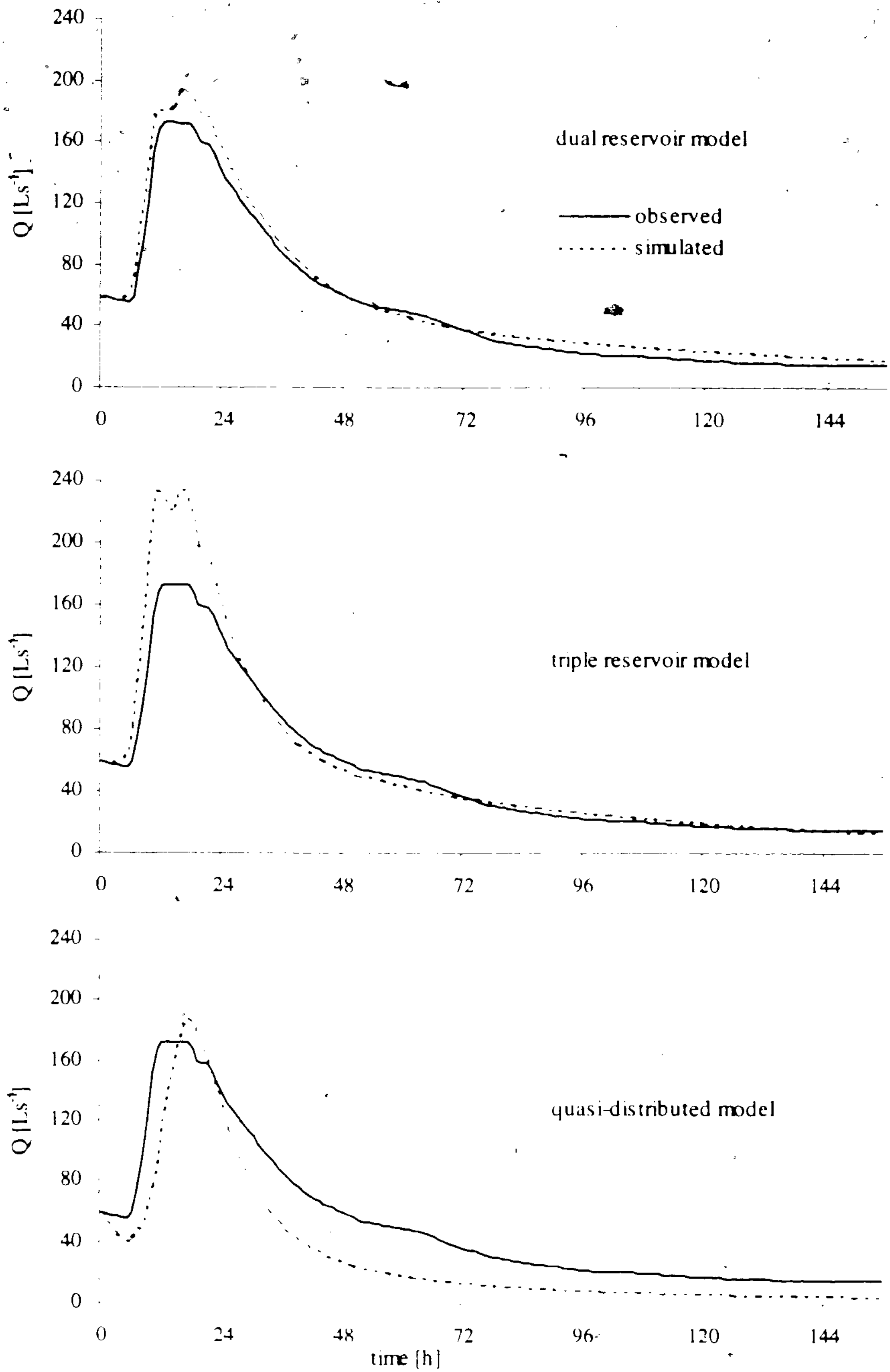



**Figure 4.8** Hydrographs of a large event (event 15), East catchment, level 2.

## 4.3 Level 3 - Differential split-sample test

### *4.3.1 Level 3 - Differential split-sample test calibration results*

For the differential split-sample test, the three models were calibrated using a dataset consisting of small volume events and then validated on a dataset of large volume events of the same catchment. The following results indicate the abilities of the models to simulate runoff events larger than what was used for calibration.

Table 4.7 shows that all three models provided comparable overall fits to the observed data, with all models calibrating to the small events of East catchment slightly better than the events of South catchment. Like the previous results, the lumped models provided better predictions of event volumes than the quasi-distributed model, while the quasi-distributed model better simulated peak discharge rates.

**Table 4.7** Calibration goodness-of-fit indices for level 3 (differential split-sample) testing showing all model efficiency Em values and percent deviation values.

| | | $E_m$ | | | | % deviations | | |
|---|---|---|---|---|---|---|---|---|
| catchment | model | $Q_r$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| south | dual reservoir | 0.792 | 0.793 | 0.457 | 0.619 | +2.5 | -16.3 | -21.5 |
| | triple reservoir | 0.846 | 0.815 | 0.648 | 0.645 | 3.7 | -14.3 | -20.5 |
| | quasi-distributed | 0.865 | 0.646 | 0.809 | 0.881 | -12.5 | -7.4 | -4.3 |
| east | dual reservoir | 0.863 | 0.853 | 0.574 | 0.894 | +0.4 | -13.7 | -8.7 |
| | triple reservoir | 0.901 | 0.833 | 0.753 | 0.905 | 2.8 | -11.8 | -7.4 |
| | quasi-distributed | 0.879 | 0.668 | 0.780 | 0.853 | -12.4 | -7.8 | +5.5 |

**Table 4.8** Calibration test statistics for level 3 (differential split-sample) testing comparing the means and standard deviation of the means of the three models to the observed values. (n=23)

a. South catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 5.8 | 52.7 | 23.3 | 1.8 | 27.8 | 9.3 |
| dual reservoir | 5.9 | 44.1 | 18.3 | 1.6 | 11.0 | 8.8 |
| triple reservoir | 6.0 | 45.2 | 18.5 | 1.7 | 15.7 | 8.8 |
| quasi-distributed | 5.0 | 48.8 | 22.3 | 1.5 | 23.4 | 9.0 |

b. East catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 13.8 | 110.0 | 18.9 | 4.1 | 47.0 | 8.5 |
| dual reservoir | 13.8 | 94.9 | 17.2 | 3.9 | 24.3 | 8.0 |
| triple reservoir | 14.2 | 97.0 | 17.5 | 4.3 | 34.5 | 7.9 |
| quasi-distributed | 12.0 | 101.4 | 20.0 | 3.6 | 52.0 | 8.6 |

### *4.3.2 Level 3 - Differential split-sample test validation results*

The quasi-distributed model was superior to both lumped models when validated on the large events of both catchments, having the highest model efficiency values ($E_m(Q_r)$) of any validation run at any level of testing (Table 4.9). The quasi-distributed model actually performed better on the larger validation events than on the smaller events used for calibration in both catchments. All three models performed well at predicting volume. Unlike other levels of testing, the quasi-distributed model performed as well as the lumped models when simulating volume, although still providing an underestimation.

The quasi-distributed model outperformed the lumped models particularly when simulating peak discharge rates. The quasi-distributed model was able to predict peak discharge well, slightly overestimating the average flow, while the lumped models markedly underestimated the peaks. The negative model efficiency values ($E_m$ ($Q_{pk}$)) in Table 4.9 indicate that the lumped model's predictions were worse than merely using the mean observed discharge rate as the estimated value. Table 4.10 shows that the quasi-distributed model, as in other levels, provided a much better estimate of the variability of peak flows than the lumped models.

Figures 4.9 to 4.12 show the flattened hydrographs of the lumped models which were unable to simulate peak flows. The figures show how well the quasi-distributed model simulated the entire observed hydrograph. The quasi-distributed model predicted the peaks, as well as the rising and falling limbs, acceptably, even for the larger events.

### *4.3.3 Level 3 - Differential split-sample test comparison summary*

For the differential split-sample test, the quasi-distributed model performed exceptionally well while the lumped models performed poorly. The quasi-distributed model was able to simulate large peak discharge rates while the lumped models underestimated the peak flows by nearly 50 percent in both catchments. These results indicate that the quasi-distributed model performed particularly well when predicting events outside of the range for which it was calibrated while the lumped models did not provide reasonable predictions.

**Table 4.9** Validation goodness-of-fit indices for level 3 (differential split-sample) testing showing all model efficiency Em values and percent deviation values.

| | | $E_m$ | | | | % deviations | | |
|---|---|---|---|---|---|---|---|---|
| catchment | model | $Q_r$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| south | dual reservoir | 0.684 | 0.837 | 0.595 | 0.919 | -10.3 | -51.8 | +4.9 |
| | triple reservoir | 0.774 | 0.832 | -0.206 | 0.921 | -9.2 | -44.4 | 3.5 |
| | quasi-distributed | 0.911 | 0.817 | 0.809 | 0.994 | -16.7 | +4.7 | -1.7 |
| east | dual reservoir | 0.760 | 0.873 | -0.120 | 0.945 | -6.2 | -45.2 | +1.0 |
| | triple reservoir | 0.810 | 0.841 | 0.165 | 0.949 | -4.1 | -37.8 | -0.4 |
| | quasi-distributed | 0.924 | 0.893 | 0.804 | 0.903 | -10.7 | +7.2 | -2.4 |

**Table 4.10** Validation test statistics for level 3 (differential split-sample) testing comparing the means and standard deviation of the means of the three models to the observed values. (n=23)

a. South catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| data | [$10^6$ L] | [$Ls^{-1}$] | [h] | [$10^6$ L] | [$Ls^{-1}$] | [h] |
| observed | 15.9 | 173.7 | 28.2 | 8.2 | 87.5 | 15.0 |
| dual reservoir | 14.3 | 83.8 | 29.6 | 6.0 | 29.7 | 17.5 |
| triple reservoir | 14.5 | 96.7 | 29.2 | 6.0 | 35.1 | 17.7 |
| quasi-distributed | 13.3 | 181.9 | 27.7 | 6.4 | 98.9 | 14.8 |

b. East catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| data | [$10^6$ L] | [$Ls^{-1}$] | [h] | [$10^6$ L] | [$Ls^{-1}$] | [h] |
| observed | 33.4 | 296.5 | 29.7 | 14.9 | 172.3 | 16.5 |
| dual reservoir | 31.4 | 162.6 | 30.0 | 11.6 | 57.9 | 17.2 |
| triple reservoir | 32.0 | 184.4 | 29.6 | 11.5 | 68.8 | 17.5 |
| quasi-distributed | 29.8 | 317.7 | 29.0 | 14.3 | 191.4 | 15.7 |

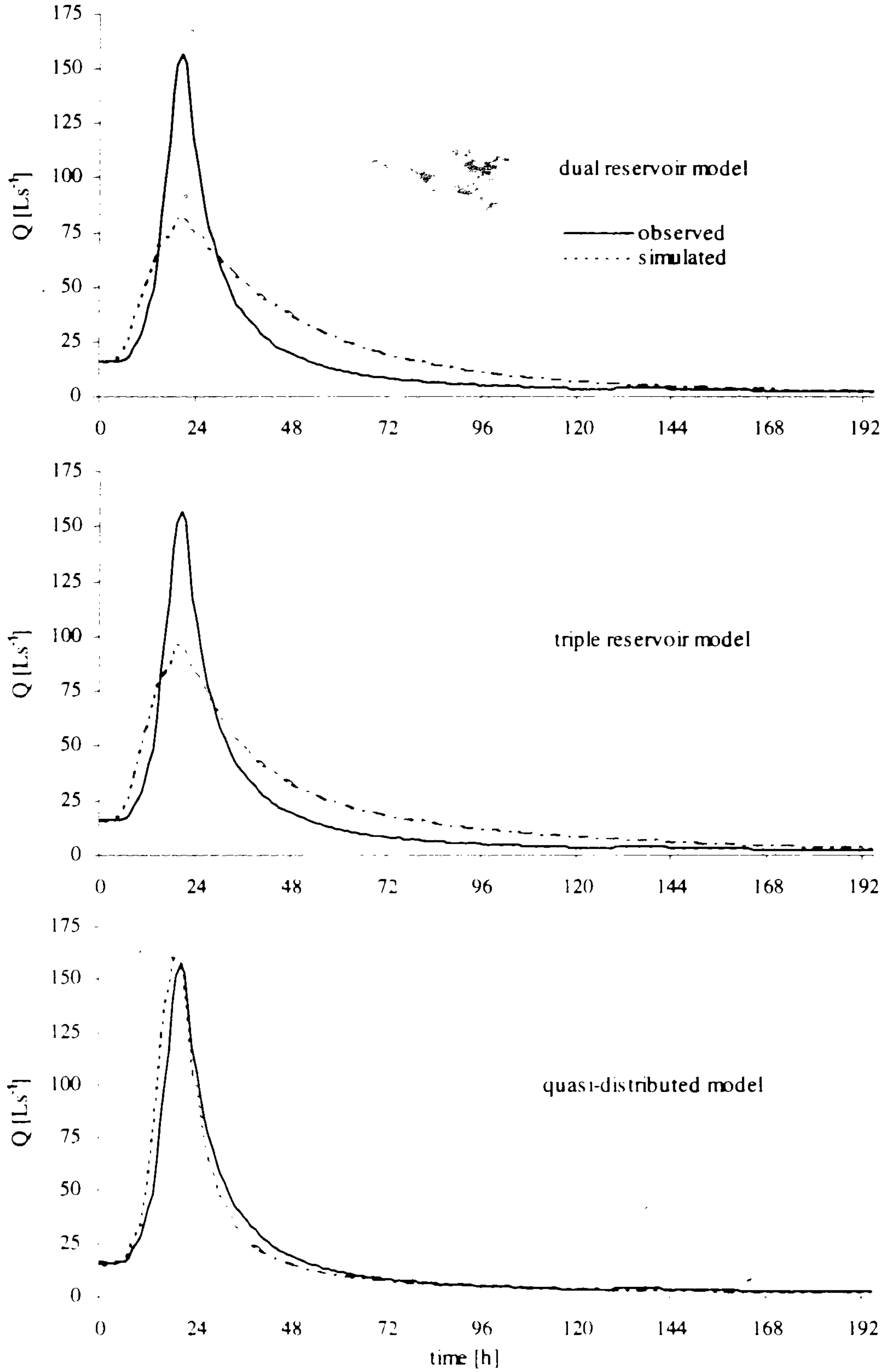



**Figure 4.9** Hydrographs of a small event (event 21), South catchment, level 3.

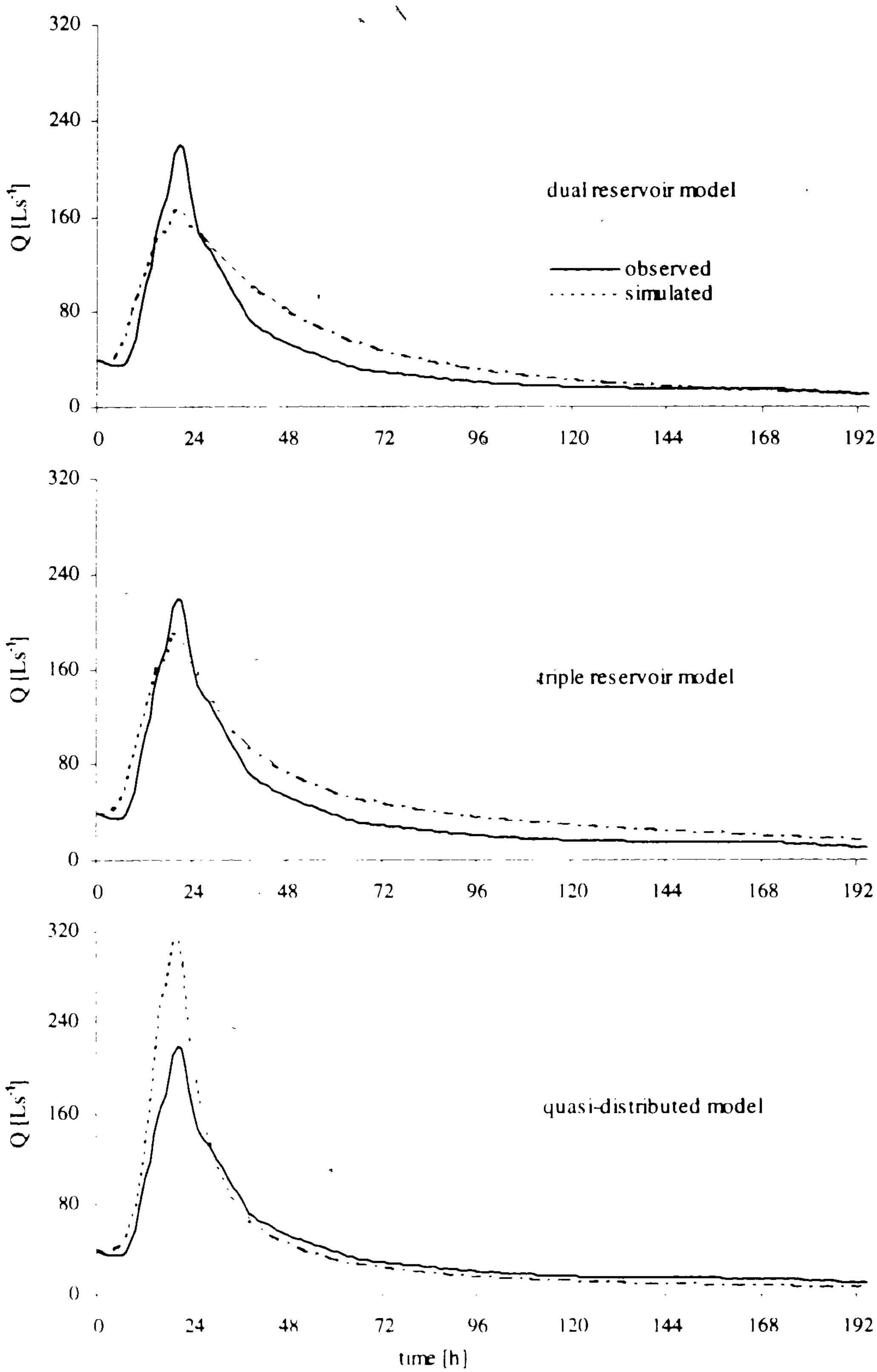



**Figure 4.10** Hydrographs of a small event (event 21), East catchment, level 3.

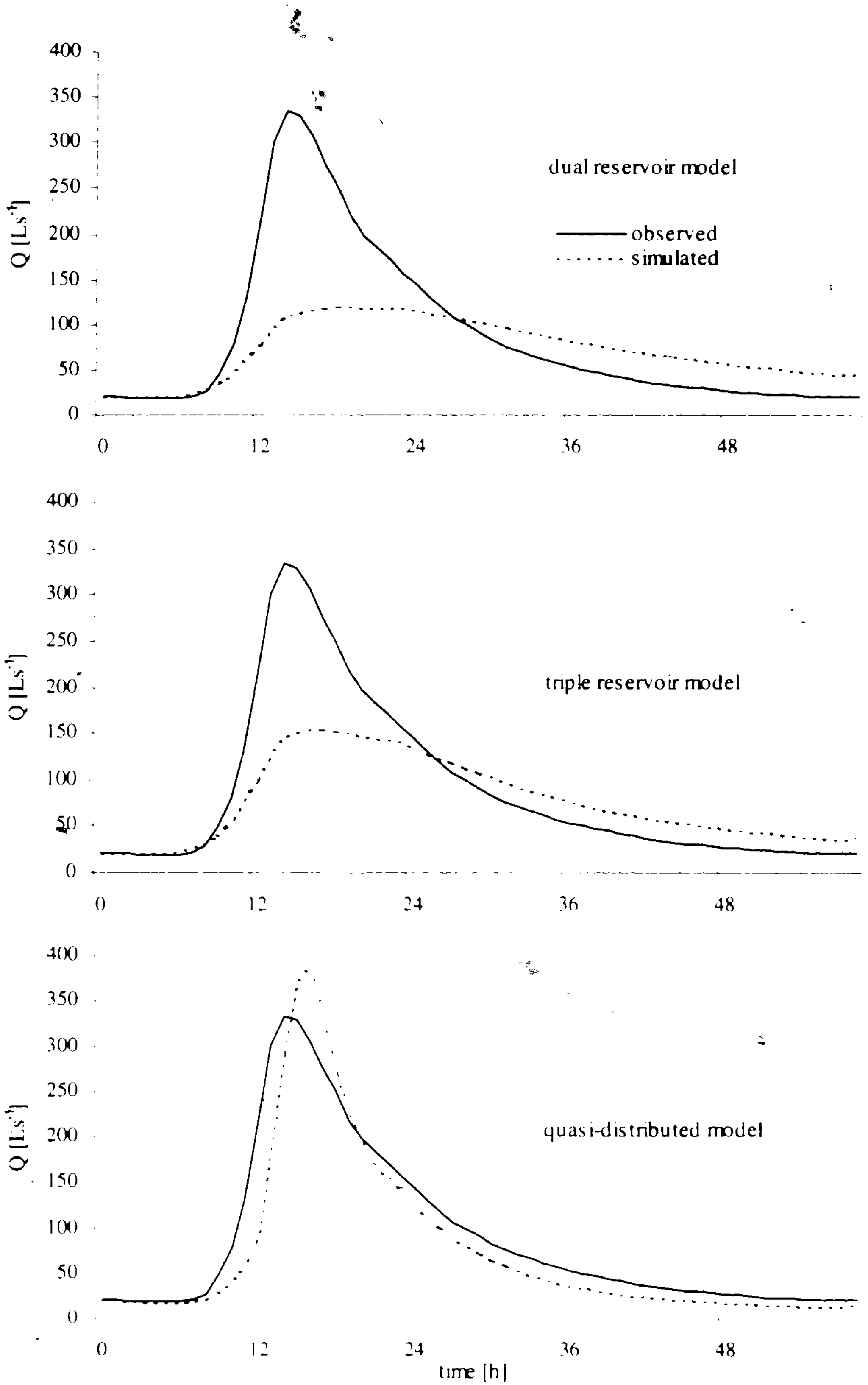



**Figure 4.11** Hydrographs of a large event (event 9), South catchment, level 3.

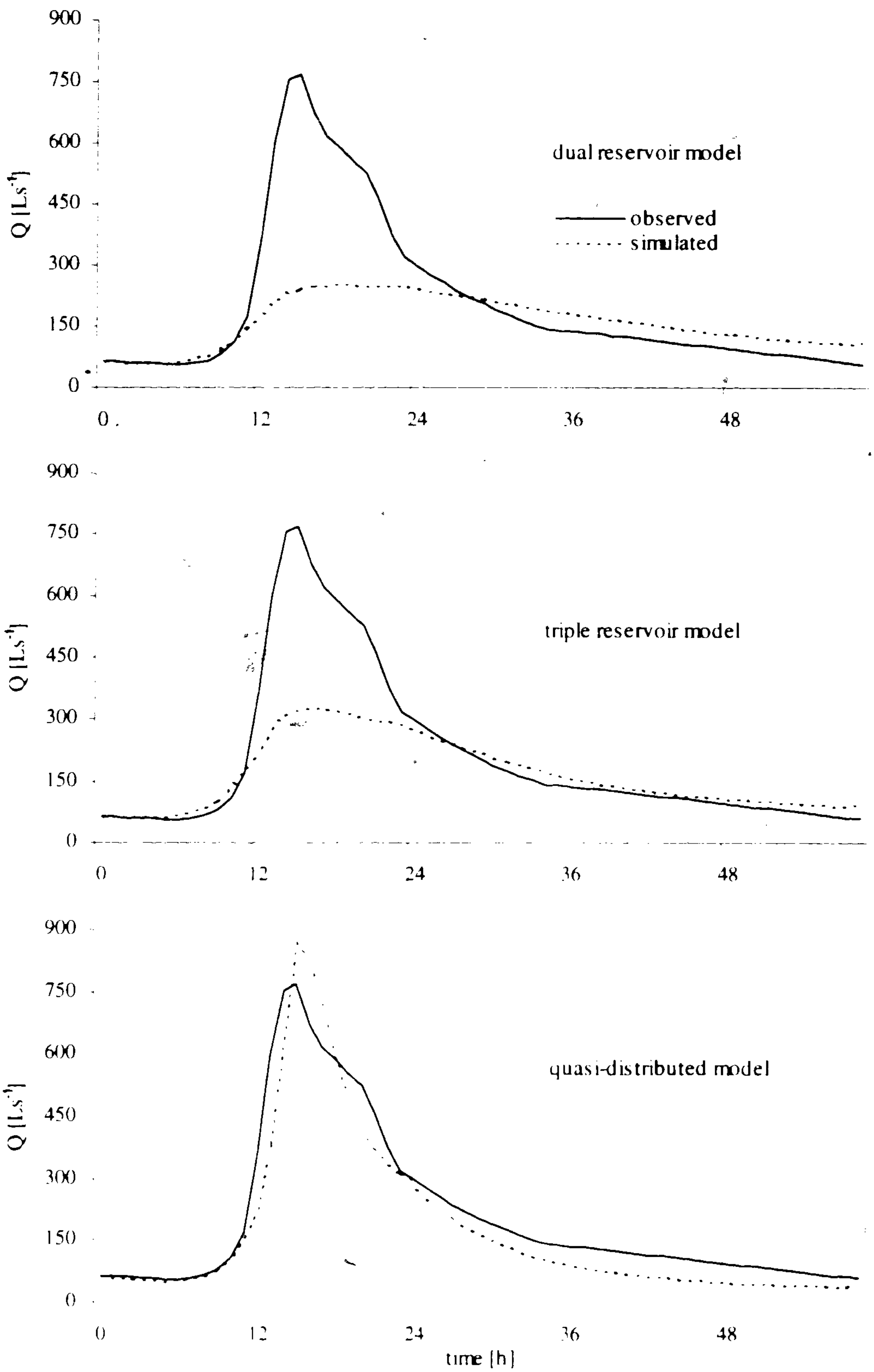



**Figure 4.12** Hydrographs of a large event (event 9), East catchment, level 3.

## 4.4 Level 4 - Proxy-basin differential split-sample test

### *4.4.1 Level 4 - Proxy-basin differential split-sample test validation results*

The proxy-basin differential split-sample test requires the models to be calibrated on a dataset of small events of one catchment and validated on the large events of a second catchment. The calibration results were presented in section 4.3.1 for each catchment. The validation results (Tables 4.11 and 4.12) show that, overall, the quasi-distributed model performed markedly better than the lumped models when transposed from East catchment to the larger events of South catchment, but performed slightly worse than the lumped models when transposed from South to East catchment. The lumped models both performed better when transposed to East catchment than they did when transposed to South catchment.

Like the overall performance, the lumped models provided better predictions of volume at East catchment than the quasi-distributed model, and worse predictions of volume at South catchment. The prediction of peak flow rates were poor by all three models. Similar to the performance of the lumped models when simulating large flows within one catchment (Level 3), the lumped models extremely underestimated peak discharge rates of the large events in the second proxy catchment, especially in South catchment where the Nash coefficients were negative values. The quasi-distributed model underestimated peaks in South catchment and overestimated peaks in East catchment by approximately a third of the observed peak flow.

Figures 4.13 to 4.16 illustrate that the lumped models generated flattened simulation hydrographs which did not reproduce the peaks nor fit the recession limbs. The quasi-distributed model is better than the lumped models at fitting the shape of the observed hydrograph and able to simulate the recession curves and predict peak discharges slightly better, especially for the larger events.

### *4.4.2 Level 4 - Proxy-basin differential split-sample test comparison summary*

For the proxy-basin differential split-sample test, as for the proxy-basin test, the quasi-distributed model performed better overall in one catchment and worse overall in the other catchment when compared to the lumped models. All three models were unable to simulate peak discharge rates.

**Table 4.11** Validation goodness-of-fit indices for level 4 (proxy-basin differential split-sample) testing showing all model efficiency Em values and percent deviation values.

| | | $E_m$ | | | | % deviations | | |
|---|---|---|---|---|---|---|---|---|
| catchment* | model | $Q_r$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ | $Q_{vol}$ | $Q_{pk}$ | $t_{pk}$ |
| East (c) | dual reservoir | 0.612 | 0.699 | -0.913 | 0.919 | -16.3 | -57.9 | +5.1 |
| South (v) | triple reservoir | 0.594 | 0.616 | -1.00 | 0.836 | -16.6 | -60.0 | +12.2 |
| | quasi-distributed | 0.846 | 0.750 | 0.532 | 0.957 | -20.8 | -29.4 | +5.4 |
| South (c) | dual reservoir | 0.781 | 0.935 | 0.188 | 0.945 | -0.3 | -37.7 | +1.0 |
| East (v) | triple reservoir | 0.874 | 0.929 | 0.543 | 0.943 | +3.3 | -19.5 | -6.0 |
| | quasi-distributed | 0.765 | 0.867 | 0.162 | 0.916 | -10.2 | +37.0 | -7.0 |

* (c) calibration catchment, (v) validation catchment

**Table 4.12** Validation test statistics for level 4 (proxy-basin differential split-sample) testing comparing the means and standard deviation of the means of the three models to the observed values. (n=23)

a. Calibrated on East catchment, validated on South catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 15.9 | 173.7 | 28.2 | 8.2 | 87.5 | 15.0 |
| dual reservoir | 13.3 | 73.2 | 29.7 | 5.4 | 25.8 | 17.5 |
| triple reservoir | 13.3 | 69.5 | 31.7 | 5.4 | 27.1 | 17.3 |
| quasi-distributed | 12.6 | 122.6 | 29.7 | 6.4 | 71.0 | 16.8 |

b. Calibrated on South catchment, validated on East catchment

| | mean | | | standard deviation of mean | | |
|---|---|---|---|---|---|---|
| data | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] | $Q_{vol}$ [$10^6$ L] | $Q_{pk}$ [$Ls^{-1}$] | $t_{pk}$ [h] |
| observed | 33.4 | 296.5 | 29.7 | 14.9 | 172.3 | 16.5 |
| dual reservoir | 33.3 | 184.9 | 30.0 | 13.1 | 66.5 | 17.2 |
| triple reservoir | 34.7 | 238.8 | 27.9 | 13.4 | 81.1 | 15.0 |
| quasi-distributed | 30.0 | 406.3 | 27.6 | 14.5 | 221.7 | 14.5 |

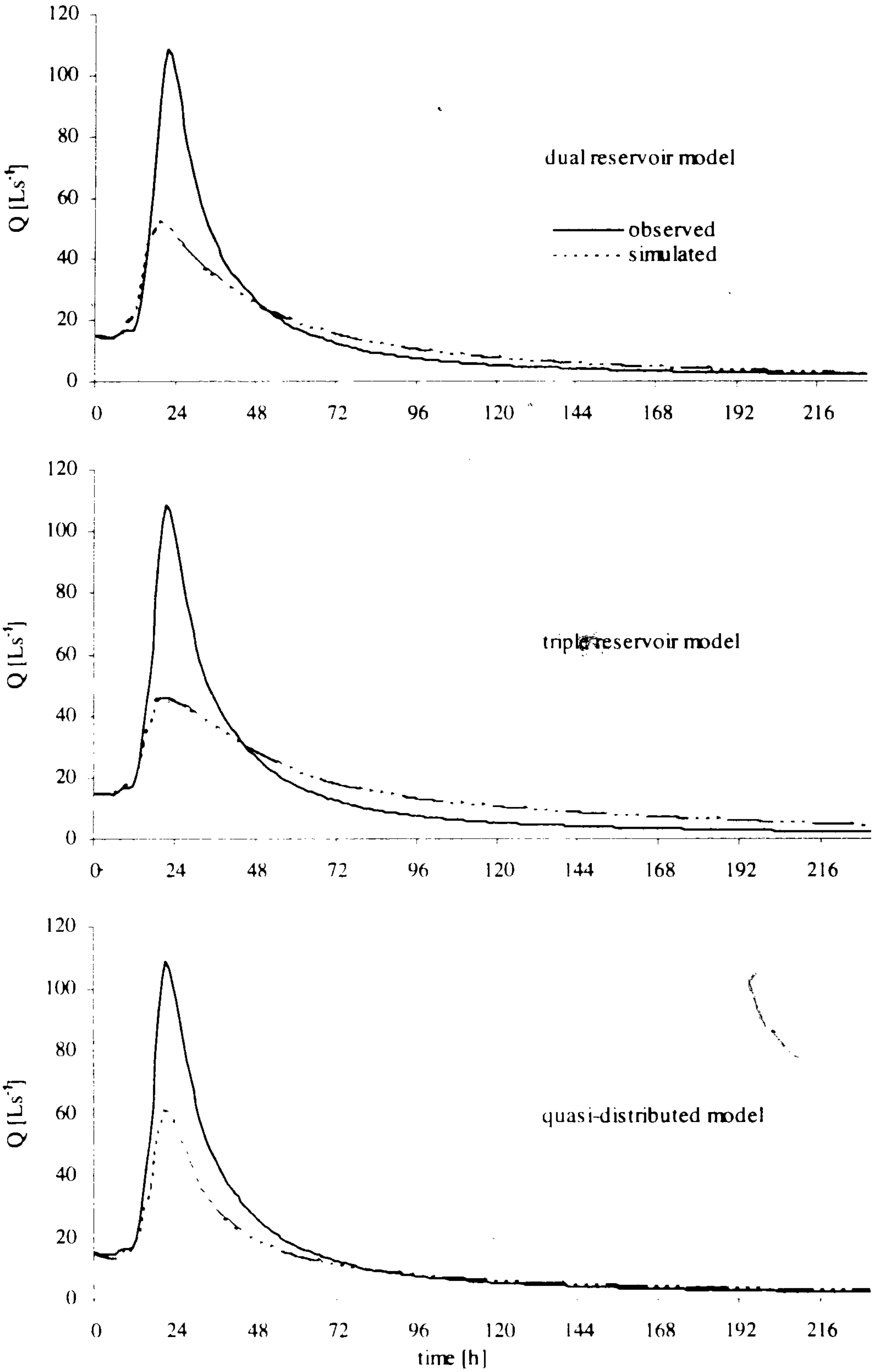



**Figure 4.13** Hydrographs of a small event (event 39), South catchment, level 4.

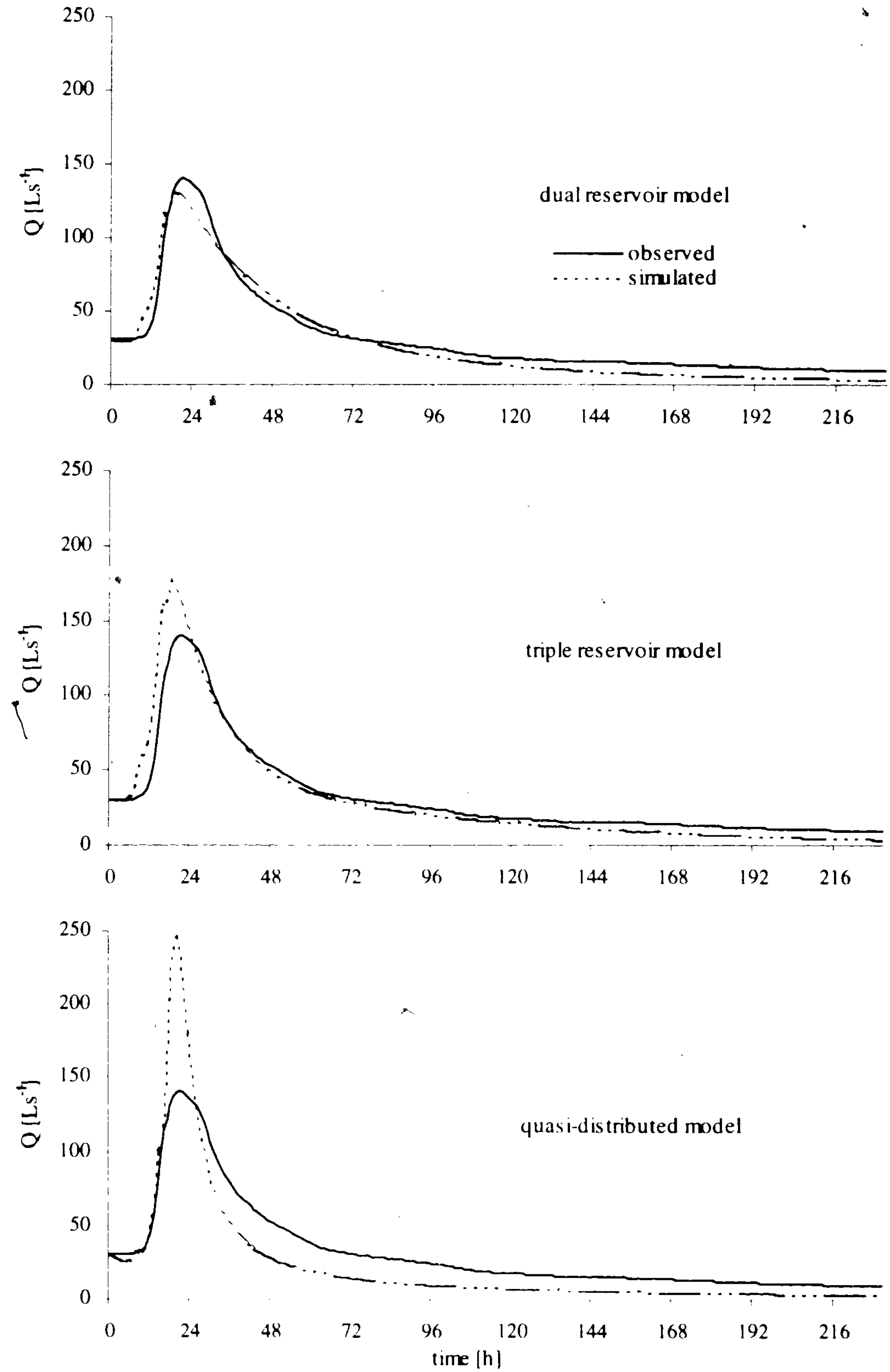



**Figure 4.14** Hydrographs of a small event (event 39), East catchment, level 4.

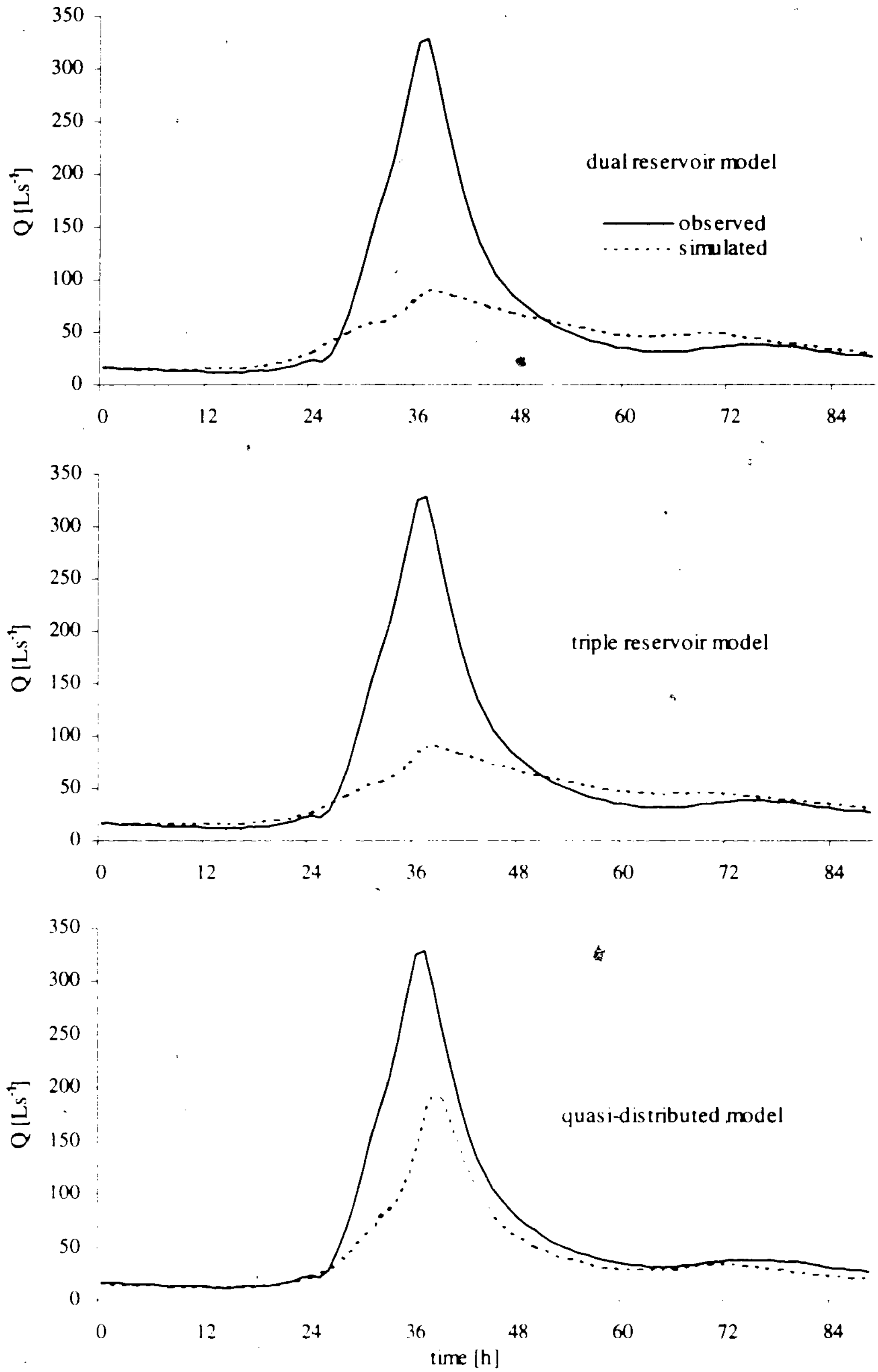



**Figure 4.15** Hydrographs of a large event (event 5), South catchment, level 4.

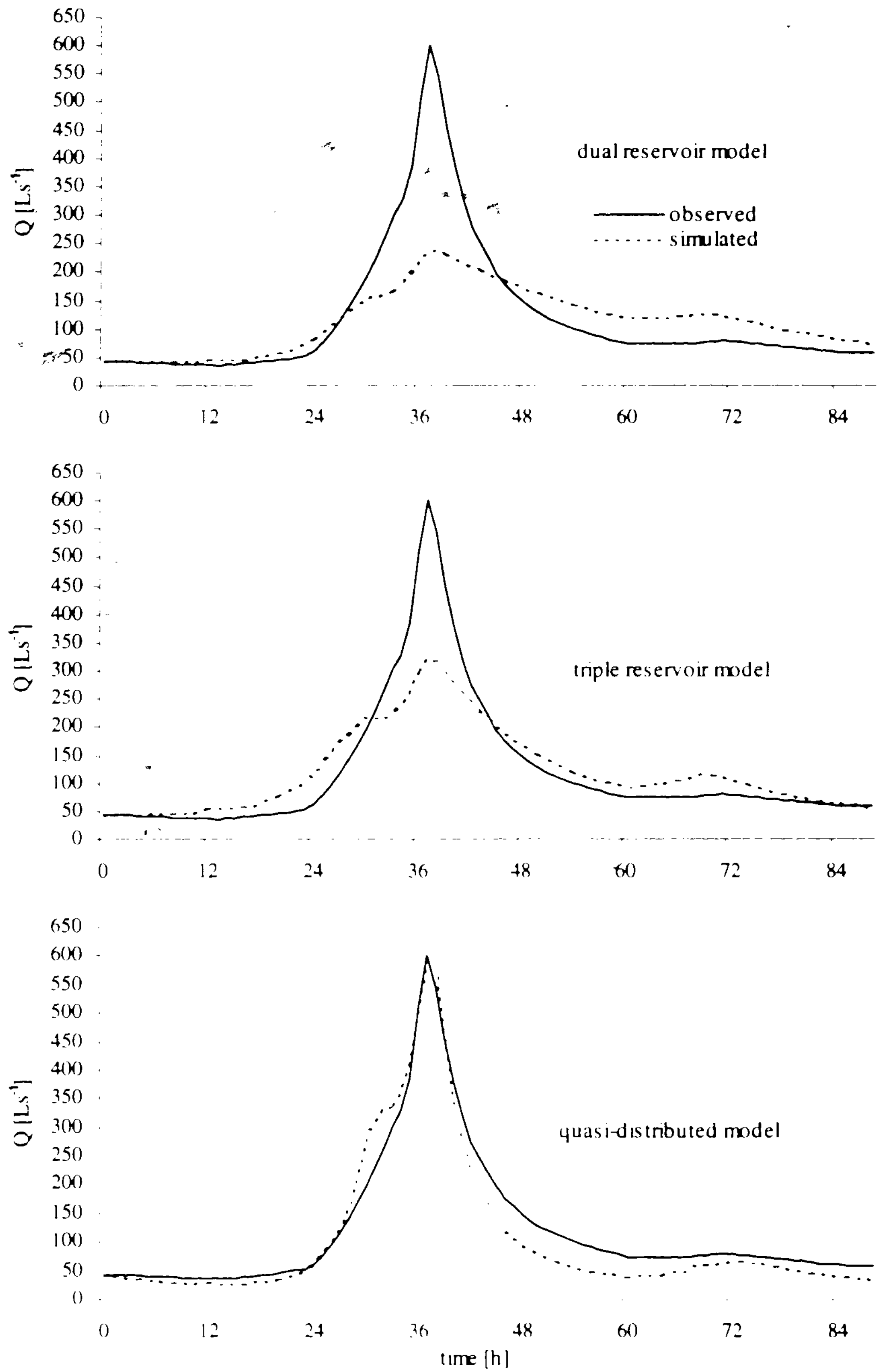



**Figure 4.16** Hydrographs of a large event (event 5), East catchment, level 4.

## 4.5 Summary of model comparisons

The model efficiency coefficient ($E_m$) of the average hourly discharge rate ($Q_r$) is the main criterion used for comparing overall model performance as it represents the ability of the model to simulate the entire hydrograph. Table 4.13 presents the validation run values for each model, within each catchment, for each level of testing.

**Table 4.13** Model efficiency values ($E_m(Q_r)$) for all validation simulations.

| | dual parallel reservoir | | triple serial reservoir | | quasi-distributed | |
|---|---|---|---|---|---|---|
| Level | South | East | South | East | South | East |
| 1 | 0.797 | 0.811 | 0.861 | 0.869 | 0.890 | 0.890 |
| 2 | 0.759 | 0.777 | 0.707 | 0.734 | 0.652 | 0.800 |
| 3 | 0.684 | 0.760 | 0.774 | 0.810 | 0.911 | 0.924 |
| 4 | 0.612 | 0.781 | 0.594 | 0.874 | 0.846 | 0.765 |

Table 4.14 summarizes which model performed the best at each level for each catchment. The quasi-distributed model demonstrated the best overall performance. However, the lumped models outperformed the quasi-distributed model in one of the two catchments in the proxy-basin and differential split-sample tests.

**Table 4.14** Top performing model at each level of testing within each catchment.

| Level | South | East |
|---|---|---|
| 1 | quasi-distributed | quasi-distributed |
| 2 | dual parallel reservoir | quasi-distributed |
| 3 | quasi-distributed | quasi-distributed |
| 4 | quasi-distributed | triple serial reservoir |

Knowing which model outperformed another at the various levels of testing is important but also rather inconclusive to a potential user trying to decide which model to apply. The apparent differences in performance among models depend not just on real differences in the models' predictive abilities, but also on the selection of events used in the testing. It is possible that the relative rankings of the models might change if a different selection of events were used. It is necessary to quantify the difference between performance results to assess if one model is really any better than another. The next and final section of this chapter addresses the issue of the statistical significance of the model comparisons.

### 4.6 Model comparison using statistical analysis

The Jackknife procedure was used to generate statistically independent, identically-distributed "pseudo-values" or estimates of the root-mean-square error (RMSE), which is essentially a re-scaled representation of the model efficiency $E_m(Q_r)$ values as described in Section 2.7.3*(ii)*. The variability of the pseudo-values reflects the sampling variability of RMSE that results from a particular sample of events used for model testing. Table 4.15 shows the mean and standard deviation of the generated RMSE values for each catchment.

**Table 4.15** Mean and standard deviation of mean of RMSE pseudo-values.

a) South catchment

| | dual parallel reservoir | | triple serial reservoir | | quasi-distributed | |
|---|---|---|---|---|---|---|
| Level | mean | std dev | mean | std dev | mean | std dev |
| 1* | 16.3 | 2.6 | 13.5 | 2.5 | 12.1 | 2.4 |
| 2* | 17.8 | 3.5 | 19.7 | 4.0 | 21.3 | 3.4 |
| 3** | 28.3 | 5.3 | 24.0 | 4.2 | 15.0 | 2.2 |
| 4** | 31.4 | 5.8 | 32.1 | 5.5 | 19.7 | 2.8 |

* n = 26 ** n = 23

b) East catchment.

| | dual parallel reservoir | | triple serial reservoir | | quasi-distributed | |
|---|---|---|---|---|---|---|
| Level | mean | std dev | mean | std dev | mean | std dev |
| 1* | 28.8 | 6.3 | 24.0 | 5.2 | 21.8 | 4.2 |
| 2* | 30.8 | 3.9 | 33.6 | 2.6 | 29.3 | 4.3 |
| 3** | 45.5 | 10.0 | 38.1 | 7.8 | 23.9 | 3.1 |
| 4** | 41.0 | 9.0 | 30.8 | 5.2 | 41.9 | 5.5 |

* n = 26 ** n = 23

Analyses of variance (ANOVAs) were used to determine if the RMSE estimates, and indirectly the $E_m(Q_r)$ coefficients, of each model's performance were significantly different from each other. Table 4.16 presents the results of the two-way ANOVAs applied to each of the four levels of testing. Using a significance level of 0.05, the statistical analysis showed that at levels 1, 2, and 4 there was no statistical difference in model performance between the three runoff models. The ANOVA results also showed, however, that there was a statistical difference (in bold type) between model performance at level 3, the differential split-sample test, indicating that one of the models performed statistically significantly better or worse than the other models at that level.

**Table 4.16** Exact probabilities resulting from ANOVA tests for each level. Probabilities significant at alpha = 0.05 are bold faced.

| | level 1 | level 2 | level 3 | level 4 |
|---|---|---|---|---|
| model | 0.189 | 0.611 | **0.003** | 0.767 |
| basin | 0.000 | 0.000 | 0.000 | 0.008 |
| model*basin | 0.943 | 0.325 | 0.971 | **0.021** |

A post-hoc Tukey test was applied to the data for level 3 to ascertain differences in RMSE distributions between the models (Table 4.17). Using a significance level of 0.05, the Tukey test results showed that there was no significant difference between model performance between the two lumped models. The results also showed that there was a significant difference between model performance of the triple reservoir model and the quasi-distributed model and an even greater significant difference between the quasi-distributed model and the dual reservoir model. The model efficiency values (Table 4.13) indicate clearly that the quasi-distributed model was a superior performer, now proven to be statistically significantly superior, over the lumped models at the differential split-sample level of testing.

**Table 4.17** Summary of pairwise comparison probabilities for the Tukey test for level 3 (differential split sample) testing. Comparisons significant at alpha = 0.05 are bold faced.

| | dual reservoir | triple reservoir | quasi-distributed |
|---|---|---|---|
| dual reservoir | 1.000 | | |
| triple reservoir | 0.585 | 1.000 | |
| quasi-distributed | **0.002** | **0.042** | 1.000 |

The ANOVA results of Table 4.16 also indicate that there was an interaction effect occurring between model and basin at level four, the proxy-basin differential split-sample test. The values of least square means generated by the analysis of variance for level 4 indicated that the two lumped models were not affected by catchment whereas the quasi-distributed model was.

In conclusion, these results show that the quasi-distributed model generally outperforms the spatially lumped models at the various levels of testing and has been shown to be statistically significantly better than the lumped models when predicting large events outside of the range for which the model was calibrated. The next chapter will discuss and expand on these results.

# CHAPTER 5

# DISCUSSION

This chapter presents a discussion of factors that can influence results of a runoff modelling study, followed by a discussion of the results themselves in order of the level of test used.

## 5.1 Sources of error

Many factors affect the accuracy of runoff simulations: input data, initial conditions, model assumptions, parameter values, runoff dynamics, and model spatial resolution. Since it is difficult to examine all of these issues properly, Loague and Freeze (1985) categorized three sources of error inherent in rainfall-runoff models: model error, input error, and parameter error. These sources are introduced in this section, with the following section providing a discussion more specific to the study and the models used.

### *5.1.1 Model error*

Model error results in the inability of a rainfall-runoff model to predict runoff accurately, even given the correct estimates and input. Model error will always be a factor since no model can represent the real system exactly. The purpose of a model comparison study is to test for the difference in model error between models.

### *5.1.2 Input error*

Input error in this study could arise from measurement error (errors involving the precipitation gauge), extrapolation error (extrapolating catchment input from the gauge location), and throughfall and stemflow estimation error (parameterization of the canopy storage model). Input error can be significant in some studies. For example, Michaud and Sorooshian (1994) found that rainfall errors were responsible for roughly half of the runoff simulation errors. Input error is not as much of a concern in this thesis research for two reasons. First, significantly smaller catchments were used as compared to those used by Michaud and Sorooshian such that errors involved in extrapolation of precipitation are minimized. Extrapolation error may be more of a concern in the larger East catchment but the results show that input error was compensated for (if compensation were necessary) since the calibration and validation runs were generally better in East catchment than South catchment. Secondly, and more importantly, any input errors would have been the same for all three runoff models being compared since the canopy storage model was calibrated independently of the runoff models.

Input error could be a factor at level 2 and level 4 testing, in which the models were validated on a catchment other than that used for calibration. If there is a greater error due to the storage canopy model estimates in either catchment, level 2 and level 4 results may not provide a valid test of geographic transposability. If the input errors are associated with the calibration catchment, parameter estimation may be erroneous and misleading. If the errors are associated with the validation catchment, the validation test becomes inapplicable.

### *5.1.3 Parameter error*

All three runoff models used in this study contain parameters that were calibrated to a particular set of events. Errors of measurement of the observed streamflow, errors in digitizing the streamflow (drift), synchronization errors (errors between the precipitation and streamflow gauge data), and errors in the stage-discharge relation may alter the obtained observed data from the actual data, resulting in different optimal parameters.

Parameter error may also result from the interdependence of model parameters for each of the models, which is the main problem with optimization. There may not be a set of unique parameter estimates that can reproduce the recorded runoff (Gan and Biftu, 1996). However, the model efficiency response surfaces generated for calibration did not indicate any multiple optimum parameter sets for the calibration data used.

In addition, subjective 'tweaking' was used after the automatic calibration process to optimize the parameters. Svensson (1977) compared subjective and automatic calibration, and concluded that subjective calibration was in some ways superior. Hence, a combination of the two types of optimization minimizes parameter error.

The differing methods of optimization between the lumped and the quasi-distributed models may have affected the results. For the lumped models, one parameter was initially calibrated using recession curves and then the other parameters were calibrated using entire events. The quasi-distributed model was calibrated by optimizing all parameters simultaneously using the entire hydrographs. This difference in calibration may be a contributing factor as to why the lumped models do not simulate the peaks as well as the quasi-distributed model since their calibration was dependent on the recession.

It may also explain why the quasi-distributed model is better able to simulate variability of peak flows. However, this difference in calibration is probably not significant since the parameter obtained from recession analysis was the one controlling low-frequency response. The parameters controlling high-frequency response (i.e. stormflow response) for the two lumped models were calibrated using the entire stormflow hydrographs.

## 5.2 Discussion of results by testing level

### *5.2.1 Level 1 - Split-sample test*

The split-sample test, in which models are calibrated and validated on a similar range of data sets within the same catchment, resulted in all three models performing similarly, with the quasi-distributed model providing the best overall fit. Although the three models did not perform statistically significantly different from one another, some reasons for the better performance of the quasi-distributed model are discussed below. The following discussion can also be applied at all four levels of testing.

The better performance of the quasi-distributed model is probably due to the more conceptual representation and the better accountability of the distribution of storage of the quasi-distributed model than the lumped models. One reason that the quasi-distributed model performed well could be a result of the routing method TOPMODEL uses. Moore (1997) showed that streamflow recession at South catchment is consistent with the exponential storage-outflow relation assumed by TOPMODEL for the first two days after a rainfall event when the catchment was wetted up. The function of the recession for East catchment has not been researched but since the models generally do better in East catchment than in South, the recession is probably consistent with the TOPMODEL assumptions in East catchment also.

The triple reservoir model probably provided a better fit than the dual reservoir model because it has more parameters. Also, the triple reservoir model has a more realistic representation of the processes, and has a better resolution of storage distribution since it is comprised of three reservoirs as opposed to two.

The lumped models were better at simulating event volumes than the quasi-distributed model, but not at simulating the shape of the hydrographs. Since all three models were calibrated to fit the entire shape of the hydrograph, the better estimates of volume are probably just a function of the poorer fit of the entire hydrograph. The lumped models responded too quickly to precipitation input and overestimated the rising limbs while underestimating the peaks, resulting in a good estimate of volume merely by coincidence by the averaging of discharge over the event. The percentage of the underestimation of peak flows by the lumped models is misleading. The peak flows of most events were adequately simulated while only a few events skewed the estimated value.

For the split sample test, all models generally exhibited similar performance. Any of the models could be used as a reliable tool for filling in gaps in streamflow records or used to extend runoff series. Considering the data requirements and efforts involved in the setup of the different models, the simplest dual model may be selected for such tasks. This conclusion is in agreement with results of other studies (e.g., Michaud and Sorooshian, 1994; Refsgaard and Knudsen, 1996).

*5.2.2 Level 2 - Proxy-basin test*

The models, when validated on the catchment for which it was not calibrated, performed differently in each catchment. The quasi-distributed model performed better than the lumped models when validated in East catchment and worse than the lumped models when validated in South catchment. These results are similar to the findings of a study by Refsgaard and Knudsen (1996), in which a quasi-distributed model did well in one catchment but not in another. Although the results for level 2 are not statistically significant, possible explanations for the differences in model performance are discussed below.

*i) Nonlinearity of TOPMODEL*

The exponential reservoir of TOPMODEL produces a non-linear response and may be an important explanation for the difference in performance of the quasi-distributed model between catchments. The nonlinearity of TOPMODEL is also important at level 3 testing (Section 5.2.3). Generally, a non-linear reservoir is more sensitive than a linear reservoir to rainfall input (Singh and Woolhiser, 1976). This sensitivity may result in a linear routing model being more accurate than a nonlinear one, even though the underlying process is actually nonlinear, resulting in the differing performance of the quasi-distributed model in the two catchments..

The sensitivity of the exponential reservoir in TOPMODEL may depend on the combination of the parameters $m$ and $T_o$. The parameter $T_o$ represents the transmissivity (linked to the hydraulic conductivity) and it affects both the interflow regime and, together with the parameter $m$, the flow exchange rate between the unsaturated and saturated zones. With these parameters being quite different in the two catchments, ($m$ for East catchment is approximately double that of South catchment), it should be expected that the quasi-distributed model would perform differently in the two catchments. Perhaps the model is less sensitive going from South to East catchment than it is going from East to South because of the specific combinations of $m$ and $T_o$ for the two catchments.

*ii) Model insensitivity to topographic index*

An important point demonstrated in level 2 testing is that, in this study, the use of a topographic index by the quasi-distributed model does not provide superior geographic transposability. In a further analysis of these results, I ran the split-sample test on each catchment using the topographic index curve of the other catchment. The resulting Nash coefficients were only slightly different than using the proper catchment index: in East catchment, the Nash value went from 0.890 to 0.883 and in South catchment the Nash value went from 0.890 to 0.895. The frequency curves also have similar shapes.

Other researchers have also found that the representation of topography does not provide a better prediction of observed events when transposed from catchment to catchment. Franchini *et al.* (1996) found TOPMODEL to be insensitive to the index curve, replacing the index with various different curves did not significantly alter the sequence of discharges. Quinn *et al.* (1991) concluded that if the frequency distributions of the topographic index have roughly the same shape, then interchangeability is possible while maintaining good hydrograph prediction. Any change in the predicted hydrographs resulting from a change in the topographic index curve is minimized by the optimization of other parameters. Iorgulescu and Jordan (1994) determined that model results combined with field investigations suggest that topography is relevant but not sufficient to override soil and geological factors in determining saturated areas.

*iii) no difference in runoff mechanisms between catchments*

Iorgulescu and Jordan (1994) found that TOPMODEL performed differently in two catchments. Iorgulescu and Jordan calculated the amount of subsurface and overland flow and determined that the runoff mechanisms were different in the two catchments used in their study. In the present study, similar percentages of subsurface (95-97%) and overland flow (3-5%) were calculated for both East and South catchments. Therefore, the runoff generating mechanisms computed by the quasi-distributed model in this study were found to be the same in both catchments and is not a factor in the differing model performance results of level 2. This finding is similar to that of Durand *et al.*(1992).

The triple reservoir model performed worse than the dual reservoir model when transposed to another catchment. This may be because the triple reservoir model has more parameters and therefore is more catchment and data specific. The dual reservoir model is more generalized so even if it did not calibrate as well, when transposed to another catchment it is not as sensitive to changes.

### *5.2.3 Level 3 - Differential split-sample test*

The statistical analysis shows that TOPMODEL is statistically superior to the lumped models when validated on events larger than the calibration set. This result agrees with the finding of Beven *et al.* (1984) that the quasi-distributed model performed significantly better on large events than small events. In contrast, the lumped models' performance worsened as the events increased in size.

The quasi-distributed model may have performed best at simulating high flow events as a result of being the only model accounting for the saturation overland flow mechanism. The saturated areas mechanism may provide a better representation of the dynamics involved when there is an increased water input to the system. Some studies have concluded that accounting for saturation overland flow is important in determining peak flows (Band *et al.*, 1993; Iorgulescu and Jordan, 1994). However, in this study, the fact that TOPMODEL accounts for saturation overland flow and the other models do not, does not seem to be important or relevant. TOPMODEL was not superior at level 2 testing, the proxy-basin test, despite its accounting for topographic effects on saturated source area dynamics. Also, TOPMODEL's predictions indicated that only a few percent of the stormflow originated as direct precipitation onto saturated areas. These points may indicate that for East and South catchments, simulating the routing of throughflow to the stream channels is probably more critical than trying to model the dynamics of the saturation overland flow source areas.

The nonlinearity of catchment response by TOPMODEL, discussed in Section 5.1.2, may be the prime reason allowing for the extrapolation to larger events better than the linear reservoirs of the lumped models. The nonlinear routing of TOPMODEL is more sensitive to precipitation than the linear reservoirs of the lumped models and therefore is better able to respond to an increase in rainfall input.

*5.2.4 Level 4 - Proxy-basin differential split-sample test*

For level 4, testing for both geographic and climatic transposability, the quasi-distributed model performed best in South catchment and worst in East catchment. This is the opposite result of level 2 testing and may indicate that although catchment type is important (as indicated in level 2), it is overridden by other factors, one of which may be the use of larger events for validation.

The ANOVA results of Table 4.17 indicate that there is an interaction effect occurring between model and basin at level 4. The values of least square means generated by the analysis of variance for level 4 indicate that the two lumped models are not affected by catchment whereas the quasi-distributed model is. This interaction effect suggests that the quasi-distributed model responds differently to different catchments and the lumped models do not (reinforced by the results of level 2). For East catchment, there is no statistically significant difference in performance amongst models. However, for

South catchment, TOPMODEL provides significantly superior predictions than the other two lumped models.

## 5.3 Statistical approach

The combination of the Jackknife method and ANOVA provides an important tool to determine the significance of model performance statistics. However, two points should be considered. First, there is the issue of significance level. Using a significance level of 0.05 is conventional but arbitrary. Since this is an exploratory study with a small sample size, a more stringent significance level would not be appropriate. The provision of the probability values (Table 4.16) does allow a user to draw inferences using alternative significance levels.

The second issue is that the ANOVA specifies the 'error variance' in the RMSE to be caused by sampling variation over the entire validation data set. An alternative design would be to treat 'event' as another effect, creating a 3-way interaction ANOVA. However, a 3-way ANOVA design would be difficult to interpret statistically. There also may be problems with the distribution of residuals (as found by Cavadias and Morin, 1985). Finally, most modellers look at model performance statistics which are aggregated over multiple events. Hence, the design in the present case conforms better to current practice in assessing model performance.

## 5.4 Generality of results

The main emphasis of this thesis was to emphasize the importance of the testing framework as opposed to the performance of the specific models used. All model comparison studies are limited in their ability to make generalized statements based on data and site specifics of the individual study. Generalizations of model performance regarding the models used in this study can only be made when based on many studies, in particular those studies which are implementing a similar framework. The results of this thesis research and the discussion are conclusive only for the specific catchments and range of conditions represented within the datasets used and cannot be extrapolated to other situations with confidence. With this caution regarding generality in mind, conclusions may be derived from the study.

## 5.5 Conclusions

The results show that the lumped models performed as well as the quasi-distributed model in similar climatic conditions in the same catchment and also when geographically transposed to a proxy catchment. However, the quasi-distributed model performs statistically significantly better than the lumped models when predicting large events outside of the range for which the model was calibrated. In conditions where models are geographically transposed and there is a significant increase in precipitation input, the quasi-distributed model performed significantly better in South catchment but not in East catchment. The conclusion is that TOPMODEL is likely to be no worse than the lumped models under various conditions and may be superior for some catchments, but is definitively better when predicting large events.

Since these are operational tests, it is not as important to understand why one model is better than another, just that one model provides superior predictions of the storm hydrographs. It is up to the user to decide at this point if superior performance warrants the additional cost of generating and analyzing DEMs.

This chapter has provided some explanations of the results of this study. The next and final chapter will summarize the objectives and findings of the research and provide suggestions for further research.

# CHAPTER 6

# CONCLUSIONS

This final chapter presents a summary of the findings and discusses the significance of the results to hydrologic modelling. The chapter concludes with suggestions for further research with respect to variations of the model testing framework and model structure.

## 6.1 Summary of main findings

The research had two main objectives: (1) to compare a quasi-distributed model to two lumped models to determine if there is a benefit associated with the increased demand for catchment data and (2) to determine if the statistical significance of differences in model performance can be quantified. These two objectives were successfully answered with the understanding that the primary intent of the research was to focus on the importance of the testing framework and that this is not a definitive test of the specific models used.

Using a significance level of 0.05, the statistical analysis shows that at levels 1 and 2 there are no statistical differences in model performance. This finding is meaningful in that it confirms there is no significant benefit in applying the more complex, quasi-distributed model and that the simpler lumped models would provide acceptably similar simulations under those conditions.

At level 3, however, the quasi-distributed model performs statistically significantly better than both lumped models in both catchments. The statistical analysis provides justification for using an advanced type of model to represent flows following a significant increase of rainfall. The statistical analysis for level 4 indicates that the quasi-distributed model performed significantly better than the lumped models at South catchment but not at East catchment. The conclusion is that the quasi-distributed model is no worse than the other two models but may perform better in certain catchments.

This research demonstrates that the ANOVA design including the Jackknife method is a workable method and could be a valuable tool for assessing statistical significance of differences in model performance. The statistical approach provides power and meaning to results to model comparison studies.

### 6.2 Significance of results to hydrologic modelling

This thesis research contributes to the state of hydrologic modelling by advocating the use of a more rigorous and standard testing framework in hydrologic modelling. A standard method of model testing and comparison will raise the level of credibility of comparisons studies and discourage exaggerated claims of model performance. The use of statistical analysis provides more definitive results of model comparison studies, minimizing any conclusions of relative model performance that may be misleading. The standard testing framework in combination with statistical analysis will provide superior information on model performance, allowing better decisions to be made with respect to operational modelling.

More specifically, the research provides further information on the relative performance of a quasi-distributed model to lumped models in forested catchments. The study also provides information on the response of the three types of models to the various levels of testing. The results, indicating which model types perform better under different conditions, give important preliminary information for further model studies.

## 6.3 Suggestions for future research

### *6.3.1 Extension of the model testing framework*

Future studies should include the evaluation of models at all four levels of testing, as opposed to just the first or second level where many researchers stop, especially when evaluating new models against existing models. Future research should extend the hierarchical testing approach with statistical testing to more than just two models, providing a better relative comparison. These models should be of varying complexity and type so that more information is collected regarding where future model development efforts could be concentrated.

In addition to increasing the number of models tested and the number of tests used; an increase in the number of catchments used in the testing would provide an improvement to model comparison studies. Testing the models on more than two catchments, if possible, would be beneficial. With regard to operational modelling, more rigorous analysis should be done on larger watersheds since most engineering hydrology decisions are on larger catchments.

The manner in which catchments are dealt with in future comparison studies can also be altered. A limitation of this study was the treatment of catchments as a fixed effect in the analysis of variance, resulting in the inability to make inferences beyond the two catchments used. More general inferences could be made regarding model performance that extend beyond the catchments used in the study if catchment was treated as a random factor in the analysis of variance.

### *6.3.2 Modifications to model structure*

This research has shown that only the quasi-distributed model, TOPMODEL, has provided statistically significantly better performance for certain conditions of the study. There has been much optimism for the potential of distributed models (Beven, 1992; Refsgaard, 1997). The approach which led to TOPMODEL is one of the most promising directions in modelling research and it deserves special consideration and effort (Iorgulescu and Jordan, 1994).

One improvement that may be made to TOPMODEL that might provide better predictions in future research is the modification of the topographic index. A different topographic index, other than that presently used in TOPMODEL, may be more suitable to a particular catchment and provide improved performance. For example, Iorgulescu and Jordan (1994) and Ambroise *et al.* (1996) found that different runoff mechanisms in two catchments required a different approach where a different topographic index function may be preferred. Woods *et al.* (1997) are starting research into a topographic index that also models the spatial variability of subsurface runoff. With increased computing power and advancements in remote sensing, digital elevation models may become easier to obtain and of better accuracy, resulting in improved topographic index distribution functions.

An increase in the quality and the quantity of calibration data will allow better predictions for all rainfall-runoff models. These would include improved streamflow and precipitation data collection. Runoff simulations are unlikely to improve until rainfall input estimations improve (Michaud and Sorooshian, 1994).

## REFERENCES


Abbot, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E., and Rasmussen, J. (1986) An introduction to the European Hydrological System - System Hydrologique Européen, SHE, 2, Structure of a physically-based, distributed modelling system. *Journal of Hydrology*, 87: 61-77.

Ambroise, B., Perrin, J., and Reutenauer, D. (1995) Multicriterion validation of a semidistributed conceptual model of the water cycle in the Fecht Catchment (Vosges Massif, France). *Water Resources Research*, 31(6): 1467-1481.

Ambroise, B., Beven, K., and Freer, J. (1996) Toward a generalization of the TOPMODEL concepts: Topographic indices of hydrological similarity. *Water Resources Research*, 32(7): 2135-2145.

Band, L.E., Patterson, P. Nemani, R., and Running, S.W. (1993) Forest ecosystem processes at the watershed scale: incorporating hillslope hydrology. *Agricultural and Forest Meteorology*, 63: 93-126.

Bergstrom, S. (1979) Spring flood forecasting by conceptual models in Sweden. In: *Proceedings of a meeting of snow cover runoff.* Colbeck, S.C. and Ray, M. (eds.) U.S. Army CRREL, Hanover, New Hampshire. pp. 397-405.

Bergstrom, S. (1991) Principles and confidence in hydrological modelling. *Nordic Hydrology*, 22:123-136.

Bergstrom, S. and Forsman, A. (1973) Development of a conceptual deterministic rainfall runoff model. *Nordic Hydrology*, 4(3).

Bernier, P.Y. (1985) Variable source areas and stormflow generation: an update of the concept and a simulation effort. *Journal of Hydrology*, 79: 195-213.

Beven, K.J. (1989) Changing ideas in hydrology - the case of physically-based models. *Journal of Hydrology*, 105: 157-172.

Beven, K.J. (1992) Editorial: Future of distributed modelling. *Hydrological Processes*, 6(3): 253-254.

Beven, K.J. (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16: 41-51.

Beven, K.J. and Binley, A. (1992) The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3): 279-298.

Beven, K.J. and Kirkby, M.J. (1979) A physically based, variable contributing area model of basin hydrology. *Hydrological Science Bulletin*, 24(1): 43-69.

Beven, K.J., Kirkby, M.J., Schofield, N., and Tagg, A.F. (1984) Testing a physically-based flood forecasting model (TOPMODEL) for three U.K. catchments. *Journal of Hydrology*, 69: 119-143.

Beven, K.J., Calver, A., and Morris, E.M. (1987) The Institute of Hydrology distributed model. *Institute of Hydrology Report No. 81*, Institute of Hydrology, Wallingford.

Burges, S.J., and Lettenmaier, D.P. (1977) Comparison of annual streamflow models. *J. Hydraul. Div. Am. Soc. Civ. Eng.*, 103: 991-1006.

Calver, A. (1988) Calibration, sensitivity and validation of a physically-based rainfall-runoff model. *Journal of Hydrology*, 103: 103-115.

Cavadias, G. and Morin, G. (1985) Approximate confidence intervals for numerical verification used in hydrologic models. INRS- Eau, Rapport Scientifique no. 166.

Cheng, J.D. (1988) Subsurface stormflows in the highly permeable forested watersheds of southwestern British Columbia. *Journal of Contaminant Hydrology*, 3: 171-191.

Chiew, F.H. and McMahon, T.A. (1994) Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. *Journal of Hydrology*, 153: 383-416.

Chiew, F.H., Stewardson, M.J., and McMahon, T.A. (1993) Comparison of six rainfall-runoff modelling approaches. *Journal of Hydrology*, 147:1-36.

Clarke, R.T. (1973) A review of some mathematical models used in hydrology, with observations on their calibration and use. *Journal of Hydrology*, 19:1-20.

Durand, P., Robson, A., and Neal, C., (1992) Modelling the hydrology of submediterranean montane catchments (Mont-Lozere, France) using TOPMODEL. *Journal of Hydrology*, 139: 1-14.

Fedora, M.A. and Beschta, R.L. (1989) Storm runoff simulation using an antecedent precipitation index (API) model. *Journal of Hydrology*, 112: 121-133.

Feller, M.C. (1988) The Haney watershed study - British Columbia. Canadian Hydrology Symposium (CHS : 88) May, 1988, Banff, Alberta, National Research Council of Canada. pp. 217-227.

Franchini, M., and Pacciani, M. (1991) Comparative analysis of several conceptual rainfall-runoff models. *Journal of Hydrology*, 122: 161-219.

Franchini, M., Wendling, J., Obled, C., and Todini, E. (1996) Physical interpretation and sensitivity analysis of the TOPMODEL. *Journal of Hydrology*, 175: 293-338.

Freeze, R.A. (1980) A stochastic-conceptual analysis of rainfall-runoff processes on a hillslope. *Water Resources Research*, 8: 1272-1283.

Gan, T.Y. and Biftu, G.F. (1996) Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure. *Water Resources Research,* 32: 3513-3524.

Gan, T.Y. and Burges, S.J. (1990) An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments. 1. Models, model properties, and experimental design. *Water Resources Research*, 26(7): 1595-1604.

Gash, J.H.C. (1979) An analytical model of rainfall interception by forest. *Quartely Journal of the Royal Meteorological Society*, 105: 43-55.

Gash, J.H.C. and Morton, A.J. (1978) An application of Rutter's model to the estimation of the interception loss from Thetford Forest. *Journal of Hydrology*, 38: 49-58.

Gash, J.H.C., Wright, I.R., and Lloyd, C.R. (1980) Comparative estimates of interception loss from three coniferous forests in Great Britain. *Journal of Hydrology*, 48: 89-105.

Grayson, R.B., Moore, I.D., and McMahon, T.A. (1992) Physically based hydrologic modelling: I. A terrain based model for investigative purposes. *Water Resources Research*, 26(10): 2639-2658.

Grayson, R.B., Bloschl, G., Barling, R.D., and Moore, I.D. (1993) Process, scale and constraints to hydrological modelling in GIS. In: *HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources,* Vienna Conference, April 1993. IAHS Publ. no. 211: 83-92.

Harr, R.D. (1977) Water flux in soil and subsoil on a steep forested slope. *Journal of Hydrology*, 33: 37-58.

Hetherington, E.D. (1976) Investigation of orographic rainfall in south coastal mountains of British Columbia. Ph.D. Thesis, University of British Columbia, Vancouver, Canada.

Hewlett, J.D. and Hibbert, A.R. (1967) Factors affecting the response of small watersheds to precipitation in humid areas. In: *Forest Hydrology*. W.E. Sopper and H.W. Lull (eds), Pergamon, New York.

Hornberger, G.M., Beven, K.J., Cosby, B.J., and Sappington, D.E. (1985) Shenandoah Watershed Study: Calibration of a topography-based, variable contributing area hydrological model to a small forested catchment. *Water Resources Research*, 21(12): 1841-1850.

Hughes, D.A. (1984) An isolated event model based upon direct runoff calculations using an implicit source area concept. *Hydrological Sciences Journal*, 29(3): 311-325.

Hughes, D.A. and Beater, A.B. (1989) The applicability of two single event models to catchments with different physical characteristics. *Hydrological Sciences Journal*, 34(1): 63-78.

Hutchinson, I. and Roberts M.C. (1981) Vertical variation in stemflow generation. *Journal of Applied Ecology*, 18: 521-527.

Hutjes, R.W., Wierda, A., and Veen, A.L. (1990) Rainfall interception in the Tai forest, Ivory Coast: Application of two simulatioin models to a humid tropical system. *Journal of Hydrology*, 114: 259-275.

Iorgulescu, I. and Jordan, J. (1994) Validation of TOPMODEL on a small Swiss catchment. *Journal of Hydrology*, 159: 255-273.

Jakeman, A.J. and Hornberger, G.M. (1993) How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, 29(8): 2637-3649.

Jensen, K.H. and Mantoglou, A. (1992) Future of distributed modelling. *Hydrological Processes*, 6(3): 255-264.

Jordan, J.-P. (1994) Spatial and temporal variability of stormflow generation processes on a Swiss catchment. *Journal of Hydrology*, 153: 357-382.

Karnieli, A.M., Diskin, M.H., and Lane, L.J. (1994) CELMOD5 - a semi-distributed cell model for conversion of rainfall into runoff in semi-arid watersheds. *Journal of Hydrology*, 157: 61-85.

Klemes, V. (1986) Operational testing of hydrologic simulation models. *Hydrological Sciences Journal*, 31(1): 13-23.

Klinka, K. and Krajina, V.J. (1986) Ecosystems of the University of British Columbia Research Forest, Haney, B.C., The U.B.C. Faculty of Forestry.

Kundzewicz, Z.W. (1986) The hydrology of tomorrow. *Hydrological Sciences Journal*, 31(2): 223-235.

Lee, R. (1980) *Forest Hydrology*. Columbia University Press, New York.

Leyton, L., Reynolds, E.R.,and Thompson, F.B. (1967) Rainfall interception in forest and moorland. In: *Forest Hydrology*, W.E. Sopper and H.W. Hull, eds. Pegamon Press, Oxford.

Loague, K.M. and Freeze, R.A. (1985) A comparison of rainfall-runoff modeling techniques on small upland catchments. *Water Resources Research*, 21(2): 229-248.

Loukas, A. and Quick, M. (1996) Spatial and temporal distribution of storm precipitation in southwestern British Columbia. *Journal of Hydrology*, 174: 37-56.

Loustau, D., Berbigier, P. and Granier, A. (1992) Interception loss, throughfall and stemflow in a maritime pine stand. II An application of Gash's analytical model of interception. *Journal of Hydrology*, 38: 469-485.

Michaud, J. and Sorooshian, S. (1994) Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. *Water Resources Research*, 30(3): 593-605.

Miller, R.G. (1974) The jackknife - a review. *Biometrika* , 61(1): 1-15.

Moore, I.D., Panuska, J.C., Grayson, R.B., and Srivastava, K.P. (1988) Application of digital topographic modeling in hydrology. *American Society of Agricultural Engineers*, ASAE Publ. no. 07-88: 447-461.

Moore, R.D., and Thompson, J.C. (1996) Are water table variations in a shallow forest soil consistent with the TOPMODEL concept? *Water Resources Research*, 32(3): 663-669.

Moore, R.D. (1997) Storage-outflow modelling of streamflow recessions, with application to a shallow-soil forested catchment. *Journal of Hydrology*, 198: 260-270.

Naef, F. (1981) Can we model the rainfall-runoff process today? *Hydrological Sciences Bulletin*, 26(3): 281-289.

Nash, J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models. *Journal of Hydrology*, 10: 282-290.

Neter, J., Wasserman, W., and Whitmore, G. (1982) *Applied statistics*, 2nd edition. Allyn and Bacon, Inc., New York.

O'Loughlin, E.M. (1986) Prediction of surface saturation zones in natural catchments by topographic analysis. *Water Resources Research*, 22(5):794-804.

Patric, J.H. (1966) Rainfall interception by mature coniferous forests of southeast Alaska. *Journal of Soil and Water Conservation*, 21: 229-331.

Pearce, A.J. and Rowe, L.K. (1981) Rainfall interception in a multi-storied, evergreen mixed forest: estimates using Gash's analytical model. *Journal of Hydrology*, 49: 341-353.

Pierson, T.C. (1980) Piezometric response to rainstorms in forested hillslope drainage depressions. *Journal of Hydrology (New Zealand)*, 19: 1-10.

Potvin, C., and Roff, D. (1993) Distribution-free and robust statistical methods: viable alternatives to parametric statistics? *Ecology*, 74(6): 1617-1628.

Power, C.A. (1984) A management information system (MIS) for the UBC Research Forest. B.S.F. Thesis, University of British Columbia, Vancouver, Canada.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1986) *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Quick, M.C. and Pipes, A. (1972) Daily and seasonal runoff forecasting with a water budget model. In: *Proceedings of the International Symposia on the Role of Snow and Ice in Hydrology*, Banff, Canada. World Meteorological Organization.

Quinn, P., Beven, K., Chevallier, P., and Planchon, O. (1991) The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes*, 5: 59-79.

Refsgaard, J.C. (1994) Model and data requirements for simulation of runoff and land surface processes in relation to global circulation models. *Global Environmental Change and Land Surface Processes in Hydrology: The Trials and Tribulations of Modeling and Measuring.* Springer-Verlag, New York, in press, cited in Michaud and Sorooshian, 1994.

Refsgaard, J.C. (1997) Parameterization, calibration and validation of distributed hydrological models. *Journal of Hydrology*, 198: 69-97.

Refsgaard, J.C. and Knudsen, J. (1996) Operational validation and intercomparison of different types of hydrological models. *Water Resources Research* 32(7): 2189-2202.

Romanowicz, R., Beven, K., Freer, J., and Moore, R. (1993) TOPMODEL as an application module with WIS. In: *HydroGIS 93: Application of Geographic Information Systems in Hydrology and Water Resources*, Vienna Conference, April 1993. IAHS Publ. no. 211: 211-223.

Rothacher, J. (1963) Net precipitation under a Douglas-fir forest. *Forest Science*, 9: 423-429.

Rykiel, E.J. (1996) Testing ecological models: the meaning of validation. *Ecological Modelling*, 90: 229-244.

Servat, E. and Dezetter, A. (1991) Selection of calibration objective functions in the context of rainfall-runoff modelling in a Sudanese savannah area. *Hydrological Sciences Journal*, 36(4): 307-330.

Singh, K.P. (1976) Unit hydrographs - a comparative study. *Water Resources Bulletin*, 12(2): 381-392.

Singh, K.P. and Woolhiser, D.A. (1976) Sensitivity of linear and nonlinear surface runoff models to input errors. *Journal of Hydrology*, 29: 243-249.

Sittner, W.T. (1976) WMO project on intercomparison of conceptual models used in hydrological forecasting. *Hydrological Sciences Bulletin*, 21(3): 203-213.

Smith, R.E. and Hebbert, R.H. (1979) A Monte-Carlo analysis of the hydrologic effects of spatial variability of infiltration. *Water Resources Research*, 15: 419-429.

Svensson, S. (1977) A statistical study for automatic calibration of a conceptual runoff model. Swedish Meteorological and Hydrological Institute, report no. RHO 10.

Thompson, J.C. (1994) Behavior and prediction of water table levels in shallow forest soils, southwestern British Columbia. M.Sc. Thesis, Simon Fraser University, Burnaby, Canada.

Tischer, E. (1986) Hydrological behavior of a forested mountain slope. M.Sc. thesis, University of British Columbia, Vancouver, Canada.

Utting, M.G. (1979) The generation of stormflow on a glaciated hillslope in coastal British Columbia. M.Sc. thesis, University of British Columbia, Vancouver, Canada.

Van Genuchten, M.T. (1991) Progress and opportunities in hydrologic research, 1987-1990, *Reviews in Geophyics*, 25: 187-192.

Weeks, W.D. and Hebbert, R.H.B. (1980) A comparison of rainfall-runoff models. *Nordic Hydrology*, 11: 7-24.

Wilcox, B.P., Rawls, W.J., Brakensiek, D.L., and Wight, J.R. (1990) Predicting runoff from rangeland catchments: a comparison of two models. *Water Resources Research*, 26(10): 2401-2410.

Willmott, C.J. (1984) On the evaluation of model performance in physical geography. In: *Spatial Statistics and Models*, G.L. Gaile and C.J. Willmott (eds.). D. Reidel Publ. Co., Dordrecht, Holland.

Wolock, D.M, and McCabe,G.J. (1995) Comparison of single and multiple direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research*, 31(5): 1315- 1331.

Woods, R.A., Sivapalan, M. ,and Robinson~~, J.S.~~ (1997) Modeling the spatial variability of subsurface runoff using a topographic index. *Water Resources Research*, 33(5): 1061-1073.

World Meteorological Organization (1975) Intercomparison of conceptual models used in operational hydrological forecasting, Operational Hydrology Report No. 7, Geneva.

World Meteorological Organization (1986) Intercomparison of models of snowmelt runoff, Operational Hydrology Report No. 23, Geneva.

Zhang, W. and Montgomery, D. (1994) Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research*, 30(4): 1019-1028.

# APPENDIX 1

# RAINFALL - RUNOFF EVENTS

**Calibration and validation events for testing levels 1 and 2.**

***calibration events***

| | start | | | | end | | | |
|---|---|---|---|---|---|---|---|---|
| **event** | **yr** | **mn** | **day** | **hr** | **yr** | **mn** | **day** | **hr** |
| ***1*** | 85 | 10 | 12 | 16 | 85 | 10 | 15 | 17 |
| ***4*** | 85 | 10 | 25 | 23 | 85 | 10 | 29 | 23 |
| ***6*** | 86 | 4 | 18 | 22 | 86 | 4 | 23 | 22 |
| ***10*** | 86 | 12 | 13 | 7 | 86 | 12 | 18 | 24 |
| ***11*** | 86 | 12 | 20 | 14 | 86 | 12 | 23 | 9 |
| ***12*** | 86 | 12 | 23 | 7 | 86 | 12 | 25 | 24 |
| ***14*** | 87 | 1 | 9 | 24 | 87 | 1 | 11 | 16 |
| ***21*** | 87 | 12 | 8 | 22 | 87 | 12 | 12 | 12 |
| ***22*** | 88 | 2 | 13 | 22 | 88 | 2 | 17 | 3 |
| ***23*** | 88 | 3 | 7 | 18 | 88 | 3 | 15 | 12 |
| ***24*** | 88 | 4 | 1 | 1 | 88 | 4 | 5 | 4 |
| ***26*** | 89 | 3 | 24 | 12 | 89 | 3 | 30 | 12 |
| ***35*** | 90 | 11 | 3 | 18 | 90 | 11 | 6 | 20 |
| ***37*** | 90 | 11 | 8 | 10 | 90 | 11 | 12 | 14 |
| ***38*** | 90 | 12 | 7 | 22 | 90 | 12 | 12 | 12 |
| ***42*** | 91 | 11 | 10 | 20 | 91 | 11 | 12 | 10 |
| ***43*** | 91 | 11 | 18 | 16 | 91 | 11 | 22 | 24 |
| ***44*** | 91 | 11 | 23 | 10 | 91 | 11 | 27 | 24 |
| ***45*** | 91 | 12 | 8 | 18 | 91 | 12 | 10 | 12 |
| ***46*** | 91 | 12 | 10 | 9 | 91 | 12 | 15 | 24 |
| ***47*** | 92 | 1 | 1 | 10 | 92 | 1 | 3 | 21 |
| ***51*** | 92 | 2 | 17 | 4 | 92 | 2 | 20 | 6 |
| ***52*** | 92 | 2 | 21 | 1 | 92 | 2 | 26 | 24 |
| ***58*** | 92 | 11 | 18 | 14 | 92 | 11 | 20 | 22 |
| ***59*** | 92 | 11 | 20 | 17 | 92 | 11 | 25 | 24 |
| ***63*** | 92 | 12 | 13 | 1 | 92 | 12 | 17 | 24 |

***validation events***

| | start | | | | end | | | |
|---|---|---|---|---|---|---|---|---|
| **event** | **yr** | **mn** | **day** | **hr** | **yr** | **mn** | **day** | **hr** |
| ***2*** | 85 | 10 | 15 | 12 | 85 | 10 | 17 | 1 |
| ***3*** | 85 | 10 | 16 | 21 | 85 | 10 | 18 | 16 |
| ***5*** | 85 | 10 | 31 | 1 | 85 | 11 | 2 | 10 |
| ***9*** | 86 | 11 | 19 | 10 | 86 | 11 | 21 | 8 |
| ***13*** | 86 | 12 | 28 | 6 | 86 | 12 | 31 | 20 |
| ***15*** | 87 | 1 | 11 | 13 | 87 | 1 | 14 | 24 |
| ***16*** | 87 | 4 | 9 | 22 | 87 | 4 | 13 | 12 |
| ***18*** | 87 | 11 | 15 | 5 | 87 | 11 | 18 | 7 |
| ***19*** | 87 | 11 | 20 | 21 | 87 | 11 | 23 | 12 |
| ***20*** | 87 | 12 | 6 | 4 | 87 | 12 | 8 | 24 |
| ***25*** | 89 | 2 | 21 | 10 | 89 | 2 | 27 | 8 |
| ***30*** | 89 | 11 | 2 | 15 | 89 | 11 | 5 | 23 |
| ***31*** | 89 | 11 | 5 | 23 | 89 | 11 | 7 | 22 |
| ***32*** | 89 | 11 | 17 | 24 | 89 | 11 | 22 | 24 |
| ***33*** | 89 | 12 | 1 | 21 | 89 | 12 | 5 | 14 |
| ***34*** | 90 | 1 | 8 | 20 | 90 | 1 | 11 | 11 |
| ***36*** | 90 | 11 | 6 | 16 | 90 | 11 | 8 | 14 |
| ***39*** | 91 | 2 | 18 | 10 | 91 | 2 | 23 | 24 |
| ***41*** | 91 | 11 | 7 | 4 | 91 | 11 | 9 | 10 |
| ***48*** | 92 | 1 | 8 | 20 | 92 | 1 | 15 | 10 |
| ***49*** | 92 | 1 | 15 | 11 | 92 | 1 | 20 | 12 |
| ***54*** | 92 | 4 | 28 | 2 | 92 | 5 | 5 | 24 |
| ***56*** | 92 | 11 | 5 | 24 | 92 | 11 | 10 | 11 |
| ***57*** | 92 | 11 | 10 | 8 | 92 | 11 | 15 | 10 |
| ***60*** | 92 | 11 | 26 | 20 | 92 | 11 | 29 | 14 |
| ***61*** | 92 | 11 | 29 | 10 | 92 | 12 | 5 | 22 |

**Calibration and validation events for testing levels 3 and 4, ranked by volume.**

**Calibration events (small volume events)**

| | **South catchment** | | **East catchment** | |
|---|---|---|---|---|
| *ranking* | *event* | *volume* | *event* | *volume* |
| *1* | 35 | 823.2 | 14 | 2276.9 |
| *2* | 14 | 842.0 | 51 | 2369.7 |
| *3* | 41 | 969.0 | 35 | 2484.0 |
| *4* | 16 | 1060.7 | 36 | 2488.9 |
| *5* | 57 | 1064.9 | 41 | 2550.0 |
| *6* | 51 | 1176.5 | 58 | 2807.2 |
| *7* | 18 | 1229.9 | 18 | 3048.6 |
| *8* | 63 | 1273.8 | 31 | 3049.8 |
| *9* | 36 | 1306.4 | 16 | 3077.5 |
| *10* | 31 | 1456.0 | 63 | 3182.9 |
| *11* | 58 | 1501.7 | 57 | 3388.7 |
| *12* | 19 | 1537.4 | 47 | 3455.5 |
| *13* | 49 | 1607.6 | 19 | 3763.6 |
| *14* | 47 | 1751.0 | 2 | 4110.2 |
| *15* | 2 | 1825.2 | 49 | 4348.8 |
| *16* | 10 | 1879.0 | 45 | 4417.0 |
| *17* | 3 | 1969.3 | 34 | 4817.9 |
| *18* | 44 | 2099.2 | 44 | 4925.4 |
| *19* | 23 | 2270.5 | 3 | 5147.8 |
| *20* | 6 | 2275.2 | 10 | 5336.7 |
| *21* | 45 | 2313.9 | 23 | 5585.5 |
| *22* | 59 | 2322.4 | 59 | 5591.2 |
| *23* | 13 | 2466.2 | 20 | 5637.9 |

* *middle 6 events deleted* *

| | | | |
|---|---|---|---|
| 12 | 2532.3 | 1 | 5835.2 |
| 1 | 2594.2 | 6 | 5879.9 |
| 22 | 2726.1 | 42 | 6070.4 |
| 25 | 2737.8 | 25 | 6083.4 |
| 34 | 2786.5 | 13 | 6257.9 |
| 20 | 2871.8 | 60 | 6363.6 |

******

**Validation events (large volume events)**

| | South catchment | | East catchment | |
|---|---|---|---|---|
| *ranking* | *event* | *volume* | *event* | *volume* |
| *1* | 15 | 3004.9 | 12 | 6440.5 |
| *2* | 24 | 3104.4 | 24 | 6629.0 |
| *3* | 56 | 3116.2 | 11 | 6772.6 |
| *4* | 60 | 3272.9 | 22 | 7191.7 |
| *5* | 42 | 3332.6 | 56 | 7326.5 |
| *6* | 11 | 3351.8 | 39 | 7349.1 |
| *7* | 39 | 3400.4 | 30 | 7454.0 |
| *8* | 21 | 3443.7 | 33 | 7549.8 |
| *9* | 48 | 3736.6 | 21 | 7906.7 |
| *10* | 54 | 3768.7 | 15 | 8071.0 |
| *11* | 61 | 3841.3 | 38 | 8285.1 |
| *12* | 30 | 3844.6 | 61 | 8370.7 |
| *13* | 46 | 3999.2 | 32 | 8640.4 |
| *14* | 52 | 4035.3 | 26 | 8658.1 |
| *15* | 33 | 4064.2 | 43 | 9155.4 |
| *16* | 26 | 4119.0 | 48 | 9201.7 |
| *17* | 4 | 4577.6 | 54 | 9418.9 |
| *18* | 43 | 4607.0 | 4 | 9423.0 |
| *19* | 38 | 4615.1 | 52 | 9473.8 |
| *20* | 32 | 4903.8 | 46 | 9616.1 |
| *21* | 9 | 5208.2 | 5 | 10736.4 |
| *22* | 5 | 5736.9 | 9 | 12206.1 |
| *23* | 37 | 14620.7 | 37 | 27687.7 |

# APPENDIX 2

# COMPUTER PROGRAM CODES

## Canopy storage model code

```
PROGRAM THRUFALL;

{ Canopy Storage Model          South catchment - Pascal code}

{  ------ constants ------     }

CONST
        gap  = 0.159;                 {free throughfall coefficient}
        Pt   = 0.02;                  {stemflow coefficient}
        Sc   = 3.30;                  {maximum canopy storage capacity mm}
        Ebar = 0.228;                 {mean evaporation rate mm/h}
        mult = 0.94;                  {mult x Pg to extrapolate to East}

{  ------ variables ------     }

VAR
        Pg              : Real;       {rainfall input  mm/h}
        Pgstep          : Real;       {rainfall per step}
        SF              : Real;       {stemflow 0.1 mm/h}
        TF              : Real;       {throughfall 0.1 mm/h}
        E               : Real;       {evaporation from canopy mm/10 min}
        Si              : Real;       {initial canopy storage  mm}
        S               : Real;       {actual canopy storage  mm}
        event           : Array [0..2000] of Integer;
        hr, i           : Integer;    {event number and hour of storm, counter}
        Infile, Outfile : Text;       {input, output files}


{  ------ start program  ------     }

Begin

  Assign (Infile, 'C:\1thesis\tf\ppt2tf.in');   Reset (Infile);
  Assign (Outfile, 'C:\1thesis\tf\ppt2tf.out');   Append (Outfile);

 event[0] := 0;
 hr := 1;

        While not eof(infile) Do
          begin
                Readln(infile, event[hr], Pg);
                Pg := Pg*mult{/10.0};

                If (event[hr] > event[hr-1]) then
                  Si :=0.0;
```

```
        end;

  { --- solve canopy water balance in 10-min time steps --- }

      Pgstep := Pg/6.0;
      TF := 0.0;                           {throughfall}
      SF := 0.0;                           {stemflow}

      For i := 1 to 6 Do                           {start 10 min step loop}
        begin
          E := (Ebar/6.0)*Si/Sc;           {evap a function of amt of saturation}
          S := Si + Pgstep - E;
              If (S < 0) then S := 0;
              If (S > Sc) then
               begin                       {TF and SF occurs when S > Sc}
                   SF := SF + Pt*Pgstep;
                   TF := TF + S - Sc;
                   S := Sc;
               end;
              Si := S;
         end;                              {end 10 min step loop}

          TF := TF*(1.0 - gap) + Pg*gap {+ SF};    {TF = drip, SF and thru gaps}

          Writeln(outfile, event[hr], Pg:5:3, TF:5:3);

          hr := hr + 1;

End;

Writeln('program is done.');

close (infile);
close (outfile);

End.   {end of program}
```

## Black-box lumped model code

```
PROGRAM DUALPAR;

{ Dual Reservoir Lumped Model        East catchment  -  Pascal code  }

USES CRT;
CONST        K1 = .0098/3600;        Area = 383000;

VAR   Infile, outfile, objfile, eventstats          : text;
      event, hr, Qalln                  : integer;
      Qobs, Qsim, TF, Qobssum           : real;
      f1, k1,K2, Q1, Q2                 : real;

TYPE  glarray = array[1..2] of real;

VAR   Store, dS        : glarray;
      Qobstot,Qobsave,Qsimsum,diff,nash,percent,rmse : real;     {objective function
variables}
      hrlast, eventlast, n,  tpko, tpks, tpkstrt, Qobsn     : integer;
      Qpko, Qpks, Qlasto, Qlasts, Qdiff, alldiff, Qave  : real;
      tobstot, tobsave, tdiff, Nashq, Nasht, allsum       : real;
      Qpkssum, tpkssum, vsum, rmseq, rmset, rmsev, percentq, percentt : real;
      Qvolo, Qvols, vobstot, vdiff, vobsave, nashv, percentv : real;
      f1min, f1max, k2min, k2max     :  real;              {optimisation loop variables}
      i,j,imax,jmax                  : integer;            {counters for loops}


PROCEDURE PARALLEL (VAR f1, K2 : real);

BEGIN

 {1. initialize storages}
 {2. for each time interval ( 1 hr), read in throughfall and stepthrough to generate
      predicted Q at 6 min time steps, then compare predicted and observed Q}

  Qobssum:=0.0;  Qalln :=0;

 WHILE NOT Eof(infile) DO
   begin
      Readln (infile, event, hr, Qobs, TF);
      Qobssum:=Qobssum + Qobs;
      Qalln := Qalln + 1;

  If (hr=0) then
   begin
      Store[1] := Qobs/K1;
```

```
      Store[2] := 0.0;
   end;

  If (f1=0) then Store[1] := 0.0;
  Qsim := 0.0;
  For i := 1 to 10 do    {integrate over 1 hr at 6 min intervals}
   begin

      Q1 := K1*Store[1];
      Q2 := K2*Store[2];
      dS[1] := tf*f1*Area - Q1;
      dS[2] := tf*(1 - f1)*Area - Q2;
      Store[1] := Store[1] + dS[1]*360;
      Store[2] := Store[2] + dS[2]*360;
      Qsim := Qsim + Store[1]*K1 + Store[2]*K2;

    end;

      Qsim := Qsim/10;
      Writeln (outfile, event:4, hr:4, Qobs:10:4, Qsim:10:4, TF);
 end;

  Writeln(outfile,' 99', ' 0' );

 {3. Replace array of initial storages with predicted
 storages at end of time interval}
 {4. Repeat (2) and (3) until end of storm event}
 {5. Repeat (1) through (4) for each storm event}

end;

{---------------begin  main program  --------------------}

 BEGIN
 ClrScr;

 assign (infile,'c:\2paralel\east\e_lgver.in'); {input event, hr, Qobs, TF}     reset (infile);
 assign (outfile,'c:\2paralel\east\s2e4ver.out'); {sim output}     rewrite (outfile);
 assign (objfile,'c:\2paralel\east\s2e4obj.out'); {obj funct output}  rewrite (objfile);
 assign (eventstats,'c:\2paralel\east\s2e4evnt.out');{event stats}   rewrite (eventstats);

 {-- loop through f1 and k2 ranges for optimisation -- }

 f1min := 0.0;
 f1max := 0.6;
 k2min := 0.01;
 k2max := 0.11;
 imax := 30;
 jmax := 50;
```

```
f1 := -0.02;

 FOR i := 1 to imax+1 DO
  BEGIN
     f1 := f1  +(f1max - f1min)/(imax);
     k2 := 0.008;
   FOR j := 1 to jmax+1 DO
    BEGIN
       k2 := k2  +(k2max - k2min)/(jmax);
       Reset (infile);
        Rewrite (outfile);
        writeln('working...');

          k2 := k2/3600;  k1 := k1/3600;

          PARALLEL (f1, k2);                    {call procedure Parallel}

          k2 := k2*3600;

{  ------------ Calculate objective functions  -----------------------}

   n := 1;  event := 1;
   Qpko := 0.0;  Qpks := 0.0;   Qvolo := 0.0; Qvols := 0.0;
   nash := 0.0;  alldiff := 0.0;   Qsimsum := 0.0;  eventlast:=0;

   reset(outfile);

 while not eof(outfile) do
   begin
      Readln (outfile, event, hr, Qobs, Qsim, tf);
      Qsimsum:=Qsimsum + Qsim;

    {  -- calc peaks, tpk, and volumes for each event -- }

  If (hr = 0) then
   Begin
       If (eventlast<>0) then
        begin
          Writeln (eventstats,eventlast:3,Qvolo:15:2,Qvols:15:2,Qpko:15:5,
                Qpks:15:5,tpko:4,tpks:4);
         Writeln ( eventlast:3, Qvolo:15:2, Qvols:15:2,Qpko:15:5, Qpks:15:5, tpko:4, tpks:4);
          tpkstrt := 0;
        end;
          Qvolo := 0.0;  Qvols := 0.0;
          Qpko := Qobs;
          Qpks := Qsim;
  End;

    If (hr <> 0) then
```

```
        Begin
          Qlasto := Qobs;
          Qlasts := Qsim;
          eventlast := event;
          hrlast := hr;
           If (tf<>0) and (tpkstrt=0) then tpkstrt := hr;
           If (Qobs > Qpko) then
             begin
                 Qpko := Qobs;
                 tpko := hr - tpkstrt;
             end;
           If (Qsim > Qpks) then
             begin
                Qpks := Qsim;
                tpks := hr - tpkstrt;
             end;
        End;
      Qvolo := Qvolo + Qobs;
      Qvols := Qvols + Qsim;

{ -- calc overall Nash for entire calibration period -- }

      If (event<99) then
        begin
          alldiff := alldiff + sqr(Qsim-Qobs);
          nash := nash + sqr(Qobs-(Qobssum/Qalln));
        end;

end;
        Writeln(eventstats,' 99');
        rmse := sqrt(1/Qalln*alldiff);
        nash := 1 - (alldiff/nash);
        percent := (Qsimsum - Qobssum)/Qobssum*100;

{ -- calc Nash and % deviations for pks, tpk, and volumes -- }

    Reset (eventstats);
        Qobstot := 0.0;   Qobsn := 0;
        tobstot := 0.0;   vobstot := 0.0;

        while not eof(eventstats) do
           begin
              Readln(eventstats, event, Qvolo, Qvols, Qpko, Qpks, tpko, tpks);
              If (event<99) then
                begin
                  writeln(event:4,Qvolo:9:1);
                  Qobstot := Qobstot + Qpko;
                  Qobsn := Qobsn + 1;
                  tobstot := tobstot + tpko;
```

```
          vobstot := vobstot + Qvolo;
        end;
    end;

   Qobsave := Qobstot/Qobsn;
   tobsave := tobstot/Qobsn;
   vobsave := vobstot/Qobsn;

   Qdiff:=0.0;  tdiff:= 0.0;   vdiff:= 0.0;
   Nashq := 0.0;  Nasht := 0.0;   Nashv := 0.0;
   Qpkssum :=0.0;  tpkssum := 0.0;    vsum := 0.0;

   Reset (eventstats);

while not eof(eventstats) do
    begin
      Readln(eventstats, event, Qvolo, Qvols, Qpko, Qpks, tpko, tpks);
       If (event<99) then
         begin
           Qdiff := Qdiff+ SQR(Qpks - Qpko);
           Nashq := Nashq + SQR(Qpko - Qobsave);
           Qpkssum := Qpkssum + Qpks;
           tdiff := tdiff+ SQR(tpks - tpko);
           Nasht := Nasht + SQR(tpko - tobsave);
           tpkssum := tpkssum + tpks;
           vdiff := vdiff+ SQR(Qvols - Qvolo);
           Nashv := Nashv + SQR(Qvolo - vobsave);
           vsum := vsum + Qvols;
         end;
    end;

    percentq := (Qpkssum - Qobstot)/Qobstot*100;
    rmseq := sqrt(1/Qobsn*(qdiff));
    Nashq := 1 - (qdiff/Nashq);

    percentt := (tpkssum - tobstot)/tobstot*100;
    rmset := sqrt(1/Qobsn*(tdiff));
    Nasht := 1 - (tdiff/Nasht);

     percentv := (vsum - vobstot)/vobstot*100;
    rmsev := sqrt(1/Qobsn*(vdiff));
    Nashv := 1 - (vdiff/Nashv);

{   append (objfile);}

   writeln;
   writeln('f1 = ',f1:8:4, '  k2 = ',k2:10:6);
   writeln('overall Nash = ', nash:8:5, '   % = ',percent:4:1);
   writeln;
```

```
      writeln ('Nash vol = ',nashv:6:3,'  Nash pks = ',Nashq:6:3,'  Nash tpk = ',Nasht:6:3);
      writeln ('%dev vol = ',percentv:6:1,'  %dev pks = ',percentq:6:1,'  %dev tpk = ',percentt:6:1);
      writeln (objfile,f1:8:4,k2:15:6,nash:11:5, percent:10:2, nashv:11:5,percentv:10:2,
            nashq:11:5,percentq:10:2,nasht:11:5,percentt:10:2);
     {rewrite(eventstats);

        END;
     END;

close (infile);
close (outfile);
close (objfile);

END.                          {END OF PROGRAM}
```

## Conceptual lumped model code

```
PROGRAM SERIAL;

{ Triple Reservoir Lumped Model        East catchment  -  Pascal code  }

 USES CRT;
 VAR  f1,phi,f1min,f1max,phimin,phimax        : real;
      i,j,imax,jmax                           : integer;  {counters for f1,phi loops}
      infile, outfile1,out2, eventstats       : text;
      event, hr, timestep, nok, nbad, Qobsn   : integer;
      Qobs, Qsim, TF, Qobstot, Qobsave, Qsimsum,diff,nash,percent,rmse: real;
      hrlast, eventlast, n,  tpko, tpks, tpkstrt   : integer;
      Qpko, Qpks, Qlasto, Qlasts, Qdiff            : real;
      tobstot, tobsave, tdiff, Nashq, Nasht        : real;
      Qpkssum, tpkssum, vsum, rmseq, rmset, rmsev, percentq, percentt : real;
      Qvolo, Qvols, vobstot, vdiff, vobsave, nashv, percentv : real;
      alldiff, Qave, allsum, Qobssum, Qalln         : real;


PROCEDURE ODE_APP (var f1,phi: real);          {start of ordinary diff eqn integration}

CONST { K3=0.0932/3600; K2=0.0513/3600; K1=0.0071/3600; {lev 3}
      K3=0.1256/3600; K2=0.0541/3600; K1=0.0074/3600; {lev 1}
      A=383000;   nvar = 3;  nstepp = 200;  eps=0.000001;

TYPE  glarray  = array[1..nvar] of real;

VAR   {f1,phi    : real;}
      y, ystart          : glarray;
      x1, x2, h1, hmin         : real;
      Kmax, kount              : integer;
      dxsav                    : real;
      Xp                       : array [1..nstepp] of real;
      Yp                       : array [1..nvar, 1..nstepp] of real;


PROCEDURE derivs (x:real;VAR y,dy:glarray);  {derivatives evaluated at each time step}

{  tf = throughfall input rate in time interval
   y = array of reservoir storages
   dy = array of time derivatives (to be evaluated)}

VAR  Q1,Q2  : real;

 BEGIN
```

```
  Q1 := K1*y[1];
  Q2 := K2*y[2];
  dy[1] := tf*f1*A - Q1;
  dy[2] := tf*(1 - f1 - phi*y[2])*A + Q1 - Q2;
  dy[3] := tf*phi*y[2]*A + Q2 - K3*y[3];

 END;
```

{*** *THE NUMERICAL RECIPES CODE BEGINS HERE* ***}

**PROCEDURE rk4** (y,dydx: glarray; n: integer; x,h: real; VAR yout: glarray);
{*Runge-Kutta routine*}

```
VAR
  i: integer; xh,hh,h6: real; dym,dyt,yt: glarray;

BEGIN
  hh := h*0.5; h6 := h/6.0; xh := x + hh;
  FOR i := 1 TO n DO yt[i] := y[i] + hh*dydx[i];
  derivs(xh,yt,dyt);
  FOR i := 1 TO n DO yt[i] := y[i] + hh*dyt[i];
  derivs(xh,yt,dym);
  FOR i := 1 TO n DO BEGIN
    yt[i] := y[i] + h*dym[i];
    dym[i] := dyt[i] + dym[i];
  END;
  derivs(x+h,yt,dyt);
  FOR i := 1 TO n DO
    yout[i] := y[i] + h6*(dydx[i] + dyt[i] + 2.0*dym[i]);
END;
```

**PROCEDURE rkqc** (VAR y,dydx: glarray; n: integer; VAR x: real;
{*Runge-Kutta quality control*}

```
              htry,eps: real; yscal: glarray; VAR hdid,hnext: real);
LABEL 1;
CONST  pgrow=-0.20; pshrnk=-0.25;
       fcor=0.06666666;
       one=1.0; safety=0.9; errcon=6.0e-4;

VAR    i : integer; xsav,hh,h,temp,errmax: real;
       dysav,ysav,ytemp: glarray;

BEGIN
  xsav := x;
  FOR i := 1 to n DO BEGIN
    ysav[i] := y[i];
    dysav[i] := dydx[i];
  END;
  h := htry;
```

```
1: hh := 0.5*h;
  rk4(ysav,dysav,n,xsav,hh,ytemp); x := xsav + hh; derivs(x,ytemp,dydx);
  rk4(ytemp,dydx,n,x,hh,y); x := xsav + h;
  IF (x = xsav) THEN BEGIN
    writeln('pause in routine RKQC');
    writeln('stepsize too small');
  END;
  rk4(ysav,dysav,n,xsav,h,ytemp); errmax := 0.0;
  FOR i := 1 TO n DO BEGIN
    ytemp[i] := y[i] - ytemp[i];
    temp := abs(ytemp[i]/yscal[i]);
    IF (errmax < temp) THEN errmax := temp;
  END;
  errmax := errmax/eps;
  IF (errmax > one) THEN BEGIN
    h := safety*h*exp(pshrnk*ln(errmax));
    GOTO 1; END;
  If (errmax <= one) then BEGIN
    hdid := h;
    IF (errmax > errcon) THEN hnext := safety*h*exp(pgrow*ln(errmax))
    ELSE hnext := 4.0*h;
  END;
  FOR i := 1 TO n DO y[i] := y[i] + ytemp[i]*fcor;
END;
```

**PROCEDURE ODEint** (VAR ystart: glarray; nvar: integer; x1,x2,eps,h1,hmin: real;
VAR nok,nbad: integer);

{*ordinary diff eqn integration*

```
LABEL 99;
CONST    maxstp=10000; two=2.0; zero=0.0; tiny=1.0e-30;
VAR    nstp,i: integer; xsav,x,hnext,hdid,h: real; yscal,y,dydx: glarray;

BEGIN
   x := x1;
   IF (x2 >= x1) THEN
     h := abs(h1)
   ELSE
     h := -abs(h1);
   nok := 0; nbad := 0; kount := 0;
   FOR i := 1 TO nvar DO y[i] := ystart[i];
   IF kmax > 0 THEN xsav := x - dxsav*two;
   FOR nstp := 1 TO maxstp DO BEGIN
     derivs(x,y,dydx);
     FOR i := 1 to nvar DO yscal[i] := abs(y[i]) + abs(dydx[i]*h) + tiny;
     IF (kmax > 0) THEN BEGIN
       IF (abs(x - xsav) > abs(dxsav)) THEN BEGIN
         IF (kount < kmax - 1) THEN BEGIN
           kount := kount + 1;
```

```
          xp[kount] := x;
          FOR i := 1 TO nvar DO yp[i,kount] := y[i];
          xsav := x;
        END;
      END;
    END;
    IF(((x + h - x2)*(X + h - x1)) > zero) THEN h := x2 - x;
    rkqc(y,dydx,nvar,x,h,eps,yscal,hdid,hnext);
    IF (hdid = h) THEN
      nok := nok + 1
    ELSE
      nbad := nbad + 1;
    IF (((x-x2)*(x2-x1)) >= zero) THEN BEGIN
      FOR i := 1 TO nvar DO ystart[i] := y[i];
      IF (kmax <> 0) THEN BEGIN
        kount := kount + 1;
        xp[kount] := x;
        FOR i := 1 TO nvar DO yp[i,kount] := y[i];
      END;
      GOTO 99;
    END;
    IF (abs(hnext) < hmin) THEN BEGIN
      writeln('pause in routine ODEint');
      writeln('stepsize too small');
  END;
    h := hnext;
  END;
  writeln('pause in routine ODEint - too many steps');
99: END;

{***  THE NUMERICAL RECIPES CODE ENDS HERE    ***}
```

**BEGIN {ODE_APP}**

```
        {0. specify parameter values K1,K2,K3,f1,phi}
        {1. initialize storages}
        {2. for each time interval (1 hr), read in throughfall and step
            through ODEint to generate predicted Q at desired
            time steps (10 min), then compare predicted and observed Q}

Qobssum:=0.0;  Qalln :=0;

WHILE NOT Eof(infile) DO
 begin
  Readln (infile, event, hr, Qobs, TF);
  Qobssum:=Qobssum + Qobs;
  Qalln := Qalln + 1;

  If (hr=0) then                         {initializing storages}
```

```
      begin
        If (f1=0) then
           ystart[1] := 0.0;
        If (f1<>0) then
           ystart[1] := Qobs/K1;
           ystart[2] := Qobs/K2;
           ystart[3] := Qobs/K3;
      end;

    Qsim := 0.0;
    X1 := 0;
    X2 := 360
    dxsav := 0.0;  kmax := 0;
    h1 := 200.0;  hmin := 0.01;

    For timestep := 1 to 10 do            {integrate over 1 hr at 6 min intervals}
      begin
        ODEint(ystart,nvar,x1,x2,eps,h1,hmin,nok,nbad);
        Qsim := Qsim + ystart[3]*K3;
      end;

      Qsim := Qsim/10;

      Writeln (outfile1, event:2, hr:4, Qobs:15:4, Qsim:15:4, TF:18:10);
  end;
    Writeln(outfile1,' 99', ' 0' );

  {3. Replace array of initial storages with predicted
   storages at end of time interval}
  {4. Repeat (2) and (3) until end of storm event}
  {5. Repeat (1) through (4) for each storm event}

end;

BEGIN{begin main}

  ClrScr;
  Assign (Infile, 'C:\3serial\east\e_cal.in');
  Reset (Infile);
  Assign (Outfile1, 'C:\3serial\east\e_cal.out');
  Rewrite (Outfile1);
  Assign (Out2, 'C:\3serial\east\e_opt.out');
  Rewrite (Out2);
  Assign (eventstats, 'C:\3serial\east\e_evnt.out');
  {append}Rewrite (eventstats);

  f1min := 0.6,                        {iterative loops for calibration}
  f1max := 1.0;
  phimin := 1.0e-9;
```

```
phimax := 1.0e-4;
imax := 5;
jmax := 8;
f1 := -0.20;

 FOR i := 1 to imax+1 DO
   BEGIN
      f1 := f1 + (f1max - f1min)/imax;
      phi := (phimax - phimin)/(jmax);

       FOR j := 1 to jmax+1 DO
        BEGIN
          phi := phi  +(phimax - phimin)/(jmax);}

                 Reset (infile);
                 Rewrite (outfile1);

       ODE_APP (f1, phi);                            {call ordinary diff eqn solver routine}

{ -------------- Calculate objective functions   --------------------------}

    n := 1;  event := 1;
    Qpko := 0.0;  Qpks := 0.0;   Qvolo := 0.0; Qvols := 0.0;
    nash := 0.0;  alldiff := 0.0;   Qsimsum := 0.0;  eventlast:=0;
    reset(outfile1);

 while not eof(outfile1) do
   begin
      Readln (outfile1, event, hr, Qobs, Qsim, tf);
      Qsimsum:=Qsimsum + Qsim;

      { -- calc peaks, tpk, and volumes for each event -- }

 {If (hr = 0) then
    Begin
      If (eventlast<>0) then
        begin
          Writeln (eventstats, eventlast:3, Qvolo:15:2, Qvols:15:2, Qpko:15:5,  Qpks:15:5,
tpko:4, tpks:4);
          Writeln ( eventlast:3, Qvolo:15:2, Qvols:15:2, Qpko:15:5,  Qpks:15:5, tpko:4, tpks:4);
          tpkstrt := 0;
        end;
         Qvolo := 0.0;  Qvols := 0.0;
         Qpko := Qobs;
         Qpks := Qsim;
       End;

      If (hr <> 0) then
        Begin
```

```
        Qlasto := Qobs;
        Qlasts := Qsim;
        eventlast := event;
        hrlast := hr;
         If (tf<>0) and (tpkstrt=0) then tpkstrt := hr;
         If (Qobs > Qpko) then
           begin
             Qpko := Qobs;
             tpko := hr - tpkstrt;
           end;
         If (Qsim > Qpks) then
           begin
             Qpks := Qsim;
             tpks := hr - tpkstrt;
           end;
      End;
    Qvolo := Qvolo + Qobs;
    Qvols := Qvols + Qsim;

{ -- calc overall Nash for entire calibration period -- }

    If (event<99) then
     begin
      alldiff := alldiff + sqr(Qsim-Qobs);
      nash := nash + sqr(Qobs-(Qobssum/Qalln));
     end;

end;
    Writeln(eventstats,' 99');
    rmse := sqrt(1/Qalln*alldiff);
    nash := 1 - (alldiff/nash);
        percent := (Qsimsum - Qobssum)/Qobssum*100;

{ -- calc Nash and % deviations for pks, tpk, and volumes -- }

      Reset (eventstats);
      Qobstot := 0.0;   Qobsn := 0;
      tobstot := 0.0;   vobstot := 0.0;

      while not eof(eventstats) do
       begin

  Readln(eventstats, event, Qvolo, Qvols, Qpko, Qpks, tpko, tpks);
  If (event<99) then
          begin
     writeln(event:4,Qvolo:9:1);
        Qobstot := Qobstot + Qpko;
        Qobsn := Qobsn + 1;
     tobstot := tobstot + tpko;
```

```
  vobstot := vobstot + Qvolo;
   end;
 end;

    Qobsave := Qobstot/Qobsn;
    tobsave := tobstot/Qobsn;
    vobsave := vobstot/Qobsn;

    Qdiff:=0.0;  tdiff:= 0.0;   vdiff:= 0.0;
    Nashq := 0.0;  Nasht := 0.0;   Nashv := 0.0;
    Qpkssum :=0.0;  tpkssum := 0.0;    vsum := 0.0;

    Reset (eventstats);

 while not eof(eventstats) do
    begin

     Readln(eventstats, event, Qvolo, Qvols, Qpko, Qpks, tpko, tpks);
  If (event<99) then
  begin

     Qdiff := Qdiff+ SQR(Qpks - Qpko);
     Nashq := Nashq + SQR(Qpko - Qobsave);
     Qpkssum := Qpkssum + Qpks;

      tdiff := tdiff+ SQR(tpks - tpko);
     Nasht := Nasht + SQR(tpko - tobsave);
     tpkssum := tpkssum + tpks;

      vdiff := vdiff+ SQR(Qvols - Qvolo);
     Nashv := Nashv + SQR(Qvolo - vobsave);
      vsum := vsum + Qvols;
  end;
    end;

    percentq := (Qpkssum - Qobstot)/Qobstot*100;
    rmseq := sqrt(1/Qobsn*(qdiff));
    Nashq := 1 - (qdiff/Nashq);

    percentt := (tpkssum - tobstot)/tobstot*100;
    rmset := sqrt(1/Qobsn*(tdiff));
    Nasht := 1 - (tdiff/Nasht);

     percentv := (vsum - vobstot)/vobstot*100;
    rmsev := sqrt(1/Qobsn*(vdiff));
    Nashv := 1 - (vdiff/Nashv);

{    append (objfile);}
```

```
        writeln('f1 = ',f1:5:3,'  phi = ',phi);
        writeln('overall Nash = ', nash:6:3);
        writeln ('Nash vol = ',nashv:6:3,'  Nash pks = ',Nashq:6:3,'  Nash tpk = ',Nasht:6:3);
        writeln ('%dev vol = ',percentv:6:1,'  %dev pks = ',percentq:6:1,'  %dev tpk =
',percentt:6:1);
        writeln (out2,f1:5:3,phi,' ',ln(phi):8:4,nash:8:5, percent:6:2);{, nashv:8:5,percentv:6:2,
            nashq:8:5,percentq:6:2,nasht:8:5,percentt:6:2);
        writeln;

 rewrite(eventstats);

 end;                                    {end of objective function calculations }
 end;

 close (infile);
 close (outfile1);
 close (out2);

 Sound(220);      { Beep Hz }
 Delay(200);      { For 200 ms }
 NoSound;         { stops beep }

END.                                {END OF PROGRAM}
```

## Quasi-distributed model code

```
PROGRAM TOP_OPT;

{ TOPMODEL (quasi-distributed) Model      East catchment  -  Pascal code  }


USES CRT;
TYPE  rarray = array[1..50] of real;

VAR   infile, outfile, topofile, objfile, eventstats  : text;

      {TOPMODEL variables}

      NAc, mi, ti, nm, nt, event, hr, Qalln        : integer;
      Qobs, Qsim, TF, Qobssum, Actf                : real;
      Ac, topo                                     : rarray;
      lambda                                       : real;
      m, T0, td, dt,Qoftot,Qbtot                   : real;

      {objective function variables}

      Qobstot, Qobsave, Qsimsum,diff,nash,percent,rmse          : real;
      hrlast, eventlast, n,  tpko, tpks, tpkstrt, Qobsn         : integer;
      Qpko, Qpks, Qlasto, Qlasts, Qdiff                         : real;
      tobstot, tobsave, tdiff, Nashq, Nasht                     : real;
      Qpkssum, tpkssum, vsum, rmseq, rmset, rmsev, percentq, percentt : real;
      Qvolo, Qvols, vobstot, vdiff, vobsave, nashv, percentv    : real;
      alldiff, Qave, allsum                                     : real;

      {optimisation loop variables}

      mmin, mmax, T0min, T0max, tdmin, tdmax                    : real;
      i,j,k, imax,jmax, kmax                    : integer;  {counters for loops}


{-----------------------------------------------------------------
   Read TOPO INDEX procedure
 -----------------------------------------------------------------}

PROCEDURE Read_Topo (var NAc:integer; var Ac,topo:rarray; var lambda: real);

VAR  i      : integer;

BEGIN
```

```
  Readln (topofile, NAc);   {read in # of index increments}

 For i := 1 to NAc do
    begin                          {read in area % and a/tanb index upper limit}
        Readln (topofile, Ac[i], topo[i]);
        writeln(Nac:5,' ', Ac[i]:8:7,' ', topo[i]:5:2);
    end;

  { -- calculate areal integral of ln(a/tanb) -- }

 lambda := 0.0;

 For i := 2 to NAc do
   lambda := lambda + Ac[i]*(topo[i] + topo[i-1])/2;
   writeln('lambda = ',lambda:5:2);
   readln;

close (topofile);

END;


{-------------------------------------------------------------
         TOPMODEL run procedure
-------------------------------------------------------------}

PROCEDURE TOPMODEL (var m, T0, td : real);

VAR  Q0, Qof, Quz, uz, Qb, Qsim  : real;           {program variables}
   Sbar, szq, olf, Acf, Acsat  : real;
   ex, sd, suz  : rarray;
   i, inc, inc2  : integer;

CONST  Area = 383000;  dt = 1.0;            {program constants}

BEGIN                                       {begin TOPMODEL procedure}

Qalln := 0;  Qobssum := 0.0;

While not eof(infile) do
   begin
      Readln (infile, event, hr, Qobs, TF);
      Qobs := Qobs/Area*3600;             {L/s to mm/hr}
      TF := TF*3600;                   {mm/s to mm/hr}
      Qof := 0.0;  Quz := 0.0;  Acsat := 0.0;

  If (hr=0) then                             {initialise variables for new event}
   begin
      Q0 := Qobs;
```

```
    szq := exp((T0+ln(dt)) - lambda);              {sat zone outflow parameter}
    Sbar := - m * ln(Q0/szq);                      {mean catchment deficit}

For inc := 1 to NAc do
  begin
    suz[inc] := 0.0;
end;
end;

 For inc := 1 to NAc do                            {loop for ln(a/tanb) increments}
  begin

    { -- calc local storage deficit -- }

     sd[inc] := Sbar + m*(lambda - topo[inc]);
     If (sd[inc]<0) then  sd[inc] := 0.0;          {if sd<0 then soil sat}

    { -- calc unsat storage -- }

    ex[inc] := 0.0;
    suz[inc] := suz[inc] +{ (1-Actf)*}TF;          {throughfall input}

    If (suz[inc]>sd[inc]) then
      begin
        ex[inc] := suz[inc] - sd[inc];             {excess of unsat zone}
        suz[inc] := sd[inc];                       {update unsat zone store}
      end;

    { -- calc drainage from unsat zone -- }

     uz :=0.0;

     If (sd[inc]>0) then
      begin
         uz := suz[inc]/(sd[inc]*td*dt);           {vertical drainage}
         If (uz>suz[inc]) then  uz := suz[inc];    {can't drain more than store}
         suz[inc] := suz[inc] - uz;                {new storage after drainage}
         If (suz[inc]<1e-7) then suz[inc]:=0.0;    {suz neglible}
         Quz := Quz + uz*Ac[inc];                  {sum Quz over catchment}
      end;

    { -- calc overland flow -- }

    olf := 0.0;

    If (inc>1) then
      begin
        inc2 := inc - 1;                           {inc2 is sat if inc is}
```

```
        If (ex[inc]>0) then                              {both limits saturated}
         begin
            Acsat := Acsat + Ac[inc];                    {sum area saturated}
            olf := Ac[inc] * (ex[inc2]+ex[inc])/2;
         end

        Else if (ex[inc2]>0) then                        {inc not sat, inc2 is}
         begin
            Acf := Ac[inc]*ex[inc2]/(ex[inc2]-ex[inc]);  {area fraction sat}
            of := Acf*ex[inc2]/2;
            Acsat := Acsat + Acf;
         end;
       end;

       Actf := Acsat;
       Qof := Qof + olf;                                 {sum Qof over catchment}
     end;                                                {end of ln(a/tanb) increment loop}

     { -- calc saturated zone drainage -- }

     Qb := szq*exp(-Sbar/m);
     Sbar := Sbar - Quz + Qb;

     Qbtot := Qbtot + Qb;
     Qoftot := Qoftot + Qof;
     Qsim := Qb + Qof;
     Qsim := Qsim*Area/3600;          {mm/hr to L/s}
     Qobs := Qobs*Area/3600;
     Qobssum:= Qobssum + Qobs;
     Qalln := Qalln + 1;
     TF := TF/3600;

     writeln(outfile, event:3, hr:4, Qobs:12:4, Qsim:12:4, TF);
end;

    writeln(outfile,99,0);            {needed for results read}

 END; {end of proc TOPMODEL}

 {---------------------------------------------------------------
    Main program
 ---------------------------------------------------------------}

 BEGIN
 ClrScr;

  assign (infile,'c:\4topmdl\east\e_ver.in'); {input event, hr, Qobs, TF}
  reset (infile);
  assign (topofile,'c:\4topmdl\east\grids\e_10top.dat');{ln(a/tanb) index}
```

```
reset (topofile);
assign (outfile,'c:\4topmdl\east\e_ver.out'); {sim output}
rewrite (outfile);
assign (objfile,'c:\4topmdl\east\e_obj.out'); {obj funct output}
rewrite (objfile);
assign (eventstats,'c:\4topmdl\east\e_evnt.out');{event stats}
rewrite (eventstats);

Read_Topo (NAc, Ac, topo, lambda);                    {read in topo index data}

 {-- write headers for obj funct file -- }

{writeln(objfile,'m   lnTo  nash   nashv  nashpk nashtpk %vol  %pks  %tpk ');

 {-- loop through m and ln(To) ranges for optimisation -- }

{ mmin := 0.0;
  mmax := 20.0;
  tdmin := 0.04;
  tdmax := 0.14;
  imax := 14;
  jmax := 20;
  kmax := 20;
  T0 := 1.1e3;
  T0 := ln(T0);

   FOR i := 1 to imax+1 DO
    BEGIN
        T0 := T0 + (14)/imax;
        m := 1;

        FOR j := 1 to jmax+1 DO
         BEGIN
            m := m  +(mmax - mmin)/(jmax);
            td := 0.035;

                FOR k := 1 to kmax+1 DO
                 BEGIN
                    td := td  +(tdmax - tdmin)/(kmax);

                      Reset (infile);
                      Rewrite (outfile);

                 m := 13.8;  td := 0.036;  T0 := 13.1;

                 T0 := ln(T0);


        TOPMODEL (m, T0, td);                          {call procedure TOPMODEL}
```

```
{  ---------   Calculate objective functions   -------------------------}

    n := 1;  event := 1;
    Qpko := 0.0;  Qpks := 0.0;   Qvolo := 0.0; Qvols := 0.0;
    nash := 0.0;  alldiff := 0.0;   Qsimsum := 0.0;  eventlast:=0;

    reset(outfile);

  while not eof(outfile) do
   begin
     Readln (outfile, event, hr, Qobs, Qsim, tf);
     Qsimsum:=Qsimsum + Qsim;

     { -- calc peaks, tpk, and volumes for each event -- }

     If (hr = 0) then
        Begin
          If (eventlast<>0) then
             begin
Writeln(eventstats,eventlast:3,Qvolo:15:2,Qvols:15:2,Qpko:15:5,Qpks:15:5,tpko:4,tpks:4);
Writeln ( eventlast:3, Qvolo:15:2, Qvols:15:2, Qpko:15:5,  Qpks:15:5, tpko:4, tpks:4);
               tpkstrt := 0;
             end;

        Qvolo := 0.0;  Qvols := 0.0;
        Qpko := Qobs;
        Qpks := Qsim;
       End;

      If (hr <> 0) then
            Begin
            Qlasto := Qobs;
            Qlasts := Qsim;
            eventlast := event;
            hrlast := hr;
            If (tf<>0) and (tpkstrt=0) then tpkstrt := hr;
            If (Qobs > Qpko) then
              begin
                  Qpko := Qobs;
                  tpko := hr - tpkstrt;
               end;
            If (Qsim > Qpks) then
              begin
                  Qpks := Qsim;
                  tpks := hr - tpkstrt;
               end;
```

```
      End;

    Qvolo := Qvolo + Qobs;
    Qvols := Qvols + Qsim;

{ -- calc overall Nash for entire calibration period -- }

    If (event<99) then
     begin
        alldiff := alldiff + sqr(Qsim-Qobs);
        nash := nash + sqr(Qobs-(Qobssum/Qalln));
     end;

end;
    Writeln(eventstats,' 99');
      rmse := sqrt(1/Qalln*alldiff);
      nash := 1 - (alldiff/nash);
      percent := (Qsimsum - Qobssum)/Qobssum*100;

{ -- calc Nash and % deviations for pks, tpk, and volumes -- }

      Reset (eventstats);
      Qobstot := 0.0;   Qobsn := 0;
      tobstot := 0.0;   vobstot := 0.0;

      while not eof(eventstats) do
         begin
         Readln(eventstats, event, Qvolo, Qvols, Qpko, Qpks, tpko, tpks);

      If (event<99) then
          begin
            writeln(event:4,Qvolo:9:1);
            Qobstot := Qobstot + Qpko;
            Qobsn := Qobsn + 1;
            tobstot := tobstot + tpko;
            vobstot := vobstot + Qvolo;
         end;
      end;

            Qobsave := Qobstot/Qobsn;
            tobsave := tobstot/Qobsn;
            vobsave := vobstot/Qobsn;

            Qdiff:=0.0;  tdiff:= 0.0;   vdiff:= 0.0;
            Nashq := 0.0;  Nasht := 0.0;   Nashv := 0.0;
            Qpkssum :=0.0;  tpkssum := 0.0;    vsum := 0.0;

      Reset (eventstats);
```

```
while not eof(eventstats) do
  begin
    Readln(eventstats, event, Qvolo, Qvols, Qpko, Qpks, tpko, tpks);
    If (event<99) then
      begin

    Qdiff := Qdiff+ SQR(Qpks - Qpko);
    Nashq := Nashq + SQR(Qpko - Qobsave);
    Qpkssum := Qpkssum + Qpks;

    tdiff := tdiff+ SQR(tpks - tpko);
    Nasht := Nasht + SQR(tpko - tobsave);
    tpkssum := tpkssum + tpks;

    vdiff := vdiff+ SQR(Qvols - Qvolo);
    Nashv := Nashv + SQR(Qvolo - vobsave);
    vsum := vsum + Qvols;
  end;
end;

    percentq := (Qpkssum - Qobstot)/Qobstot*100;
    rmseq := sqrt(1/Qobsn*(qdiff));
    Nashq := 1 - (qdiff/Nashq);

    percentt := (tpkssum - tobstot)/tobstot*100;
    rmset := sqrt(1/Qobsn*(tdiff));
    Nasht := 1 - (tdiff/Nasht);

    percentv := (vsum - vobstot)/vobstot*100;
    rmsev := sqrt(1/Qobsn*(vdiff));
    Nashv := 1 - (vdiff/Nashv);

{   append (objfile);}

  writeln('T0 = ',T0:6:2, '  m = ',m:4:1,'  td = ',td:5:3);
  writeln('overall Nash = ', nash:8:5, '    % = ',percent:4:1);
  writeln;
  writeln ('Nash vol = ',nashv:6:3,'  Nash pks = ',Nashq:6:3,'  Nash tpk = ',Nasht:6:3);
  writeln ('%dev vol = ',percentv:6:1,'  %dev pks = ',percentq:6:1,'  %dev tpk =
',percentt:6:1);
  writeln (objfile,T0:8:3,m:8:3,td:8:3,nash:11:5, percent:10:2, nashv:11:5,percentv:10:2,
     nashq:11:5,percentq:10:2,nasht:11:5,percentt:10:2);

  rewrite(eventstats);

{end of procedure results}

    END;
  END;
```

```
    END;

close (infile);
close (outfile);
close (objfile);

Sound(220);       { Beep Hz }
Delay(200);       { For 200 ms }
NoSound;          { stops beep }
readln;

END.                              {END OF PROGRAM}
```

# APPENDIX 3

# SCATTERPLOTS

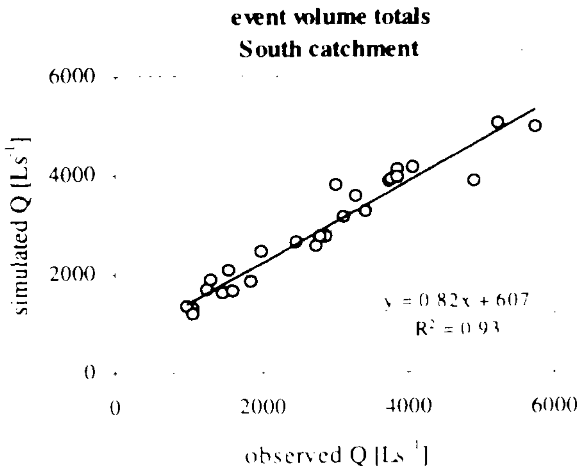
# OF VERIFICATION DATA SETS

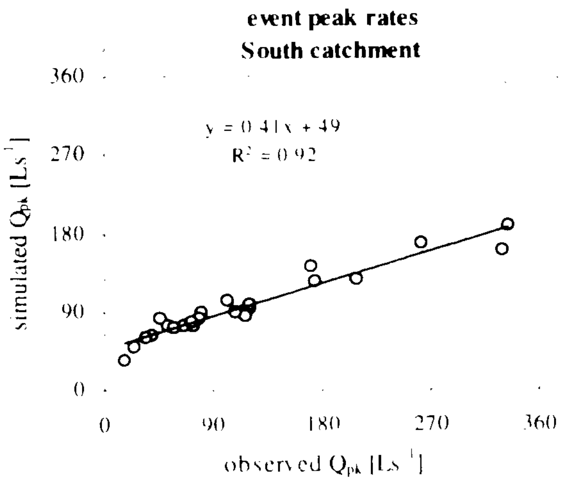
# SCATTERPLOTS

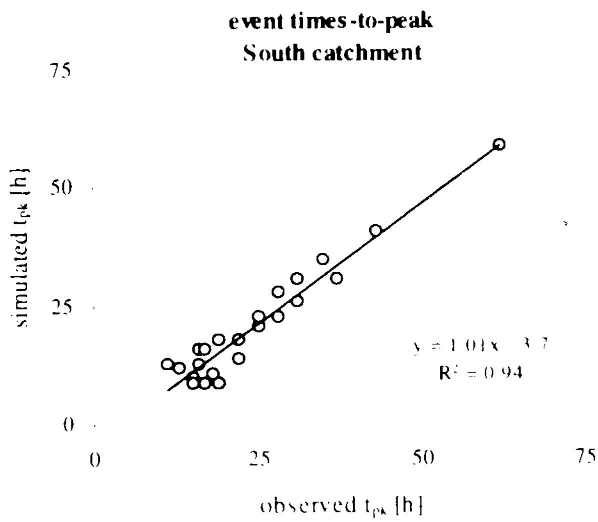
# OF VERIFICATION DATA SETS

## Level 1 Split-sample test

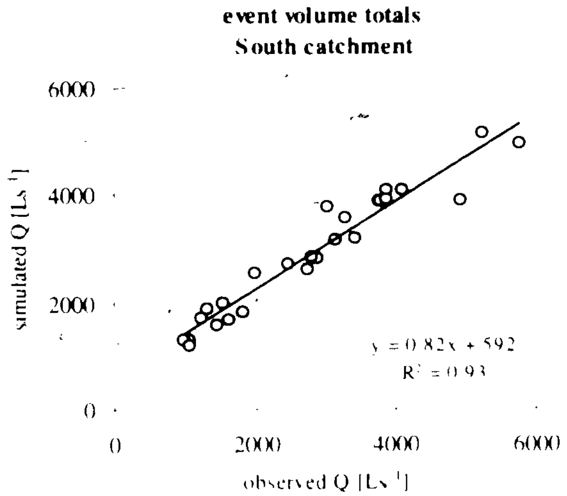
## Level 1  Dual parallel reservoir lumped model

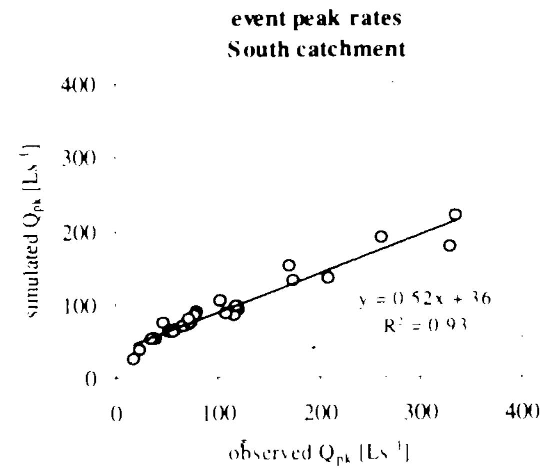


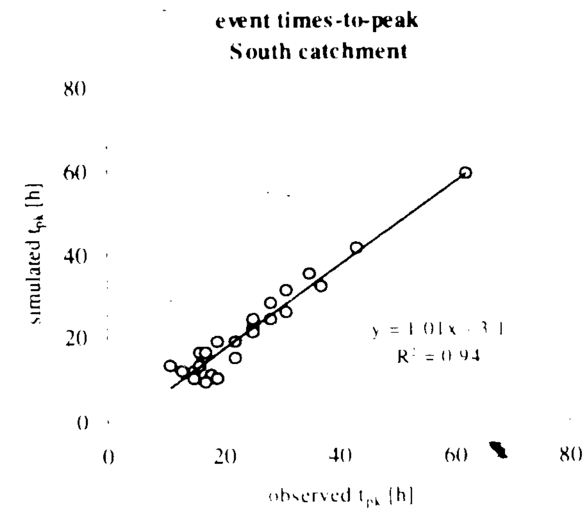



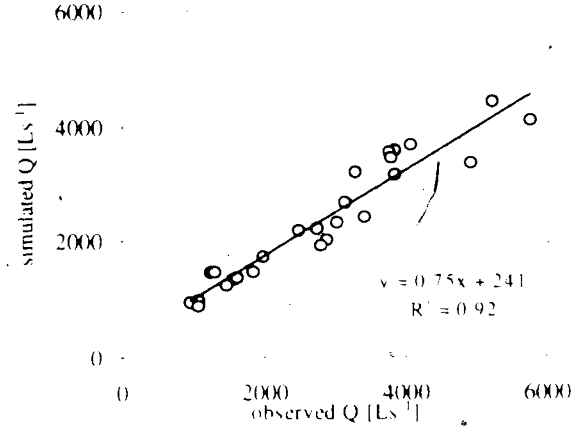
Level 1 Triple serial reservoir lumped model



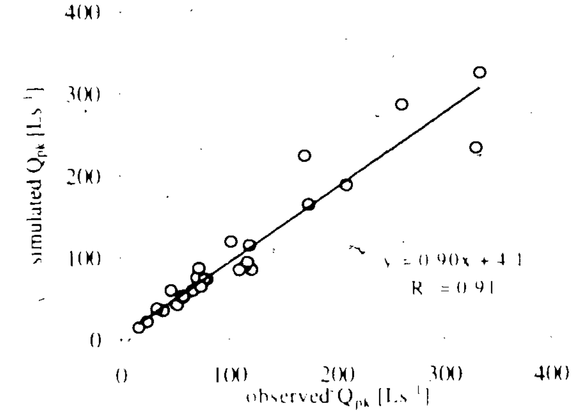



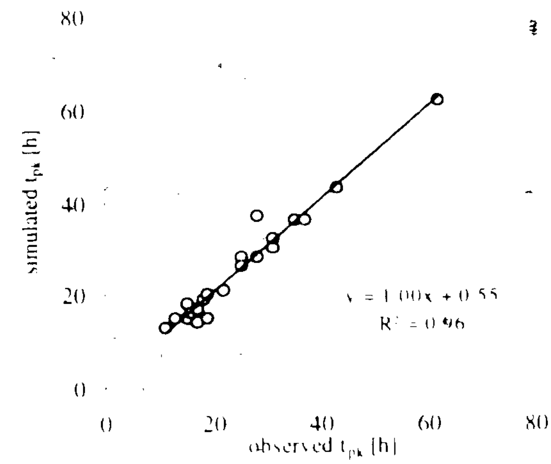
**event volume totals**
**South catchment**



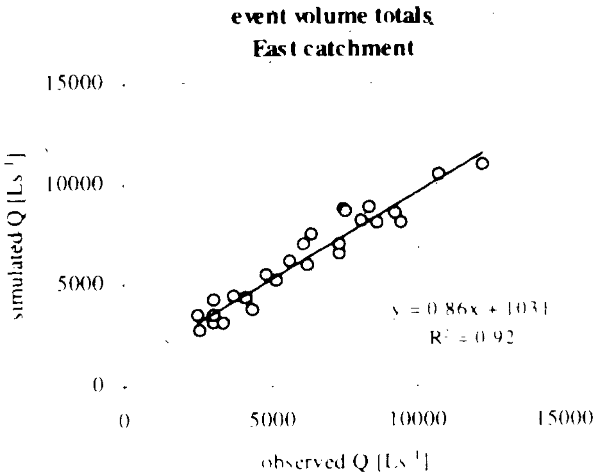

**event peak rates**
**South catchment**

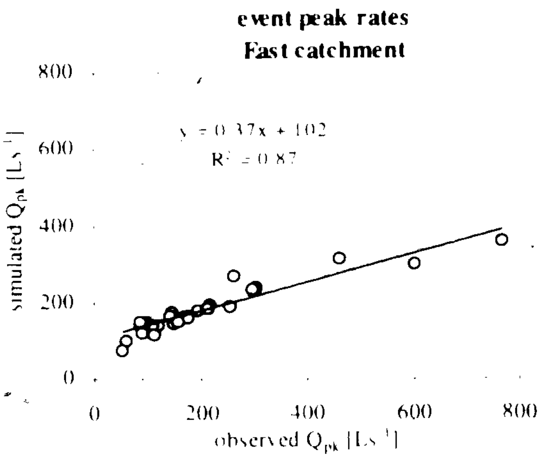


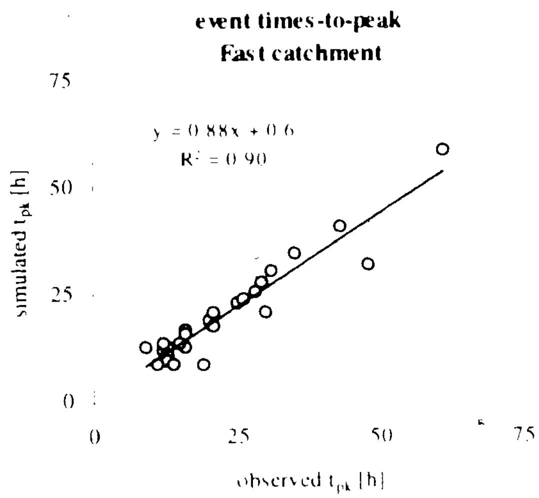

**event times-to-peak**
**South catchment**



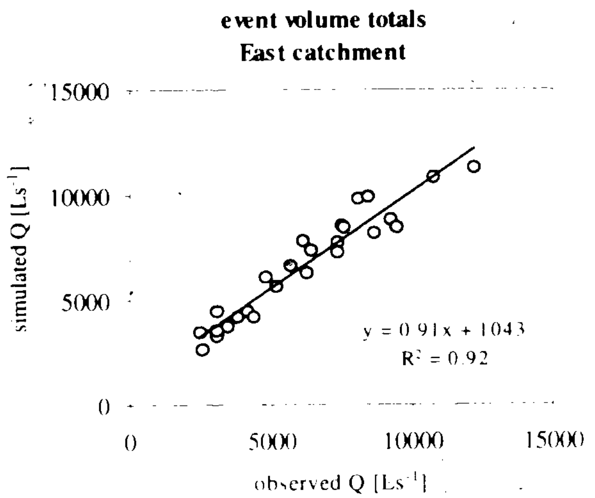
## Level 1 Dual parallel reservoir lumped model

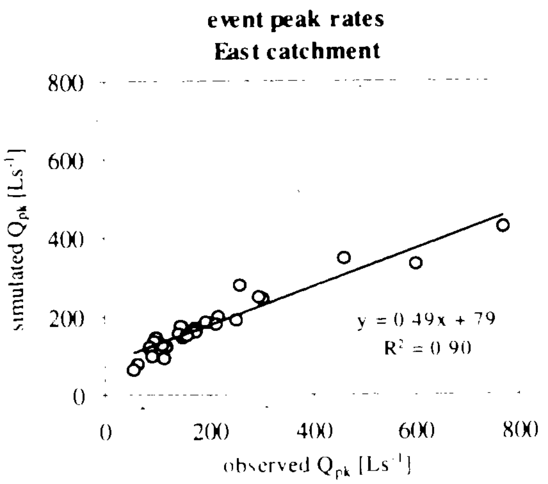


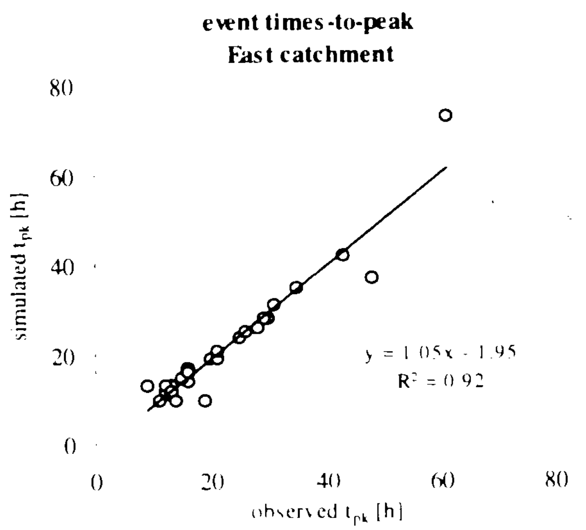



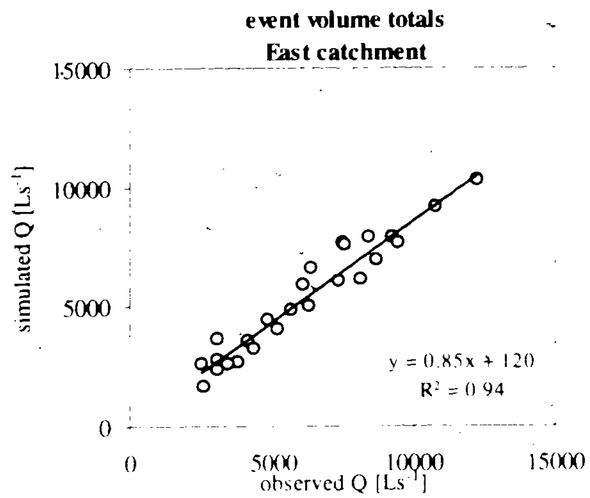
Level 1 Triple serial reservoir lumped model

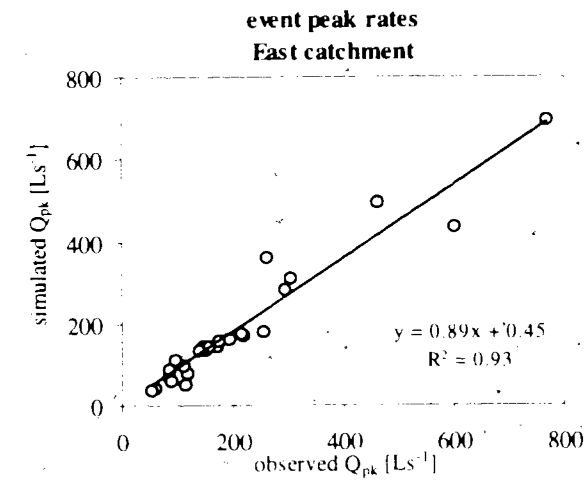


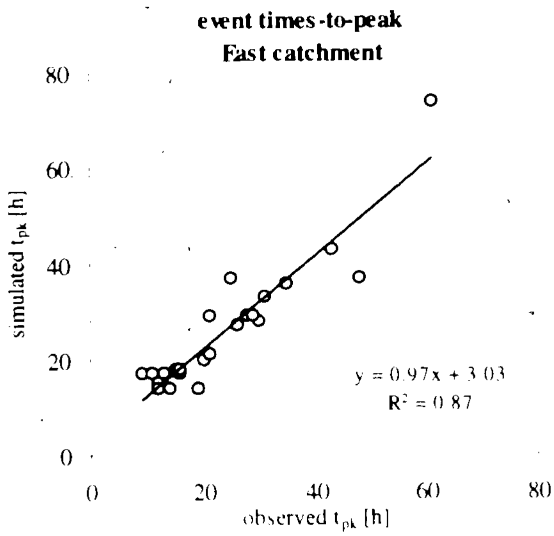



Level 1 Quasi-distributed model



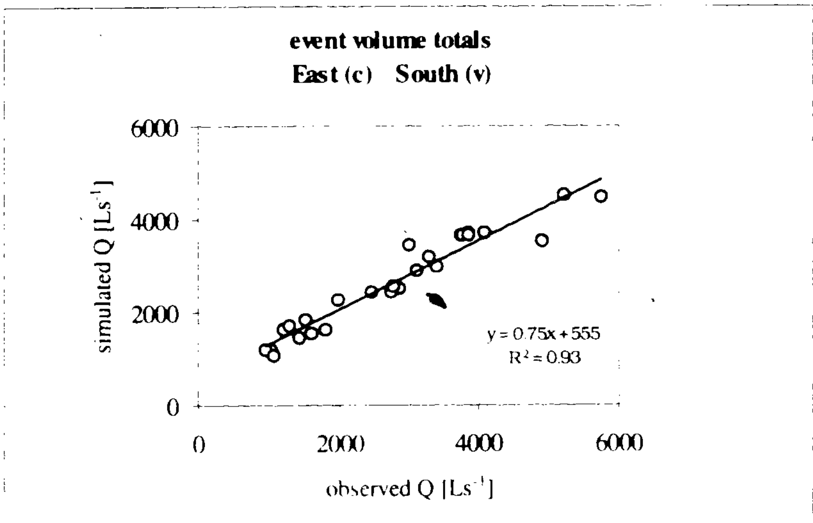



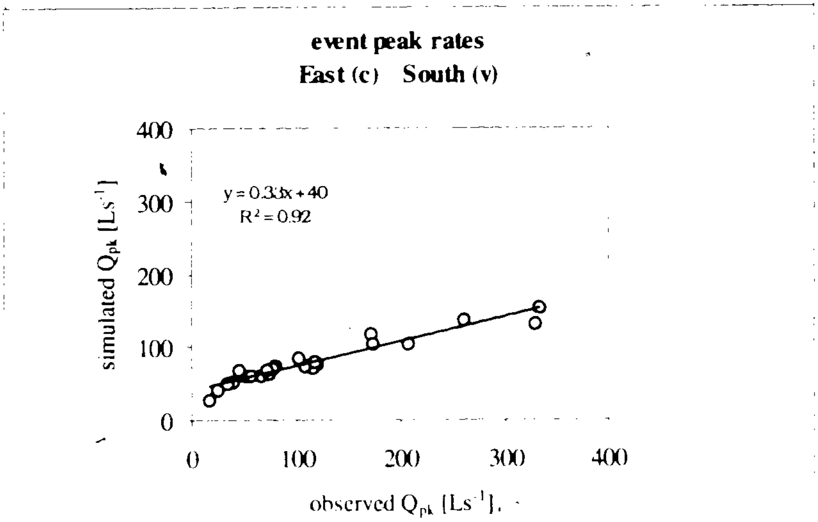
## SCATTERPLOTS

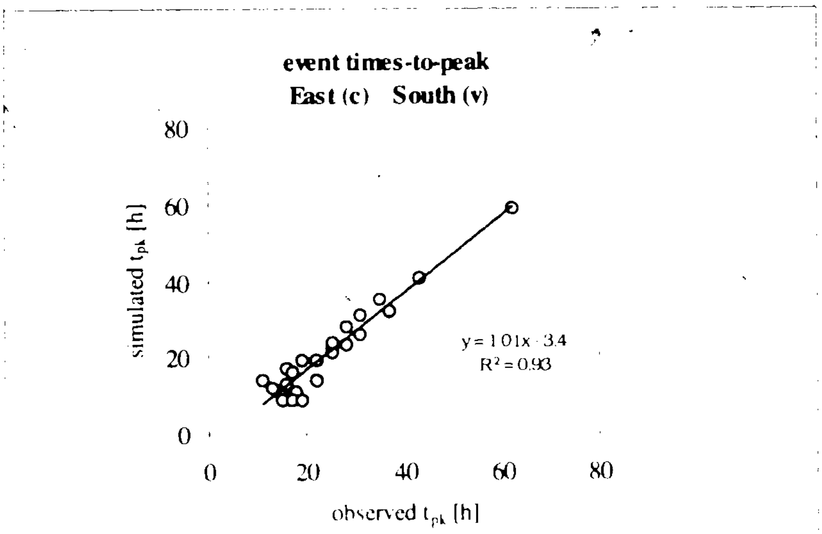
## OF VERIFICATION DATA SETS

### Level 2 Proxy-basin test

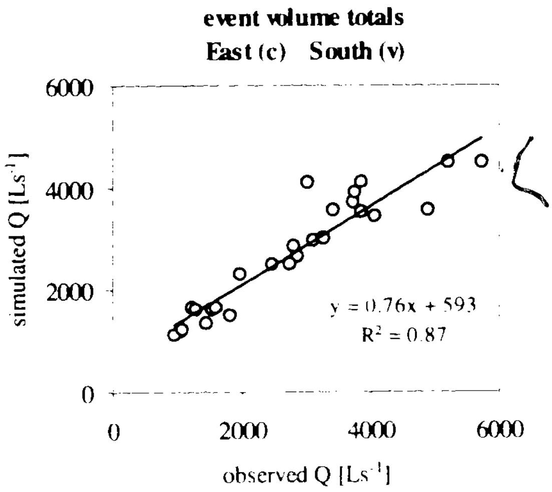
Level 2 Dual parallel reservoir lumped model

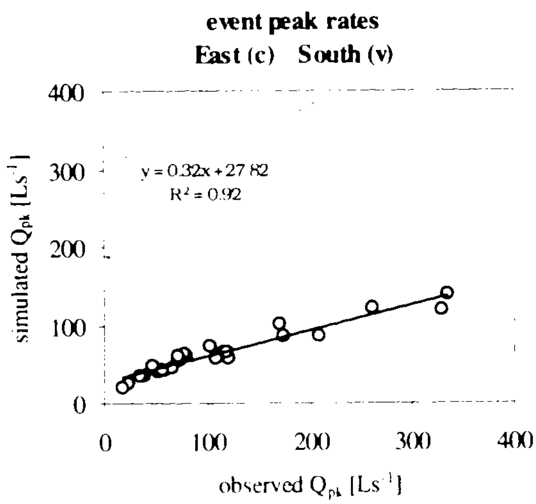


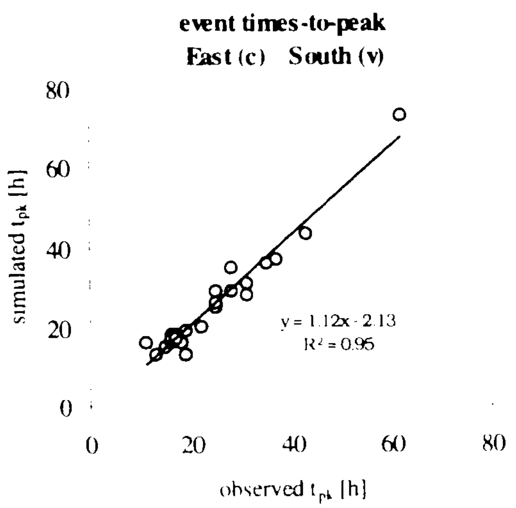



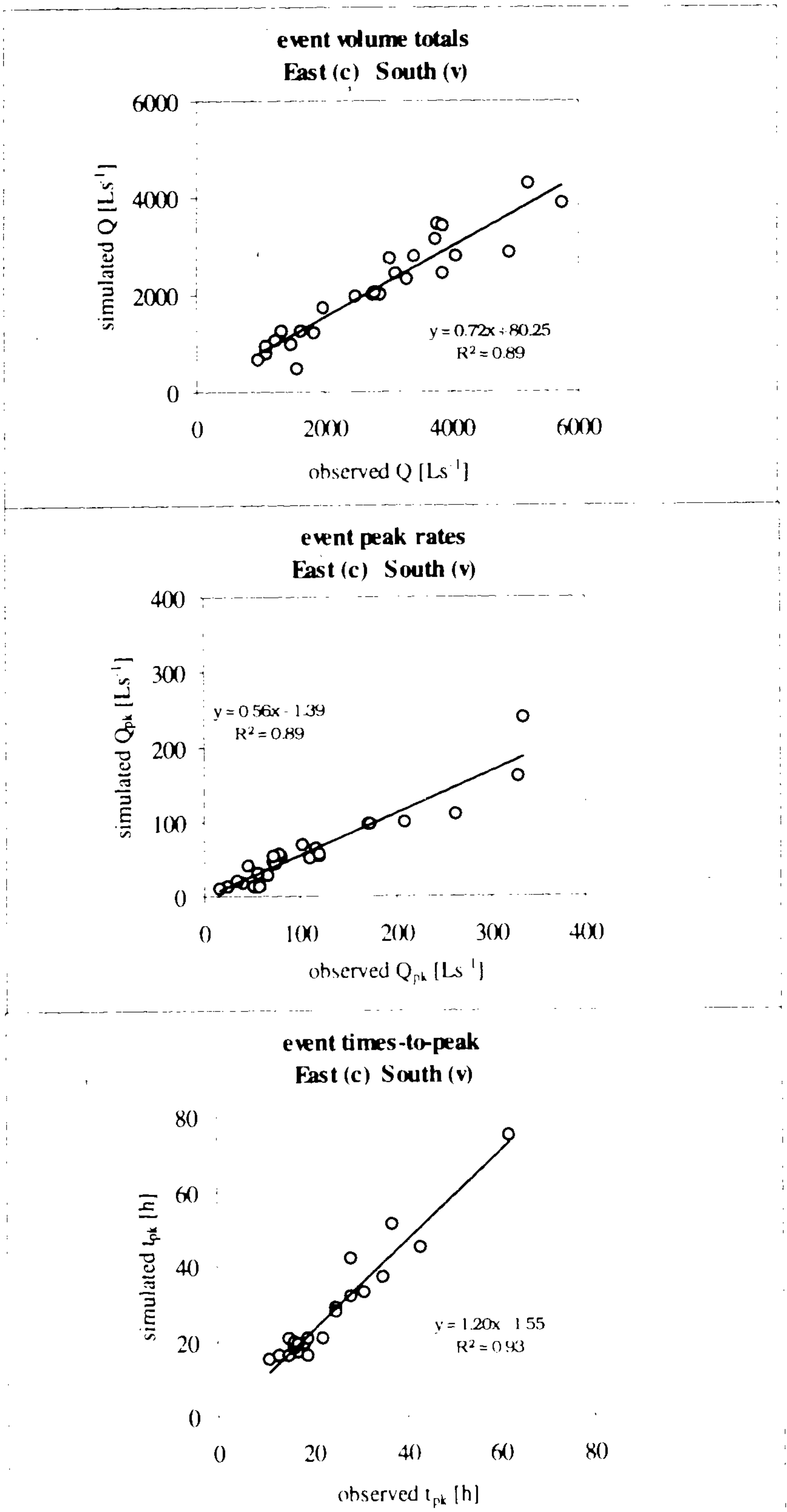



Level 2 Triple serial reservoir lumped model

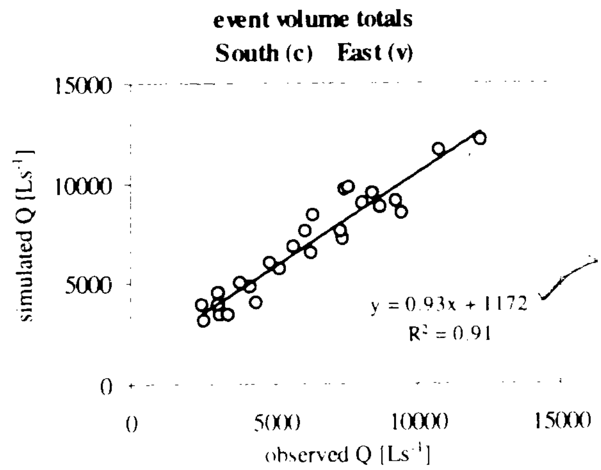


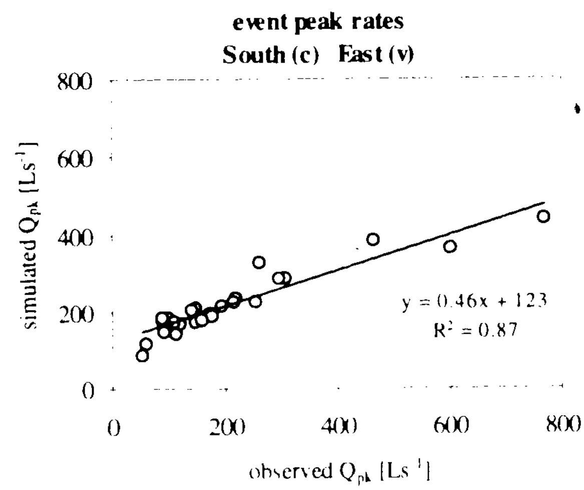



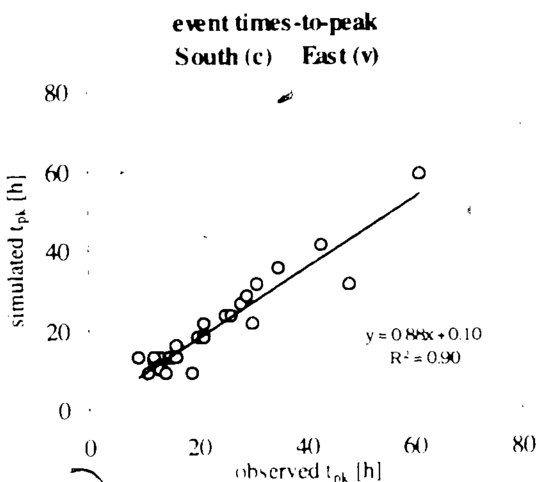




event volume totals
East (c) South (v)
simulated Q [Ls⁻¹]
observed Q [Ls⁻¹]
y = 0.72x + 80.25
R² = 0.89
0
2000
4000
6000
event peak rates
East (c) South (v)
simulated Qpk [Ls⁻¹]
observed Qpk [Ls⁻¹]
y = 0.56x - 1.39
R² = 0.89
0
100
200
300
400
event times-to-peak
East (c) South (v)
simulated tpk [h]
observed tpk [h]
y = 1.20x - 1.55
R² = 0.93
0
20
40
60
80

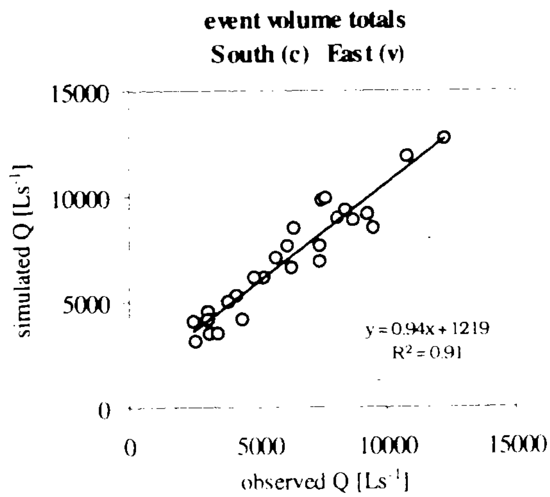
Level 2 Dual parallel reservoir lumped model

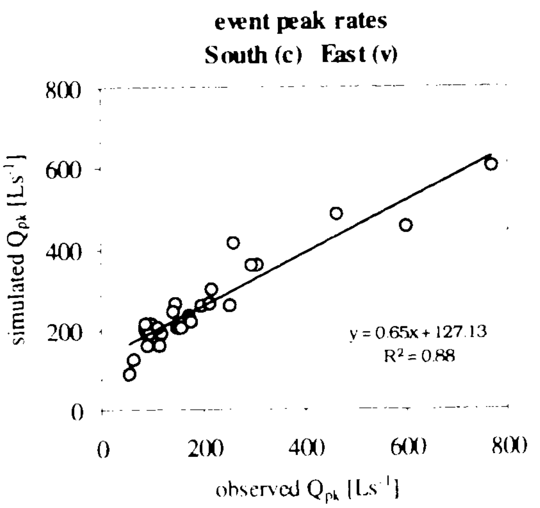


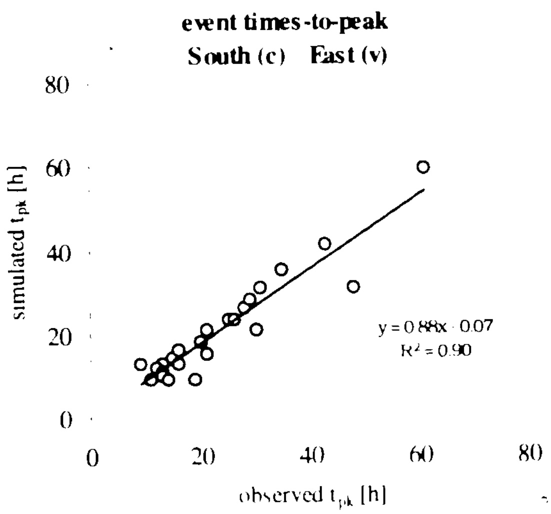



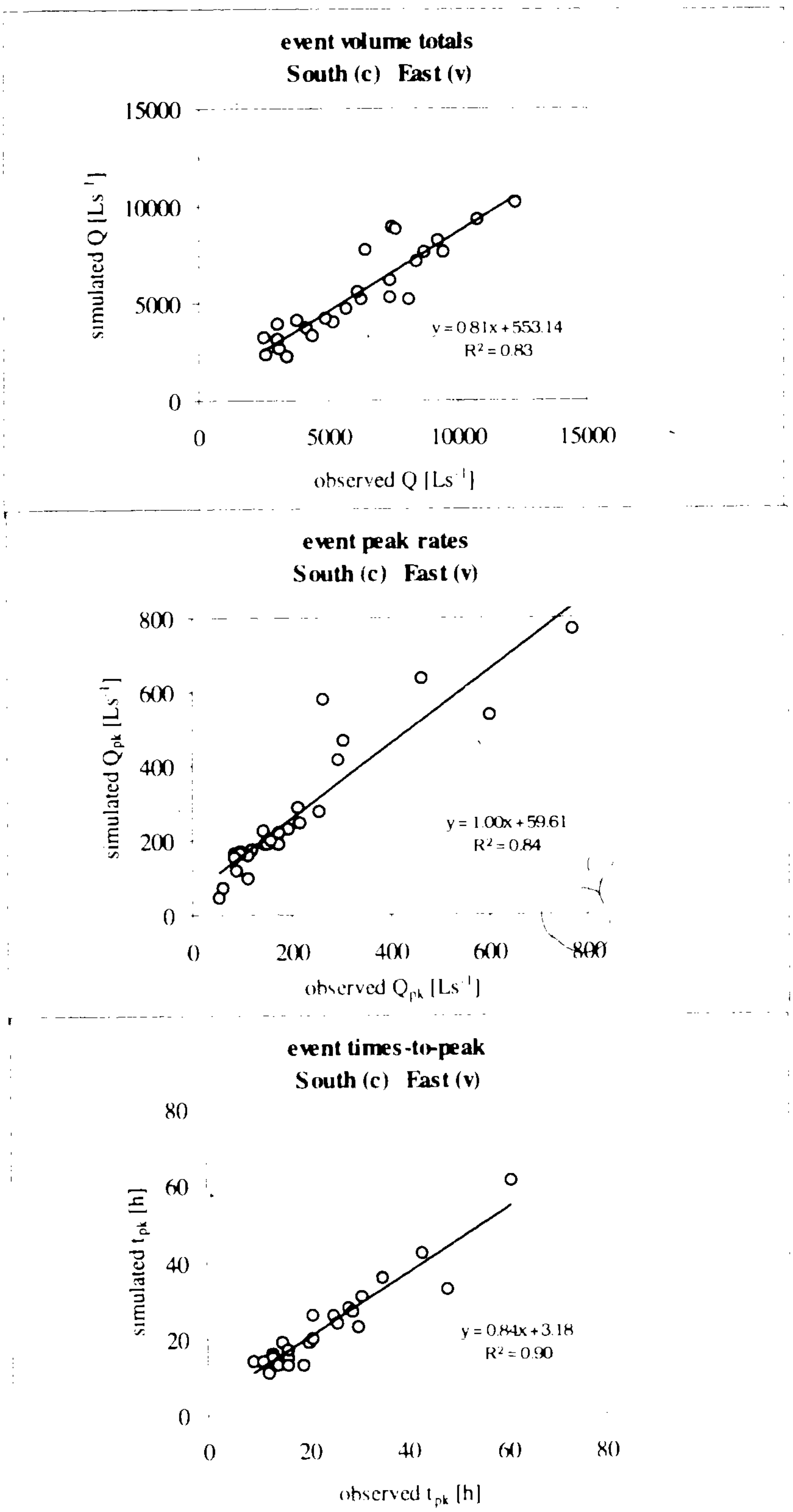



Level 2   Triple serial reservoir lumped model



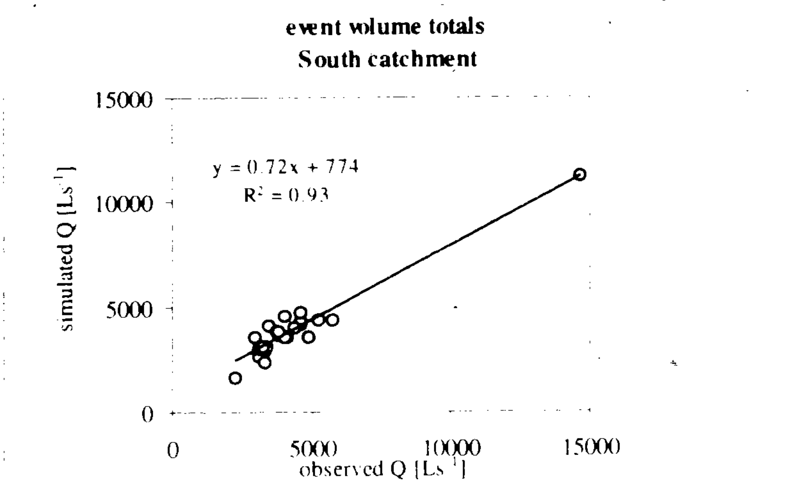

event volume totals
South (c) Fast (v)
simulated Q [Ls⁻¹]
observed Q [Ls⁻¹]
y = 0.81x + 553.14
R² = 0.83
0
5000
10000
15000
event peak rates
South (c) Fast (v)
simulated Qpk [Ls⁻¹]
observed Qpk [Ls⁻¹]
y = 1.00x + 59.61
R² = 0.84
0
200
400
600
800
event times-to-peak
South (c) Fast (v)
simulated tpk [h]
observed tpk [h]
y = 0.84x + 3.18
R² = 0.90
0
20
40
60
80

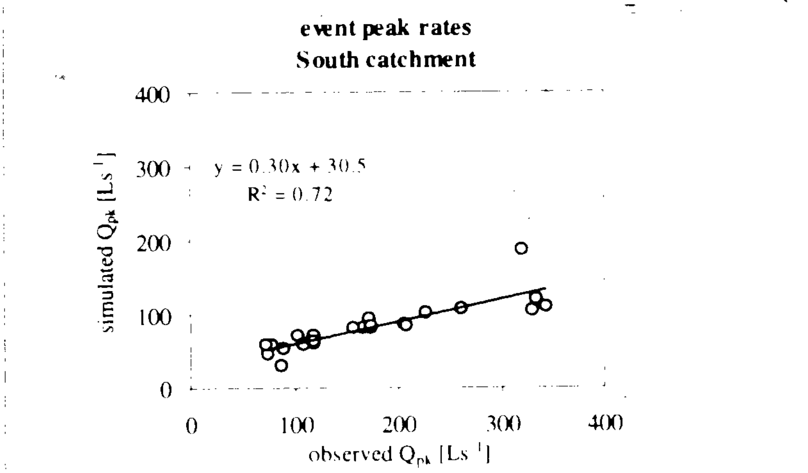
SCATTERPLOTS

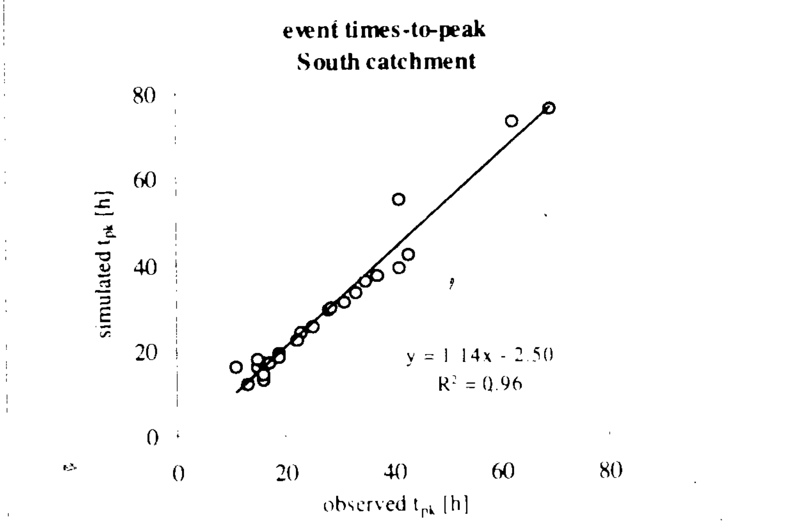
OF VERIFICATION DATA SETS

Level 3 Differential split-sample test

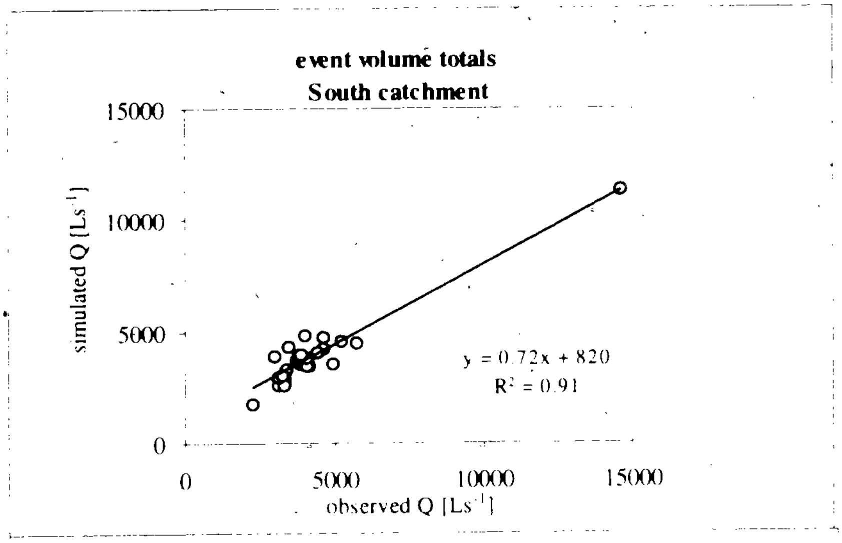
Level 3 Dual parallel reservoir lumped model

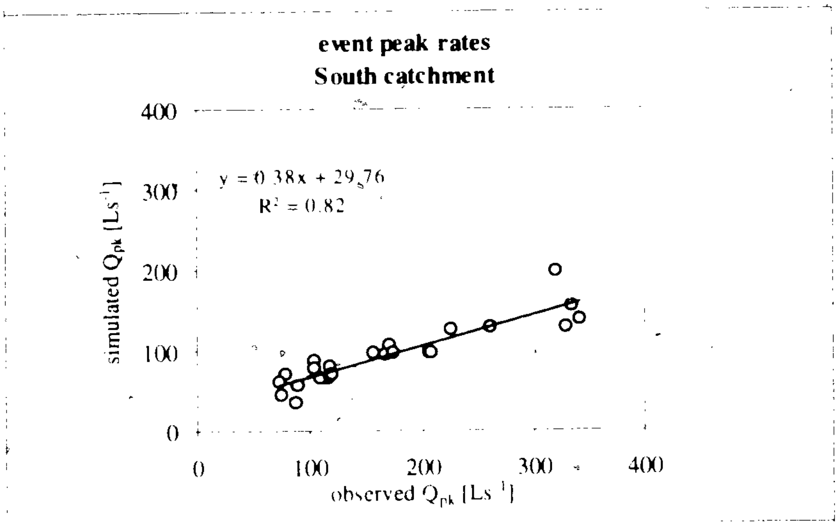


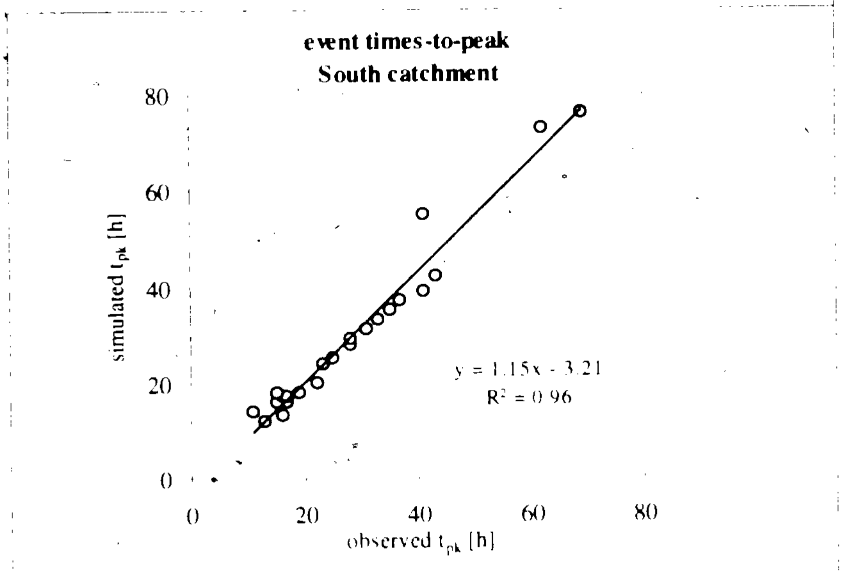



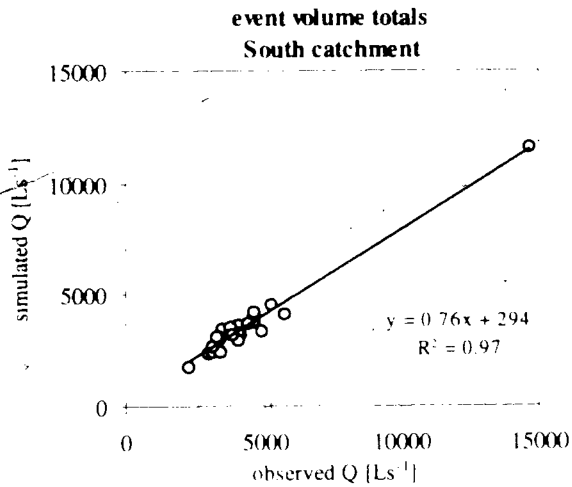
Level 3 Triple serial reservoir lumped model

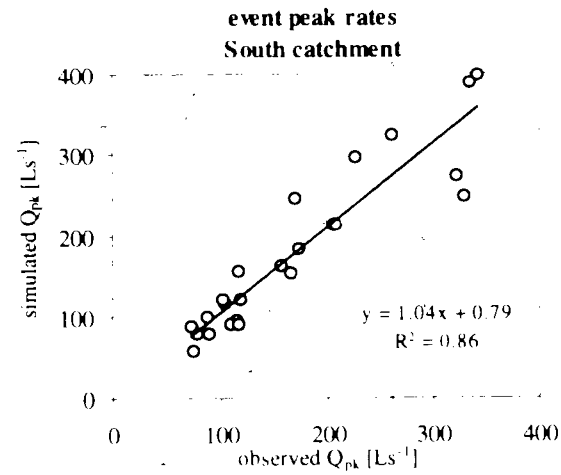


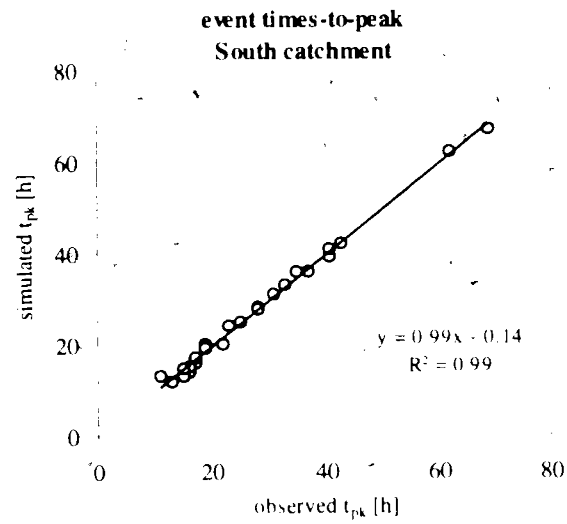



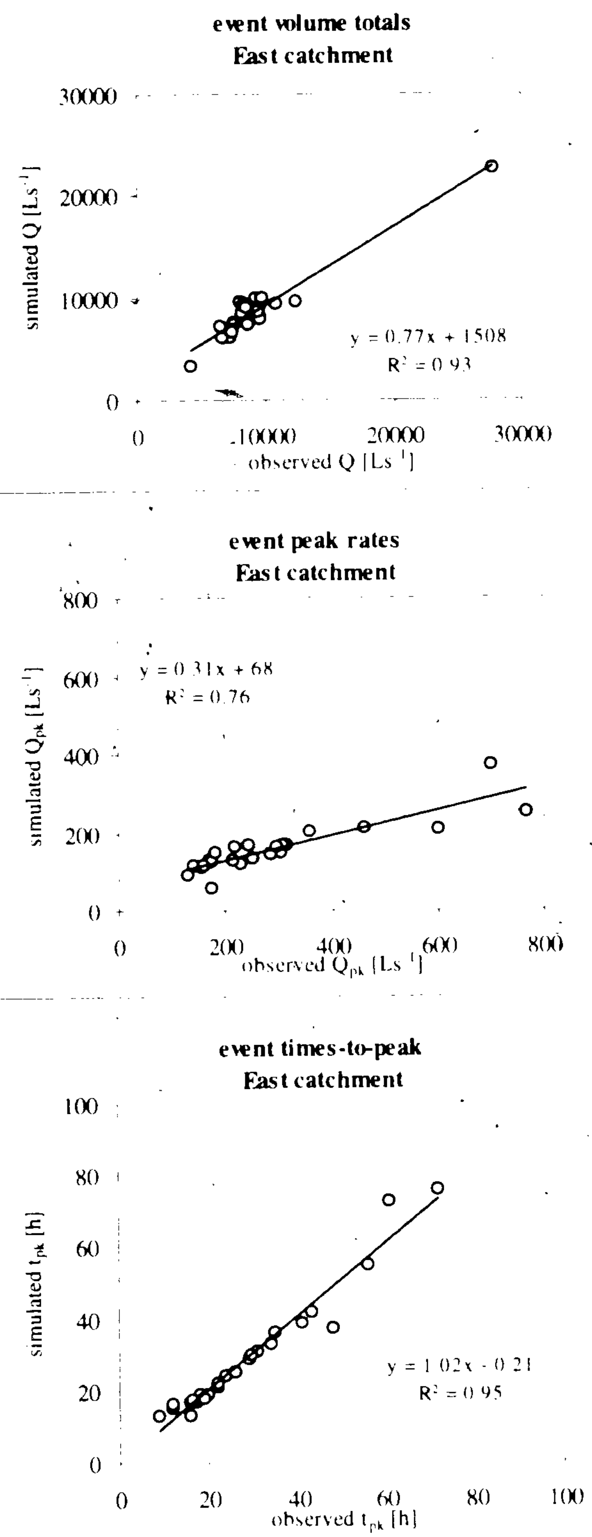
Level 3 Quasi-distributed model



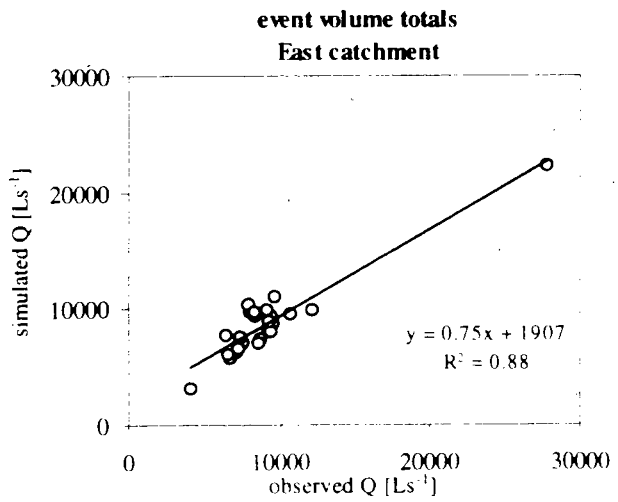



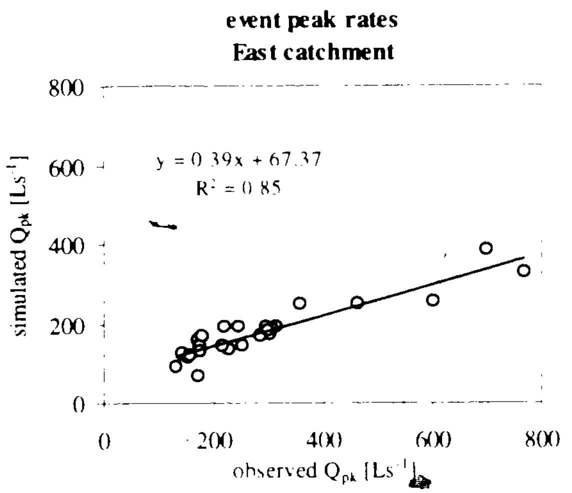



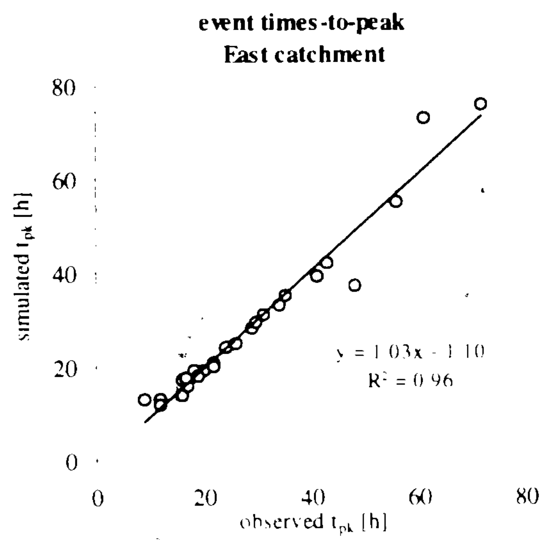
Level 3 Dual parallel reservoir lumped model



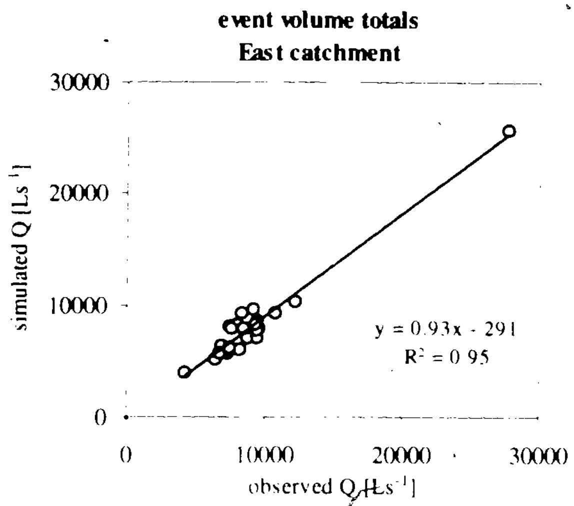
Level 3 Triple serial reservoir lumped model

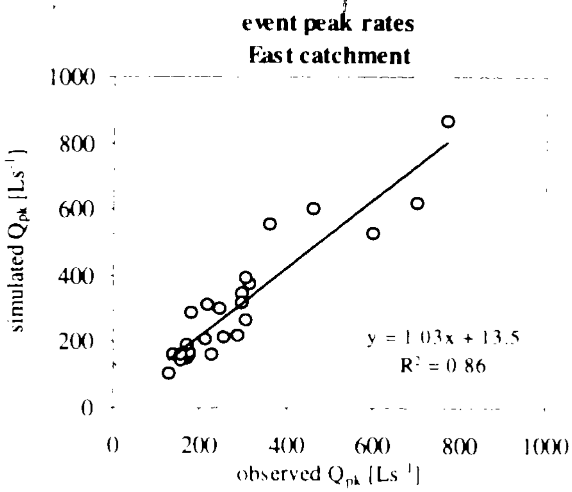


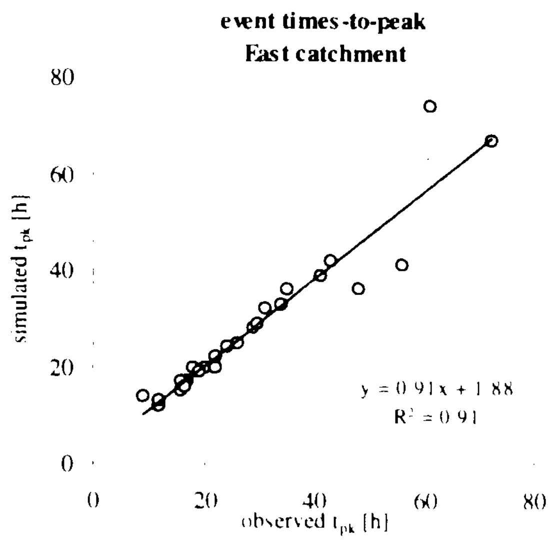



Level 3 Quasi-distributed model



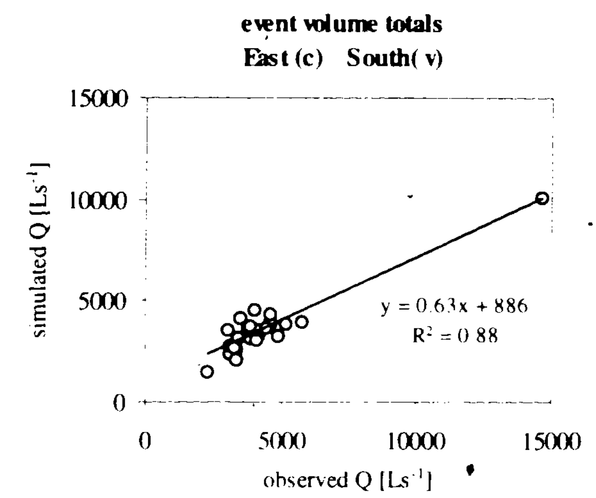



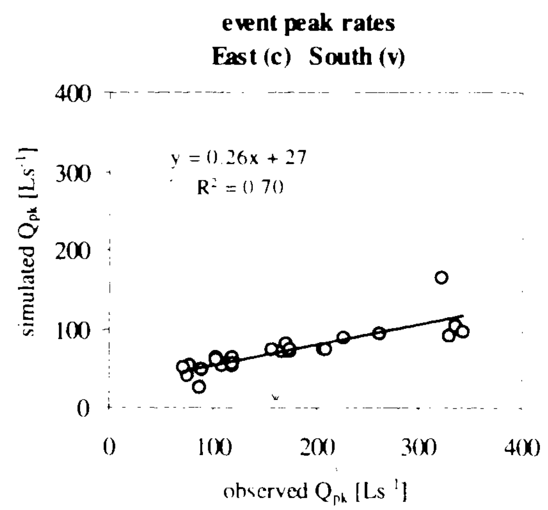
## SCATTERPLOTS

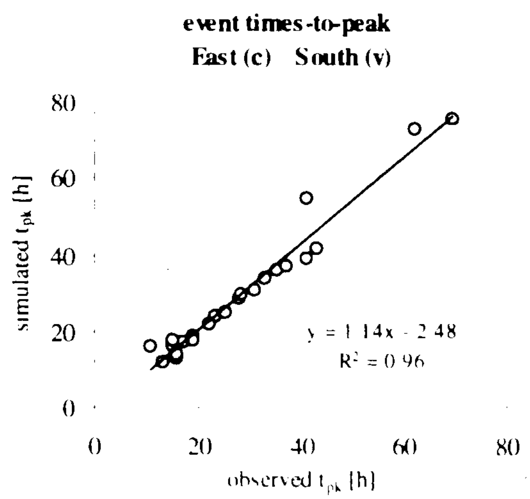
## OF VERIFICATION DATA SETS

**Level 4 Proxy-basin differential split-sample test**

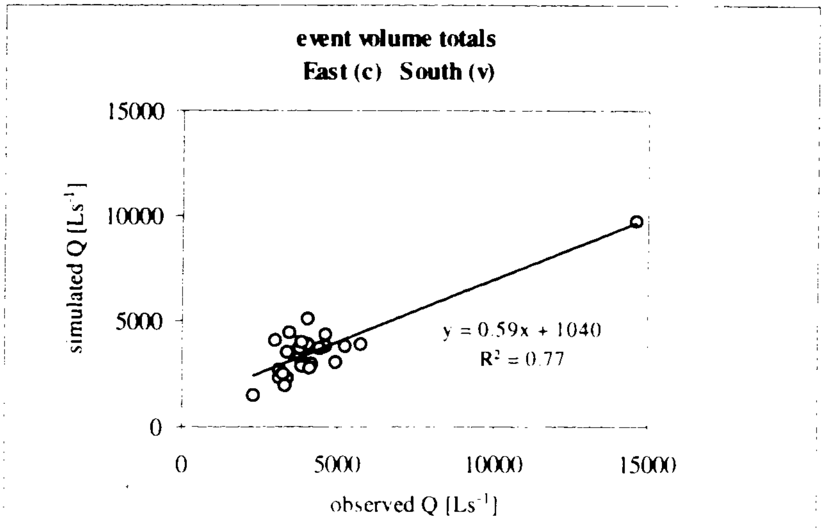
Level 4 Dual parallel reservoir lumped model

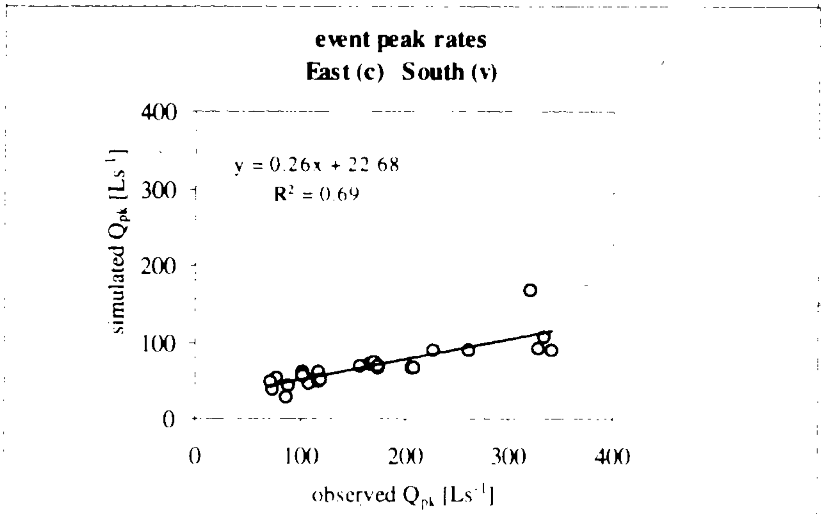


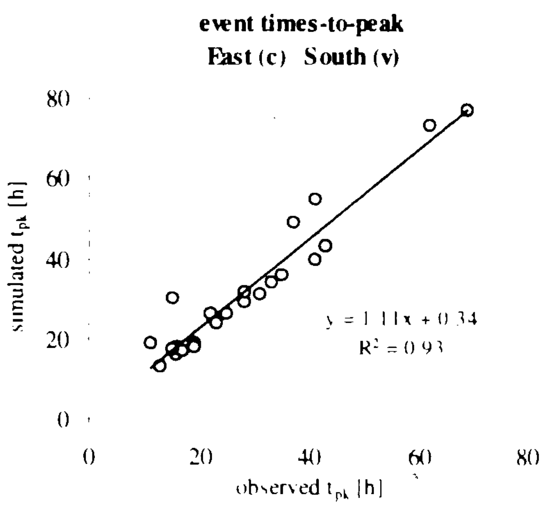



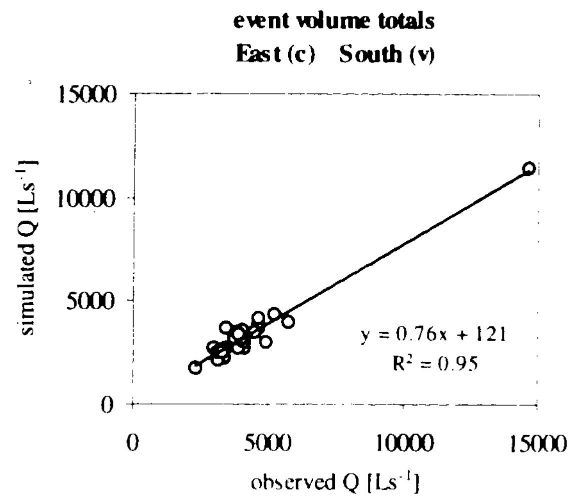
Level 4 Triple serial reservoir lumped model

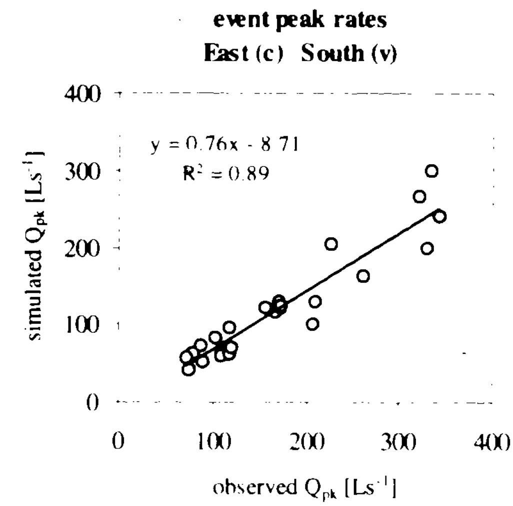


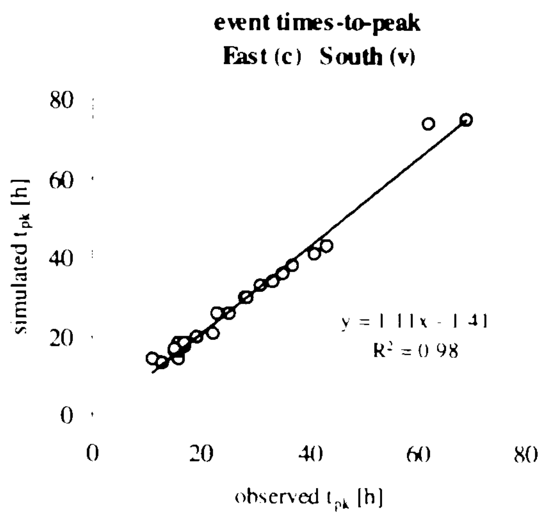



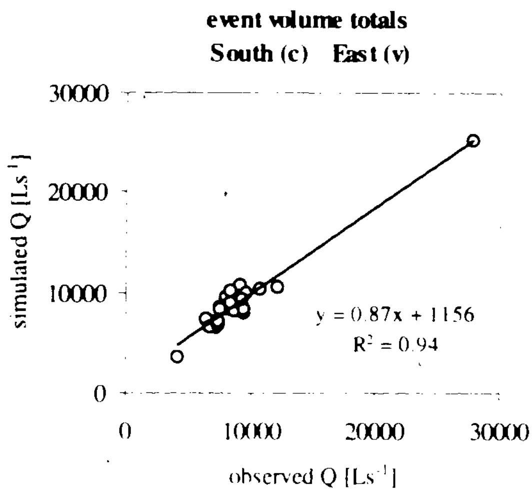
Level 4 Quasi-distributed model

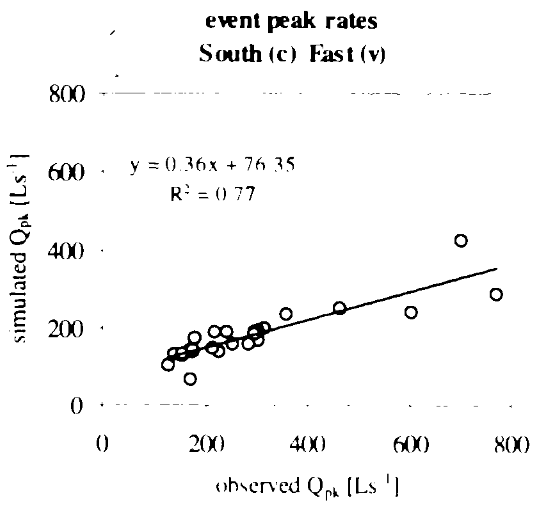


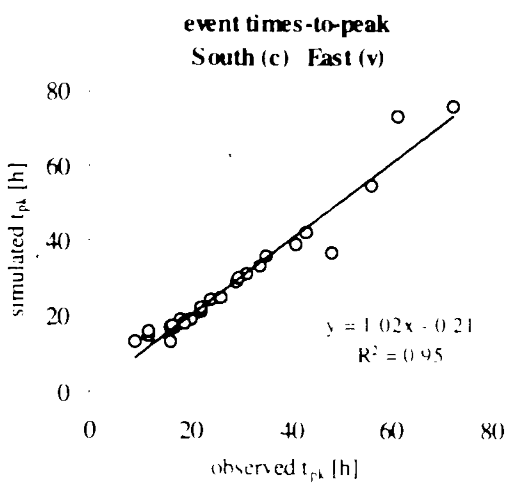



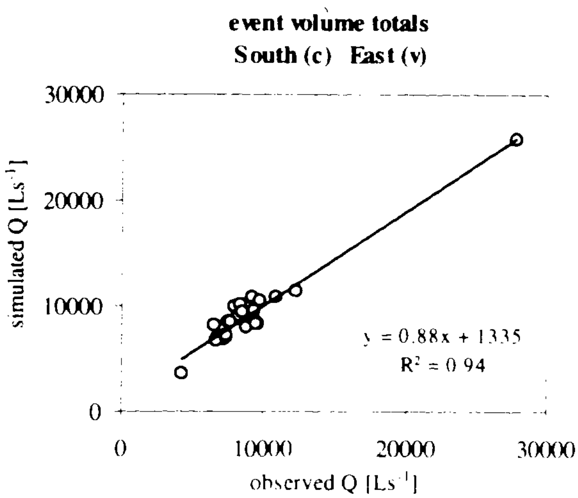
Level 4 Dual parallel reservoir lumped model

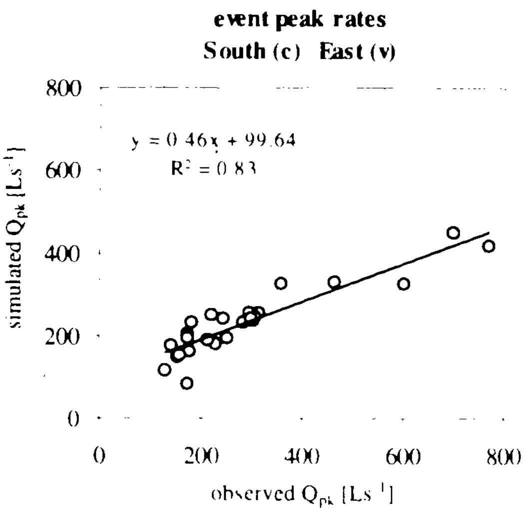


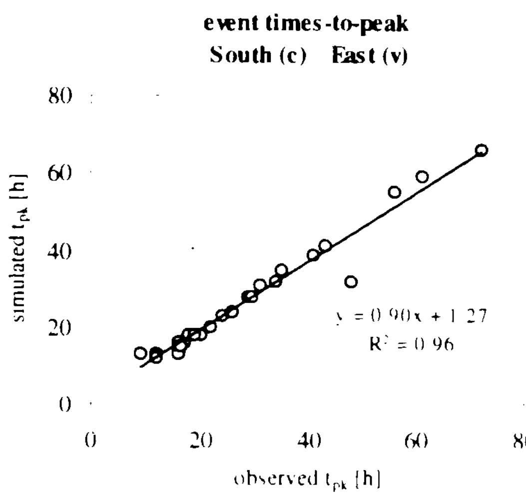



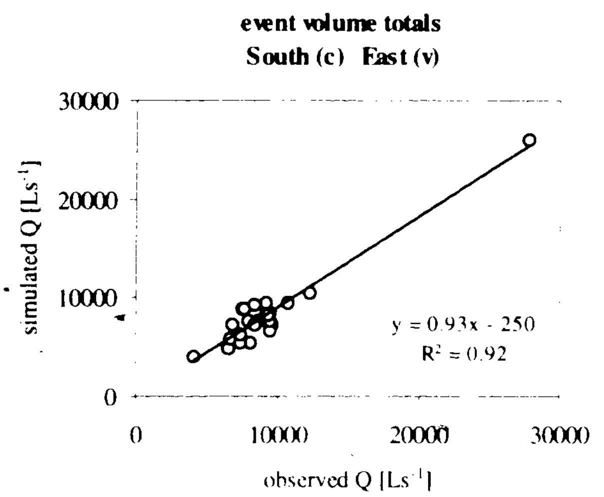



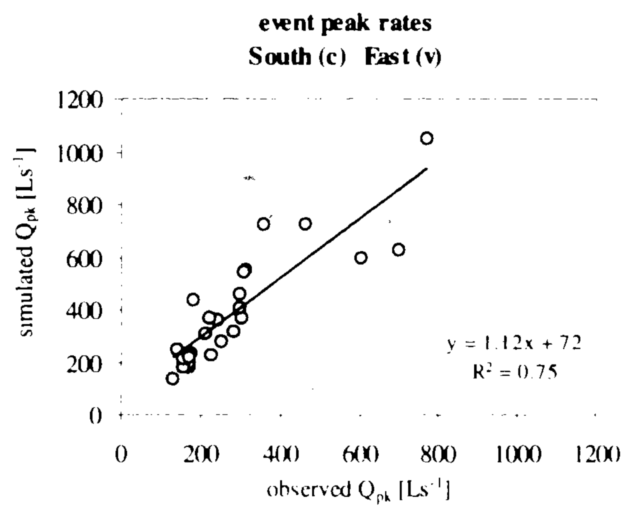
Level 4 Triple serial reservoir lumped model

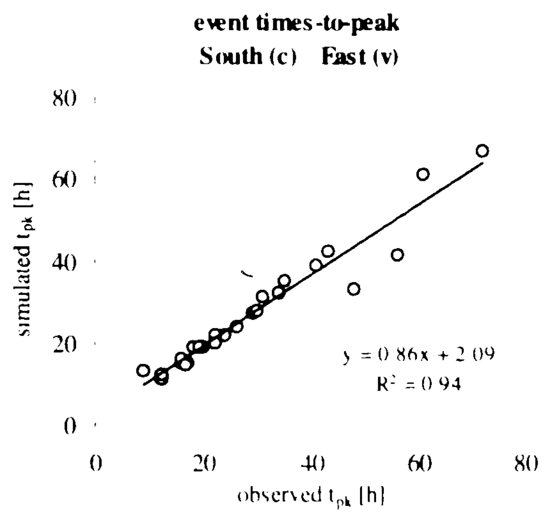



event volume totals
South (c) Fast (v)
simulated Q [Ls-1]
observed Q [Ls-1]
0
10000
20000
30000
y = 0.93x - 250
R² = 0.92


event peak rates
South (c) Fast (v)
simulated Qpk [Ls-1]
observed Qpk [Ls-1]
0
200
400
600
800
1000
1200
y = 1.12x + 72
R² = 0.75


event times-to-peak
South (c) Fast (v)
simulated tpk [h]
observed tpk [h]
0
20
40
60
80
y = 0.86x + 2.09
R² = 0.94