Search for the Optimal Selection Formulas When Calibrating Separately for Flood and Dry Season

Bachelor Thesis Timor Post

 15^{th} of April 2013



Search for the Optimal Selection Formulas When Calibrating Separately for Flood and Dry Season

Bachelor Thesis Timor Post

 $15^{\mbox{th}}$ of April 2013

Author:

Timor M.I. Post S0174165

Organizations:

University of Twente China Three Gorges University

Supervisor in The Netherlands: Supervisor in China: Dr. Ir. Martijn J. Booij Dr. Xiaohua Dong



UNIVERSITY OF TWENTE.

Preface

In the beginning of 2012 I started thinking about interesting places to go in order to write my bachelor thesis. Since I always liked travelling and getting out of my comfort-zone, I quickly decided that I should leave my home country, the Netherlands. After looking through many different assignments in many countries, an assignment in the country that I was most interested in did not show up anywhere. The country that I wanted to go to most, was China.

Part of the reason why I wanted to travel to China is that it is a booming economy, experiencing percentages of growth that no Western country can compete with. This has led to many people thinking, or at least saying, that China will end up owning half the world and become the biggest economy. I wanted to experience these statements for myself to see how the country is really like.

When I told professor Martijn Booij about this he immediately thought of professor Xiaohua Dong who was a PHD student at the University of Twente, partially supervised by Martijn Booij. Xiaohua Dong had shown interest a few years before in setting up an exchange program between the University of Twente and the China Three Gorges University (CTGU), where professor Dong had started working. Personally, I really enjoyed the thought of being a pioneer in setting up this program. After the first contact with professor Dong, I was already very welcome to visit and write my thesis at the CTGU in Yichang.

What unfortunately complicated things slightly was that while writing my preparatory report it turned out that the original focus of the research had already been researched, that the result of the research was very predictable and would not add any knowledge to the field. Therefore it was necessary to switch the focus of the research but luckily it was possible to use most parts of the preparatory work.

My visit to China was a great experience and this was helped greatly by my local friend Sun Li. It was very difficult to travel alone because of the language barrier but I was lucky enough that he always organized little trips and we often went out for lunch or dinner.

Another person that I would like to mention here is Yu Dan. She helped me out by translating some Chinese textbook pages and with understanding the Xinanjiang model at some stages of the process. Yu Dan also went on daytrips with me which were a lot of fun and I am very grateful for this.

The process of writing my bachelor thesis, including travelling to Yichang has taught me more than I could have expected on both a personal and professional level. Special thanks go to my two professors, Xiaohua Dong and Martijn Booij, both for their time and professional knowledge. It turned out to be an unforgettable experience and I would like to thank everyone that has been involved at any part of the process.

Summary

This research has calibrated the Xinanjiang model with the Genetic Algorithm using different selection formulas. The Xinanjiang model is a rainfall-runoff model which requires precipitation and evaporation as input and generates the river discharge in cubic meters per second at a certain location and for a certain moment in time. The Genetic Algorithm was used in combination with collected data in the Qingjiang river basin in China. Using the precipitation, evaporation and river discharge, the proper settings (parameter values) were found for the Xinanjiang model so it would function in the Qingjiang river basin. Performing this multiple times with different selection formulas means that it was possible to make a comparison between these formulas in order to find which one works best. Because the model could not function accurately for an entire year with one set of parameters, different parameter values were needed for the wet season and the dry season which has shown in other research (Muleta, 2012) to increase the accuracy of the predictions. Therefore, the goal of the research is to find the selection formulas, which are used in the Genetic Algorithm, that perform best for each season.

The conclusion drawn is that for the flood seasons the Nash-Sutcliffe formula should be used. The Relative Mean Absolute Error has performed satisfactory as well, but there seems to be no reason to prefer this formula over the Nash-Sutcliffe. For the dry season, the best selection formula is the Relative Mean Absolute Error. On a close second place is the RSR formula, which is the Nash-Sutcliffe formula with a square root in the numerator and denominator.

This research has been performed as a bachelor thesis which is requirement for finishing the bachelor study in Civil Engineering at the University of Twente. The research has mostly taken place at the China Three Gorges University in Yichang between the end of 2012 and early 2013. The data of the Qingjiang river basin was already available at the university of which the years 1990 and 1991 were used.

Table of Contents

| Preface |
|-------------------------------|
| Summary |
| 1. Introduction |
| 1.1 Background |
| 1.2 Research Outline |
| 1.3 Studied area7 |
| 2. Xinanjiang Model |
| 2.1 Introduction |
| 2.2 Structure |
| 3. Genetic Algorithm |
| 3.1 Introduction14 |
| 3.2 Structure |
| 3.3 Application15 |
| 4. Results |
| 4.1 Calibration |
| 4.2 Validation |
| 5. Discussion |
| 6. Conclusion |
| 7. Recommendations |
| References |
| Appendix |
| A. Calibrated Results |
| B. Xinanjiang Model in Matlab |

1. Introduction

1.1 Background

Since several thousand years, man has been interested in predicting floods and droughts. This dates back to the ancient Egyptians who were relatively successful at predicting the timing of the flooding of the Nile using a construction known as the Nilometer. By using historical data for their predictions, their predictions were quite accurate (Sivapalan, 2003). This was very important to them since they relied on the annual flooding to fertilize their land. Predictions and techniques have been evolving since then, and a technique developed in 1980 was the Xinanjiang model. The Xinanjiang model is used to determine the river discharge at a certain point in a river and can be used to predict the size, timing and duration of floods and droughts. The Xinanjiang model was chosen since good results have already been achieved on the same river basin using the model. To do this, the model consists of many formulas with parameters that need to fit the local situation in order to predict the correct discharge. The Genetic Algorithm is used to find these values for the Xinanjiang model since it has shown to perform well when calibrating hydrological models in many previous researches (Franchini & Galeati, 1997; Wang, 1991).

The difference between droughts and floods is that in the wet and flood prone months, hydrologists are mainly interested in the timing, size and duration of floods while in the dry months they are more interested in the exact values of river flow in order to distribute the available water to farms, factories and to houses as drinking water. Floods can cause serious damage that leads to loss of resources, money and human lives. Nevertheless, droughts lead to an even higher loss of resources like vegetation, and have caused 42% of all deaths occurring in natural disasters between 1991 and 2000 versus 15% for floods (Unesco Publishing, 2003). This shows the importance of an accurate prediction during the dry months, since it is at least as important as predictions done for the wet months.

Separate predictions for the two different seasons will lead to more accurate predictions (Muleta, 2012). However, the most used and advised formula is the Nash-Sutcliffe formula (Hall, 2001; Vaze et al., 2011). This is a formula that performs especially well in flood predictions, and is also advised for dry seasons (Barma & Varley, 2012) even though the formula has obvious flaws which will be discussed during the report. The aim of this research is to find a better formula for the dry and the wet season to see whether or not the Nash-Sutcliffe deserves the reputation it has.

1.2 Research Outline

Goal

The goal of the research is to give a recommendation for which selection formula to use in the dry season and which selection formula to use in the wet season by comparing different selection formulas for these seasons.

Research Questions

To achieve this goal, it is important to formulate a main research question and the corresponding subquestions.

Main Research Question

Which selection formula should be used when calibrating for the dry season and which selection formula should be used for the wet season?

The answer to the main research question will be answered in Chapter 6. Conclusion

Sub-questions

- 1. Which selection formulas will be used for calibration?
- 2. What is the performance of the calibrated results?
- 3. Which selection formula has resulted in the best calibrated result?
- 4. How do the calibrated results perform when validating using the month following the calibration period?
- 5. Which selection formula has resulted in the best validated result?

Research answers

The answers to the sub-questions can be found in the following chapters:

Table 1 Location of answers to the research sub-questions

| Sub-question | Location of the answer |
|--------------|-------------------------|
| 1 | Chapter 3.3 Application |
| 2 | Chapter 4.1 Calibration |
| 3 | Chapter 4.1 Calibration |
| 4 | Chapter 4.2 Validation |
| 5 | Chapter 4.2 Validation |

1.3 Studied area

For the research, data from a part of the Qingjiang river basin is used. The river basin is located in the Chinese province of Hubei and flows through the Geheyan dam into the Yangtze river. The river runs for a length of 423 km and is mostly banked with steep valleys with their height ranging from 200 to about 1000 meters, which causes the discharge in the river to respond very quick to precipitation. The river discharge data was collected at the Yuxiakou measurement station which means that about 70% of the river basin, shown in grey in Figure 1, is used for the research. Some characteristics of the river basin are stated in Table 2 (Dong, Liu, & Xuan, 2009).

Table 2 Values for the Qingjiang river basin, upstream of the Yuxiakou measurement station(Dong et al., 2009)

| Qingjiang river basin | | | | | | | | |
|--------------------------------|-----------|--|--|--|--|--|--|--|
| Area | 12209 km² | | | | | | | |
| Mean discharge | 464 m³/s | | | | | | | |
| Annual Mean Precipitation | 1400 mm | | | | | | | |
| Annual Mean Evapotranspiration | 820 mm | | | | | | | |



Figure 1: Qingjiang river basin, grey area was used for research (Dong et al., 2009)

2. Xinanjiang Model

2.1 Introduction

The three component Xinanjiang model as it is known today was first published in 1980 (Zhao, Zhang, Fang, Liu, & Zhang, 1980). Preliminary development of the two-component Xinanjiang model forms the base of the current model and was finished in 1973. Inspired by the research done in 1978 by M.J. Kirkby, a third runoff component was added to the model, improving its performance. The term 'three component' refers to the surface, groundwater and interflow that are calculated and combined to predict the river discharge. The model is called a rainfall-runoff basin model and is meant to be used semi-humid to humid areas. Assumed is that precipitation is stored in the pervious soil and will only lead to surface runoff when the storage capacity of the soil is reached. Net precipitation on impervious soil is added to surface runoff. The two other components that form the runoff take place in the soil. The groundwater runoff originates from the deeper and saturated soil whereas the interflow runoff solely takes place in the top parts of the soil. However, when calculating the interflow runoff in the Xinanjiang model, the location of this flow is not important.

Besides projects in China, where the model is most widely implemented, also projects in Nepal, Sri Lanka and the United States have achieved positive results using the Xinanjiang model (Yuanyuan, Xuegang, & Zhijia, 2012).



Figure 2 Xinanjiang model flowchart

2.2 Structure

Input

The model requires two sets of data as input: P [mm] and E [mm]. These values are either measured or predicted and stand for: precipitation and pan evaporation, or open water evaporation. These values are often measured at multiple locations in an area to create an average value for a region.

Division of precipitation on pervious & impervious areas

Precipitation on pervious soil will cause infiltration until the storage capacity of the soil is reached, after which the precipitation will cause surface runoff. This makes precipitation on pervious areas behave differently than precipitation on impervious areas where the water will not infiltrate and always cause surface runoff. Since the Three Layer Soil Model calculates the soil evaporation and runoff after infiltration it is important to make this division between pervious and impervious precipitation. The dimensionless IM factor that is multiplied with the precipitation-values to calculate the direct runoff resulting from the impervious area, has a theoretic value between 0 and 1 but in reality is likely to be $IM \le 0, 15$. The precipitation on the pervious area $P_p[mm]$ is the difference between the measured precipitation and $P_i[mm]$.

| Precipitation impervious area | $P_i = P * IM = P * Impervious Area / Total Area$ | Eq. 1 |
|-------------------------------|---|-------|
| Precipitation pervious area | $P_p = P - P_i = P * (1 - IM)$ | Eq. 2 |

Three layer soil model

The area of the basin that is pervious will allow for infiltration of water. Runoff is generated when the total storage capacity of the pervious surface is reached.

The current storage of water in the soil consists of three layers: the Upper Layer Storage ULS [mm], Lower Layer Storage LLS [mm] and Deep Layer Storage DLS [mm] (see Figure 3). Infiltration occurs when the precipitation is greater than the evaporation: $P_p - E > 0$. The order in which this occurs can be seen in Figure 4. Replenishment of a layer is possible when the deeper layers have reached their capacity values, this process is shown on the left side of the figure. The moisture capacities of the three layers are the parameters ULSC [mm], LLSC [mm] and DLSC [mm]. These three values must be found by calibration, a list of all values that need calibration is shown in chapter 3, Table



Figure 3 Visualization of the three soil layers

Once the upper layer is filled to capacity, any remaining precipitation causes runoff R [mm], also known as the net precipitation.



Figure 4 Three Layer Soil Moisture model flowchart with the processes following net precipitation on the left and net evaporation right

Evaporation of water in the soil layers occurs when the evaporation is greater than the precipitation: $P_p - E < 0$ (right side of Figure 4). If $P_p - E > 0$ then the evaporation can be subtracted from the precipitation directly. The net evaporation is calculated through the Three-Layer Soil Moisture model depending on the capacity values for the three layers, the ratio of pan evaporation to potential evaporation K [-] and C [-], the value for LLS when the evaporation continues into the Deep Layer Storage.

The evaporation reduces ULS [mm] to below 0 mm. If the Upper Layer Storage becomes negative, the remaining evaporation (negative value of ULS) continues into the Lower Layer Storage, LLS [mm], at a reduced rate:

| Lower Layer Storage | $LLS = LLS_0 - ELL$ | Eq. 3 |
|-------------------------|----------------------|-------|
| With: | | |
| Evaporation Lower layer | ELL = ULS * LLS/LLSC | Eq. 4 |

If the Lower Layer Storage reaches C [mm], the evaporation continues at a further reduced rate in the deep layer:

| Deep Layer Storage | $DLS = DLS_0 - EDL$ | Eq. 5 |
|------------------------|-----------------------------------|-------|
| With: | | |
| Evaporation Deep layer | $EDL = C * (K * E - ULS) - LLS_0$ | Eq. 6 |

3.

Runoff Division

The runoff R [mm] which resulted from the Three Layer Soil model is partially stored as free water in a reservoir S [mm] and the remaining amount generates the Surface Runoff, RS [mm]. The reservoir is emptied by Interflow Runoff RI [mm] and Groundwater Runoff RG [mm] which are both dependent on linear equations consisting of the stored amount S which is dependent on the surface of pervious area currently producing runoff FR [km²]. f [km²] is the portion of the basin area for



Figure 5 Visualization of RS, RI and RG production

which the free water storage is lower than BU [mm] which is a summation of the previous R-values minus the outflows, and parameters KI [-] and KG [-] are for the determination of the Interflow and Groundwater Runoff, respectively.

| Interflow Runoff | RI(t) = S(t) * FR(t) * KI | Eq. 7 |
|-----------------------|---------------------------|-------|
| Groundwater Runoff | RG(t) = S(t) * FR(t) * KG | Eq. 8 |
| Runoff Producing Area | $FR(t+1) = R(t)/P_p(t)$ | Eq. 9 |

Linear Reservoirs

The Interflow Runoff and Groundwater Runoff are the input values for two separate linear reservoirs, SI and SG, that both have one input and one output. The outputs TI [mm] and TG [mm] are dependent on the current stored amount and are calculated with Eq. 11 and Eq. 12. The Surface Runoff is summed with the precipitation on the impervious area P_i to generate TS. Because the Lag and Route method will require a total basin outflow Q_T [m³/s], the values of TS [mm], TI [mm] and TG [mm] must be summed. However, because their values still represent millimeters and the Q_T value must be in cubic meters per second, the result is multiplied with a value U [km²/hr] (Eq. 14) with A being the total basin area in km² and Δt being the time step between measurements in hours.

| Surface Outflow | $TS(t) = RS(t) + P_i(t)$ | Eq. 10 |
|--|--|------------------|
| Interflow Outflow | TI(t) = TI(t-1) * CI + RI(t) * (1 - CI) | Eq. 11 |
| Groundwater Outflow | TG(t) = TG(t-1) * CG + RG(t) * (1 - CG) | Eq. 12 |
| Total Basin Outflow Converter Value | $Q_{T} = (TS + TI + TG) * U$ $U = A/(18 * \Delta t)$ | Eq. 13 Eq. 14 |

Slope Routing Delay

There are different methods to approximate the actual behavior of the river and the 'lag' that occurs between the rainfall and the river's response. Probably the most well-known methods are the Unit Hydrograph and Lag and Route. The second technique is used in this research since in order to use the Unit Hydrograph method, the response of the basin area has to be approximated using several isolated storms and these were not be found in the runoff data. Also, the Lag and Route method (Eq. 15) is very user-friendly. It requires two parameters: a recession constant CR [-] and lag in time L [6 hr]. The effect of this routing can be seen in Figure 6.

Lag and Route $Q_{Final}(t) = Q_{Final}(t-1) * CR + Q_T(t-L) * (1-CR)$ Eq. 15



Figure 6 Lag and Route method, upstream versus downstream

3. Genetic Algorithm

3.1 Introduction

The basic theory behind the Genetic Algorithm (GA) was first developed in the mid 1950's by Nils Aall Barricelli and Alex Fraser (Chen, Davis, Jiang, & Novobilski, 2011) but their ideas were still meant for the simulation of the evolutionary process since both researchers were geneticists. Widespread knowledge and use arose around 1970 after the algorithm was recognized to perform well in evolutionary simulations. The first books on the subject were published in the 1960's but it was Goldberg who first used the Genetic Algorithm for problem-solving when he solved travelling salesmen problems in 1989 (Goldberg, 1989).

The Genetic Algorithm creates a population of individuals (or: solutions) that all have their own combination of genes (or: parameter values). These genes determine how well the individual is able to handle a certain problem. The population starts out with many different possible combinations inside the limitations of the species (these limitations are the lower and upper bounds of the parameters). However, as the problem arises, the individuals with the best genes will have a higher chance to produce offspring than the individuals with lesser genes. The offspring will have a combination of their parents' genes since the parents create combinations of their own successful genes with a process known as 'crossover'. This process takes place in every generation until the group has grown to the original amount of individuals with, hopefully, an even better combination of genes to survive the problem. Finally, after many generations of this process, with the best individuals having the highest chance to survive each generation, the individuals are generated that are most suitable for the current problem. The problem in this situation is to match the output of the Xinanjiang model to the measured river flow values.

3.2 Structure

The structure of the Genetic Algorithm can be seen in the flowchart shown in Figure 7. The different steps in the process are explained below.

Generate Initial Population

The initial population consists of solutions that comprise of randomly chosen parameter values within the given boundaries. A greater initial population will lead to better results as a larger area of the solution space is covered. However, in the case of a hydrological model consisting of many parameters it is impossible to cover all the possible solutions. To cover most of the solution space it is important to start with a high number of solutions, perhaps 100 to several thousands to calibrate since this will have a great effect on the quality of the outcome when the complexity of the subject is large.

Selection

The 'problem' in the case of the Xinanjiang model is the accuracy of the



Figure 7 Genetic Algorithm flowchart

predicted runoff. This accuracy can be assessed using many different performance calculations but some widely used formulas are stated in the next subchapter Application, including some that are proposed as an alternative to the Nash-Sutcliffe formula which is probably the most widely used (Hall, 2001; Vaze et al., 2011). All the formulas measure the error between the predictions and the measured runoff. Solutions that score well in the performance calculation in the Selection stage have a higher chance to be selected to reproduce than lesser solutions (Figure 7).

Mating

The selected solutions will reproduce to create new solutions that hopefully improves the result of the calibration. This is the core-process that allows the GA to reach the best solution. This reproduction can happen in slightly different ways but the final outcome is that an exchange of parameters occurs and thus creates new solutions. This process is called 'Crossover'.

Mutation

An important factor in the generation of new solutions is mutation. Since 'Crossover' only generates combinations of existing solutions, mutations are important to ensure diversity since they randomly change parameter values and allow for truly new solutions to be generated. This process should occur rarely since mutations are often destructive for the individual solution but some can lead to unexplored parts of the solution space which might hold the global optimum as can be seen in the extremely simplified situation in Figure 8.



Stopping Criteria

There are several different stopping criteria to control the model, for example, the change of the best solution over the last 10 generations, or the total amount of generations. Personally, I have found that the amount of generations is a good predictor for the quality of the result. After using the GA calibration several times it will become quite clear where the best individual stops improving. However, generations are a crucial part of the GA as it allows for the solutions to iterate towards the optimal combination of parameters so it is important to allow for enough generations in order to compare results between calibrations. Using too many generations will only result in a longer calculation time without improving the result.

3.3 Application

The input that was used for the Genetic Algorithm has a great influence on the results. The input and settings for the GA that was used can be found below. Literature has two different beliefs on the optimal

settings. Either a researcher should use a large population, several hundreds or thousands, with a small amount of generations, or a small population, around 50, with a large amount of generations in combination with a high chance on mutations. Since the literature is inconclusive about this subject, the research has chosen the first option (Gotshall & Rylander, 2000).

Population

For this research, a population of 2000 solutions was chosen. This was due to the long calculation time involved with an even higher population. Since it is advised to choose a population of 100 to 1000 times the length of a solution string, which is equal to the amount of parameters in the Xinanjiang Model, 2000 is an acceptable amount for 17 parameters. The parameters with their corresponding lower and upper boundaries are shown in the table below and were found in (Dong et al., 2009) and by trial and error.

Table 3 Calibrated values with lower and upper bounds

| Symbol | Units | Description | Range |
|--------|-------|--|----------|
| К | - | Value to approach actual evaporation | 0,3-0,99 |
| IM | % | Percentage of the impervious surface | 0,01-0,1 |
| ULSC | mm | Soil moisture storage capacity of the upper layer | 1-250 |
| LLSC | mm | Soil moisture storage capacity of the lower layer | 40-250 |
| DLSC | mm | Soil moisture storage capacity of the deep layer | 20-250 |
| В | - | Attenuation coefficient of surface runoff routing | 0,1-0,99 |
| C | mm | Lower value for Lower Layer Storage to determine Deep Layer Evaporation | 0,1-0,7 |
| SM | mm | Areal mean free water capacity of the surface soil layer, representing the | 5-100 |
| | | maximum possible deficit of free water storage | |
| Ex | - | Exponent of the free water capacity curve influencing the development of | 1-7 |
| | | the saturated area | |
| KG | - | Outflow coefficient of the free water storage to groundwater relationships | 0,01-0,3 |
| KI | - | Outflow coefficient of the free water storage to interflow relationships | 0,01-0,3 |
| CG | - | Recession constants of the groundwater storage | 0,9-0,99 |
| CI | - | Recession constants of the interflow storage | 0,3-0,99 |
| CR | - | Recession constant in the lag and route method for routing | 0,05-0,9 |
| L | 6 hr | Lag in time | 2-4 |
| RIO | mm | Start value for interflow runoff | 0-30 |
| RG0 | mm | Start value for groundwater runoff | 0-30 |

Selection formulas

This section contains the chosen selection formulas and, as noted in the previous subchapter, since the Nash-Sutcliffe formula is widely used, it will serve as a comparison for the performance of the other selection formula'.

Nash-Sutcliffe
$$NS = 1 - \frac{\sum_{t=1}^{N} (Q_{c,t} - Q_{o,t})^2}{\sum_{t=1}^{N} (Q_{o,t} - \overline{Q_o})^2}$$
 Eq. 16

In the Nash-Sutcliffe formula, $Q_{c,t}$ is the calculated runoff at point in time t and $Q_{o,t}$ is the observed (or measured) value. $\overline{Q_o}$ is the mean of all observed runoff values. The problem with this formula is that it

favors solving large differences while leaving smaller but systematic differences unsolved. For an example of this problem, see Figure 9. The Nash-Sutcliffe will prefer solution B over solution C because of the effect the square in the numerator has even though the net error is larger for B than for C. Equations 17 through 19 will prefer solution C over B, minimizing the total error. In wet periods, hydrologists are mostly interested in the timing and size of floods for which the Nash-Sutcliffe is perfect. However, the exact value of the runoff over a long period of time during a drought can be very beneficial as well. To find the effects that the choice of selection-formula has, the different formulas will be applied to both the wet and the dry period during the calibration and validation process in the next chapters.



Figure 9 Calibrated results, blue line are the Target Values and the black line the Generated Values. Situation A could be the result from the first iteration, B is an example of the second iteration when using Eq. 16 or Eq. 20, C is an example of the second iteration when using Eq. 17, Eq. 18 or Eq. 19

The Relative Mean Absolute Error (RMAE) is a commonly used formula to calculate the performance (Eq. 17). It will be used in this research because the formula will prefer situation C over B in Figure 9. For this reason a more extreme version of this formula has been created too. This is done by taking the square root of the numerator (Eq. 18) to make small differences between the calculated runoff and the observed runoff more important to the calibration process than in the RMAE.

Relative Mean Absolute Error
$$RMAE = \frac{\sum_{t=1}^{N} |Q_{c,t} - Q_{o,t}|}{\sum_{t=1}^{N} Q_{o,t}}$$
Eq. 17Relative Mean Root Absolute Error $RMRAE = \frac{\sum_{t=1}^{N} \sqrt{|Q_{c,t} - Q_{o,t}|}}{\sum_{t=1}^{N} Q_{o,t}}$ Eq. 18

The Equal Volume formula is the most experimental of the performance formulas. It has an obvious flaw since when the total volumes are equal over the selected timeframe, this means nothing for individual moments in time. While no great results are expected from this formula, it might give an interesting result. The RSR formula is proposed by (Moriasi & Arnold, 2007) as a formula which can be used to compare hydrological models, since the value of the widely used Nash-Sutcliffe coefficient is highly dependent on the situation in which it is used and can therefore not be compared to other results.

Equal Volume
$$EV = 1 - \frac{\sum_{t=1}^{n} Q_{c,t}}{\sum_{t=1}^{n} Q_{o,t}}$$
 Eq. 19

RSR
$$RSR = \frac{\sqrt{\sum_{t=1}^{N} (Q_{c,t} - Q_{o,t})^2}}{\sqrt{\sum_{t=1}^{N} (Q_{o,t} - \overline{Q_o})^2}}$$
 Eq. 20

Stopping Criteria

The stopping criteria are also of influence on the final result. When the criteria do not allow for enough iterations, the GA will not be able to function properly. After much experimenting with this, the results did not improve significantly after about generation 25 but to allow for positive surprises the maximum amount of generation was chosen at 40. See the figure below as an example of the calibration process.



Figure 10 This example of the calibration process shows clearly that the last 15 generations do not improve the result significantly but do cause the calibration to take about 60% longer than was necessary

4. Results

4.1 Calibration

All the calibrations for the Xinanjiang model have been executed 5 times using the Genetic Algorithm, after which the best performing calibration was selected for each selection formula, this means a total of 50 times. The settings and structure for this process was discussed in the previous chapters. There are two different months used for calibration, the first was July during the flood season of 1991 and the second was in August during the start of the dry season of 1990. The validations, in chapter 4.2, have used the calibrated values and applied those to the months following the calibrations.

Flood

The optimizations A through E used the measurements taken in July of 1991 which is during the wet season. Optimization A was performed with the Nash-Sutcliffe selection formula, B with RMAE and so on. This is visualized by the bordered cells. The most interesting about these results is that calibration A, using the Nash-Sutcliffe formula, does not only have the lowest N-S score but also performs best when the performance is measured in RSR. Naturally, the best performance for the RSR calculation was expected of calibration E since this optimization was focused on minimizing the RSR value. However, when comparing the N-S formula to the RSR formula it is clear that these are basically the same except for one detail which can be found in the square root decreasing both the numerator and the denominator. Therefore, a minimal N-S value will lead to a minimal RSR value. The reason why the RSR score for calibration A is lower than for calibration E is purely coincidental and is caused by random characteristics in the Genetic Algorithm. The solution given by calibration E is therefore definitely a local optimum in the solution space with A being either a different local optimum or the global optimum.

| | Selection formulas | | | | | | | | | | |
|--------------|--------------------|---------------|---------------|--------------|---------------|--|--|--|--|--|--|
| Calibrations | Nash-Sutcliffe | RMAE | RMRAE | Equal Volume | RSR | | | | | | |
| А | <u>0,9477</u> | 0,1593 | 0,0113 | 2,22% | <u>0,2286</u> | | | | | | |
| В | 0,9380 | <u>0,1273</u> | 0,0090 | 3,14% | 0,2490 | | | | | | |
| С | 0,8571 | 0,1468 | <u>0,0084</u> | 0,51% | 0,3780 | | | | | | |
| D | 0,8953 | 0,1651 | 0,0103 | <u>0,00%</u> | 0,3235 | | | | | | |
| E | 0,9470 | 0,1521 | 0,0109 | 1,23% | 0,2302 | | | | | | |

Table 4 Calibration results for July 1991 during the wet season



Figure 11 Optimization A: wet period using Nash-Sutcliffe



Figure 12 Optimization B: wet period using RMAE



Figure 13 Optimization C: wet period using RMRAE







Figure 15 Optimization E: wet period using RSR

Visual inspection

The results of optimizations A through E are all satisfactory since all calibrations show near-identical behavior to the observed runoff. However, all the optimizations have a common mistake. The peak occurring around 7/3/1991 seems to have a very quick response if we look at the observed runoff and the rainfall-data. The constant lag in time L seems unable to deal with this peak since this calculated runoff peak is delayed in all calibrations. An explanation for this error is the fact that the rainfall data is averaged, resulting in a single precipitation value for an entire river basin. Local storms occurring close to the basin outlet can cause very quick responses that cannot be handled by the lumped Xinanjiang model.

An interesting difference between the figures is that optimizations C and D have a very high peek at the flood starting at 7/1/1991 but estimate the low values between 7/21/1991 and 8/1/1991 with a greater accuracy than optimizations A, B and E.

We will see how these optimizations actually perform when predicting the next month, August 1991, in the chapter Validation.

Drought

The optimizations F through J used the measurements taken in August of 1990 which is during the start of the dry season that year. The results of this set of calibrations is quite surprising since F scored best on three different criteria and very well on the remaining two. Calibration I however has very bad scores except for the one it is calibrated for. This is surprisingly different than the performance shown in the flood calibration by the Equal Volume calibration.

| Selection formulas | | | | | | | | | | | |
|--------------------|----------------|---------------|---------------|--------------|---------------|--|--|--|--|--|--|
| Calibrations | Nash-Sutcliffe | RMAE | RMRAE | Equal Volume | RSR | | | | | | |
| F | <u>0,9642</u> | <u>0,0458</u> | 0,0194 | 0,10% | <u>0,1892</u> | | | | | | |
| G | 0,9563 | <u>0,0459</u> | <u>0,0190</u> | 0,80% | 0,2066 | | | | | | |
| Н | 0,9371 | 0,0519 | <u>0,0190</u> | 0,50% | 0,2509 | | | | | | |
| 1 | 0,0494 | 0,2156 | 0,0409 | <u>0,00%</u> | 0,9750 | | | | | | |
| J | 0,9591 | 0,0499 | 0,0202 | 0,30% | 0,2024 | | | | | | |

Table 5 Calibration results for August 1990 during the dry season



Figure 16 Optimization F: dry period using Nash-Sutcliffe



Figure 17 Optimization G: dry period using RMAE



Figure 18 Optimization H: dry period using RMRAE



Figure 19 Optimization I: dry period using Equal Volume



Figure 20 Optimization J: dry period using RSR

Visual inspection

All calibrations were unable to accurately calculate the small flood occurring at 8/9/1990 or show the small fluctuations occurring in the second part of the month. Instead, the second part of the month looks more like a linear equation. This is partially explained by the absence of measured rainfall so there is no data for the model to respond to. Again, this fluctuation of the observed runoff with no measured rainfall can be explained by measurement errors and local storms (Dong et al., 2009).

Interesting about these figures is that only optimization I shows a high response to rainfall. It is obviously a very bad result with the core of the problem lying in the way the formula is set up since positive errors compensate negative errors but the validation will make the final judgment.

Comparison

The average values between the calibrated flood and drought season differ significantly at some parameters (see Table 6), mainly for the Upper Layer Soil Capacity ULSC, the routing coefficient B, the groundwater outflow coefficient KG, the lag and route recession constant CR and the start value for the groundwater runoff RG0. To take a more detailed look into these large differences, see Figure 21 Figure 22. In these boxplots, with all values divided by the average of the other calibrations for that parameter it is clear that some have a high degree of diversity and some are practically constant over all calibrations. For the parameters ULSC, B, KG and RG0 we also see a high degree of variation within the seasonal calibrations. The only parameter that does not vary a lot per calibration but only per season, is CR (see Eq. 15). This is possibly an important parameter that determines the difference between these seasons.

| Average parameter values | | | | | | | | | | | | | | | | | |
|--------------------------|------|------|-------|--------|-------|------|------|-------|------|------|------|------|------|------|------|------|------|
| | К | IM | ULSC | LLSC | DLSC | В | С | SM | Ex | KG | KI | CG | CI | CR | L | RIO | RG0 |
| Drought | 0,81 | 0,02 | 74,88 | 102,12 | 75,38 | 0,35 | 0,46 | 67,73 | 1,84 | 0,12 | 0,15 | 0,99 | 0,84 | 0,84 | 3,40 | 1,62 | 2,98 |
| Flood | 0,75 | 0,03 | 45,97 | 91,63 | 83,33 | 0,18 | 0,41 | 63,03 | 2,03 | 0,25 | 0,16 | 0,94 | 0,55 | 0,45 | 3,00 | 1,51 | 1,03 |
| % Difference | 8% | 8% | 63% | 11% | 10% | 100% | 10% | 7% | 9% | 54% | 1% | 6% | 52% | 87% | 13% | 7% | 189% |

Table 6 Average parameter values per calibrated season



Figure 21 Boxplot of the calibrated values for the flood season, values are divided by the average value



Figure 22 Boxplot of the calibrated values for the dry season, values are divided by the average value

4.2 Validation

All calibrations as seen above are subjected to the following month. This is the true test since this means that the performances below are actual predictions and not the best attempt of the Genetic Algorithm to approach a set of discharge values.

Flood

The month following the calibrated month was used for validation, this means that for the wet season the month of August 1991 was used. The scores after validation have increased which means that the accuracy has gone down. However, this is normal for validations. Important is that the accuracy does not get too low but more importantly is the visual inspection to make sure that the timing and size of the floods is correct. Interesting is that validation A has the lowest score on three of the criteria and that the RMAE of validation C is lowest by far even though it was not calibrated for this criteria.

| | Selection formulas | | | | | |
|-------------|--------------------|---------------|---------------|---------------|---------------|--|
| Validations | Nash-Sutcliffe | RMAE | RMRAE | Equal Volume | RSR | |
| А | <u>0,8995</u> | 0,3035 | 0,0196 | <u>16,91%</u> | <u>0,3170</u> | |
| В | 0,8988 | 0,3016 | 0,0192 | 24,35% | 0,3182 | |
| С | 0,8918 | <u>0,2115</u> | <u>0,0149</u> | 17,34% | 0,3288 | |
| D | 0,8641 | 0,3658 | 0,0213 | 31,39% | 0,3687 | |
| E | 0,8637 | 0,3452 | 0,0204 | 23,84% | 0,3692 | |

Table 7 Validation results for August 1990 during the wet season







Figure 24 Validation B: wet period using RMAE



Figure 25 Validation C: wet period using RMRAE



Figure 26 Validation D: wet period using Equal Volume



Figure 27 Validation E: wet period using RSR

Visual inspection

Visual inspection tells us that validation C is the only result that seriously overshoots the two big floods but does follow the observed runoff closer after the floods than the other four results. Validation A,B,D and E have similar results but combined with the numerical results shown in Table 7, the N-S and RMAE functions have performed slightly better than the others.

Drought

The month following the calibrated month was used for validation, this means that for the dry season the month of September 1990 was used. This is where the model encounters problems. All the

performance calculations are unacceptably high. The Nash-Sutcliffe scores are negative, which means that the average value of the measurements is closer to reality than the calculated discharges produced by the Xinanjiang model. The RMAE validation, while very inaccurate, seems to be the best of the worst. Possible explanations for the bad predictions will be given in the chapter Discussion, after the visual inspection below.

| | Selection formulas | | | | | | |
|-------------|--------------------|---------------|---------------|---------------|---------------|--|--|
| Validations | Nash-Sutcliffe | RMAE | RMRAE | Equal Volume | RSR | | |
| F | -2,6686 | 0,1735 | 0,0507 | <u>9,574%</u> | 1,9153 | | |
| G | <u>-0,6268</u> | <u>0,1308</u> | <u>0,0448</u> | 10,41% | <u>1,2755</u> | | |
| Н | -14,3742 | 0,2785 | 0,0595 | 27,76% | 3,9210 | | |
| 1 | -89,4851 | 0,6502 | 0,0910 | <u>9,515%</u> | 9,5124 | | |
| J | -2,4104 | 0,1895 | 0,0539 | 14,56% | 1,8466 | | |

Table 8 Validation results for September 1990 during the dry season



Figure 28 Validation F: dry period using Nash-Sutcliffe



Figure 29 Validation G: dry period using RMAE



Figure 30 Validation H: dry period using RMRAE



Figure 31 Validation I: dry period using Equal Volume



Figure 32 Validation J: dry period using RSR

Visual inspection

Visual inspection seems to tell us that calibration F performs well and that G and J perform best. The figures are all drawn at the same scale but when zoomed in, there are many moments where even validations G and J are off by about 20%. Validations H and I have extreme peaks where the other validations only show a small reaction, for example around 9/23/1990. When looking at the parameters belonging to H and I it becomes clear why this happens. Because these validations have an IM value of, respectively, 0,03 and 0,06 which mean an impervious area for these situations of 3 and 6%, they respond too extremely on rainfall. Evidence for this is the fact that the reaction to rainfall is much more intense for validation I, than H which corresponds with the IM values. This can be seen in Table 9 below, the other values for the dry season are only 0,01 and have a smaller reaction to rainfall.

Table 9 IM values for the dry season validations

| | Nash-Sutcliffe (F) | RMAE (G) | RMRAE (H) | Equal Volume (I) | RSR (J) |
|----|--------------------|----------|-----------|------------------|---------|
| IM | 0,01 | 0,01 | 0,03 | 0,06 | 0,01 |

5. Discussion

The Xinanjiang model has performed well during the research. The model is based on observations made in nature which makes the steps that it follows logical. However, all the formulas and relationships within the model are linear and because of this, it lacks some complexity for it to approximate droughts properly when rainfall somehow behaves in a way which the model cannot properly handle. Also, the documentation concerning this model is very limited. The useful literature in English is either in expensive books or in a handful of articles. This lack of information did cause some confusion but in combination with a Chinese textbook, the full Xinanjiang model was available.

The Genetic Algorithm has performed excellently. The only comment is that results gained with the GA are non-repeatable because of the random processes going on in the algorithm. However, because of the way it is built into Matlab, it was very easy to use. It is possible to use it via the *optimtool* which makes it easy to use for a first time user. However, there is also a command to start the Genetic Algorithm of which the most basic form is: *ga(fitnessfcn,nvars)*. This makes it possible to use the algorithm directly from an m-file, which makes the process a lot easier to repeat and to keep a log of the previous runs. This is something that I can recommend for future users and something that I will certainly do in any future use.

The accuracy of the results gained after calibration are problematic. This is because all the results are based on the single calibration which scored best for that selection formula. What would have given a more useful result is a comparison between multiple, well-scoring calibrations per selection formula to judge the selection formulas on an average of these calibrations. However, due to a lack of time, this addition was not possible in the timeframe of this research.

Using the Xinanjiang model in combination with the Genetic Algorithm works very well. The model shows the best performance during the wet season and mainly in flood prediction. However, when the runoff values before and after floods is the subject of a research, the RMRAE formula is a better option than the Nash-Sutcliffe. This is because the weight of several small errors is greater than a single large error which means that the calculated values will be more accurate.

The Nash-Sutcliffe formula is very popular in hydrological modeling and has proven to perform the best in this research for predicting floods. Although the peeks of the floods were almost perfectly predicted, the calculated discharge following a flood was overstated.

The Root Mean Average Error has performed nearly as good as the Nash-Sutcliffe in predicting discharge during the wet season but there is no reason to prefer it over the Nash-Sutcliffe in flood prediction. However, it did perform best in predicting the river discharge for the dry season.

The performance of the Relative Mean Root Absolute Error formula was excellent for the wet season in tracking the discharge closely. It overstates the size of the floods but has a better prediction overall than the other formulas. However, during wet seasons hydrologists are not very interested in this information. The reason that this formula was chosen is that it does not prefer an exact flood prediction but that instead it is more sensitive to the 'power of the masses'. Therefore it has resulted in a close approximation for most of the values (Figure 25) excluding the peeks.

The expectation for the Equal Volume formula was that it would only lead to results like Figure 19. Surprisingly, in the wet season it has performed quite well (Figure 14). Still, this formula is likely to cause problems with calibrations so it is not advised.

Although the RSR formula weighs errors in exactly the same way as Nash-Sutcliffe, it has performed considerably less in the validation part of the research. This is likely due to coincidence which is a large part of the Genetic Algorithm.

6. Conclusion

As stated in the Research Outline, the goal of the research is to give a recommendation for which selection formula to use in the dry season and which selection formula to use in the wet season by comparing different selection formulas for these seasons. Several formulas have been compared in their calibration and validation results, with some clearly outperforming others.

For calibration of the flood season, the Nash-Sutcliffe selection formula has performed best with the Relative Mean Absolute Error at a good second place. In both cases, the peaks were near-perfectly predicted in both size, timing and duration. It is likely that this conclusion is valid for every type of rainfall-to-runoff model with every type of calibration technique. However, since the Nash-Sutcliffe has outperformed the RMAE slightly and is preferred in most researches, there is no reason to prefer the RMAE over the Nash-Sutcliffe.

For calibration of the dry season, the same formulas show up in the top two. Figure 29 and Figure 32 seem identical but the numerical evidence of Table 8 shows that the RMAE has outperformed RSR. Although both Nash-Sutcliffe scores are negative, which tells us that the predicted values were less accurate than the average of the measured values, this score should be disregarded because the low discharge numbers require a very high accuracy to get a good Nash-Sutcliffe score and because the month that was used for validation does not show a lot of variation. This means that the average of all the measured values is in itself actually a good approximation.

7. Recommendations

For future research which uses the Xinanjiang model it is important that the knowledge gap between the English and Chinese literature is lessened. There are many Chinese books available that discuss and explain the model but there is a very limited amount of sources for the non-Chinese.

Interesting research can be done by repeating this research with a different rainfall-runoff model containing more non-linear aspects, calibrated with the Genetic Algorithm. Then by producing multiple calibrations per selection formula and analyzing them in the same way this research has done, the conclusion can be drawn whether or not the conclusions drawn in this research are based on coincidence.

References

- Barma, D., & Varley, I. (2012). *Hydrological modelling practices for estimating low flows guidelines*. Canberra. doi:978-1-921853-83-8
- Chen, S., Davis, S., Jiang, H., & Novobilski, A. (2011). CUDA-Based Genetic Algorithm on Traveling Salesman Problem. *Computer and Information Science*, 241–252.
- Dong, X., Liu, J., & Xuan, Y. (2009). Automatic calibration of a lumped Xinanjiang hydrological model by genetic algorithm. *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, 211–217. doi:10.1109/TIC-STH.2009.5444504
- Franchini, M., & Galeati, G. (1997). Comparing several genetic algorithm schemes for the calibration of conceptual rainfall-runoff models. *Risorgimento*, *42*(June), 357–380.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. (Addison-Wesley, Ed.)*Addison Wesley* (Vol. Addison-We, p. 432). Addison-Wesley. doi:10.1007/s10589-009-9261-6
- Gotshall, S., & Rylander, B. (2000). Optimal Population Size and the Genetic Algorithm 2 Test Parameters. *Proc On Genetic And Evolutionary Computation Conference*.

Hall, M. J. (2001). How well does your model fit the data? Journal Of Hydroinformatics, 3(1), 49–55.

- Klopfenstein, L. C. (2009). Genetic Algorithms. Retrieved April 14, 2013, from http://www.klopfenstein.net/lorenz.aspx/genetic-algorithms
- Moriasi, D., & Arnold, J. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ..., 50*(3), 885–900. Retrieved from ftp://ftp.brc.tamus.edu/pub/outgoing/bkomar/windows/DUET-H_WQ/MoriasiModelEval.pdf
- Muleta, M. K. (2012). Improving Model Performance Using Season-Based Evaluation. *Journal Of Hydrologic Engineering Asce, 17*(JANUARY), 191–200. doi:10.1061/(ASCE)HE.1943-5584.0000421.
- Sivapalan, M. (2003). Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrological Processes*, *17*(15), 3163–3170. doi:10.1002/hyp.5155
- Unesco Publishing. (2003). The United Nations World Water Development Report. *Water for People Water for Life*, 272–290.
- Vaze, J., Jordan, P., Beecham, R., Frost, A., Summerell, G., & eWater Cooperative Research Centre. (2011). *Guidelines for rainfall-runoff modelling: Towards best practice model application*. doi:978-1-921543-51-7
- Wang, Q. J. (1991). The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models. *Water Resources Research*, 27(9), 2467–2471. doi:10.1029/91WR01305

- Yuanyuan, M., Xuegang, Z., & Zhijia, L. (2012). On the Coupled Simulation of Xinanjiang Model with MODFLOW. *Journal of Hydrologic Engineering*. doi:10.1061/(ASCE)HE.1943-5584.0000706
- Zhao, R. J., Zhang, Y. L., Fang, L. R., Liu, X. R., & Zhang, Q. S. (1980). The Xinanjiang model. *Hydrological Forecasting Proceedings Oxford Symposium*, (129), 351–356.

Appendix

A. Calibrated Results

Table 10 Calibrated results for the wet season per selection formula, and their average per parameters

| Wet Season | Nash-Sutcliffe | RMAE | RMRAE | Equal Volume | RSR | Average |
|------------|----------------|--------|-------|--------------|----------------|---------|
| К | 0,83 | 0,65 | 0,92 | 0,68 | 0,68 | 0,75 |
| IM | 0,01 | 0,03 | 0,01 | 0,06 | 0,01 | 0,03 |
| ULSC | 12,78 | 5,39 | 36,72 | 145,52 | 29,44 | 45,97 |
| LLSC | 40,02 | 128,06 | 63,62 | 185,58 | 40,88 | 91,63 |
| DLSC | 32,64 | 165,74 | 92,10 | 66,93 | 59,26 | 83,33 |
| В | 0,10 | 0,12 | 0,14 | 0,41 | 0,12 | 0,18 |
| C | 0,18 | 0,12 | 0,61 | 0,49 | 0,67 | 0,41 |
| SM | 85,35 | 38,03 | 21,77 | 70,15 | 99 <i>,</i> 86 | 63,03 |
| Ex | 3,35 | 2,20 | 1,45 | 1,60 | 1,55 | 2,03 |
| KG | 0,30 | 0,26 | 0,16 | 0,28 | 0,26 | 0,25 |
| KI | 0,15 | 0,08 | 0,27 | 0,14 | 0,14 | 0,16 |
| CG | 0,92 | 0,93 | 0,98 | 0,94 | 0,92 | 0,94 |
| CI | 0,40 | 0,34 | 0,88 | 0,74 | 0,40 | 0,55 |
| CR | 0,47 | 0,60 | 0,53 | 0,29 | 0,37 | 0,45 |
| L | 3,00 | 3,00 | 3,00 | 3,00 | 3,00 | 3,00 |
| RIO | 2,24 | 0,79 | 0,39 | 1,69 | 2,45 | 1,51 |
| RGO | 0,97 | 0,71 | 1,18 | 0,01 | 2,29 | 1,03 |

Table 11 Calibrated results for the dry season per selection formula, and their average per parameters

| Dry Season | Nash-Sutcliffe | RMAE | RMRAE | Equal Volume | RSR | Average |
|------------|----------------|-------|--------|--------------|--------|---------|
| К | 0,86 | 0,86 | 0,65 | 0,83 | 0,84 | 0,81 |
| IM | 0,01 | 0,01 | 0,03 | 0,06 | 0,01 | 0,02 |
| ULSC | 43,57 | 49,72 | 101,95 | 142,98 | 36,18 | 74,88 |
| LLSC | 49,08 | 45,27 | 159,68 | 131,98 | 124,56 | 102,12 |
| DLSC | 55,14 | 27,61 | 175,09 | 99,04 | 20,03 | 75,38 |
| В | 0,19 | 0,11 | 0,16 | 0,74 | 0,58 | 0,35 |
| C | 0,43 | 0,27 | 0,65 | 0,34 | 0,59 | 0,46 |
| SM | 85,00 | 84,97 | 62,49 | 6,34 | 99,83 | 67,73 |
| Ex | 1,54 | 1,54 | 1,29 | 3,51 | 1,35 | 1,84 |
| KG | 0,07 | 0,09 | 0,09 | 0,27 | 0,06 | 0,12 |
| KI | 0,09 | 0,24 | 0,12 | 0,22 | 0,10 | 0,15 |
| CG | 0,99 | 0,99 | 1,00 | 0,99 | 0,99 | 0,99 |
| CI | 0,93 | 0,86 | 0,94 | 0,63 | 0,86 | 0,84 |
| CR | 0,83 | 0,89 | 0,80 | 0,81 | 0,89 | 0,84 |
| L | 4,00 | 3,00 | 4,00 | 3,00 | 3,00 | 3,40 |
| RIO | 3,32 | 0,20 | 2,68 | 0,52 | 1,40 | 1,62 |
| RG0 | 1,57 | 0,50 | 0,28 | 0,98 | 11,55 | 2,98 |

B. Xinanjiang Model in Matlab

```
%Xinanjiang Model with calibrated values A(1:17)
A = \begin{bmatrix} 0.862369745 & 0.01408829 & 43.57346342 & 49.07708639 \end{bmatrix}
55.1356702 0.188093387 0.427643567 85.00099673 1.537894452
0.068908208 0.093422777 0.994607191 0.926641728 0.829516292
4
    3.319457186 1.570220339]; %Example values
K = A(1);
                %Parameter for E to actual evaporation
IM = A(2);
                %Percentage Impervious Area
WUM = A(3);
               %Upper layer capacity
                %Lower layer capacity
WLM = A(4);
              %Deep layer capacity
WDM = A(5);
                %Coefficient of surface runoff routing
B = A(6);
C = A(7);
               %Constant needed for Deep evaporation
SM = A(8); %Areal mean free water capacity
Ex = A(9);
               %Exponent of the free water capacity curve
influencing the development of the saturated area
             %Parameters for Groundwaterflow
KG = A(10);
KI = A(11);
                %Parameter for interflow
CG = A(12); %Recession constant of groundwater storage
               %Recession constant of interflow storage
CI = A(13);
CR = A(14); %Recession constant in the lag and route
method
L = round(A(15)); %Lag in time
RIO = A(16); %Startvalue for Interflow Runoff
RG0 = A(17); %Startvalue for Groundwater Runoff
[Q calculated] = Xinanjiang
```

(K, IM, WUM, WLM, WDM, B, C, SM, Ex, KG, KI, CS, CG, CI, CR, L, RIO, RGO);

```
function [Q_Final] = Xinanjiang
(K,IM,WUM,WLM,WDM,B,C,SM,Ex,KG,KI,CS,CG,CI,CR,L,RI0,RG0)
%Load Inputs
P(1:4384) = 0; %P is the average precipitation per 6 hrs
P(1:1460) = importdata('year_1_p.txt',' ');
P(1461:2920) = importdata('year_2_p.txt',' ');
P(2921:4384) = importdata('year_3_p.txt',' ');
E(1:4384) = 0; %E is the average pan evaporation per 6 hrs
E(1:1460) = importdata('year_1_e.txt',' ');
```

```
E(1461:2920) = importdata('year 2 e.txt', ' ');
E(2921:4384) = importdata('year 3 e.txt',' ');
8-----8
%Values needed in functions
E(1:length(P)) = 0;
                            %Preallocation
WM = WUM+WLM+WDM;
                            %Total capacity of all layers
R(1:length(P)) = 0;R = R'; %Preallocation
WU(1:length(P)) = 0;
                           %Preallocation upper storage
WL(1:length(P)) = 0;
                            %Preallocation lower storage
                            %Preallocation deep storage
WD(1:length(P)) = 0;
WU(1) = WUM/2;
                            %Startingvalue upper layer
                            %Startingvalue lower layer
WL(1) = WLM;
                            %Startingvalue deep layer
WD(1) = WDM;
W(1:length(P)) = 0;
                            %Preallocation
W(1) = 0.5*WUM+WLM+WDM; %Startingcapacity of combined layers
S0 = 0;
                            %Startvalue
FR0 = 0;
                            %Startvalue
A = 12209 - 12209 \times IM;
                            %Pervious surface in km^2
delta t = 6;
                      %Time between measured values in hours
U = A/(3.6*delta t);
                            %To transfer mm to m^3/s
```

8-----8

[P_pervious,P_impervious] = perv_imperv(P,IM,K,E); %Dividing the precipitation into P_pervious and P_impervious for the Pervious and Impervious areas

for m=1:length(P)

```
if P_pervious(m) > 0 %if there is any rainfall then the
water infiltrates and runoff can be produced at point m in
time
[R(m)] = calculate_runoff(W(m),WM,B,P_pervious(m),IM);
%Rainfall infiltrates and R(m)>0 if the water storage
capacity is reached in the pervious soil
end
```

```
[WU(m+1),WL(m+1),WD(m+1),W(m+1),E(m)] =
moisture_content_soil_and_evap(WU(m),WL(m),WD(m),E(m),K,C,P_
pervious(m),WUM,WLM,WDM); %Evaporation is subtracted from
the soil water storage
```

end

[RS,RI,RG] =
net_rainfall_division(R,P_pervious,S0,FR0,SM,KG,KI,Ex);
%Dividing the runoff into Surface, Interflow and Groundwater
Runoff
[T] =
net_rainfall_to_runoff_with_reservoir(RS,RI,RG,CI,CG,RI0,RG0,U,P_impervious); %The different runoffs are routed through
linear reservoirs and combined to create the "sub"-basin
inflow
[Q_Final] = runoff_to_river_flow(T,CR,L); %Lag and Route
method is used to change the "sub"-basin inflow to
compensate for the location of the measurements
end

function [P_pervious,P_impervious] =
perv_imperv(Precipitation,Impervious,K,E)
%Precipitation is divided between Impervious and Pervious in
this function

```
P_pervious = (Precipitation*(1-Impervious))-K*E;
P_impervious = Precipitation*Impervious;
P_pervious = max(0,P_pervious); %Negative numbers become 0
end
```

```
function [R] = calculate_runoff(W,WM,B,PE,IM)
%Runoff (R) Generation on Pervious Areas
Wmm = WM*((1+B)/(1-IM)); %Wmm is the maximum areal mean
tension water capacity
AU = Wmm*(1-(1-W/WM)^(1/(1+B)));
if (PE+AU) <Wmm
        R = (PE-WM+W+WM*(1-(PE+AU)/Wmm)^(B+1));
else
        R = PE-(WM-W);
end
R = R';
end</pre>
```

```
function [WUnew,WLnew,WDnew,W,E] =
moisture content soil and evap(WU,WL,WD,E,K,C,PE,WUM,WLM,WDM
)
%Three soil layer model, infiltration and evaporation
WUnew = WU;
WLnew = WL;
WDnew = WD;
W1=WU+WL+WD; %W1 is the current situation
%REFILLING THE LAYERS IF PE>0
if PE > 0 && WD < WDM
   WDnew = WD+PE;
   if WDnew > WDM
       PE = WDnew-WDM;
       WDnew = WDM;
       WLnew = WL + PE;
       if WLnew > WLM
           PE = WLnew-WLM;
           WLnew = WLM;
           WUnew = WU+PE;
       end
   end
elseif PE > 0 && WL < WLM
    WLnew = WL+PE;
    if WLnew > WLM
        PE = WLnew-WLM;
        WLnew = WLM;
        WUnew = WU + PE;
        if WUnew > WUM
            WUnew = WUM;
        end
    end
elseif PE > 0 && WU < WUM
    WUnew = WU + PE;
    if WUnew > WUM
        WUnew = WUM;
    end
end
Evaporation occurs when PE = 0
if PE == 0 && WU > 0
EU = K * E;
WUnew = WU-EU;
    if WUnew <0
```

```
EL = (K*E+WUnew) *WL/WLM;
        WUnew = 0;
        WLnew = WL-EL;
    end
elseif PE == 0 && WU == 0
    WUnew = 0;
    EU = K * E;
    EL = (K*E) *WL/WLM;
    WLnew = WL-EL;
    if WL < C || WLnew < C
        if WLnew < 0
            WLnew = 0;
        end
        ED = C*(K*E-EU)-EL;
        WDnew = WD-ED;
        if WDnew < 0
           WDnew = 0;
        end
    end
end
W = WUnew+WLnew+WDnew; %W is the new situation
E = W1-W; \&E is the negative change of the total storage in
the layers, which is the evaporation from the layers when PE
= 0
if E > 0
    E = 0;
end
end
```

```
function [RS,RSS,RG] =
net_rainfall_division(R,PE,S0,FR0,SM,KG,KI,Ex)
%RUNOFF_DIVISION Summary of this function goes here
% Detailed explanation goes here
% Detailed explanation goes here
% (1:length(R)+1) = 0;
% (1:length(R)+1) = 0;
% (1:length(R)) = 0;
%
```

```
RS(n) = 0;
        RSS(n) = S(n) * KI * FR(n);
         RG(n) = S(n) * KG * FR(n);
         S(n+1) = S(n) - (RG(n) + RSS(n)) / FR(n);
         if isnan(S(n+1)) == 1;
             S(n+1) = 0;
         end
    elseif R(n) > 0
         FR(n+1) = R(n) / PE(n);
         S(n+1) = S(n) * FR(n) / FR(n+1);
         if isnan(S(n+1)) == 1;
             S(n+1) = 0;
         end
         SMM = SM*(1+Ex);
         SMMF = SMM^{(1-(1-FR(n+1))^{(1/Ex)})};
         SMF = SMMF / (1+Ex);
         AU = SMMF* (1 - (1 - S(n+1) / SMF)^{(1/(1+Ex))});
         if (PE(n)+AU)<=0
             RS(n) = 0;
             RSS(n) = 0;
             RG(n) = 0;
         elseif (PE(n)+AU)>=SMMF
             RS(n) = (PE(n) + S(n+1) - SMF) * FR(n+1);
             RSS(n) = SMF*KI*FR(n+1);
             RG(n) = SMF*KG*FR(n+1);
             S(n+1) = SMF - (RSS(n) + RG(n)) / FR(n+1);
         elseif (PE(n)+AU) > 0 \& (PE(n)+AU) < SMMF
             RS(n) = (PE(n) - SMF + S(n+1) + SMF * (1 - 
(PE(n)+AU)/SMMF)^(Ex+1))*FR(n+1);
             RSS(n) = (PE(n) + S(n+1) -
RS(n)/FR(n+1))*KI*FR(n+1);
             RG(n) = (PE(n)+S(n+1)-RS(n)/FR(n+1))*KG*FR(n+1);
             S(n+1) = S(n+1) + PE(n) -
(RS(n) + RSS(n) + RG(n)) / FR(n+1);
        end
    end
    if S(n+1) > SM
         S(n+1) = SM;
    end
end
end
```

```
function [T] =
net rainfall to runoff with reservoir (RS,RI,RG,CI,CG,RI0,RG0
,U,Pimp)
%T is the outflow of the basin determined by the three
runoffs
TS in mm(1:length(RS)) = 0;
TI in mm(1:length(RI)) = 0;
TG in mm(1:length(RG)) = 0;
for n = 1:length(RS)
     if n==1
     RI(n) = RI0;
     RG(n) = RG0;
     end
     if n>1
     RI(n) = RI(n) + RI(n-1) - TI in mm(n-1);
     RG(n) = RG(n) + RG(n-1) - TG in mm(n-1);
     end
    TS in mm(n) = RS(n) + Pimp(n);
    TI in mm(n) = TI in mm(n) * CI + RI(n) * (1 - CI);
    TG in mm(n) = TG in mm(n) *CG+RG(n) * (1-CG);
end
TS = TS in mm^*U;
TI = TI in mm*U;
TG = TG in mm * U;
T=TS+TI+TG;
end
```

```
function [Q_Final] = runoff_to_river_flow(T,CS,L)
%With the Lag and Route method, the basin outflow T is
changed into the
%actual river flow Q_Final

Q_Final(L:length(T)+L-1) = 0;
TR(1:length(T)+L) = 0;
TR(L:length(T)+L-1) = T;
for t = L+1:length(T)+L
        Q_Final(t) = Q_Final(t-1)*CS+(1-CS)*TR(t-L);
end

Q_Final = Q_Final(L+1:length(T)+L);
end
```