Global Sensitivity Analysis of the WetSpa model for the Ve river in Vietnam

Tom Doldersum

Global Sensitivity Analysis of the WetSpa model for the Ve river in Vietnam

Bachelor Thesis
University of Twente
Enschede - The Netherlands

Executed at Hanoi University of Science
Hanoi – Vietnam

By T. Doldersum
July 2009

Supervisors :
Hanoi University of Science: N.T. Giang (PhD, deputy director)
University of Twente: Dr. ir. M.j. Booij

Frontpagepicture: The beautiful nature of Vietnam at the Perfume Pagode

Summary

Global climate change has been widely perceived as one of the main reasons leading to an increase in frequency and magnitude of hydro-meteorological extreme events. These extreme events can lead to flooding. Flooding hinders the socio-economic development on both national and global scale. Flood prediction is an important instrument. The Hanoi University of Science's goal is to contribute to raise the degree of accuracy in flood forecasting for the Ve River Basin.

This research is done within the framework of the Bachelor Thesis. The assignment was to calibrate only for floods the WetSpa model. The WetSpa model is a GIS based-distributed model for flood prediction and water balance simulation on catchment scale. The study area for this thesis is the Ve river basin located in central Vietnam near Quang Ngai. First there is done a preliminary calibration to get a feeling with the model and the study area. Furthermore this calibration has to facilitate determining the ranges for the sensitivity and uncertainty analysis. The second topic of the research was to execute a sensitivity analysis with the method introduced by Morris (1991) to find out the most influential inputs and parameters with regard to the simulated discharges.

For calibrating and validating the model three datasets were available. Of which two datasets have been used for calibrating the model. The results for the calibration were rather good, both datasets resulted in a Nash-Sutcliffe coefficient of 0.88. The preliminary verification of the model has been done with the split sample test, by using the third dataset. The result was rather poor, the Nash-Sutcliffe coefficient was equal to 0.57. Moreover the hydrograph of the simulated discharges was also rather poor, only the pattern of the simulated discharges was nearly the same. But the peak values of the simulated flood did not make sense at all, comparing them with the observed values. There are a couple of reasons for this poor verification. First some Vietnamese data is converted into Arcview standards, for instance in Arcview there are 14 land use classes used but in the Vietnamese data there were just 7 classes. Second the warming-up time could be too short. Third there were some struggles with Arcview resulting in not changeable initial conditions, which could have a negative effect on the results. Fourth the Ve river basin becomes an open-basin during an extreme rainfall event, this effect could not be taken into account within this version of the WetSpa model.

The method introduced by Morris (1991) is a sensitivity analysis which takes each run one parameter or input variable into account. The main advantage of this method is the relatively low computational cost, the number of runs are needed is equal to a linear function of the number of examined factors. The main disadvantage is that the method cannot estimate individual interactions among factors.

The most important results of the sensitivity analysis were that the WetSpa model is not sensitive for the time till the peak discharge is reached, which was also found during the calibration process. Further the most sensitive parameter for the value of the peak discharge and the total volume of the flood is the groundwater recession coefficient.

The most important conclusions were that for better results it is important to have more data and more hourly data. For further research it is also recommended to integrate the steps of Arcview and the WetSpa model into one new program. So the whole program can run automatically which increase the calibrating process. Moreover there could be done a sensitivity analysis which takes all parameters and input variables into account.

Preface

At 20 March I arrived at the Noibai Airport in Hanoi. This was the start of a 18-week experience in Vietnam, during the first 12 weeks I have done my Bachelor Thesis. Therefore I want to thank first Mr. Giang for the opportunity to work in this country on a very exciting topic. Also I want to thank Mr. Liu for his quick answers and his advice about the WetSpa model. At least I want to thank Mr. Booij for helping me with searching for a suitable place to do my thesis, his quick answers and his comments during my stay.

I would like to thank Mrs. lieu, Mr. Tuyet and Hoai for washing my clothes, introducing me to the Vietnamese culture and the Vietnamese kitchen. I never forget that Mrs. Lieu took care that I clean my room and I will always remember the friendly 'Goodmorning' from Mr. Tuyet.

Also special thanks to Chi, who helped me and Daniel to change the fortrancode of the WetSpa model to make it usable for more than one run. Beside the project I want to thank her for inviting me to go with the fieldtrip of the faculty, which was a great experience. Further she arranged that I and Daniel could enjoy an engagement party and several weddings. Moreover she shows us the beautiful landscape around the Perfume Pagoda. Another special moment to remember is the trip to her hometown in the rural areas in the Northern of Vietnam where her family lived and where we enjoyed the Vietnamese kindness.

Next to these Vietnamese people, I want to thank Daniel for his support. Together we discussed a lot about the data, the model, the progress of both reports and about the problems we met. Although I did not know Daniel well before, in Vietnam we became real friends. Together we played Badminton against the Vietnamese students, visited many places in Hanoi , drove motorbike trough Hanoi and we hope to explore Vietnam together after we finished our reports.

At least I want to thank my homefront. First I want to thank my parents for their listening ear and making this experience possible. Also I want to thank my brothers for the conversations we had.

Hanoi , 12 June 2009
Tom

List of abbreviations

CR	Criterion
DEM	Digital Eelevation Model
GIS	Geo Information System
NS	Nash-Sutcliffe
OAT	One At a Time
PET	Potential EvapoTranspiration
WetSpa	Water and Energy Transfer between Soil, Plants and Atmosphere
HUS	Hanoi University of Science

List of symbols

B	Parameter within the WetSpa model; controls the shape of the variation curve for the interception capacity. (-)
d_j	Elementary effect
G_0	Initial groundwater storage (mm)
G_{max}	Maximum groundwater storage (mm)
i	Number of time step
k	Number of parameters
K_g	Groundwater recession coefficient (m^2/s)
K_i	Scaling factor for interflow computation (-)
K_{run}	Surface runoff exponent for a near zero rainfall intensity (-)
K_{ss}	Initial soil moisture (-)
m	m=k+1
M	Parameter within the WetSpa model; is an exponent for calculating the groundwater flow at the subcatchment outlet
Model1	Semi-distributed model
Model2	Fully-distributed model
N	Number of time steps over the simulation period
P	Number of levels
Parinput	Matrix with the values for the different parameters and input variables for the sensitivity analysis.
P_{max}	Threshold rainfall intensity (mm/hour)
Qo_i	Observed stream flow (m^3/s)
Qs_i	Simulated stream flow (m^3/s)
r	Number of runs
Rain	Input rainfall data for the WetSpa (mm)
Set	Possible base values (-)
x*	Base values (-)
y	Output of the WetSpa model (m^3/s)
Δ	Difference between rows in the B* matrix
μ	Mean
$\overline{Qo}$	Mean observed stream flow (m^3/s)
σ	Standard deviation

Table of contents

List of figures

List of tables

1 Introduction

Global climate change has been widely perceived as one of the main reasons leading to an increase in frequency and magnitude of hydro-meteorological extreme events as is shown by Karl et al. (1995), Metz et al. 2007 and Tsonis (1996). These extreme events can lead to flooding. Flooding hinders the socio-economic development on both national and global scale. Flood prediction is an important instrument to inform habitants about a coming flood, which causes that the damaging effect of flood events will reduce. But according to Krzysztofowicz (2001) flood prediction remains far from perfect, and falls short of society's expectations for timely and reliable warnings.

The flood forecasting project for the Ve River Basin of the Hanoi University of Science (HUS) contributes to the field of disaster prevention. The main goal of the project is to raise the degree of accuracy in flood forecasting of the Ve River Basin. Therefore a model which supply valuable information is necessary. To achieve this goal first an accurate flood forecasting model is needed and second a sensitivity and an uncertainty analysis have to be done.

There are many hydrological models which are suitable for flood forecasting, for instance the CASCD2 (Downer, et al, 2002) and HYDROTEL (Fortin, et al, 2001). However within this research the WetSpa model (Liu & De Smedt, 2005) will be utilized, because this model is used before by the other members of the research group. Furthermore the WetSpa model is suitable for this study area. Before the uncertainty analysis can be done the most influential parameters and input variables on the discharge have to be known. In this research the sensitiveness of the parameters and input variables of the distributed hydrological model WetSpa to the simulated discharges are determined, using a global method. This is done to facilitate the fast calibration and verification of the model in practical application.

The objective of this study is to find out the most influential parameters and inputs of the WetSpa model with regard to the simulated discharges, to facilitate the flood forecasting application of the model to Ve river basin.

The steps which have to be done before the model can be utilized in practice are shown in Figure 1. Since there is a time limit for this research only the grey part of the scheme has been done. There was data available, but before starting with calibrating an inventory of these data has to be made, because not all data was correct or useful. Then the preliminary calibration and verification of the model has been done to get a feeling with the model. Afterwards the ranges of the parameters and inputs for the sensitivity analysis are determined.

The second part of this thesis is to execute the global sensitivity method introduced by Morris (1991) for the WetSpa model. A sensitivity analysis has been done before by Bahremand en De Smedt (2008) but they executed the PEST method for the WetSpa model which is a local method. One of their reccomendations was:" For further studies, using a global method for the sensitivity analysis is recommended. Local methods, such as the PEST method, use only local information that might not represent the entire parameter space, therefore, the results yield only local sensitivities or uncertainties." (Bahremand & De Smedt, 2008)

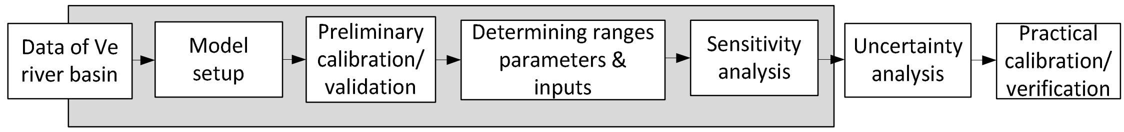


Figure 1: Scheme before the discharges could be predict

The WetSpa model is utilized within this research. Actually the WetSpa extension is meant. But for practical reason WetSpa extension is called WetSpa model or just WetSpa. The aim of the extension is not only at predicting flood, but also investigating the reasons behind, especially the spatial distribution of topography, land use and soil type (Liu & De Smedt, 2004). For the improvements of the model comparing to the original WetSpa model interested readers are referred to the user manual written by Liu and De Smedt (2004).

In chapter 2 a brief description of the study area will be presented. Afterwards the WetSpa model, data and the calibration process are described in chapter 3. In chapter 4 the sensitivity analysis method, introduced by Morris (1991), is presented. In chapter 5 the results will be discussed and finally in the chapter 6 the conclusions and some recommendations for further research are described.

2 Study area

The study area of this research is the upstream part of the Ve river basin. The Ve river is located in the central coast region of Vietnam. This part of the report gives information about Vietnam and the study area.

2.1 Vietnam

Vietnam is situated in South-East Asia, its official name is the Socialist Republic of Vietnam. It is bounded by Laos to the west, Cambodia to the southwest, China to the north and the East Sea to the east. This information is shown in Figure 2.

Figure 2: Location of Vietnam in Asia after (Wikimedia Commons) and between its neighboring countries (NCBUY, 2009)

The capital Hanoi is located in the north of Vietnam. The other large city of Vietnam is located in the South, Ho Chi Minh City, previous called Saigon. Some general information is listed in Table 1. Vietnam consists of 63 Provinces. The study area of this research, the Ve river basin, is located in the Quang Ngai province.

Table 1: General information about Vietnam (Wikipedia, 2009)

Vietnam	
Capital city	Hanoi
Official language	Vietnamese
Surface area	331,690 km^2
Water	1,3 %
Mid-year estimation of the population (2008)	86.116.559
Density	253/km^2

2.2 Study area

The Quang Ngai province is in the south central coast region of Vietnam. It is located 883 km south from Hanoi and 838 km north of Ho Chi Minh City. The Ve river is located south in the Quang Ngai province, shown in Figure 3. The total Ve river basin has a surface area of 1300 km^2; the main stream is 91 km long. Within this project only the upstream part from An Chi is taken into account, which has a surface of 757,32 km^2. The Ve River rises from the mountainous region Truong Son in the south and leaves the study area at An Chi. The study area is shown in the right part of Figure 3. The study area is a semi-open basin. Which means that during extreme circumstances water can flow in and out the Ve river basin, this does not occur during normal circumstances.

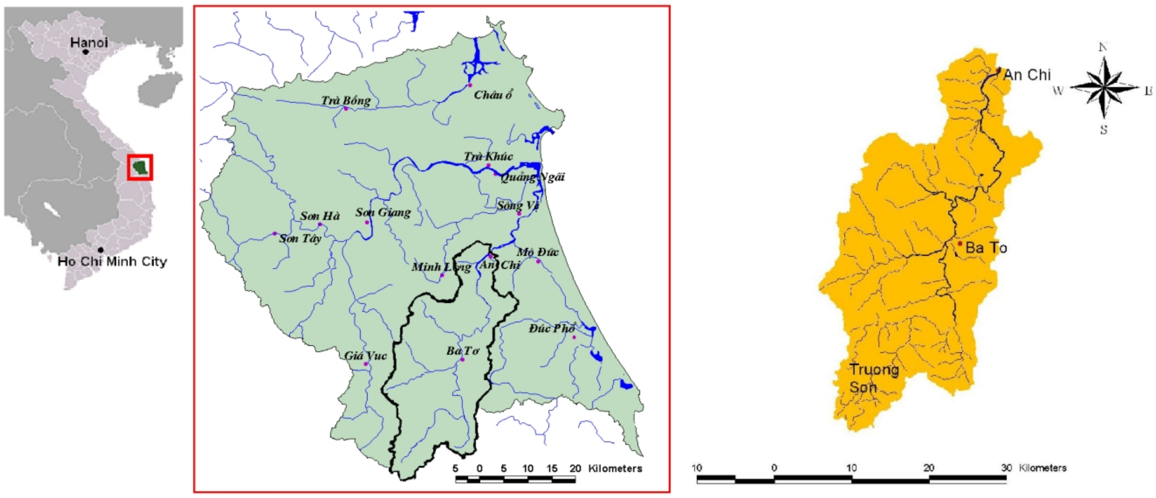


Figure 3: Location of the Quang Ngai Province in Vietnam (Wikipedia, 2006), and the Ve river upstream of An Chi inside Quang Ngai Province (Son, 2008)

For the Ve river basin, two problems in flood forecasting have priority. The degree of accuracy is very poor at the moment and the lead time of predicting the water level has to be improved (Son, 2008). In the next part characteristics of the study area are described.

2.2.1 Lithological characteristics

The study area consists of many lithological structures. The most conspicuous lithological characteristic of Ve river basin is a rapid change in topographical gradient in profile from the south to the north, shown in the DEM (Digital Elevation Model) in left picture in Figure 4. The right picture in Figure 4 shows the soil of the river basin. There are six different types of soil. In the mountainous region, sandy loam is the most common soil type and in the plain, sandy clay loam is the most common soil type (Son, 2008).

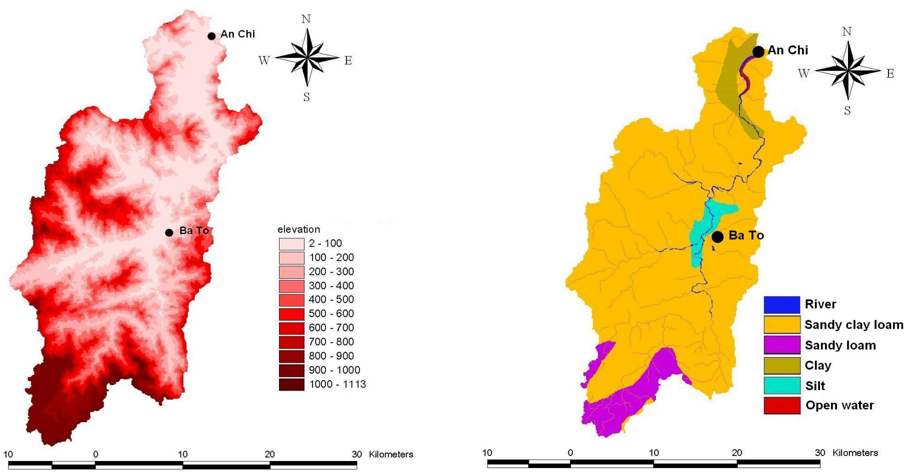


Figure 4: DEM (Left) and Soil type (Right) of the study area

2.2.2 Land use

The dominant land use of the study is deciduous shrub. In the mountainous regions in the south evergreen broad leaf tree cover the surface. There is also a substantial amount of irrigated crop in the study area. An overview of the land use is shown in Figure 5.

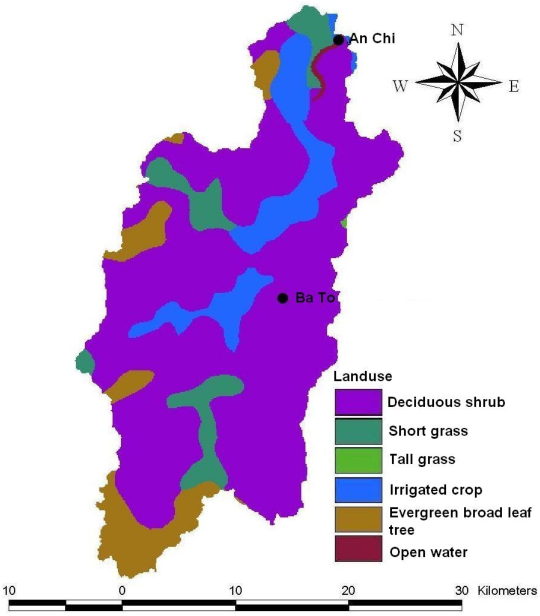


Figure 5: Land use map of the study area

2.2.3 Climatic conditions

The Ve river basin is situated to the south of the Hai Van pass, which separates the two main climate regions of Vietnam. South of the Van Hai pass, there is a moderate tropical climate. In this region of Vietnam the average annual temperature is about 26^{0}C.

The precipitation in the plain is about 2000-2200 mm yearly, upstream it exceeds 3000 mm. During the year there are approximately 140 rainy days. The rainy season starts in September and ends in December. The amount of rainfall during this rainy season is 65-85% of the total amount of annual precipitation. (Son, 2008)

3 Preliminary calibration of the WetSpa model

Before the sensitivity analysis is executed, a preliminary calibration and verification is done for the Ve river to get feeling with the model and to determine the ranges for the sensitivity analysis. Therefore the preliminary calibration process will be described in this chapter, the results will be presented in section 5.1. In the first section of this chapter a description of WetSpa is given, followed by a description of the data. In section 3.3 the preliminary calibration process is presented. Next the correlation between the different parameters and input variables will be determined in section 3.4. Afterwards in section 3.5 the criteria which will be used to evaluate the model are described. Followed by a description of the ranges of the different parameters and inputs to facilitate the preliminary calibration. Then the start values in Arcview for calibrating the model will be presented in section 3.7. Finally the manner how the model will be verified is described in section 3.8.

3.1 Model

The model used within this research is the WetSpa model. Therefore this model is presented in the first section of this chapter. Next the different parameters and input variables are described in section o. In section 3.1.3 the most sensitive parameter for model calibration will be described.

3.1.1 WetSpa model

"The WetSpa (extension) model is a GIS based-distributed hydrological model for flood prediction and water balance simulation on catchment scale" (Bahremand & De Smedt, 2008). WetSpa is an acronym for "Water and Energy Transfer between Soil, Plants and Atmosphere". It is a physically based model and the hydrological processes considered in the WetSpa model are precipitation, depression storage, snowmelt, surface runoff, infiltration, evapotranspiration, percolation, interflow, and groundwater flow. For detailed information about the formulas used to describe these processes, the User manual of Liu and De Smedt (2004) can be read. These processes and formulas are not described within this research.

WetSpa consists of two models: a semi-distributed model, and a fully-distributed model. The fully-distributed model has a large processing time. Therefore, for calibration the simpler semi-distributed model can be used but this one is less accurate. For calculating the results the complex fully-distributed model is used within this research.

Section 3.1.1.1 describes the necessary input for the WetSpa model. After this section 3.1.1.2 will give a brief description of the processes in the grid cells. Furthermore a few assumptions and limitations of the model will be discussed in section 3.1.1.3.

3.1.1.1 Arcview and WetSpa

The WetSpa model is a GIS (Geo Information System) based model, and therefore consist of two parts. The first part, ArcView, is used to read the geo-information data. This must be done before the second part of the model, the calculation with the WetSpa model, can be used. The process of loading the data in ArcView is time-consuming, because the model has to save all the data of the study area. The maps loaded in Arcview are used to calculate the values for new maps that are built in ArcView. This process is also time-consuming, because all steps must be taken manually. For example, the map of the Manning coefficient has to be loaded in Arcview. Afterwards this map is used for determining the velocities of the water in every grid cell.

During this loading process a few input values have to be set. These different input variables will be described in section o. It is important to choose these carefully because these maps are the basis for all calculations for the study area.

3.1.1.2 Grid cell

The model calculates the different types of discharges and the evapotranspiration for every grid cell separately. In Figure 6 the structure is presented at grid cell level. A short description of this process is given in the next part of this paragraph.

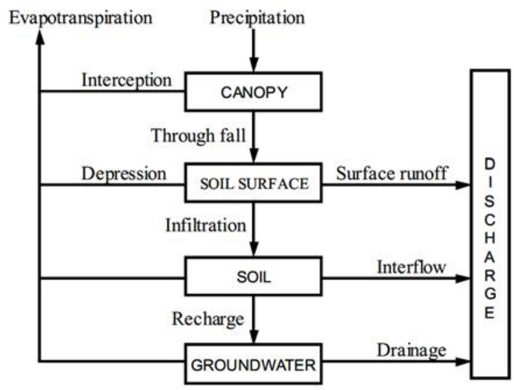


Figure 6: Structure of WetSpa Extension at a pixel cell level (Liu & De Smedt, 2004)

Incidental rainfall first encounters the plant canopy, which intercepts part or all of the rainfall till the interception storage capacity is reached. The rest of the water reaches the soil surface, where three different processes can take place. The water can infiltrate into the soil zone, it can enter the depression storage or it can divert as surface runoff. The depression storage is subject to evaporation and further infiltration. The initial losses at the beginning of a storm consist of the interception and depression storage. But what is more important within this thesis is that those initial losses do not contribute to the storm flow. When the water infiltrates into the soil layer a fraction percolates to the groundwater and some diverts by interflow. Furthermore, the soil layer is subject to evapotranspiration and the available soil moisture. The groundwater discharges are dependent on the recession coefficient and the amount of groundwater storage. The total discharge of a grid cell is the summation of drainage, interflow and surface runoff.

Each layer of a grid cell produces some evapotranspiration. Which depends on several factors:

- **Soil and vegetation:** The calculation of this amount is based on the relationship as a function of potential evapotranspiration, vegetation type, stage of growth and soil moisture content.
- **Surface layer:** The actual evapotranspiration is computed as the area-weighted mean of the land use percentage. There is transpiration from the vegetated parts and evaporation from the soil, but there is no evaporation from impervious areas.
- **Groundwater:** A small portion of the total amount of evapotranspiration originates from the groundwater reservoir. This amount will be determined as a proportion of the groundwater storage.

The total amount of evapotranspiration is calculated as the sum of evaporation from interception storage, depression storage and the evapotranspiration from soil and groundwater storage.

The routing of overland and channel flow is implemented by the method of the diffusive wave approximation. For the technical details see the user manual (Liu & De Smedt, 2004). An approximate solution is used to route water from each grid cell to the basin outlet or a selected convergent point in het basin. The flow path response function at the outlet of the basin or any other downstream convergent point is calculated by convoluting the responses of all cells located within the drainage area in the form of the probability density function. This routing response serves as an instantaneous unit hydrograph and the total discharge is obtained by convolution of the flow responses from all spatially distributed precipitation excesses generated in the grid cells. (Bahremand & De Smedt, 2008)

3.1.1.3 Limitations and assumptions

Liu and De Smedt (2004) describe twelve important assumptions and ten important limitations of the WetSpa model. In this section the most relevant limitations and assumptions are discussed. Limitations and assumptions are considered relevant when they are interesting for a flood forecasting case, the topic of this research.

3.1.1.3.1 Assumptions

- Soil characteristics are isotropic and homogeneous for a single raster cell
- Precipitation is spatially homogeneous within a raster cell
  - It is important to be aware of these assumptions, it is clear that in fact reality is sometimes different. Furthermore, it is important to remember that there are a few meteorological stations used to generate the rainfall data. Within this case this assumption looks reliable.
- Evapotranspiration does not occur during a rainstorm or when the soil moisture is lower than residual soil moisture
  - It sounds like a reliable assumption because during a rainstorm there cannot be a lot of evapotranspiration.
- Water flows along its pathway from one cell to another, and cannot be partitioned to more than one adjacent raster cell.
  - This is a major assumption because this proves that the model would not take into account an upstream flood.

3.1.1.3.2 Model limitations

- The WetSpa model runs with a continuous input of data.
  - The importance of this limitation can be explained by an example. In the case of the Ve river, there are four rainfall stations. One of them has hourly measured data, and the other six-hourly. Hourly data will produce a more accurate result, so it is preferable to use these. However, this means that the six-hourly data must be adapted. Furthermore, it is clear that the results will be influenced negatively when only few data are available.
- Values assigned to any raster or grid cell represent an average value over the area of each cell.
  - This limitation discusses the same point as the first two assumptions. There will be an error in the results, but this problem or limitation cannot not be changed.
- The impervious fractions for urban areas are set subjectively depending upon cell size, since no detailed measurements are available.
  - These fractions may cause an error in the model results, since these fractions may not reflect reality.
- The land use categories are grouped, for which some of the categories might be somewhat ambiguous.
  - For instance, the category agriculture may include farmsteads, disturbed areas and other land uses which are not identifiable as another land use category. Therefore it is important to appropriate the land use classes carefully, this fact may cause directly an error in the simulated results.

3.1.2 Parameters and input variables

The WetSpa model is a complex hydrological model and there are many input variables and parameters. In the next part of this report an overview of the parameters and input variables of the model is given. Furthermore, for every parameter and input variable a short description is given about how this parameter is incorporated into the model. First an overview of the input data is given, secondly the values that are requested during the loading of the WetSpa model are described and finally the global parameters are described.

3.1.2.1 *Meteorological and geographical data*

The model is GIS-based and it needs five input maps. These are DEM, soil type, land use, the location of meteorological stations and the stream network. Apart from geographical information, hydro-meteorological information is also needed. Hydro-meteorological information consists of rainfall, PET (Potential EvapoTranspiration) and discharges. Temperature information is optional; it is only needed when snow occurs within the study area, which is not the case. The data used within this research are described in section 3.2.

3.1.2.2 *Input variables for model set-up*

As written in section 3.1.1.1 it is necessary to load the data of the study area in Arcview, before the WetSpa model can be utilized. During this set-up time in Arcview, some input variables have to be given. An overview of these input variables is given in Table 2, a more detailed description of these input variables will be presented in appendix A.1.

Table 2: Input variables during set-up time

Variable	Description
Cell threshold for stream networks	Threshold value for creating a stream network
Threshold for minimum slope	Minimum slope
Setting a flood frequency	Choose a flood frequency
Cell threshold for the watershed	Determining the watersheds in the study area
A minimum ratio reflecting the moisture condition	Setting the initial moisture condition
Choosing a way to determine the Manning's coefficient	Choosing from three options the way Manning's coefficient has to be determined.
Percentage for urban area	Set a value for the percentage of urban area.
Setting a flow limit	Choose whether to set a flow limit or not

3.1.2.3 *Global parameters*

In the WetSpa model, twelve global parameters are compiled by the designers to simplify the calibration process. These parameters have physical interpretations. They are important in controlling runoff production and hydrographs at the basin outlet, but difficult to assign properly on a grid scale. Therefore, it is preferable to calibrate these parameters against observed runoff data in addition to the adjustment of the spatial distributed model parameters. (Liu (1) & De Smedt, 2004)

Table 3 gives an overview of these different parameters and appendix A.2. presents a more detailed description of the global parameters.

Table 3: Overview of the global parameters

Parameter	Description	Unit
Dt(h)	Time step in hours.	H
K_i	Scaling factor for interflow computation.	-
K_g	Groundwater recession coefficient	-
K_{ss}	Initial soil moisture	-

K_{ep}	Correction factor for potential evapotranspiration.	-
G_0	Initial groundwater storage in water depth	mm
G_{max}	Maximum groundwater storage in water depth	mm
T_0	Base temperature for snow melting.	°C
K_{snow}	Degree-day coefficient for calculating snowmelt.	mm/°C/day
K_{rain}	Rainfall degree-day coefficient for estimating snowmelt.	mm/mm/°C/day
K_{run}	Surface runoff exponent when the rainfall intensity is very small	-
P_{max}	The threshold rainfall intensity (mm/d or mm/hour; depending on the timestep)	mm

((Liu & De Smedt, Documentation and User Manual, 2004)& (Liu & Corluy, 2005))

3.1.3 Preliminary sensitivity analysis

In the user manual of Liu and De Smedt (2004) a preliminary sensitivity analysis for model calibration is already done. Their results are shown in appendix A.3. This Sensitivity analysis presents just an indication of the sensitivity analysis, because each time just one parameter is changed and after one round of changing the results were presented. Therefore it could be really useful for the calibrationprocess. Further, Nurmohamed, et al. (2006) found that the most sensitive global parameters are K_i, K_g, K_{ss} and G_0. This will be really useful during the manual calibration because only the global parameters are calibrated, after the model set-up in Arcview.

3.2 Data

For this project data are available from two sources. The first source is KBR, a non-profit organization which receives funds from the Australian government. They set up the Quang Ngai Disaster Mitigation Project. The aim of the project was to mitigate the impact of natural disasters in the Quang Ngai province. The total funds received from the Australian government was $ 13.5 million. (Aid Activities, 2008) The second source of data is the HUS (Hanoi University of Science). The description of data is divided into three parts: hydro-meteorological data, map data and tables.

3.2.1 Hydro-meteorological data

The hydro-meteorological data are provided by KBR. The main source for KBR for hydro-meteorological data is the Hydro Meteorological Service. Data are available for three floods, which took place in November 1999, December 1999 and October 2003. The hydro-meteorological data are divided into streamflow data, rainfall data and potential evapotranspiration (PET). These are described in the next sections of the report.

3.2.1.1 Streamflow data

The streamflow data are measured at An Chi, where the Ve River leaves the study area. The discharge is measured hourly for November 1999 and December 1999. During the October 2003 flood not hourly discharges were measured, but hourly water level data. For 15 measurements discharges were also available. Shown in Figure 7.

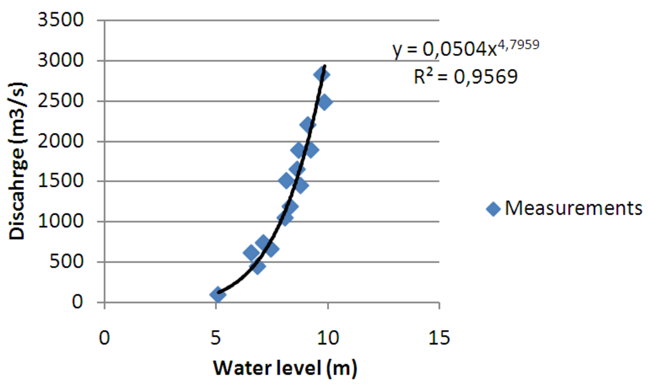


Figure 7: The relation between the discharge and waterlevel

To convert water level data to discharges, a trendline was added. This power-function had a R^2 of 0,9569, which indicates a good fit. The formula of the trendline was used to create discharges from the water level data. The result was checked with the fifteen original measurements in Figure 8, showing a good fit.

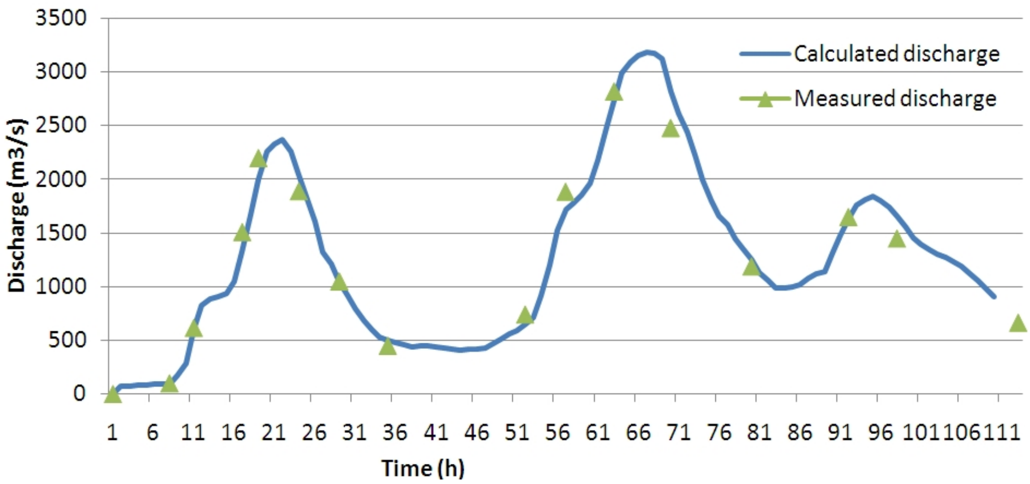


Figure 8: Calculated and measured discharge of October flood

3.2.1.2 Rainfall data

The rainfall data are provided by KBR, who got the data from the Hydro Meteorological Service, and also from the Hydro Meteorological Forecasting Centre. For the rainfall five stations are be taken into account, because they cover the study area. However, one station lacks data, so it is not taken into account. The coverage of this station is very small, about 0,02 % of the study area. So the effect of eliminating this station on the model output is very small. Figure 9 shows how the other four stations cover the study area.

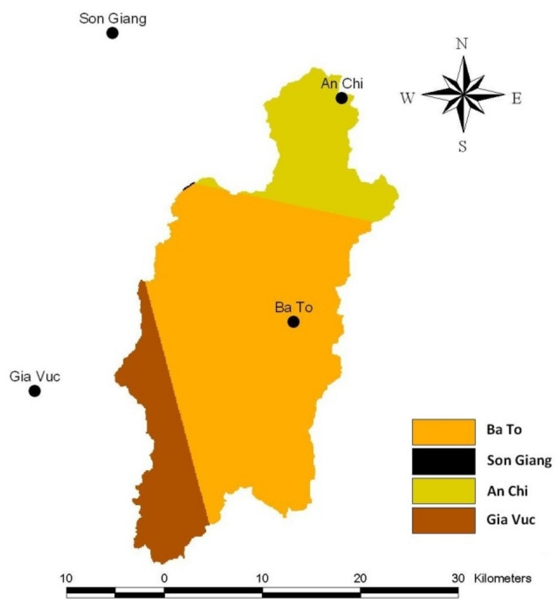


Figure 9: Meteo Stations covering the study area

At three stations (An Chi, Son Giang and Gia Vuc) the rainfall was measured with a six-hourly time step, at one station (Ba To) it was measured one-hourly. The data must be in accordance with the other ones, and therefore the data of the three six-hourly stations are changed into one-hourly data. The temporal (one-hourly) rainfall pattern of Ba To is used as a format for the temporal pattern of the three other rainfall stations. To determine a hourly pattern for the other stations, their 6 hourly data is multiplied with the corresponding rainfall of 1 hour in Ba To which is divided by the summation of the 6 corresponding hours of Ba To.

In reality the temporal patterns of rainfall at the four stations are probably not exactly the same. To compensate this, a random factor could be implemented. However, the result of this can model the reality better or worse. Therefore no random factor is implemented within this research.

3.2.1.3 PET (potential evapotranspiration)

PET-data were not available within this research. However, PET is so small during floods that it is almost negligible. (Gash & Stewart, 1977) Therefore it is reasonable to use a PET of 0 during the flood period. So global parameter K_{ep} does not affect the model output.

3.2.2 Maps

There are five digital maps available for this project. They are provided by the HUS. These maps deal with DEM (Figure 4), land use (Figure 5), soil type (Figure 4), measurement locations and the stream network. The DEM, land use and soil type were available on a 90 m by 90 m grid cell size. Some improvement of the available data needed to be made, before using them in de model. The improvements made are described in the next part.

3.2.2.1 Boundaries

The original files of DEM, land use and soil type covered a square around the study area. But the WetSpa model does not work when an area bigger than the study area is implemented. Therefore the maps were initially clipped by a boundary, also given by the HUS. However, this boundary was drawn in straight lines. This does not correspond with reality, because a watershed is a natural phenomenon.

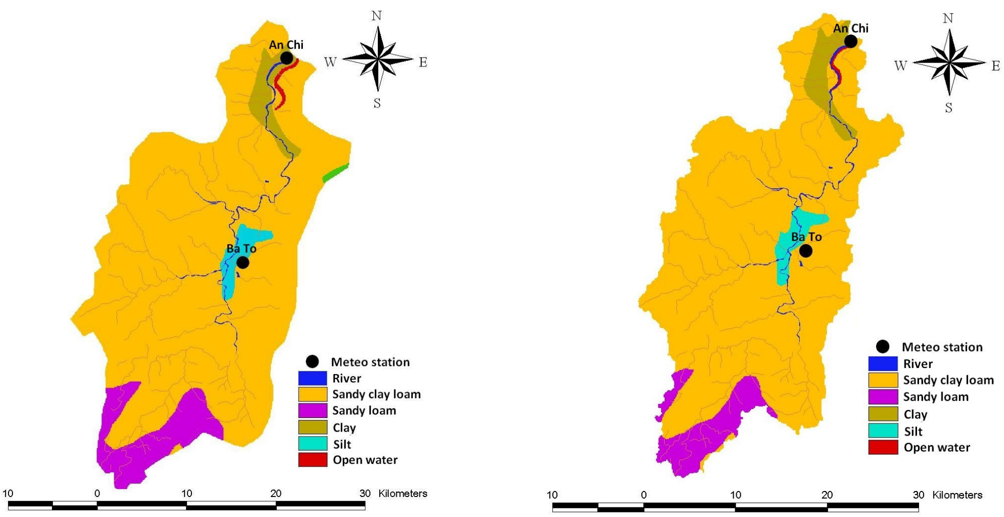


Figure 10: Left: Original soil type map; Right: Referenced soil type map with the right boundaries

Therefore a second option is used to calculate the boundary. This is done by a function in ArcView, to calculate the boundary of a watershed from a DEM-map. This boundary is used to clip every map.

The left picture of Figure 10 shows an example of the original boundary from the soil type map, the right picture is the new soil type map.

3.2.2.2 Georeference
The data were not georeferenced in the same way. The difference between the maps was approximately 10-15 grid cells. The left picture of Figure 10 is an example of a wrongly referenced map. The maps were not a correct input for the model, and therefore not useful. Therefore the maps were referenced on the location of the river at An Chi. This is because the river at An Chi can be seen clearly on all maps. In Figure 10 the river at An Chi is red, classified as ‘open water’. The right picture is referenced correct, because the river and the open water fall together. The referencing is done in Arcview.

3.2.2.3 Classification
The maps of land use and soil type are related to tables in the WetSpa model. The (Vietnamese) maps are classified in a different way than the tables used within WetSpa. Therefore some Vietnamese students translated the Vietnamese classes into the WetSpa classes as well as they could. However, the classes of the model and the maps cannot be translated fully correctly, because the same classes did not exist. This translation can have impact on the output of the model, but it is impossible to measure this impact.

3.3 Calibration process
As described in section 3.1.1.1 the model actually exists of two parts. Before running the model, all maps have to be loaded in Arcview. For example for determining the real river with Arcview, Arcview need a threshold value. This value has to be determined manually to get the best result, this is an important part of the calibration process therefore these input values will be described in section 3.6.1. Since there are three datasets available, there is chosen to use two of them for calibration and the model is verified on the third flood. Two datasets are used for the calibration because there was lack of data and now all available data is used which should improve the results. An advantage of calibrating with two datasets is that the parameter set is an average of two datasets. Therefore the verification results has to be better then when only one dataset is used for calibrating, surely in the case when there is just a small amount of data available. But on the other hand calibrating the model now is more time consuming because each time the model has to be run twice instead of one time. During the whole calibration process for all parameters are utilized only uniform distributions.

As written before the WetSpa model uses two models. Model1 is a semi-distributed model and can be used for the first part of the calibration process. Then Model2, the fully-distributed model, can be used when there is found a good fit for model1. So the number of runs for Model2 will be smaller than for model1. The calibration process is described in three steps.

3.3.1 Manual calibration
During the first calibration a semi-automatic model of WetSpa was not available, therefore there is started with a manually calibration. After the set-up time in Arcview the real WetSpa model could be run and produces some results. These results has been analyzed with the criteria who are described in section 3.5 and compared with the hydrograph of the observed values. Then the values of the global parameters were changed to improve the results. This process is repeated till the target values for the criteria were achieved, the target values will be discussed in section 3.5. Since two datasets are used to calibrate the model, the calibration process is more time consuming. The purpose of the

calibration is to find a parameter set that produces the same results for both the November and the October dataset. At the end there has to be found one parameter set that produces nearly the same results for the criteria for both datasets.

3.3.2 Random Sampling

After the manual calibration the results were pretty good, see Table 4 and appendix B.1. for the graphs. Unfortunately the verification results with the split sample test (see section 3.8 for an explanation) were really poor, see also Table 4 and figure 2 of appendix B.1. for the graph. The conclusion was that the model is calibrated on a wrong optimum. Therefore 9000 random samples with the Random Sampling method were evaluated for all three datasets. Because this sampling method is suitable for simple models, with a small processing time (Saltelli, et al., 2000). After running model1 for the first 1000 random samples, the produced results were evaluated with dotty plots to generate new ranges for a new random sample which should fit better with one of the datasets. The ranges of those datasets are shown in appendix B.2.

Table 4: Calibration and verification results

	Calibration		Validation
Criterion*	November	October	December
Model confidence	0.872	1.094	1.824
Nash-Sutcliffe	0.856	0.852	0.440
Adapted version of NS	0.874	0.869	0.389

*The criteria are described in section 3.5.

For this random sampling only the NS was calculated as criterion to judge each parameter set, this is done for practical reasons. Nevertheless the result with the samples of the Random Sampling method was a good fit for November and October, but December shows poor results for the same parameter sets like it did for the verification after manual calibration. Nevertheless it was found that there is a good fit possible (NS=0.88) for December in Model1 but for a different parameter set, the results is shown in appendix B.3. Also for this good fit, the results for the peak discharges are poor. The peak discharges are fairly good predicted, however the pattern of the simulated graph fit not really well with the pattern of the observed graph for the high discharges. A reason for this could be that the Ve River Basin becomes a open basin during an extreme rainfall event.

The main problem consists in the parameter Kg. The dotty plots in appendix B.4. show that the maximum NS for December is in the opposite direction of the maximum NS for October. The difference with November is smaller but also this graph shows a preference for a small value for Kg. It is not surprising that Kg is an important parameter, because in the preliminary sensitivity analysis (presented in section 3.1.3) it was also found as an important parameter.

3.3.3 Latin Hypercube Sampling

For the final calibration round the calibration set has to be evaluated with model2, this model has a larger processing time. Therefore a more efficient random sampling method is necessary, also argued by Uhlenbrook and Sieber (2005).

Latin Hypercube Sampling (LHS) is a stratified sampling approach that efficiently estimates the statistics of an output. The probability distribution of each parameter is subdivided into N ranges with an equal probability of occurrence (1/N). Random values of the parameters are simulated such that each range is sampled just once. After there is chosen a random value for each range this process is repeated till the number of random parameter sets is achieved. The order of selecting the ranges is randomized and the model is executed N times with a random combination of parameter values from each prior defined range. (Yu, et al., 2001)

Within this research N is set to five, and in the first round 400 parameter sets were evaluated for defining the parameter ranges. Afterwards the dotty plots were analyzed for November and October and new ranges were set for the final calibration round of 1000 model simulations, the ranges of the parameters are presented in appendix B.2. The processing time of the fully-distributed model for 1000 model simulations took approximately eight hours. The final result is presented in section 5.1.

3.4 Correlation between global parameters

For manual calibration it is useful to know what the correlations are between different parameters. However the Morris method which will be used for the sensitivity analysis cannot calculate the correlation between the different parameters. Moreover the main disadvantage of the Morris method is that individual interactions among factors cannot be estimated. (Saltelli, Chan, & Scott, 2000)

Bahremend and De Smedt (2008) argued in their article about the WetSpa model, that there is no significant relationship between the global parameters. Except for one combination, between K_g and K_ss is a rather high negative correlation found. During the manual calibration those parameters will be changed carefully.

3.5 Criteria

To give a judgment about the accuracy of the model, the model will be evaluated qualitatively by visual comparison of the simulated and observed data and quantitatively by using three criteria who are implemented in the WetSpa model. The “Model Confidence”, “Nash-Sutcliffe” and the “Adapted version of Nash-Sutcliffe efficiency for high flow evaluation” will be utilized. Andersen, et al. (2001) have set some performance criteria to the possible results of the Nash-Sutcliffe coefficient, see Table 5. Therefore the target value during the calibration process for the Nash-Sutcliffe coefficient is set to 0.85. This target value applies also for the “Adapted version of Nash-Sutcliffe efficiency for high flow evaluation”. For the model confidence the idea of Andersen, et al. will be used, but this criterion could be higher than 1 therefore the performance “good” will be set on 0.85-0.90 and 1.05-1.15. Actually there are five criteria implemented in the WetSpa model, but for this research only three of them are useful. Nevertheless, a description of all criteria is presented in Table 6.

Table 5: Performance criteria (Andersen, et al., 2001)

Performance	Nash-Sutcliffe
Very Good	> 0.95
Good	0.85-0.95
Fair	0.70-0.85
Poor	< 0.70

Table 6: Description of the calibration criteria (Liu & De Smedt, 2004)

Criterion	Description
Model confidence:	Model confidence represents the proportion of the variance in the observed discharges that are explained by the simulated discharges. A value close to 1 indicating a high level of model confidence.

$CR2 = \frac{\sum_{i=1}^{N}\left(Qs_i - \overline{Qo}\right)^2}{\sum_{i=1}^{N}\left(Qo_i - \overline{Qo}\right)^2}$	
Nash-Sutcliffe efficiency: $CR3 = 1 - \frac{\sum_{i=1}^{N}\left(Qs_i - Qo_i\right)^2}{\sum_{i=1}^{N}\left(Qo_i - \overline{Qo}\right)^2}$	The Nash-Sutcliffe efficiency varies from a negative value to 1, with 1 indicating a perfect fit between observed and simulated hydrographs. When CR3 is below zero, it indicates that average measured stream flow would have been as good a predictor as the modeled stream flow.
Adapted version of Nash-Sutcliffe efficiency for high flow evaluation: $CR5 = 1 - \frac{\sum_{i=1}^{N}\left[\left(Qo_i + \overline{Qo}\right)\left(Qs_i - Qo_i\right)\right]^2}{\sum_{i=1}^{N}\left[\left(Qo_i + \overline{Qo}\right)\left(Qo_i - \overline{Qo}\right)\right]^2}$	In contrast with the previous criterion this criterion gives a value which emphasize the quality for high flow simulations. Therefore more weight is given to the high simulated discharges as can be seen in the formula. The best value for CR5 is 1.

For calculating the criteria the warming-up period would not be taken into account. Since there is lack of data the warming-up period for the simulated graph will be very short and is set to the start of the first peak. In practice this mean that the warming-up period is just a couple of hours, in Table 7 an overview of the different periods is shown.

Table 7: Boundaries for model calibration/verification

Dataset	Start calculating criteria (DDMMYYYY, Time)	End calculating criteria (DDMMYYYY, Time)
November	01-11-1999, 12:00	07-11-1999, 00:00
December	02-12-1999, 23:00	08-12-1999, 01:00
October	15-10-2003, 07:00	10-19-2003, 13:00

3.6 Ranges of the different parameters and input variables

In this section the ranges of the parameters and input variables that are utilized for the calibration will be described. In section 3.6.1 the ranges of the input variables during set-up time of the model are shown and in section 3.6.2 the ranges of the global parameters.

Determining the ranges is an important part of the calibration process. However determining ranges is a subjective activity argued by Beven and Binley (1992). Therefore it is important to choose wide ranges during the setup time, but such that the values are still useful. For the sensitivity analysis the ranges are determined again because with the results of the preliminary calibration the ranges could be set more accurately which will improves the results.

3.6.1 Input variables Arcview

In Table 8 the ranges of the different input variables are shown and if necessary the range will be discussed. A description of these input variables is already given in section 3.1.2.2 therefore only the ranges will be shown here.

Table 8: Ranges of the input variables during set up time

Variable	Range	Additional information
Stream network	0-1000	
Minimum slope	0.001-0.1	
Flood return period	2, 10 or 100 years return period	
Watersheds	1000-8000	It has to be a multiply of the stream network threshold value.
Saturation	0.0-1.0	
Manning	3 options	- Interpolation among different stream orders - Remain the constant as in the lookup table - Change to another constant
Percentage urban are	0-100%	
Flow limits	No or Yes	Yes à The limits has to be between 0 and 3 m/s

3.6.2 Global parameters

Table 9 presents the ranges which are utilized during the manual calibration. For the first sample set for the Random sampling the ranges are based on the results of the manual calibration. The table with the ranges is presented in appendix B.2.

Table 9: Ranges op the global parameters

Parameter	Range	Additional information
Ki	0-10	Generally greater than 1*
Kg	0-1.5	
K_ss	0-3	
G0	0-200	
G_max	0-200	This value cannot be greater than G0.
K_run	0-10	This value is generally less than 3 according to the previous applications.*
P_max	0-150	

* (Liu & De Smedt, Documentation and User Manual, 2004)

3.7 Start values in Arcview

In Table 10 the values for the input variables during the setup time in Arcview are presented. As written in section 3.1.2.2 the value for stream network is determined by observing the real and the calculated river, the result is shown in Figure 11.

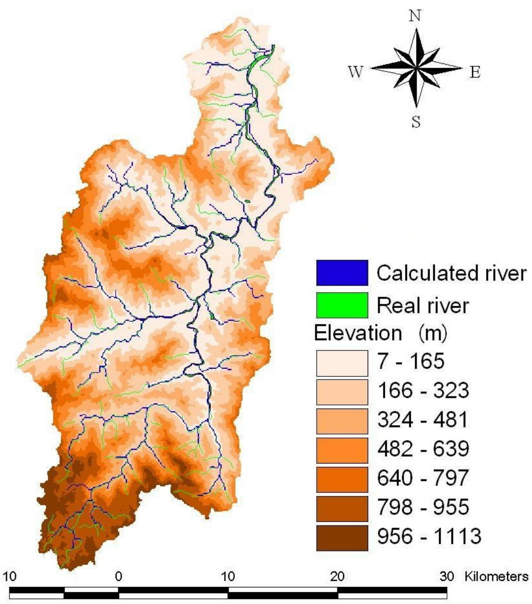


Figure 11: Real and calculated river

After the poor manual calibration results, it was likely to change the initial conditions (the conditions which are created in Arcview and used for the simulations with WetSpa) to improve the results. The preferred parameters were: initial moisture, the number of watersheds and the manning coefficient. However after changing the initial conditions in Arcview and running the WetSpa model again the produced graph does not make sense at all, the result is shown in appendix B.5. Therefore this problem is further investigated. However, due to time limitation the solution is not found. Therefore only the original initial conditions are taken into account.

Table 10: Input variables in Arcview

Input variable	Result	Clarification
Stream network	400	After trying several times this threshold value produces the most realistic stream network.
Minimum slope	0.01%	Standard value and there isn't a reason to change it.
Flood return period	T2	Standard value and there isn't a reason to change it.
Watersheds	4000	It has to be a multiply of the stream network threshold value and therefore it is set to 4000 and produces 13 subwatersheds.
Saturation	0.8	The model will be utilized for flood prediction and there is no start up time therefore this value is set to 0.8.
Manning	Use lookup tables	This option is chosen because it seems to generate the best results.
Percentage urban are	30%	Standard value and there were no arguments available to change this value.
Flow limits	No	Because without limits Arcview produces good results.

3.8 Preliminary verification

To verify the model results of the calibration, which are described in section 5.1., a split sample test is executed. For this preliminary verification the dataset of December is used for the split sample test. The December flood is an independent flood of the November flood. The split sample test shows how well the calibration succeeded, therefore the same criteria as for calibration are used.

4 Sensitivity analysis

The objective of this research was to do a sensitivity analysis for the WetSpa model using the method introduced by Morris (1991). Therefore in the first section an outline of this method is described. In section 4.2 the parameters and input variables which will be examined within the sensitivity analysis are presented. Then the parameter ranges are determined in section 4.3. Next in section 4.4 the model and the used dataset are determined. Finally the most informative output will be discussed in section 4.5.

4.1 Morris method[1]

The main reason why Morris introduced his One-At-a-Time (OAT) design method in 1991 was to improve the economy of an sensitivity analysis. “The economy of a design will be defined to be the number of elementary effects it produces divided by the number of experimental runs.” (Morris, 1991) The method he introduced has a economy of $\frac{k}{(k+1)}$; where k is the number of parameters.

This design is based on the construction of a *B** matrix with rows that represent input vectors x, for which the corresponding experiment provides k elementary effects from k+1 runs. To make things more clear all steps of the method will be described in this section.

4.1.1 Input

The first step of the Morris method is to determine the number of parameters (*k*) and the number of levels (*p*); for the more economical design *p* has to be even. Then the set for the possible values (base values) of *x** have to be made, this has to be done like equation (2).

To make the method a bit more clear, an example *B** matrix will be created. Therefore *p* and *k* are set to respectively 4 and 3[2].

$$\Delta = \frac{p}{[2(p-1)]} = \frac{2}{3} \tag{1}$$

$$Set = \left\{0, \frac{1}{(p-1)}, \frac{2}{(p-1)}, \ldots, 1-\Delta\right\} = \left\{0, \frac{1}{3}\right\} \tag{2}$$

$$m = k + 1 = 4 \tag{3}$$

From the “Set” a base value is randomly chosen for each parameter (Every column in the matrix presents one parameter). For the example *x** happen to be:

$$\mathrm{x}^* = \left(0, \frac{1}{3}, \frac{1}{3}\right)$$

4.1.2 Create B* matrix

To build a B* matrix, the first step is the selection of a $m \times k$ matrix *B* with elements that are 0s and 1s, such that for every column there are two rows of B that differ in only one element. The simplest way to create this matrix, is making a triangular of 1s starting at the second row. Further a $J_{m,k}$ matrix

---

[1] For writing this section the following sources are used: (Morris, 1991) and (Saltelli, et al., 2000)

[2] The results for this example are shown after the formula is given. The calculations steps will not be described

of 1s and a k-dimensional *D** matrix with elements either +1 or -1 with equal probability has to be build. At least a $k \times k$ dimensional *P** matrix has to be build which is a random permutation matrix and contains in each column one element equal to 1 and all others to 0, such that no two columns have 1s in the same position. *B*, *J*, *D**, *P** happen to be:

$$B_{m,k} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \qquad J_{m,k} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad D^*_{k,k} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad P^*_{k,k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Now all matrices and values that needed for the *B** matrix are available, the *B** matrix would be given by equation (4).

$$B^* = \left(J_{m,1} \times x^*\right) + \left(\left(\frac{\Delta}{2}\right)\left[\left(2B_{m,k} - J_{m,k}\right)D^*_{k,k} + J_{m,k}\right]\right) \times P^*_{k,k} = \begin{bmatrix} 2/3 & 1/3 & 1 \\ 0 & 1/3 & 1 \\ 0 & 1/3 & 1/3 \\ 0 & 1 & 1/3 \end{bmatrix} \tag{4}$$

To utilize the B* matrix for the sensitivity analysis it has to be multiplied by the ranges of the different parameters. Since every column in the B* matrix is standing for a parameter and every row gives a random number for this parameter. Every column has to be multiplied by the interval of the corresponding parameter and add the minimum value of the interval to this result. This can be expressed by equation (5).

$$parinput^1 = B^*_{m,1} \times range + \min imum\ value\ parameter \mathrm{int} erval \tag{5}$$

[1]Parinput: is the *B** matrix with values within the parameter ranges for the different parameters. This input will be used for calculating the elementary effects.

4.1.3 Elementary effects

Every row in B* differs only in one column from the row below, moreover this difference is always equal to –Δ or Δ. Therefore during the simulation of the “parinput” *n* runs provide *n*-1 elementary effects. After running the model with the “parinput” files the values of the elementary effects can be calculated. Therefore the following statement is used with corresponding formula presented in equations (6) and (7).

If $B^*_j - B^*_{j+1} < 0$ then:

$$d_j\left(parinput_j\right) = \frac{y(parinput_{j+1}) - y(parinput_j)}{\Delta} \tag{6}$$

Else $\Delta > 0$ then

$$d_j\left(parinput_j\right) = \frac{y(parinput_j) - y(parinput_{j+1})}{\Delta} \tag{7}$$

j = number of a row (to calculate all elementary effects of one input file it has te repeat k times)
Δ = equal to equation (1)
y = output of the WetSpa model

d_j = elementary effect

After simulation of the first “parinput” this action will be repeated “*r*” (=#runs) times. So in total there are needed $r \times m$ runs of the WetSpa model, where *r* could be chosen relatively small since every run simulate 1 randomly chosen elementary effect.

4.1.4 Means and standard deviations

For all parameters and inputs there are “r” elementary effects (d_j) calculated. With these results the means and the standard deviations of the different parameters and inputs can be calculated, the used formulas are presented in equation (8) & (9), the results will be presented in a graph and the most sensitive parameters can be extracted from this graph.

$$s = \sqrt{\left(\frac{1}{1-n} \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2 \right)} \quad with \;\; \bar{x} = \frac{1}{n} \sum_{i=n}^{n} x_i \tag{8}$$

$$m = \frac{1}{r} \cdot \sum_{j=1}^{r} d_j \tag{9}$$

s =standard deviation

m =mean

In this graph three kind of outputs can be distinguished, the meaning of these outputs is explained by Nguyen and De Kok (2006). “F*or instance, a combination of a relatively small mean m_i with a small standard deviation* s_i *indicates a negligible effect of the input x_i on the output. A large mean m*i *and a large standard deviation s_i indicate a strong non-linear effect or strong interaction with other inputs. A large mean mi and a small standard deviation s_i indicate a strong linear and additive effect.”*

4.2 Parameters and input variables

There are many parameters in the WetSpa model, which are shown in the previous chapters. To start with making an overview of all variables in the WetSpa model, the preliminary sensitivity analysis done by Liu and De Smedt (2004), described in section o., is used. Besides of these variables there are some parameters which are handled in the WetSpa model self. The relevant parameters will be discussed in this section. In section 4.2.1 the purpose of the sensitivity analysis will be discussed. Followed by a description of the issues to deal with when using Arcview, with regard to the sensitivity analysis. To end up with the parameter ranges for the sensitivity analysis presented in section 4.2.3.

4.2.1 Purpose of the sensitivity analysis

The aim of this thesis is to find to most influential parameters and inputs of the WetSpa model, with regard to the simulated discharge. Therefore the purpose of this section was to make an overview of all input variables and parameters of the WetSpa model. Since WetSpa receives its initial conditions from Arcview also these parameters and inputs are taken into account, an overview of the number of parameters and input variables is shown in Table 11. However there were 2 main problems with Arcview which will be described in the following section. Because of this there are just 10 input variables/parameters which will be taken into account, these are described in section 4.2.3.

Table 11: Number of parameters/inputs which are preferred to take into account in the sensitivity analysis

Program	What	#parameters	Taken into account
Arcview	All tables	15	-
Arcview	Different input variables during setup time	7	-
WetSpa	Global parameters	7	√
WetSpa	Model parameters	2	√
WetSpa	Input data	1	√
	Summation	**32**	**10**

4.2.2 Problems with Arcview with regard to the sensitivity analysis

As written before Arcview produces the input for the WetSpa model. This is a problem for this sensitivity analysis. Because it is not possible to run the set-up in Arcview automatically without changing the avenue scripts, changing these scripts is not possible within this research because of time limitation. Obviously it is too time consuming to change every run the set-up in Arcview, since it take approximately 10 minutes to run once the set-up.

To solve this problem the output files, which have a table format, of Arcview could be changed with a factor, like equation **Fout! Verwijzingsbron niet gevonden.**(10) (see section 4.3.2). For instance the produced maps with the manning coefficient. The manning coefficients map is produced by replacing the land use categories with the corresponding manning coefficients. Then this map could be multiplied by a factor to get different manning coefficient. This would be done to take the manning coefficient into account in the sensitivity analysis. But this does not solve the problem, since Arcview use created maps to calculate some other maps. An overview in which order the different maps should be calculated is shown in appendix C.1.. To go on with the previous example, the map with manning coefficients is also used to calculate the velocity. Resulting in that when the map with manning coefficients will be changed (changed with a factor) the velocity should be recalculated. Which is too time consuming since Arcview cannot run automatically.

There is one other option, take only the last produced map in Arcview into account like equation (10). But based on the illustration in Figure 12 it is decided to do not take any map into account in that way. Because the change of an individual parameter can have a bigger effect than changing the result which will be taken into account in the sensitivity analysis or the other way around.

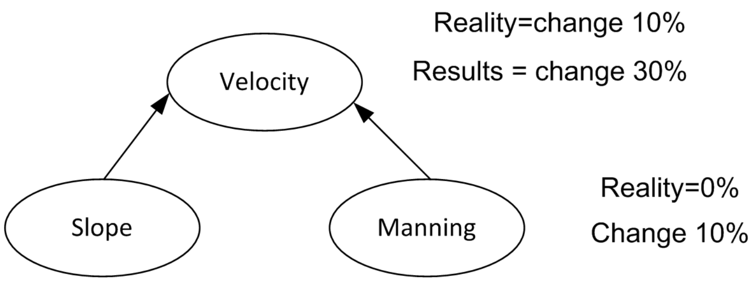


Figure 12: The change of an individual parameter could be larger than changing the result

4.2.3 Parameters

As shown in Table 11 there are 10 parameters taken into account in the sensitivity analysis. These parameters are shown in Table 12. Now follows an explanation why the rainfall is taken into account and a description of the parameters in WetSpa. The other 7 parameters are self-evident, since it are global parameters.

Table 12: 10 Parameters for the sensitivity analysis

	Data	Global parameters							Parameters WetSpa	
Name	Rain	Ki	Kg	K_ss	G0	G_max	K_run	P_max	B	M

4.2.3.1 Rainfall
The rainfall is taken into account to investigate what is the sensitivity of a small measurement error. Therefore the rainfall for the whole area will be increased or decreased with a factor, which will be explained in section 4.3.2. There are 4 measuring stations used for predicting the rainfall in the study area. However the rainfall is considered as 1 rainfall amount. Because the question is not which measuring station is the most sensitive because that is the one with the biggest area, but how sensitive is the rain to the discharge. Moreover the measuring method of the stations will be the same, therefore the stations should have the same amount of error. Therefore it make sense to change the rainfall with the same factor for all stations.

4.2.3.2 Parameters WetSpa
There are two parameters in the WetSpa model taken into account. The parameter B controls the shape of the variation curve for the interception storage. In the model it is set to 1.35. It is one of a few parameters which is set in the WetSpa model, therefore it is taken into account. The other parameter M, is put in the groundwater flow equation. For m= 1 the reservoir is linear and m=2 the reservoir is non-linear. This parameter can vary between 1 and 2. (Liu & De Smedt, 2004)

4.3 Parameter ranges
As said in section 3.6 the ranges will be determined again for the sensitivity analysis to make the results of the analysis more useful. In the first section the determination of the ranges of the global parameters are described and in section 4.3.2 the process for the other parameters is described.

4.3.1 Global parameters
For determining the ranges of the global parameters both calibration datasets, November and October, are used. To explain the way how the ranges are determined, the process will be described for the global parameter *Kg* (groundwater recession coefficient).

The two dotty plots in figure 7 of appendix C.2. show that there is an optimum for Kg around 0.015 for both datasets. The dotty plots present in figure 8 in the same appendix show only the results for *NS*>0.7, for both November as well as October. Comparing the results for these 4 graphs with each other, the range for *Kg* is set to 0.002 as minimum and 0.06 as maximum. That is the range were both datasets produce results above *NS*>0.7

4.3.2 Other parameters
The other parameters are rainfall and the parameters *B* and *M*. The parameter B and the rainfall will changed with a factor, this will be done as described in equation (10). The range for the rainfall is for all stations equal as described in section 4.2.3.1, since the factor represents the measurement error. Liu and De Smedt (2004) do not determine a range for *B*. Since this parameter controls the shape of the variation curve of the interception storage capacitity it seems to be fair to change this parameter

within the same range as the global parameters. For the parameter *M* the range is set from 1 to 2, because in the user manual these limits are described.

$$rain = rain \times factor \tag{10}$$

4.3.3 Ranges

In Table 13 the ranges of all parameters are presented and also a small description is given.

Table 13: Description parameters used for the Sensitivity Analysis

Name	Minimum	Maximum	Unit	More detailed description
Rain	0.9	1.1	%	A relatively small range is chosen because in this survey measured data is used which has normally a small error.
K_i	2	11	-	-
K_g	0.002	0.06	-	-
K_{ss}	0	1.5	-	-
G0	0	50	mm	In all dotty plots there was no reason to make a smaller range for these parameters. Therefore their original ranges are used for the sensitivity analysis. To estimate this statement the graph G0 is shown in appendix C.3.
G_{max}	50	150	mm	
K_{run}	0	10	-	
P_{max}	0	500	mm/day	
B	0.4	1.6	%	This range is based on the ranges of Ki, Kg and Kss, because only for those three the ranges are changed for the sensitivity analysis based on the made dotty plots. Then the percentage difference with respect to the calibrated datasets is determined for each parameter. Afterwards the average of those differences are determined.
M	1	2	-	This is the total range of the parameter M.

4.4 Model and dataset

As written before the fully-distributed model will be used for calculating the results within this research. However the first steps of the calibration process are normally to calibrate the WetSpa model with the semi-distributed model to produce relatively fast results. Therefore it seems useful to check if the semi-distributed and fully-distributed model have the same sensitive parameters.

For the sensitivity analysis it should make no difference which dataset will be used for the calculations, therefore the dataset of November will be utilized. However during the calibration time it was obvious that the results of the same parameter set not always produce the same results for the criteria. Which indicates that both floods have different hydrological properties. Therefore both calibrated parameter sets will compared with each other within this sensitivity analysis.

4.5 Most informative output of WetSpa model

The aim of this thesis is to find the most influential parameters and input variables, to facilitate the flood forecasting application of the model to Ve river basin. Therefore it is concluded that there are three informative kinds of output for decision makers: the peak discharge, time between the rainfall and the peak discharge and the total volume of the flood. For this thesis all three options are investigated within this sensitivity analysis. In the first section the used time serie is described. Followed by an explanation of the manner how the different outputs are calculated.

4.5.1 Time series

Within this analysis with flood is meant the highest flood during the time of the November flood. Therefore for November the last flood is taken into account for the sensitivity analysis and for October the second one. The start and end time of the floods are presented in Table 14 and their graphs in appendix C.4.

Table 14: Time series November and October for the Sensitivity Analysis

Dataset	Start calculating criteria (DDMMYYYY, Time)	End calculating criteria (DDMMYYYY, Time)
November	05-11-1999, 08:00	06-11-1999, 19:00
October	16-10-2003, 19:00	18-10-2003, 11:00

4.5.2 Calculating elementary effects for each output

In Table 15 a description is given, how the different outputs will be calculated. The elementary effect for each output can be calculated by taking the difference between consecutive runs.

Table 15: Description of calculating the different outputs

Output	How to calculate
Peak Discharge	The sensitivity of the peak discharge will be determined on the basis of the maximum discharge.
Time till peak discharge	This output determine the time difference between the peak rainfall and the peak discharge. Where the peak rainfall is the highest rainfall amount before the highest flood.
Total volume	The total volume will be calculated by integrating the whole flood following this equation: $Volume = \int_{LF}^{RF} Q(t)\ dt$

5 Results and discussion

In this part of the report the results are presented and discussed. In section 5.1 the results of the preliminary calibration will be shown, followed by the results of the preliminary verification in section 5.2. In the last section the results of the sensitivity analysis with the method introduced by Morris (1991) will be discussed.

5.1 Calibration result

After the manual calibration a poor verification was the result. Therefore the semi-distributed model is calibrated with the Random Sampling method, this process is shown in section 3.3. Then the LHS method is used for making samples for the final calibration with the fully-distributed model. These results are presented in this section.

Table 16: Calibration results with NS>0.88

Description	1st set	2nd set
NS November	0.890	0.884
NS October	0.883	0.889
Ki	3.663	3.565
Kg	0.018	0.014
K_ss	0.686	0.907
G0	8.384	30.247
Gmax	70.942	51.448
Krun	7.047	1.328
Pmax	0.583	300.638

After running the LHS method for the fully-distributed model, the results for the criterion NS were pretty good for both November and October. However the amount of good fits with the same parameter set for both was much smaller. Nevertheless there were two parameter sets which generate a NS>0.88 for November as well as October. These two are presented in Table 16. For two reason the second set is chosen as the best calibration set. First the difference between the NS for November and October is the smallest, even though the difference between both sets is very small. The second reason which is more important, after a visual comparison of both graphs it is concluded that the first dataset gives a higher simulated discharge for both peak discharges. Which is good for November but poor for October.

In Figure 13 the graphs for November and October are shown, both with a good fit. Beside the NS criteria during the manual calibration two other criteria were used, their values are presented in Table 17. First the graph and the results of the criteria for November will be discussed and afterwards the graph of October.

Table 17: Calibration results

Criterion*	November	October
Model confidence:	0.648	0.841
Nash-Sutcliffe:	0.880	0.889
Adapted version of NS:	0.887	0.908

The graph of November shows the same pattern as the observed values which indicates a good fit. Till the 31th hour a poor simulated graph is shown with regard to their observed values. This can be attributed to the short warming-up time. Another conspicuous point of the graph is the peak around the 111th hour. This part of the graph is constantly below the observed values, this is the consequence of calibrating on the two floods. The Nash-Sutcliffe coeffcient and the adapted version of NS have reached their target values, set in section 3.5. Only the model confidence has a poor result for the calibration. However this is accepted because during the manual calibration it is found that the model confidence fluctuates rapidly. Therefore if the visual result and the other parameters have a sufficient result it is accepted that the model confidence has a relatively poor result.

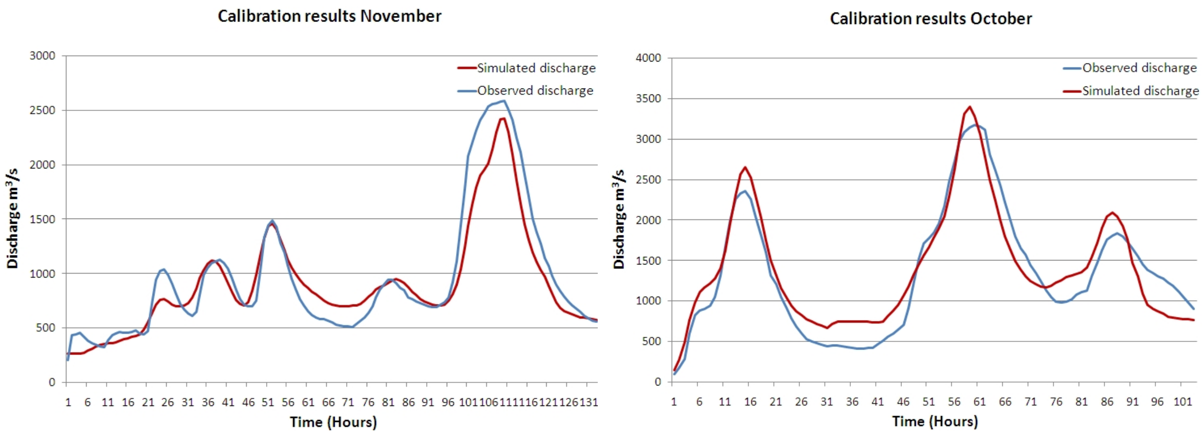


Figure 13: Preliminary calibration results for October and November

Furthermore, after the last calibration round the produced results with the LHS method were accepted as the best possible fit. Even though it is expected that the results can be more optimalized, since the used samples are just randomly generated.

The graph for October shows a really good fit with their observed values. The first part of the graph is interesting because this part shows rather good initial conditions for October. On the other hand the first part for the graph of November shows a poor fit. During the calibration process it is found that it is more difficult to get a good fit for October then for November, therefore the initial conditions for October are rather good and the first part of the graph for November is poor. To improve these initial conditions using a warmin-up time should improve these conditions, however since their is lack of data this is not possible.Furthermore there is a really good fit of the 3 peak discharges, but on the other hand the minimum discharges are constantly slightly overestimated. A reason for this could be that it is more difficult for the global parameters to find a good fit for the high discharges, the consequence in this case is a more poor result for the low discharges. However, in the case of flood forecasting it is more important that the high discharges are well predicted, so a possible prediction is more reliable. Moreover the criteria shown in Table 17 are really good for October, all criteria achieve their target values criteria>0.85 (see section 3.5.).

5.2 Verification result

As written in 3.3.2 a good fit for the same parameter set as November and October cannot be found for the December dataset. This is still the problem after running LHS for the fully-distributed model. The graph which presents the results is shown in Figure 14 and the results for the criteria are presented in Table 18.

Table 18: Split sample test results

Criterion	December
Model Confidence:	1.501
Nash-Sutcliffe:	0.573
Adapted version of NS:	0.534

The graph shows in general the same pattern as the observed values. But the problem is that the details like the 66th hour are not the same as the observed values. A possible explanation for this is the fact that the warming-up time is too short, which cannot be changed since there is a lack of data. Another option is that the initial conditions in Arcview are wrong, which cannot be controlled since

Arcview produce unusual results which is explained in section 3.7. Further every part of the graph is rather over- or underestimated. A possible explanation for this is that the study area becomes an open basin during an extreme rainfall event. In this case it means that when the simulated discharges are overestimated in reality water will flow in another river basin or the other way around for underestimation. Another explanation is that the characteristics of this flood are different from the floods of November and October. This can be founded by a visual comparison of the floods of November and October with December. Because it is obvious that the peak of December takes much more time than the peaks during the floods of November and October. Besides these explanations it is necessary to be mentioned that the reliability of the input data is uncertain.

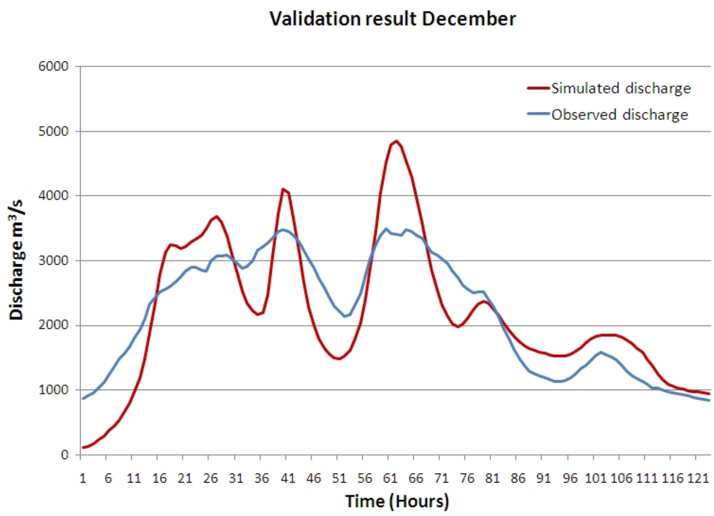


Figure 14: Verification result

5.3 Sensitivity analysis result

The aim of this bachelor thesis is: *"to find out the most influential parameters and inputs of the WetSpa model, to facilitate the flood forecasting application of the model to Ve river basin."* Before presenting the results of the sensitivity analysis it has to be mentioned that a lot of spatial parameters are not taken into account because of the difficulties with Arcview, which are explained in section 4.2.2.

For the parameters which are taken into account *p* (level) is set to 4 and the number of runs is equal to 10. In the next sections the different results will be discussed, in the first section the results for the comparison between model1 and 2 will be discussed. Followed by an comparison in section 5.3.2. between the November and October dataset. In the sections 5.3.3, 5.3.4 and 5.3.5 the results for respectively the sensitivity of the parameters and input to the peak discharge, rainfall to peak discharge time and the total volume will be presented. In the last section the physical meaning of the parameters and inputs with regard to the outcomes of the sensitivity analysis of the total volume and peak discharge will be discussed.

5.3.1 Model1 and Model2

All three graphs of both models are presented in appendix D.1. There are two interesting differences between the results of model1 and 2. In the first place it is surprising that parameter m (#10) is less sensitive for model1 than 2 with regard to the sensitivity of the peak discharge. In the second place there is a difference established for the graphs of "rainfall to peak discharge time". This difference will be explained in section 5.3.4.

5.3.2 November and October

A comparison of the results of those two datasets show that both November and October presents the same pattern for the Peak discharge and the Total Volume, the graphs are presented in appendix D.2. The graphs show also another interesting point, most parameters and input variables have a higher standard deviation and mean for the dataset of October. This difference was also noticed by making the dotty plots during the calibration process. Because the slopes of the dotty plots were higher than the slopes of the dotty plots of November, which can be verified in appendix C.2. Since the Morris method gives only a quantitative judgement of the parameters it is no problem when the mean and standard deviations differs.

5.3.3 Sensitivity of model inputs and parameters with regard to peak discharge

The results of this sensitivity analysis are shown in Figure 15. The inputs K_{run} (7), *M* (10), K_g (3), Rain (1) and K_i (2) are clearly separated from the other inputs. The inputs *Rain* (1) and k_i (2) have a relatively high mean and a relatively small standard deviation, which indicates a strong linear and additive effect. Conclude that the parameters K_{run} (7), *M* (10), K_g (3), *Rain* (1) and K_i (2) are important, and that of these K_{run} (7), *M* (10) and K_g (3) appear to have effects that involve either curvature or interactions. The parameters K_{ss} (4), G_0 (5), G_{max} (6), P_{max} (8) and *B* (9) have a negligible sensitivity on the peak discharge.

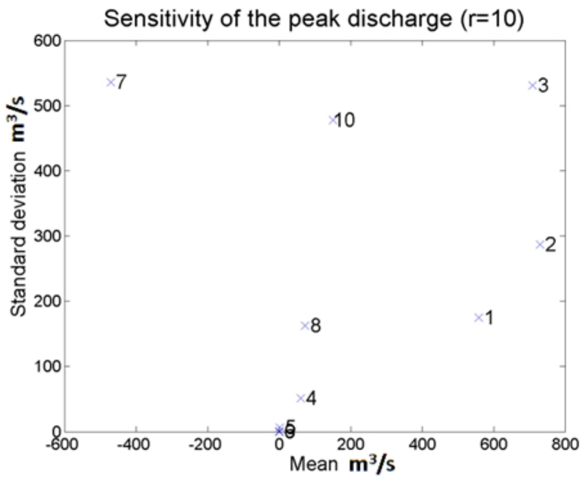


Figure 15: Sensitivity of model inputs and parameters with regard to peak discharge (November)

5.3.4 Sensitivity of model inputs and parameters from the rainfall to the peak discharge time

Before presenting the results it was found during the preliminary calibration that the model is not sensitive for the time between the rainfall and the peak discharge. Because during the preliminary calibration process the peak does not make large moves. This fact is also shown in the results produced by the Morris method, see Figure 16. Only parameter K_{run} (7) has a mean which is greater than 1, which indicate that the average move of the peak is more than 1 hour. All the other parameters have a smaller mean than 1 hour. Furthermore the minimum change of the peak is 1 hour because there is only hourly data available.

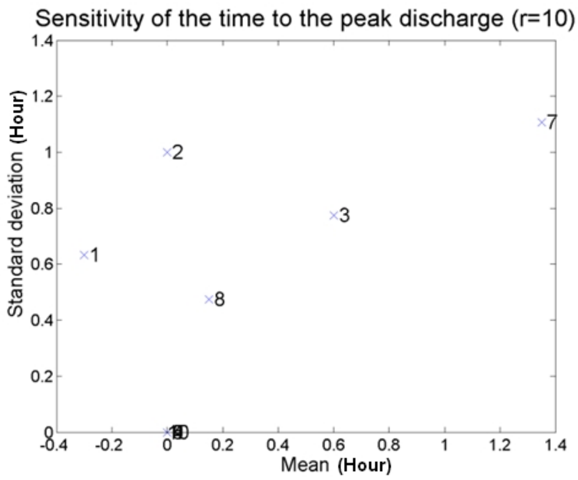


Figure 16: Sensitivity of the parameters and input of the model from the rainfall to the peak discharge time (November)

5.3.5 Sensitivity of the model inputs and parameters with regard to the total volume

In Figure 17 the graph of the sensitivity of the model inputs and parameters to the total volume is presented. There is shown that the parameters K_g (3) and K_{run} (7) are separated from the other parameters and both parameters have a relatively high standard deviation and mean. Further parameters *Rain* (1) and K_i (2) have a relatively high mean which indicate a strong linear and additive effect.

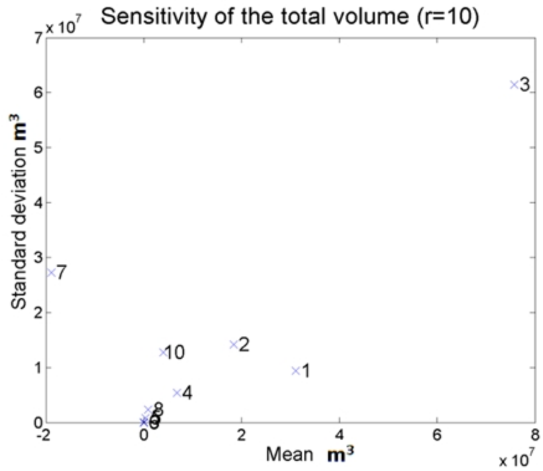


Figure 17: Sensitivity of the model inputs and parameters with regard to the total volume (November)

5.3.6 Discussion about the physical meaning of the parameters and inputs

The parameters K_{ss} (4), G_0 (5), G_{max} (6), P_{max} (8) and *B* (9) are not sensitive to the peak discharge and the total volume. For the global parameters (4-8) it was expected because during the preliminary calibration they were also not sensitive. Parameter *B* (9) is a parameter of the WetSpa model and calculates the initial interception loss for every time step. Therefore it is not surprising that this parameter is not sensitive within this flood forecasting case to the peak discharge and the total volume.

The relatively high mean of the parameters *Rain* (1) and K_i (2) could be expected with regard to their physical meaning. For parameter K_i (2) the high sensitivity was expected for this dataset, because it was a relatively short dataset. Since Ki (2) is a initial parameter it has just a large influence on the first part of the graph. Therefore the parameter will be less sensitive when there is a greater amount of data available. For parameter *Rain* (1) the relatively high sensitivity was also expected since it is the only way water comes into the study area.

The most sensitive parameter to the peak discharge and the total volume is K_g(3), because it has each time the highest mean and standard deviation. This was expected because during the preliminary calibration it was also found that *Kg* was the most important global parameter. With regard to its physical meaning within the model. “The lower the K_g the flatter the curves of groundwater flow.” (Nurmohamed, Naipal, & De Smedt, 2006) It was expected that K_g has a relatively high value during calibration (0.014) because for flood forecasting the graph has to be more curvature.

Beside this there are 2 another interesting outcomes, first the relatively high sensitivity of K_{run} (7). This was surprising because during the calibration time this parameter did not seem very sensitive. However on the basis of its physical meaning it could be expected that it should be a sensitive parameter. Because this parameter reflects the effect of rainfall intensity on the surface runoff for when the rainfall intensity is very small. Second the outcome of parameter *m* (10) is interesting, because this parameter is sensitive to the peak discharge but not to the total volume. This can be explained with the physical meaning of this parameter within the model. Because the parameter calculates the groundwater flow which has a large influence on the total discharge, but it has more influence on the high discharges than the lower discharges. Therefore the parameter is only sensitive for the peak discharge and not for the total volume since this takes a longer time serie into account.

6 Conclusions and Recommendations

In the previous chapter the results of this thesis are presented and discussed. In this chapter the conclusions and recommendations are described. In the first section the conclusions of the preliminary calibration are presented, followed by the conclusions of the preliminary verification in section 6.2. Finally the conclusions and recommendations of the sensitivity analysis are presented in section 6.3. At the end of each section a list with specific recommendations for that part of the thesis is presented.

The objective of this study was to find out the most influential parameters and inputs of the WetSpa model with regard to the simulated discharges. In conclusion it has to be said that the objective is achieved and the most influential parameter is the groundwater recession coefficient (K_g). Nevertheless an important comment on this result is that the input variables within the Arcview program could not take into account.

6.1 Preliminary calibration

In section 5.1. the good results of the calibration are shown. Contrary to the stated reasons for the more poor parts of the simulated discharges and the large uncertainty in the inputdata. It was likely to take the parameters and inputs of Arcview into the sensitivity analysis to get an overview which parameter or input variable is the most sensitive one. However, because of some problems with Arcview not all parameters and input variables are taken into account. These uncertainties are important to remember by judging the results. However all things considered it has to be concluded that the calibration results are rather good.

- Within this research there are four rainfall measuring stations used, but just one of them has hourly measurements and the others have six-hourly measurements (described in section 3.2.1.2). Therefore it is recommended to make all measuring stations suitable to measure hourly, so researches that will be done in the future have more reliable data.
- Doing a research on the characteristics of Vietnamese land use and soil type characteristics. Such that the tables in Arcview could be changed and the right values are taken into account.

6.2 Preliminary verification

The conclusion for the manual calibration has to be that there is not calibrated on a wrong optimum. The poor verification could be explained by one of the following reasons: the flood of December is different from October and November, the initial conditions have to be less important which can be achieved by using more data so the warming-up period can be longer or in the most extremely case it has to be concluded that the WetSpa model is not suitable for this study area. For the last explanation more research needs to be done before this can be argued.

- Changing the initial conditions in Arcview and improve in this way the verification results.
- There should be taken longer data series into account to improve the initial conditions if changing the conditions in Arcview does not make sense
- There should be done a research to the characteristics of the study area when it behaves itself like an open basin. On the basis of this result there should be determined when the WetSpa model is suitable and when it is not.

6.3 Sensitivity analysis

In the first place a comparison between the semi-distributed model and the fully-distributed model has been done. The main conclusion for this comparison is that there are no differences between the global parameters and the different sensitivities for parameter *M* could not be explained. After comparing the two datasets of November and October the conclusion is that there is no notable difference between the Morris sensitivity analysis results of both sets. Therefore as expected it make no sense which graph is used for the sensitivity analysis.

Since the means of the graph presented in Figure 16 are mostly smaller than one hour. It is concluded that the parameters and inputs of the model are not sensitive for the time between the rainfall and the discharge peak.

Since the groundwater recession coefficient (K_g) has the highest standard deviation and mean for the graphs of Total volume and the Peak discharge, it is concluded that this is the most sensitive parameter. This conclusion can be estimated by the findings during the calibration process, were K_g by far was the most important parameter.

Also it was shown that parameter *M*, which is a parameter in the WetSpa model, was a relative sensitive parameter. This shows that it is important to do a sensitivity analysis for more parameters than the global ones and results emphasize that there has to be done a sensitivity analysis which also takes into account the parameters within Arcview.

- Integrate the steps of Arcview and the WetSpa model into one new program which has the big advantage that a sensitivity analysis for all parameters and input variables could be done and the program can run automatically.
- Before starting with calibrating the practical model, execute a sensitivity analysis which takes all parameters and inputs (also from Arcview) into account. Therefore the WetSpa model should run as one program.

References

Aid Activities. (2008). Retrieved May 14, 2009, from AusAID:The Australian Governement's overseas aid program: http://www.ausaid.gov.au/vietnam/projects/quangngai.cfm

Andersen, J., Refsgaard, J., & Jensen, K. (2001). Distributed hydrological modelling of the Senegal River Basin - model construction and validation. *Journal of Hydrology* , Volume 247, 200-214.

Bahremand, A., & De Smedt, F. (2008, 27 march). Distributed Hydrological Modeling and Sensitvity Analysis in Torysa Watershed, Slovakia. *Water Resources Management* , Volume 22, 393-408.

Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty predicition. *Hydrological processes* , Volume 6, 279-298.

Downer, C., Ogden, F., Martin, W., & Harmon, R. (2002). Theory, development, and applicability of the survace water hydrologic model CASC2D. *Hydrological processes* , Volume 26, 255-275.

Fortin, J., Turcotte, R., Massicotte, S., Moussa, R., Fitzback, J., & Villeneuve, J. (2001). A distributed watershed model compatible with remote sensing and GIS data, I:Description of the model. *Journal of Hydrology* , Volume 6, 91-99.

Gash, J., & Stewart, J. (1977). The evaporation from Thetford Forest during 1975. *Journal of Hydrology* , Volume 35, 385-396.

Karl, T., Knight, R., & Plummer, N. (1995). Trends in high-frequency variability in the twentieth century. *Nature* , Volume 377, 217-220.

Krzysztofowicz, R. (2001). The case for probabilistic forcasting in hydrology. *Journal of Hydrology* , Volume 249, 2-9.

Liu, Y., & Corluy, J. (2005). *Steps of running WETSPA* . Brussel: Vrije Universiteit Brussel; Department of Hydrology and Hydraulic Engineering.

Liu, Y., & De Smedt, F. (2004). Documentation and User Manual. *WetSpa Extension; A GIS-based Hydorlogic Model for Flood Prediction and Watershed Management* . Vrije Universiteit Brussel; Department of Hydrology and Hydraulic Engineering.

Liu, Y., & De Smedt, F. (2005). Flood Modeling for Complex Terrain Using GIS and Remote Sensed Information. *Water resource management* , Volume 19, 605-624.

Liu, Y., Gebremeskel, S., De Smedt, F., Hoffmann, L., & Pfister, L. (2006). Predicting storm runoff from different land-use classes using a geographical information system-based distributed model. *Hydrological processes* , Volume 20, 533-548.

Metz, B., Davidson, O., Bosch, P., Dave, R., & Meyer (eds), L. (2007). *Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge university press.

Morris, M. (1991). Factorial sampling plans for prelimenary computational experiments. *Technometrics* , Volume 33, 161-174.

NCBUY. (2009). Retrieved may 7, 2009, from http://dl.ncbuy.com/imp/vm-map/jpg

Nguyen, T. G., & De Kok, J. (2006). Systematic testing of an integrated systems model for coastal zone management using sensitivity and uncertainty analyses. *Environmental Modelling & Software* , Volume 22, 1572-1587.

Nurmohamed, R., Naipal, S., & De Smedt, F. (2006). Hydrologic modeling of the Upper Suriname River basin using WetSpa and ArcView GIS. *Journal of spatial Hydrology* , Volume 6, 1-17.

Safari, A., De Smedt, F., & Moreda, F. (2009). WetSpa model application in the Distributed Model Intercomparison Project (DMIP2). *Journal of Hydrology* , Article in press.

Saltelli, A., Chan, K., & Scott, E. (2000). *Sensitivity Analysis.* Chichester: John Wiley and Sons Ltd.

Son, N. (2008). Research on simulating rainfall runoff process in order to use water and land resources in some Midland headwater catchments sensibility. Vietnam: Hanoi University of Science.

Tsonis, A. (1996). Widespread increases in low-frequency variability of precipation over the last century. *Nature* , Volume 382, 700-702.

Uhlenbrook, S., & Sieber, A. (2005). On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model. *Environmental modelling and software* , Volume 20, 19-32.

Wikimedia Commons. (n.d.). Retrieved may 5, 2009, from http://commons.wikimedia.org/wiki/File:Blank-map-world-reverse.png

Wikipedia. (2006). Retrieved April 15, 2009, from http://en.wikipedia.org/wiki/File:LocationVietnamQuangNgai.png

Wikipedia. (2009, 30 March). Retrieved 2 April, 2009, from Wikipedia: http://nl.wikipedia.org/wiki/Vietnam

Yu, P., Yang, Y., & Chen, S. (2001). Comparison of uncertainty analysis methods for a distributed rainfall-runoff model. *Journal of Hydrology* , Volume 244, 43-59.

Appendices

Table of contents

List of figures

List of tables

A. Studyarea and the WetSpa model

A.1. Input variables during setup time

Table 1: Input variables during setup time

Variable	Description
Cell threshold for stream networks	When the number increased the stream network will be less detailed. During a kind of calibration it is desired to make the stream network similar to the real river.
Threshold for minimum slope	This value has to be set to give the model a minimum slope for slopes how are (nearly) equal to zero. In this way the model could calculate with these values and the discharge could be determined.
Setting a flood frequency	There could be chosen from three options: 2, 10 and 100 year return period. Normally a 2-year return frequency period is chosen.
Cell threshold for the watershed	This value represents how accurate the watersheds will be determined, by adjusting the threshold value the number of watersheds will be changed. Normally it is a multiply of the threshold for stream networks.
A minimum ratio reflecting the moisture condition	Setting a initial moisture condition and for an event based flood modeling (that will be done in this thesis) a very important input value.
Choose a way to determine the manning's coefficient	There are 3 options: interpolation among different stream orders, remain the default constant as in the lookup table or change to another constant. For this thesis the second option will be used, so the lookup tables will be utilized.
Percentage for urban area	This values represents the urban areas in the study area. The default percentage for urban areas is 30%.
Setting a flow limit or not	There has to be chosen a minimum and maximum velocity or there could be chosen to do it not.

A.2. Global parameters

The parameters for calculating the snow melting wouldn't be described because they would not take into account.

Table 2: Global parameters

Parameter	Description
Dt(h)	Time step in hours. Set the time step 24 hour in case of a daily scale. It 's also possible that the timescale is less than one, e.g. set the time step to 0,5 when the model runs on time scale of half an hour. Note: it is necessary that all the meteorological data has the same time step as the one specified for Dt. (Liu & Corluy, 2005)
K_i	Scaling factor for interflow computation. In the WetSpa model interflow is assumed to occur when soil moisture exceeds the field capacity and there is sufficient hydraulic gradient to move the water. Then Darcy's law is used for the simulation of interflow. Especially for the areas with sloping landscapes and well-vegetated cover this is a important parameter.
K_g	Groundwater recession coefficient This parameter takes the area effect into account and effects the groundwater recession regime for an averaged subcatchment area (total area divided by the number of subcatchments). The program will calculate the actual recession coefficient for each subcatchment based upon its area.
K_{ss}	Initial soil moisture It is a key element in the model controlling the hydrological processes of surface runoff production, evapotranspiration, percolation and interflow. A well calibrated initial soil moisture provides a much more realistic starting point for the predictions. For long-term flow simulation this parameter is less important as it effects the long-flow simulation only in the initial part of the simulation. Therefore an assumption of uniform initial moisture distribution can be made in this case with modeling purpose of flood prediction under present condition.
G_0	Initial groundwater storage in water depth(mm) The groundwater balance is calculated on subcatchment scale. But this value is set up in the input parameter file for all subcatchment.
G_{max}	Maximum groundwater storage in water depth (mm) This parameter effects the amount of water extracted from the groundwater storage for evapotranspiration.
K_{run}	Surface runoff exponent when te rainfall intensity is very small Rainfall intensity has a big influence in controlling the proportion of surface runoff and infiltration. For this reason an empirical exponent is introduced in the model. The concept is that the proportion of surface runoff is very small, or even nil, under de condition of very small rainfall intensity, and the proportion increases along with the increase of rainfall intensity. When the parameter is equal to 1, the actual runoff coefficient is a linear function of the relative soil moisture content, and the effect of rainfall intensity on the runoff coefficient is not taken into account.
P_{max}	The threshold rainfall intensity (mm/d or mm/hour; depending on the timestep) This value describes when the precipitation intensity is higher than P_max, a linear relationship is used to the surface runoff parameter and to achieve this K_run is set to 1.

A.3. Preliminary sensitivity analysis

Table 3: Parameter sensitivity for model calibration (Liu (1) & De Smedt, 2004)

Parameter	Relative sensitivity	Major effects	Calibration priority	Independent evaluation
Precipitation/Evapotranspiration				
Station weight	High	Runoff volume	1	√
Correction factor	High	Runoff volume	1	
Vegetation fraction	High	Runoff volume	2	
Vertical precipitation gradient	Medium	Runoff volume	2	√
Vertical PET gradient	Medium	Runoff volume	2	√
Maximum groundwater storage	Medium	Low flow shape	2	
Snowmelt				
Base temperature	High	Snowmelt	1	√
Temperature degree-day factor	High	Snowmelt	1	√
Rainfall degree-day factor	High	Snowmelt	2	√
Temperature lapse rate	High	Snowmelt	2	√
Runoff distribution				
Potential runoff coefficient	High	Volume, high flow shape	1	
Surface runoff exponent	High	Volume, peak discharge	1	
Threshold rainfall intensity	High	Volume, peak discharge	1	
Impervious fraction	High	Volume, high flow shape	1	√
Interception capacity	Medium	Runoff volume	2	√
Depression capacity	Medium	Runoff volume	2	√
Flow routing				
Surface roughness coefficient	Medium	High flow shape	2	√
Channel roughness coefficient	High	High flow shape	2	√
Hydraulic radius	High	High flow shape	2	
Threshold of minimum slope	Medium	High flow shape	3	
Threshold of stream network	Medium	High flow shape	3	
Interflow scaling factor	High	Volume, flow shape	1	
Baseflow recession coefficient	High	Low flow shape	1	
Number of subcatchments	Medium	Low flow shape	3	
Soil properties				
Hydraulic conductivity	Medium	Runoff volume	3	√
Porosity	Low	Runoff volume	3	√
Field capacity	Low	Runoff volume	3	√
Wilting point	Low	Runoff volume	3	√
Residual moisture content	Low	Runoff volume	3	√
Pore size distribution index	Low	Runoff volume	3	√
Root depth	Medium	Runoff volume	3	√
Initial conditions				
Soil moisture	Low	Flow shape	3	√
Groundwater storage	Low	Flow shape	3	√
Interception storage	Low	Flow shape	3	√
Depression storage	Low	Flow shape	3	√
Initial baseflow	Low	Flow shape	3	√

B. Calibration

B.1. Results manual calibration

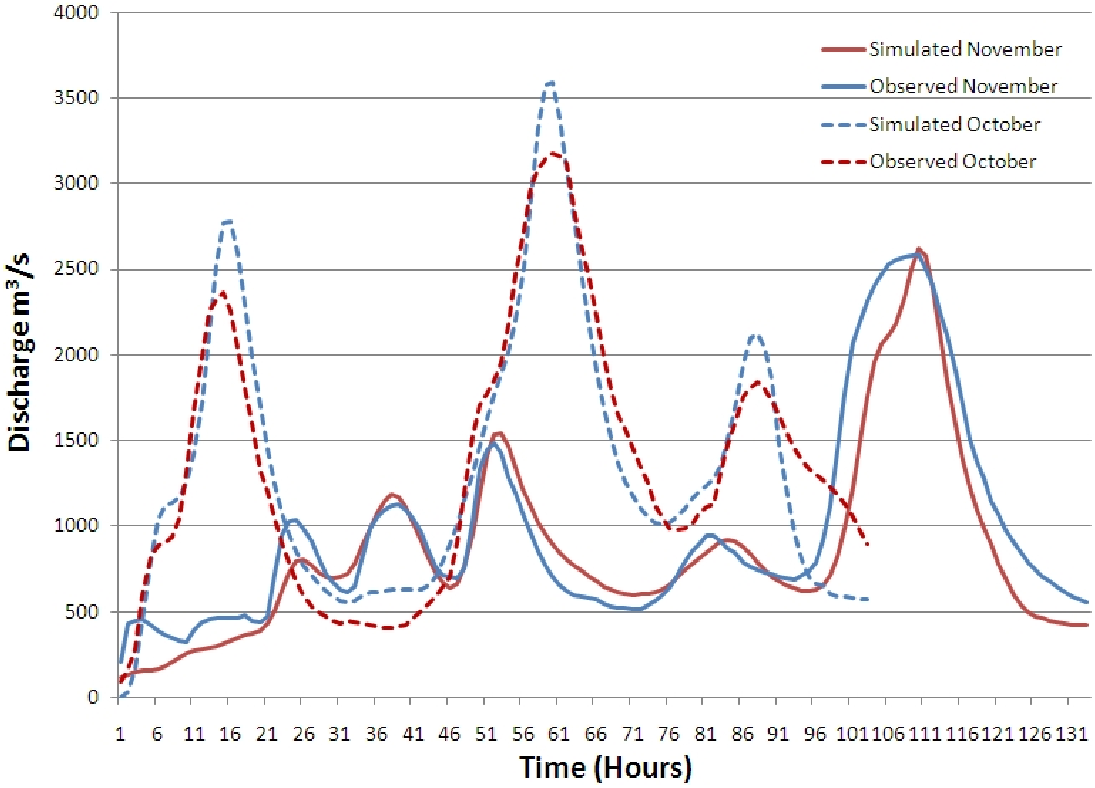


Figure 1: Calibration results after manual calibration

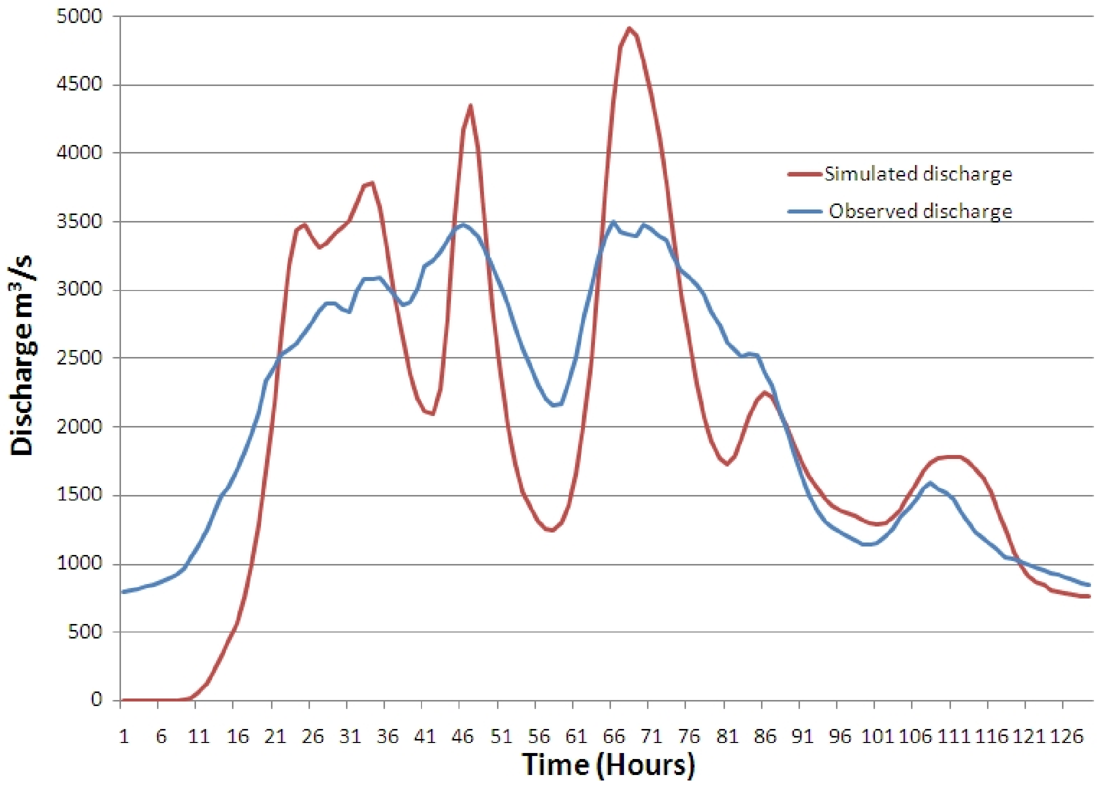


Figure 2: Validation result after manual calibration

B.2. Ranges parametersets

Table 4: Parametersets for Random Sampling and Latin Hypercube Sampling

Parameterset	1	2	3	4	5	6	7	8	9	LHS(1)	LHS(2)
Ki_min	0	3	0	3	2	2	2	2	2	0	3
Ki_max	12	8	8	5	6	6	6	10	10	10	11
Kg_min	0	0	0.01	0.03	0	0	0	0	0	0	0
Kg_max	0.5	0.1	0.06	0.03	0.04	0.025	0.025	0.03	0.03	0.5	0.1
K_ss_min	0	0	0	0	0.5	0.5	0.5	0	0	0	0
Kss_max	2	2	3	3	2.5	2.5	2.5	2.5	2.5	2.5	1.5
G0_min	1	0	0	0	0	0	0	0	0	0	0
G0_max	100	60	40	40	40	40	40	40	40	50	50
Gmax_min	25	0	0	0	0	0	0	0	0	0	50
Gmax_max	125	100	100	100	100	100	100	100	100	200	150
Krun_min	0	0	0	0	0	0	0	0	0	0	0
Krun_max	12	12	12	12	12	12	12	12	12	12	12
Pmax_min	0	0	0	0	0	0	0	0	0	0	0
Pmax_max	500	500	500	500	500	500	500	500	1000	2000	500

B.3. Best graph for December

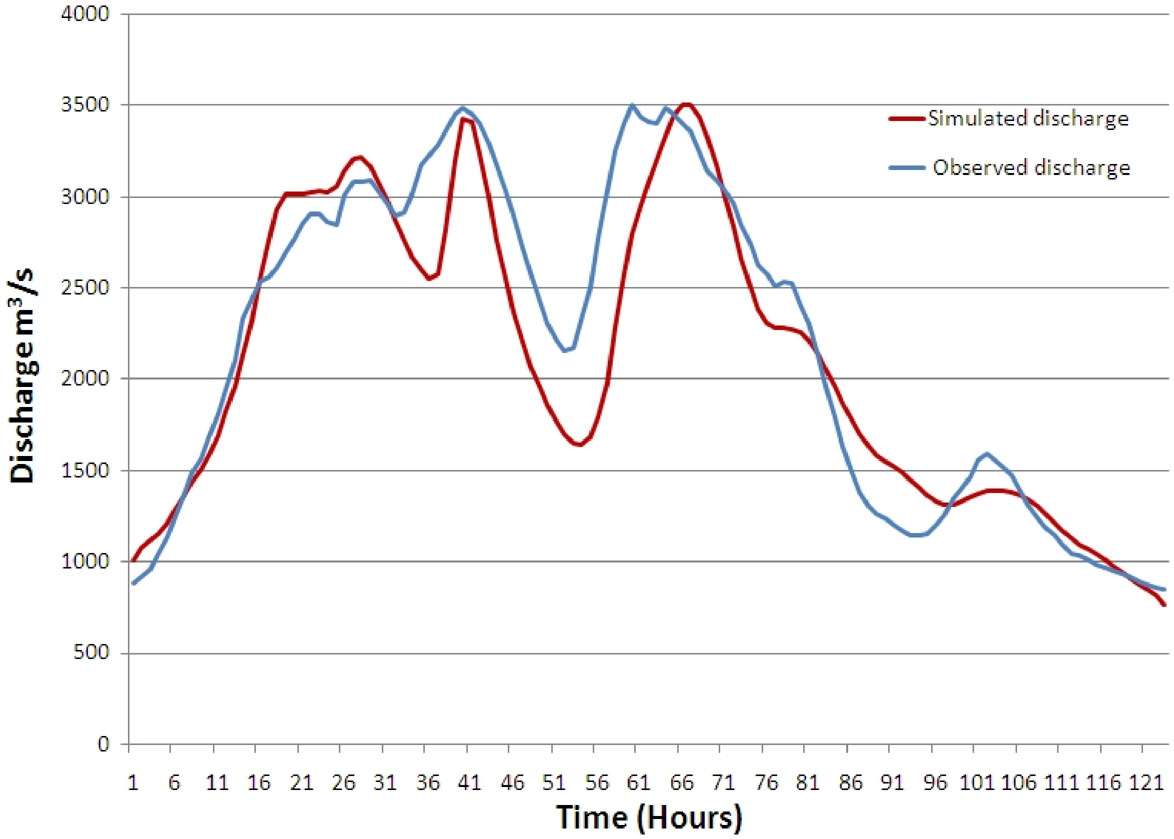


Figure 3: Best Hydrograph for December (NS=0.88)

B.4. Plots after calculating with Random samples

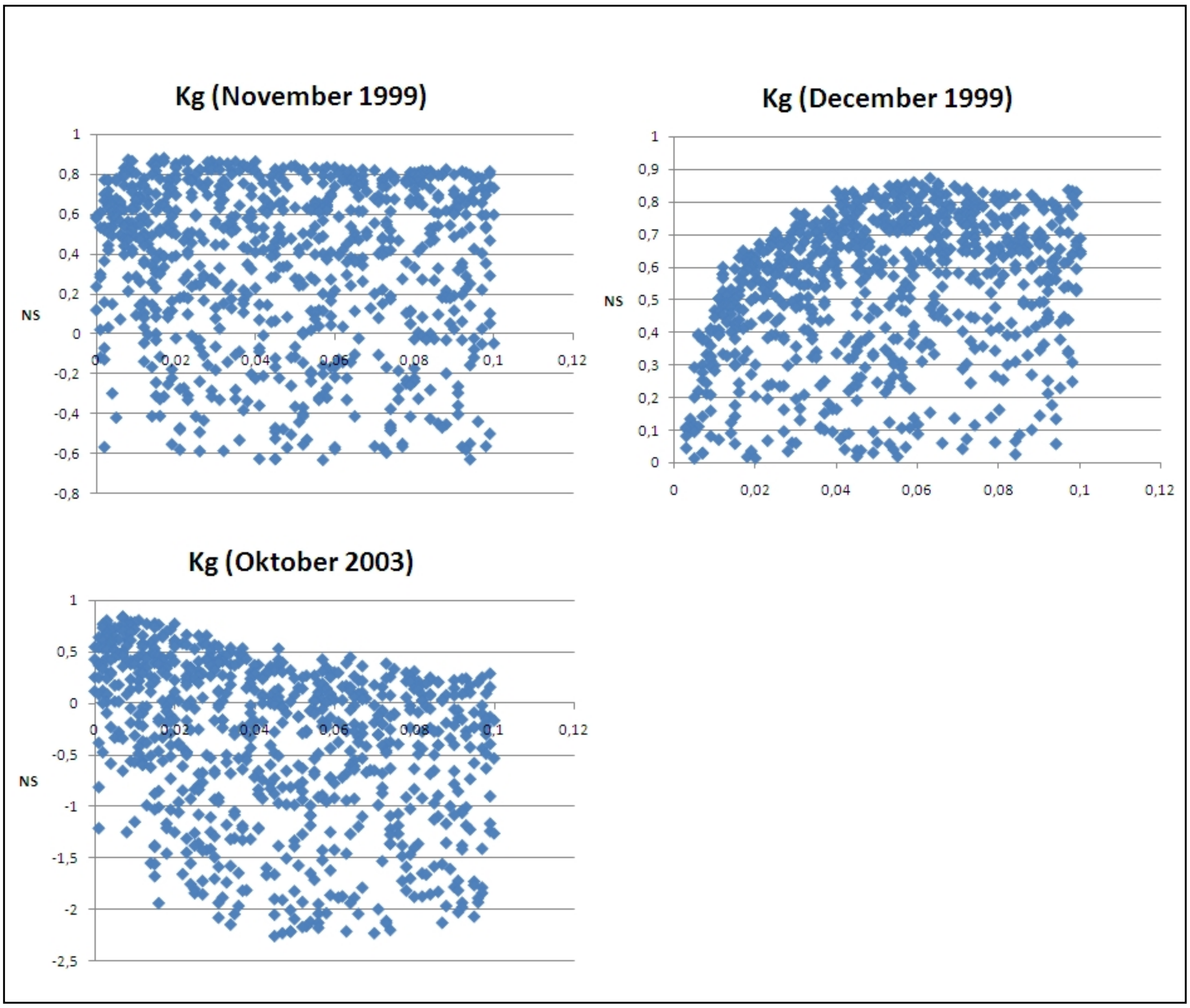


Figure 4: Dotty plots for Kg

B.5. Output after changing Arcview

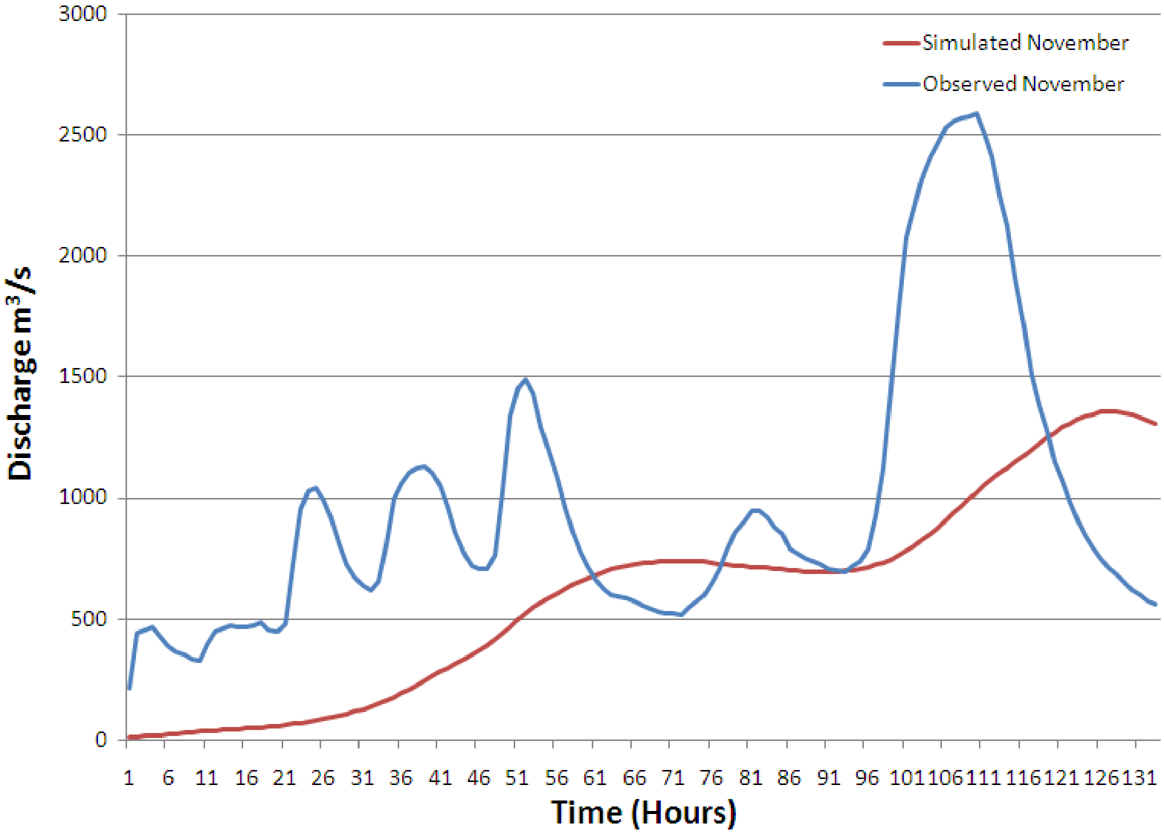


Figure 5: Discharges after trying to change the initial parameter

C. Sensitivity Analysis

C.1. Overview loading maps in Arcview

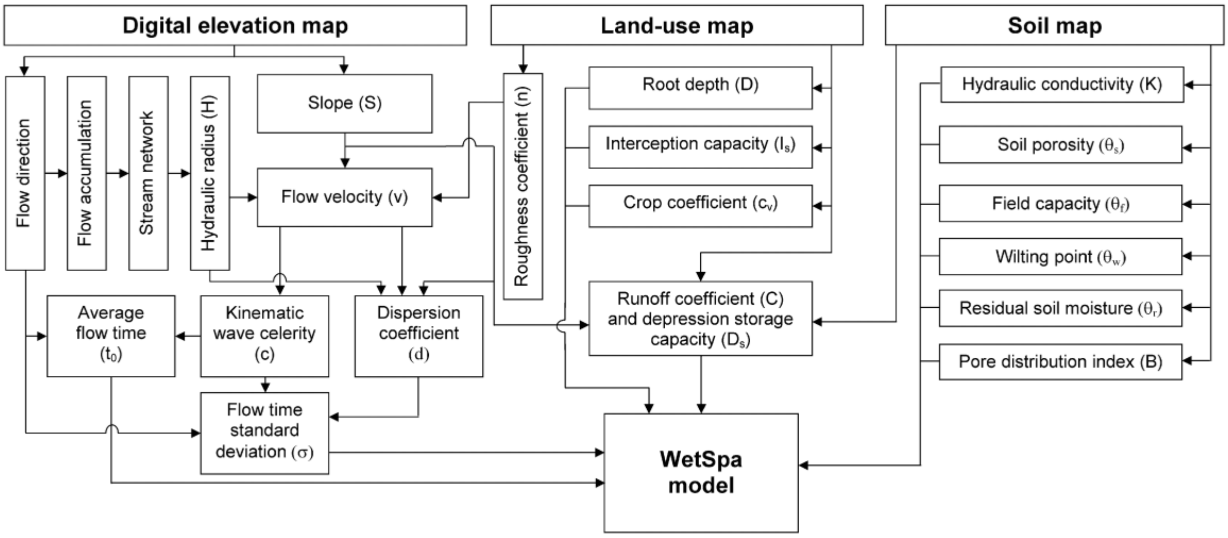
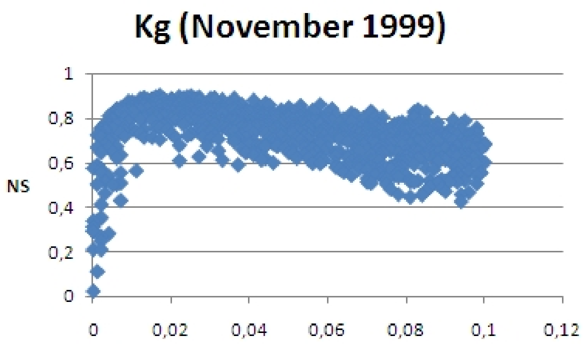


Figure 6: Overview of GIS distributed model parameter derivation on the basis of three basic input maps of elevation, landuse and soiltype. (Safari, De Smedt, & Moreda, 2009)

C.2. Dotty plots Kg



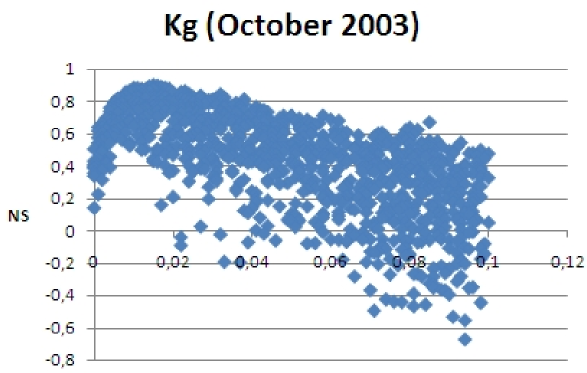
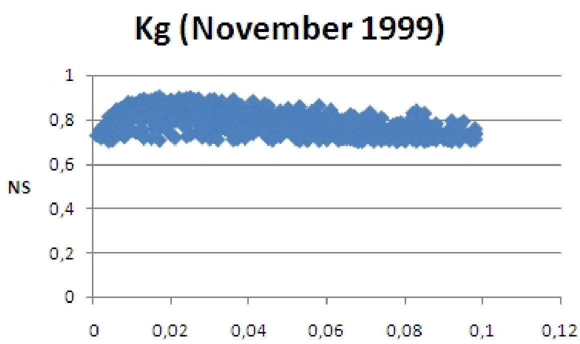


Figure 7: Dotty plots of LHS (R=1000) for both datasets



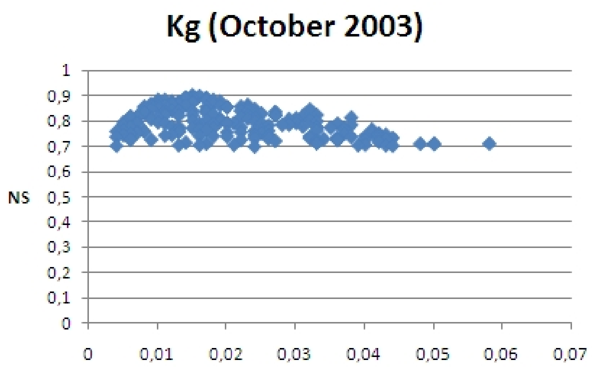


Figure 8: Dotty plot for values with NS>0.7 for both datasets

C.3. Dotty plot G0

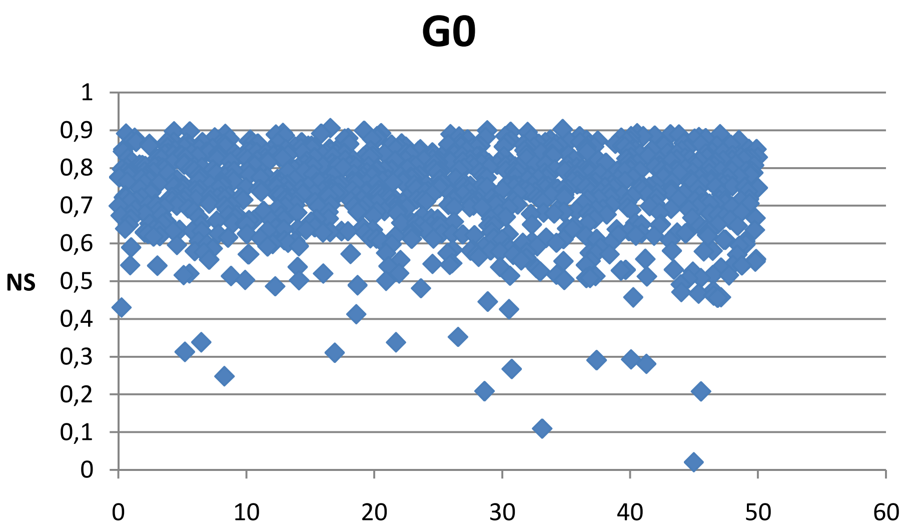


Figure 9: Dotty plot of G0

C.4. Floods for sensitivity analysis

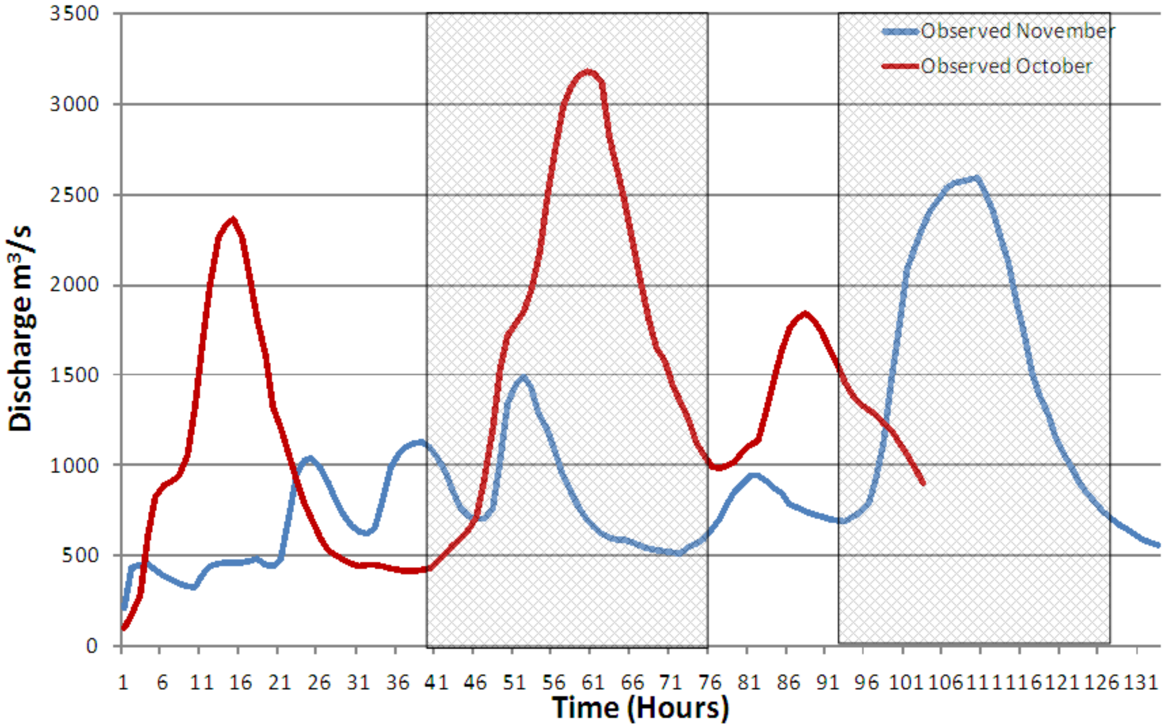


Figure 10: Floods which are taken into account for the sensitivity analysis

D. Results

D.1. Model1 versus Model2

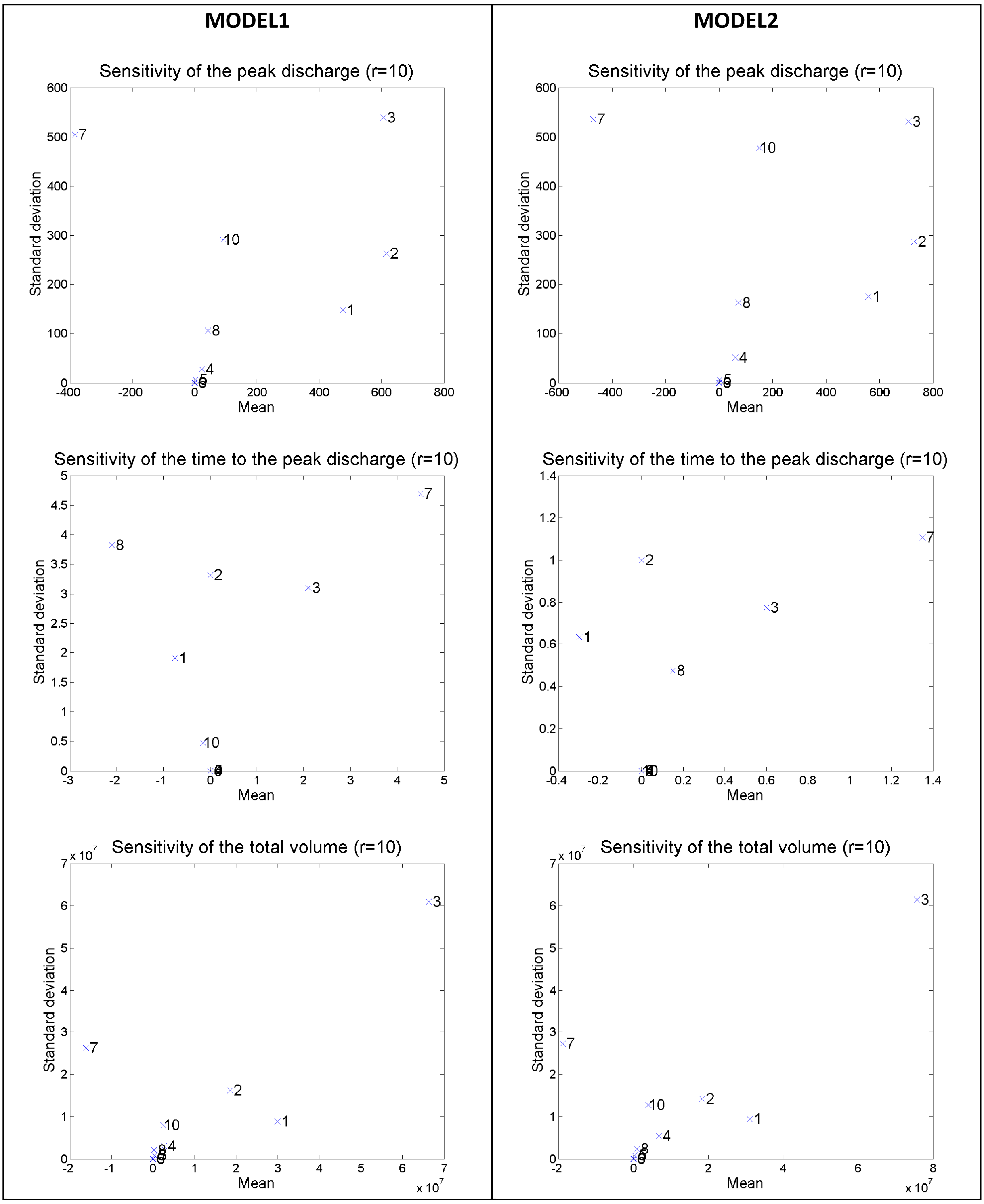


Figure 11: Overview sensitivity of model1 versus model2

D.2. November versus October

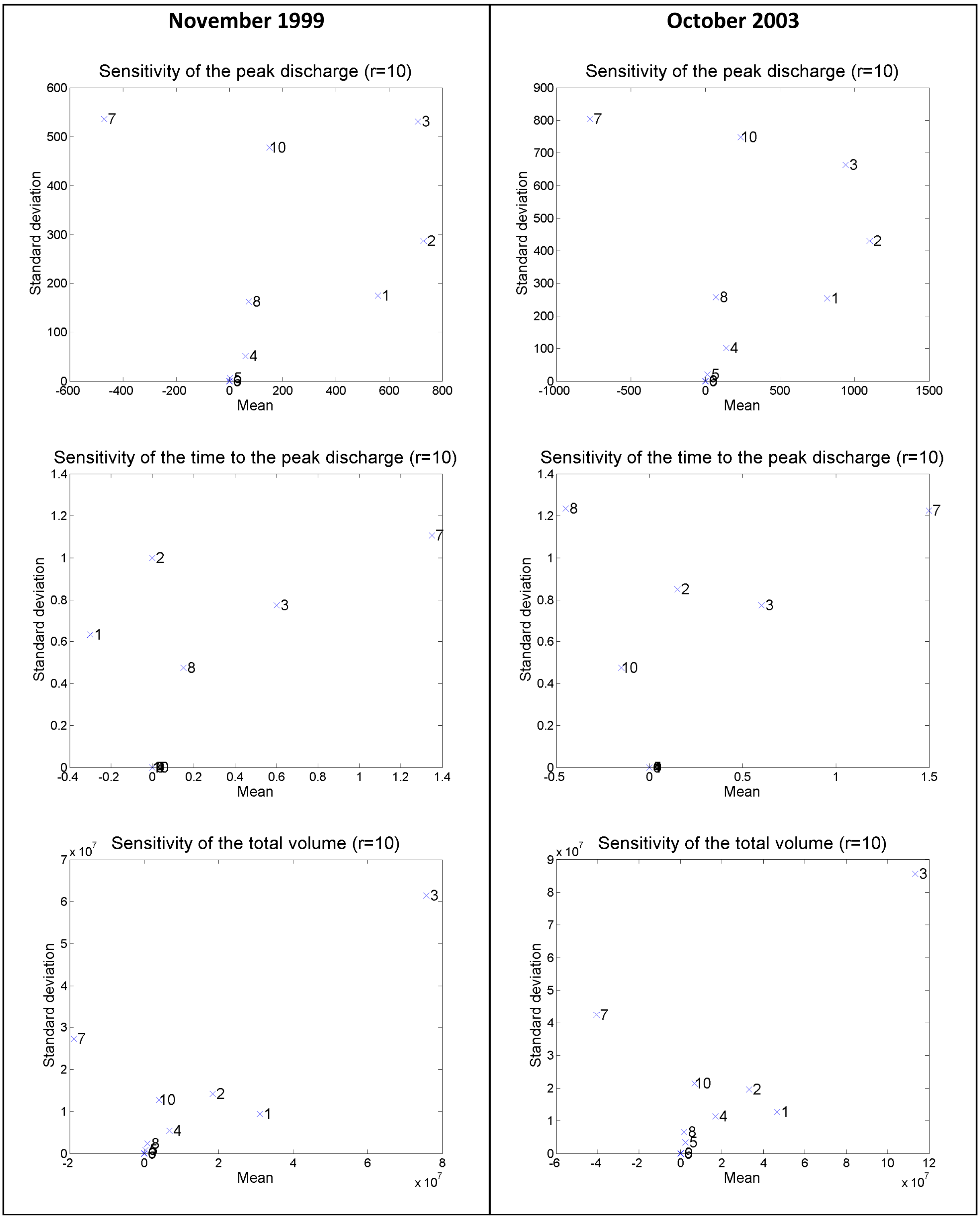


Figure 12: Overview sensitivity of November versus October