UPPSALA UNIVERSITET

UPTEC W 15043
Examensarbete 30 hp
November 2015

Modelling Hydrological Impacts of Forest Clearcutting through Parameter Regionalization

Modellering av hydrologisk påverkan från skogsavverkning genom parameterregionalisering

---

Benjamin Selling

ABSTRACT

Modelling Hydrological Impacts of Forest Clearcutting through Parameter Regionalization

Benjamin Selling

The aim of this thesis was to test and evaluate whether parameter regionalization of a hydrological model can be used to model the impact of forest clearcutting on streamflow in Sweden. This is an important task to be able to perform water management and impact assessments adequately. The HBV conceptual rainfall-runoff model was applied for 218 Swedish catchments of different sizes that were spread across the country and covered a wide range of different forest cover percentages. The modelling approach included calibration of the model for each catchment using a genetic algorithm and then associating the resulting optimal parameter values with the percentage of forest cover. The obtained relationship between different model parameters and forest cover was validated with help of a paired catchment study site in northern Sweden where a clear cut was done in 2006: calibrated optimal parameter sets of pre- and post-clearcutting conditions were compared to parameter sets obtained from the Sweden-wide analysis.

Correlations were found for about half of the fifteen hydrological model parameters, but the validation with the paired catchment study site could only partially confirm these obtained relationships. The results suggest that the adopted parameter regionalization approach is too basic. However, some of the results seem promising and emphasize the need for further research and development of the approach to provide a more reasonable method to model the impact of forest clearcutting on streamflow.

Keywords: Rainfall-runoff modeling, Forest clear cut, HBV model, Parameter regionalization, Impact assessment.

Department of Earth Sciences. Program for Air, Water and Landscape Science, Uppsala University. Villavägen 16, SE-752 36, UPPSALA, ISSN 1401-5765.

REFERAT

Modellering av hydrologisk påverkan från skogsavverkning genom parameter-regionalisering
Benjamin Selling

Det huvudsakliga målet med detta examensarbete var att testa och utvärdera om parameterregionalisering av en hydrologisk modell kan vara en lämplig metod för att modellera och kvantifiera påverkan från skogsavverkning på vattenbalansen i Sverige. Detta är en viktig uppgift för att kunna hantera våra vattenresurser och utföra konsekvensanalyser på ett tillfredsställande sätt. En konceptuell hydrologisk modell tillämpades på 218 avrinningsområden av olika storlekar och som var geografiskt utspridda i hela Sverige där även andelen skog i avrinningsområdena hade ett brett spektrum. Den använda modelleringsmetoden innefattade kalibrering av varje avrinningsområde genom att använda en genetisk algoritm, varefter de optimala parametervärdeana korrelerades mot andelen skog i avrinningsområdet. Idén med denna metod är att använda dessa potentiella samband för att justera modellparametrarna och därmed simulera en skogsavverkning. De erhållna sambanden mellan modellparametrarna och skogstäcket validerades med hjälp av data från en försöksstudie i norra Sverige där en skogsavverkning gjordes under 2006. Skillnaden mellan de bäst fungerande parametervärdena före och efter skogsavverkningen jämfördes med de tidigare sambanden från andra avrinningsområden i Sverige.

Signifikant korrelation hittades för ungefär hälften av de 15 hydrologiska modellparametrarna, men valideringen mot den riktiga skogsavverkningen kunde bara delvis bekräfta de erhållna sambanden. Resultaten visar att detta sätt att använda parameterregionalisering antagligen är för grundläggande. Vissa resultat är ändå lovande och fortsatt forskning och utvidgning av metoden är nödvändig för att kunna tillhandahålla en rimlig metod för att kvantifiera en skogsavverknings effekter på vattenbalansen.

Nyckelord: Hydrologisk modellering, Skogsavverkning, HBV modellen, Parameterregionalisering, konsekvensanalys.

Institutionen för geovetenskaper, Luft-, vatten-, och landskapslära, Uppsala universitet
Villavägen 16, SE-752 36, UPPSALA, ISSN 1401-5765

PREFACE

This project of 30 credits is the finishing part of the Master of Science program in Water and Environmental Engineering at Uppsala University and the Swedish University of Agricultural Sciences. The project was initiated and supervised by Claudia Teutschbein, post-doctoral researcher. Subject reviewer was Thomas Grabs, senior lecturer. Examiner was Allan Rodhe, senior professor. All three mentioned above are active at the Department of Earth Sciences, Program for Air, Water and Landscape Sciences at Uppsala University.

I would like to express my gratitude to everyone who has supported me during this work. First, I want to say thank you to my supervisor Claudia for all the helpful comments, guidance and interesting discussions we have had during this work. It has been tremendously helpful. I would also like to thank Sofia Hedberg for your helpful thoughts and discussions and for the successful cooperation we have had.

Uppsala, September 2015
Benjamin Selling

 UPTEC W 15043, ISSN 1401-5765 Published digitally at the Department of Earth Sciences, Uppsala University, Uppsala, 2015.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Modellering av hydrologisk påverkan från skogsavverkning genom parameter-regionalisering

Benjamin Selling

Skogen har stor inverkan på vattenbalansen och innehar en nyckelroll i att kontrollera om vattnet från nederbörden ska avdunsta eller hamna i våra åar och älvar, vilket främst sker via grundvattnet men också till en viss del genom strömning på markytan. Detta samband mellan skog och vattenflöden har länge varit ett stort intresse bland hydrologer. Under slutet på 80-talet skapade en tvåårsperiod av kraftiga flöden i Sverige en debatt om att det kunde vara skogsavverkningar som bidrog till de förödande översvämningarna. Flera forskningsprojekt startades då för att studera detta ämne och idag är det allmänt känt att skogsavverkningar påverkar både mängden avrinning och tidpunkten för vårfloden. Men att bedöma storleken på denna påverkan är fortfarande svårt och detta har varit huvudsyftet med detta examensarbete. Att kvantifiera skogens påverkan på vattenbalansen är viktigt för att veta hur stor del av eventuella förändringar i vattenbalansen som beror på skogsavverkningar eller andra källor som till exempel klimatförändringen. Denna kunskap skulle även kunna användas för att undersöka påverkan på vattenresurshantering, ekosystemtjänster och den kan även användas för att utföra livscykelanalyser.

När en skogsavverkning sker påverkar det främst hur mycket vatten som kan avdunsta och transpireras i området samt att ytans råhet och infiltrationskapacitet påverkas då träden tas bort. Vad som också påverkas är markens albedo, som beskriver markytans benägenhet att reflektera solens strålar. En ändring av albedo ger en skillnad i hur mycket snö som lägger sig på marken och när den smälter, vilket även påverkar vattenbalansen. Det traditionella sättet att undersöka skogens påverkan på vattenbalansen är genom att använda två liknande avrinningsområden. Det ena avverkas samtidigt som det andra lämnas orört varpå skogsavverkningens påverkan kan studeras. I detta examensarbete har ett annat tillvägagångssätt som använder modellering för att kunna studera påverkan i större avrinningsområden använts.

Syftet med examensarbetet har varit att testa om modellering med parameter regionalisering kan vara en möjlig metod att kvantifiera påverkan från skogsavverkningar. Metoden innebär att den hydrologiska modellen HBV används för att simulera vattenflöden i avrinningsområden, vilket är en enkel och beprövad hydrologisk modell från SMHI. Modellen består av 15 modellparametrar som oftast bestäms genom att modellen kalibreras mot uppmätt data. Denna kalibrering gjordes för 218 olika avrinningsområden i Sverige vilket resulterade i ett stort antal parametervärden för varje parameter och avrinningsområde. Sedan gjordes en jämförelse mellan dessa parametervärden och procentandelen skogstäcke i avrinningsområdet. Om det hittas ett samband mellan dessa parametrar är tanken att det sambandet kan utnyttjas för att justera parametervärden för att sedan kunna simulera en skogsavverkning och på så sätt modellera en skogsavverkning och kvantifiera dess påverkan.

De 218 avrinningsområdena, av olika storlekar samt utspridda i hela landet, gav signifikanta samband för cirka hälften av modellens parametrar medan några parametrar inte visade några samband alls. Sambanden ansågs inte vara tillräckligt starka, så en gruppering av några mindre avrinningsområden gjordes i norra Sverige för att undersöka om sambanden blev starkare med en mindre variation i geografisk utspridning och storlek. Korrelation mellan den mindre gruppen av avrinningsområden och skogstäcket hittades för 6 parametrar och visade på lite starkare korrelation, men signifikant korrelation hittades för färre parametrar än när samma sak gjordes för hela Sverige. Resultaten stämmer bra överens med tidigare studier med samma tillvägagångssätt men även ytterligare korrelationer fanns i detta arbete för parametrar som tidigare inte haft korrelation.

Den mindre gruppen av avrinningsområden i norra Sverige jämfördes sedan med tre avrinningsområden där en verklig skogsavverkning har ägt rum och där flödesmätningar har gjorts. Dessa avrinningsområden kunde då fungera som en validering för om metoden kan tänkas vara användbar för syftet. Valideringen mot de avverkade avrinningsområdena visade liknande korrelation för 3 av dessa modellparametrar. Det visar att vissa samband finns men för att kunna använda denna metod måste den utvecklas vidare från denna alltför enkla metod. Det verkar också vara så att modellens parametrar har interna samband i modellens struktur. Detta innebär att det inte är så enkelt som att bara ändra på någon parameter, utan en parameter kan hänga ihop med någon annan och bli kompenserad av en tredje, vilket gör att det är inte rimligt att hantera parametervärden var för sig. Sambandet mellan parametrar bör studeras och parametrarna bör hanteras i grupp i stället för enskilt.

Sammanfattningsvis hittades tydliga samband mellan modellens parametrar och skogstäcket i avrinningsområdena men ett mer avancerat tillvägagångssätt behövs för att kunna göra rimliga kvantifieringar av skogsavverkningens påverkan.

LIST OF ABBREVIATIONS

CC	Clear cut
CLI	Command line interface
GAP	Genetic Algorithm and Powell optimization
GUI	Graphical User Interface
HBV	A rainfall-runoff model, Hydrologiska Byråns Vattenbalansavdelning.
PET	Potential evapotranspiration
$\mathbf{R_{eff}}$	The Nash Sutcliff model efficiency
$\mathbf{r_s}$	Spearman's correlation coefficient
SMHI	Swedish Meteorological and Hydrological Institute
SVAR	Svenskt Vatten ARkiv, Swedish Water Archive

Table of contents

1 INTRODUCTION

1.1 Background

Humans continuously interact with nature and influence the environment since the beginning of human history. But since the industrial revolution these interactions have increased dramatically and are now also showing great environmental consequences (Liu et al., 2007). Today, the human society has a big influence on land and water resources: 60% of the global freshwater runoff is used and 83% of the global land area is influenced by human activities (Sanderson et al., 2002). We are now experiencing major global environmental changes such as shift in land use and land cover patterns (Vitousek, 1994) that are mainly driven by the exponential population growth and the growing rate of resource consumption.

The major factors of land use changes that have an effect on hydrology include deforestation and afforestation, drainage of wetlands, urbanization and intensified agriculture (Calder, 1993). The changing land cover affects the hydrology in many different ways. One obvious factor is the evapotranspiration process, but also surface roughness, infiltration capacity and albedo are highly affected and influence the water balance (De Roo et al., 2001).

In Sweden 57% of the land is used for productive forest (Christiansen et al., 2014). The forest plays an important role in the hydrologic cycle, and directly affects runoff and the ecosystem services related to streams and lakes. During 2014 the forestry industry contributed with 11% of the total Swedish export and 2.2% of Domestic Gross Product showing the important economic influence of this industry. Trying to quantify the impact from a changed land use such as a forest clearcutting has therefore been an interest for hydrological researchers for many years, but the complex processes involved are still not fully understood (Ellison et al., 2012). In Sweden, a 2-year period with extreme autumn flood events started a discussion on the hydrological effects of forest clearcutting on a national level in the late 1980's (Brandt et al., 1988). Understanding the complex nature of forest-water relationships is important for managing water and forest resources, ecosystem services and to be able to perform reasonable impact assessments, such as calculating the water footprint for forests and forest based products (Zhang et al., 2012).

Generally, forest vegetation plays a key role in controlling the water balance and forest harvesting is therefore likely to influence the timing and amount of runoff (Sørensen et al., 2009). The traditional method of studying of the impact of forest clearcutting on streamflow has been to use paired catchments: two small and nearby catchments with similar characteristics are chosen and then one of the catchments is left undisturbed as a reference catchment while the other one is harvested. By measuring streamflow or other parameters such as nutrients or chemicals at the catchments outlets, the impact of the forest clearcutting can be investigated. This is a common approach and has been done in many studies (e.g., Jones & Post 2004; Stednick 1996; Schelker et al. 2013). These studies present some conclusions that are nowadays well known effects from forest clearcutting:

- Increased total runoff and more pronounced peak flows.
- Strong local influence on flood risk, especially in small basins and when the percentage of forest harvesting is high.
- The effect from a 10-15 percentage of forest harvested cannot be compared to the effect from extreme weather conditions.
- Increased snowpack influences the timing and amount of spring flood, which can be persisted up to 35 years after clearcutting.

The major drawback of paired catchment studies is that they can only be done for small catchments and that the conclusions are only valid for a limited region. The method can neither be used for modelling the potential impacts in other catchments, which is desirable for impact assessments (Seibert and McDonnell, 2010).

Therefore more recent studies have tried different modelling approaches. For example, Seibert and McDonnell (2010) use a model-based change-detection approach, Zegre (2011) uses a simple rainfall-runoff modelling approach and Brath et al. (2006) try a spatially distributed rainfall-runoff model to generate different synthetic river flow series for different land use scenarios. Another interesting approach to quantify the impact from land use change is parameter regionalization. The idea with parameter regionalization is to relate calibrated and optimised model parameter sets to physical catchment characteristics. This method has mostly been used in studies aiming at modelling ungauged catchments, because it can be assumed that catchments with similar characteristics show similarities in their hydrological behaviour and can, thus, be simulated based on similar model parameter sets (Bárdossy, 2007). Only few studies have tried to quantify the effect of land use on streamflow through parameter regionalization. For instance, Wooldridge et al. (2001) combined parameter regionalization with a semi-distributed model to simulate four Australian catchments, while Hundecha and Bárdossy (2004) used a conceptual rainfall-runoff model to regionalize parameters for 30 catchments in Germany. A parameter regionalization approach to quantify forest clearcutting impacts on streamflow has not yet been done in Sweden, although the huge amount of available runoff data in Sweden provided by the Swedish Meteorological and Hydrological Institute (SMHI) is promising for this approach.

This study is largely based on findings by Seibert (1999), who did a parameter regionalization for 11 catchments in southern Sweden and provided promising results as he was able to find significant correlations between forest cover percentage and 4 parameters of a relatively simple lumped rainfall-runoff model. One general issue with regionalization that has to be considered, is that many catchments need to be included for a good statistical framework but with larger regions an uncertainty is added, because of differences in climate and other conditions (Seibert, 1999).

1.2 Hypothesis and Objectives

The main hypothesis of this thesis is that a parameter regionalization of a conceptual lumped rainfall-runoff model for a large number of catchments is a suitable method for quantifying the hydrological effects of deforestation in Sweden.

The main objective of this thesis is to perform a parameter regionalization of a conceptual lumped runoff model (HBV model) for a large number of catchments. This implies finding optimal parameter sets through model calibration for each catchment and correlating these parameters to forest cover percentages of the catchments. Significant correlations would support the idea of tweaking the model parameters to simulate a change in the forest cover and to quantify the hydrological impacts of forest clearcutting.

The second objective is to validate the parameter regionalization approach with help of a ‘real-world’ paired catchment experiment to see if it can be used to model observed effects of forest harvesting.

2 MATERIALS AND METHODS

2.1 Area of study

2.1.1 Swedish catchments

In the first major analysis, 324 catchments all over Sweden were used (Figure 1). Catchment sizes varied considerably from just a few km^2 up to almost 50 000 km^2, but the major part being about 100-300 km^2. Since the very big catchments are not really feasible to model (human impact becomes too large), a limit was set at 16 000 km^2, which roughly equals the mean catchment size plus two standard deviations. 15 catchments that were bigger than the specified limit were therefore excluded in the further analysis.

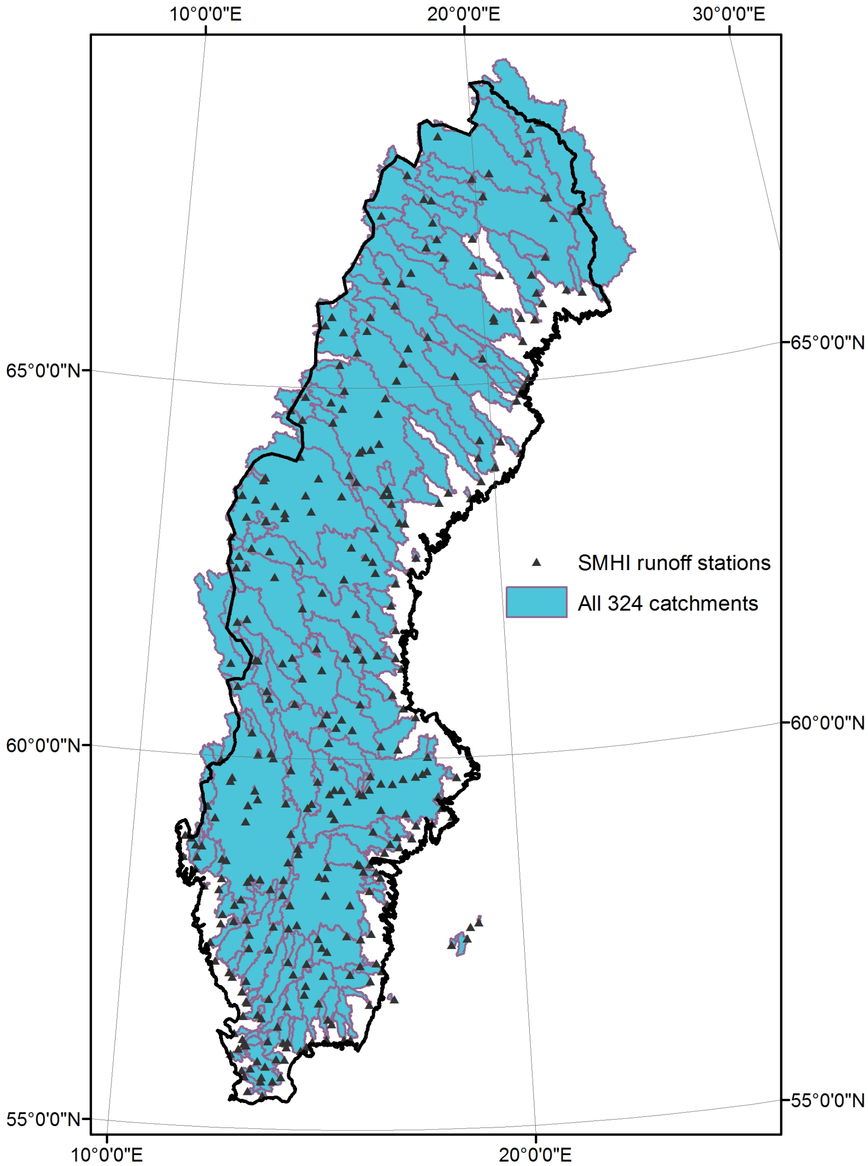

Figure 1 Boundaries of all 324 catchments and the corresponding locations of runoff stations (source: SMHI's SVAR database (Henestål et al., 2015)

2.1.2 Balsjö catchments

In northern Sweden about 70 km west of the city Umeå there are three catchments (Figure 2, Table 1) that were part of a paired catchment project called "277 Balsjö experiment" (Schelker et al., 2013). Earlier studies have focused on the effect of forest clearcutting on water chemistry and the Methyl-Mercury in surface water (Eklöf et al., 2014) but also the hydrological effects have been studied for these catchments (Schelker et al., 2013). As part of this paired catchment experiment, streamflow measurements were taken at three stations which makes further hydrological water balance analyses feasible.

The idea of the Balsjö experiment was to perform different forest treatments in the three small catchments in May 2006 to study the effect of forest clearcutting. The northernmost catchment ('RefS', Figure 2) was left almost untreated (Table 1) and was meant to be a reference catchment. 'RefS' is a subcatchment of catchment 'North' (Figure 2), of which only the part downstream of 'RefS' was harvested. In the southermost catchment ('CC', Figure 2), the major part of the forest was harvested (Table 1). The vegetation in the area is typical for this boreal region with forest of Scots pine dominating the drier soils and Norway spruce in lower regions. The catchments also have about 2-10% wetland soils. The catchment areas are about 0.4 km^2 ha for catchments 'CC' and 'North', while the smaller untreated 'RefS' is about 0.24 km^2 (Schelker et al., 2013).

Table 1 Percentage of forest cover in the three Balsjö catchments before and after the forest harvesting event in 2006 (Source: SLU Forest Map, Dept. of Forest Resource Management, Swedish University of Agricultural Sciences).

Percentage of forest cover in catchment	CC	North	RefS
Before CC 2006	96	81	73
After CC 2006	34	50	74

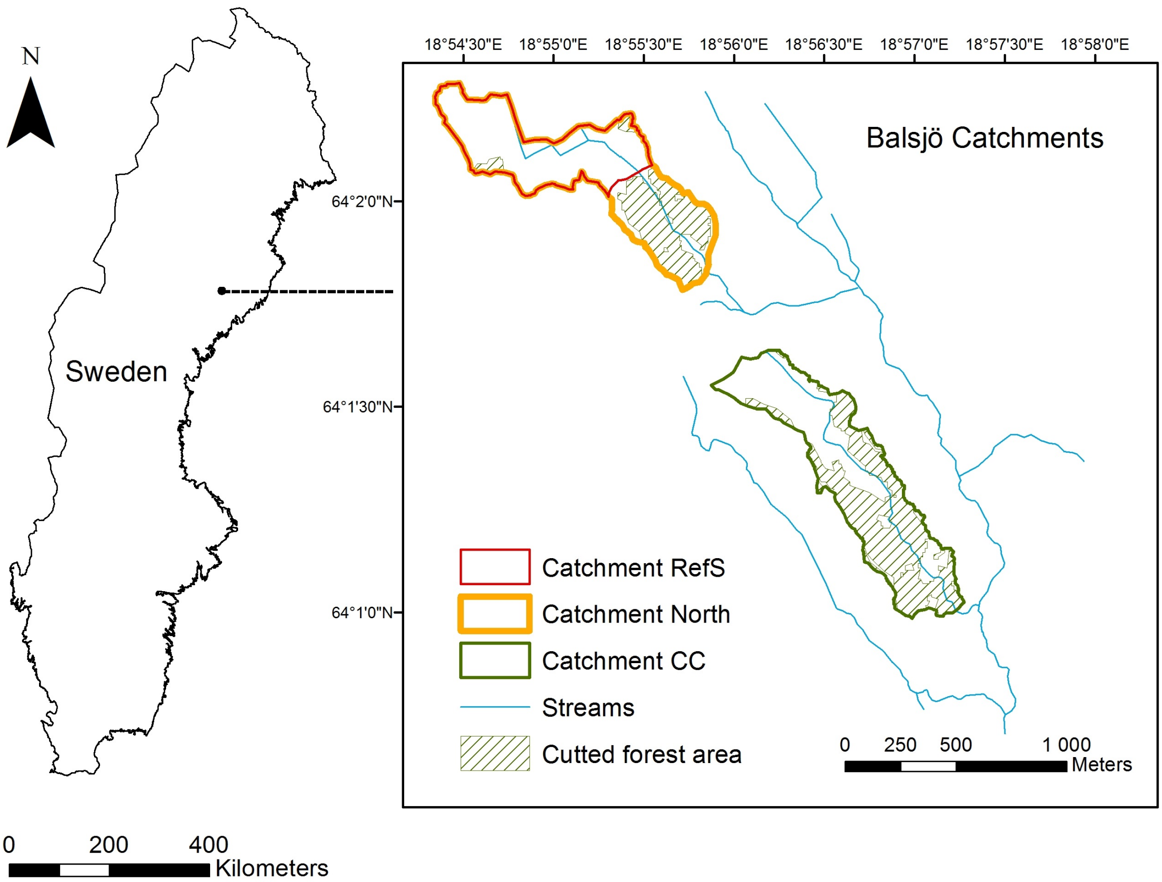

Figure 2 A map of the location and extent of the Balsjö catchments as well as the area of forest that was harvested in 2006.

Observed values of discharge were available for just one year before the clear cut. This is not optimal and a few years of data is normally preferred but the analysis was done anyway. To quantify the impact from this uncertainty, multiple calibrations were done with different length of the dataset, for the period after the clear cut. The different calibrations were done with data for one, three and all (5.5) years after the clear cut. This can demonstrate how the length of data set affects the resulting parameter set.

2.2 Hydrological modelling

2.2.1 HBV model description

The hydrological model used in this study is the HBV model (Bergström, 1976), which was developed at SMHI in the beginning of the 1970's and has continuously been improved since (Bergström, 1992; Lindstrom et al., 1997). It has been used in many countries as a standard tool for runoff simulations, flood forecasting and other research projects (Booij, 2005; Jia and Sun, 2012; Primožič et al., 2008). The HBV model is a relatively simple conceptual rainfall-runoff model and the HBV model software used in this work is "HBV-light 4.0.0.9" (Seibert and Vis, 2012).

Historically, the HBV model has often been found to perform well, especially for Swedish conditions (e.g., Harlin, 1992; Seibert and Vis, 2012) and does not need as much input data as more complex models. Also computational effort is not very high and it can easily be run and calibrated for many catchments. The parameters in the lumped version of the HBV model are not physically based so there is a need for a calibration algorithm. However, the parameters are partly based on physical concepts, which means relationships between model parameters and the land cover in the catchment can be expected (Hundecha and Bárdossy, 2004).

HBV is normally used to simulate daily streamflow and the needed input variables are daily temperature, daily precipitation and monthly mean values of potential evapotranspiration. The model consists of four main routines: a snow routine, a soil routine, a groundwater routine and a routing routine (Figure 3), which are controlled by 15 model parameters (Table 2). It is furthermore based on a simple water balance (Equation 1) that is trying to include all the needed hydrological processes (SMHI, 2014).

$$P - E - Q = \frac{d}{dt}[SP + SM + UZ + LZ] \qquad (1)$$

Where:
P = Precipitation
E = Evapotranspiration
Q = Discharge
SP = Snow pack
SM = Soil moisture
UZ = Upper groundwater zone
LZ = Lower groundwater zone

The snow routine describes the snowfall and snowmelt with a degree-day method (CFMAX), if there is any snow, but the model is also applicable for regions without snowfall. Whether precipitation is falling as snow or rain is defined by a parameter describing a threshold temperature (TT). In the soil routine the groundwater recharge and the evaporation are simulated as a function of actual soil water storage. This is computed by three parameters: FC that describes the maximum soil moisture, BETA that determines the relative contribution to runoff and LP that describes the shape of the reduction curve for potential evapotranspiration saying that for soil moisture values below LP the actual evapotranspiration will start to decrease. The groundwater routine transforms the water from the soil routine to discharge or to groundwater storage in the groundwater boxes. The routine is described by two groundwater reservoirs and the flows are described by the parameters for recession coefficients (K_0, K_1, K_2), a threshold value (UZL) and a constant percolation value (PERC). As a last part of the model, the routing routine describes the generated runoff and includes a filter that uses a triangular weighting function that has one parameter called MAXBAS (Bergström, 1992).

For more information about the HBV model and a more explicit description, the reader is referred to Bergström (1995) and Seibert (1999).

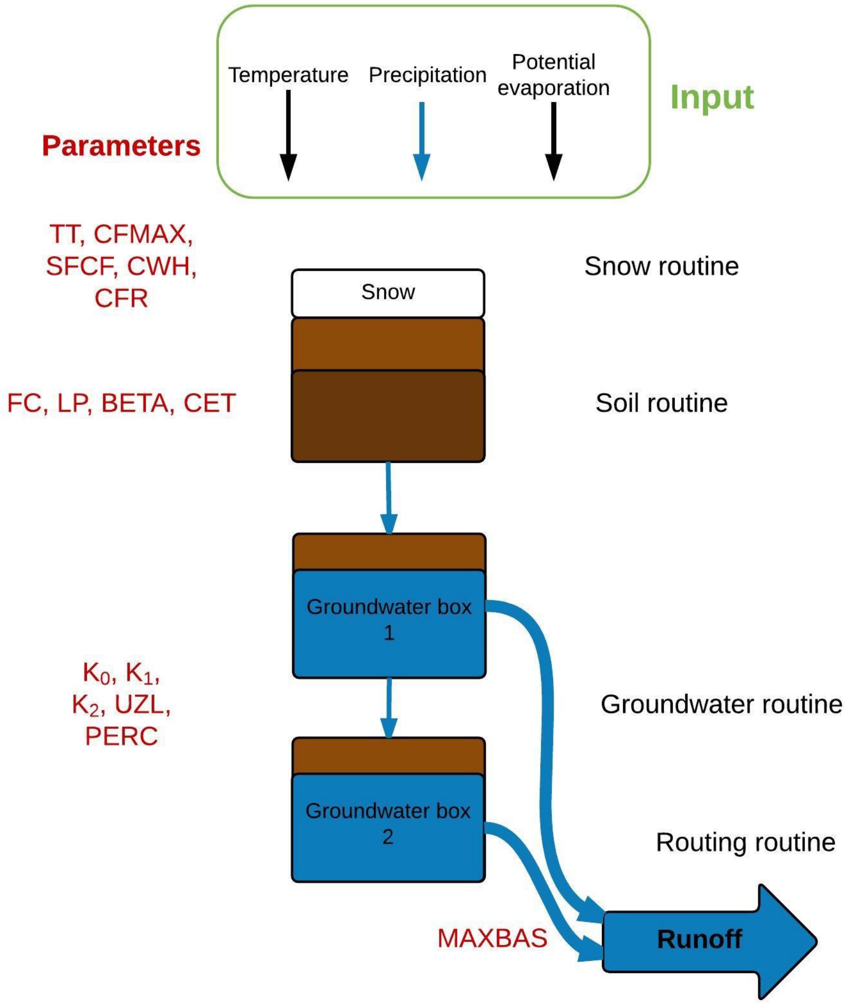

Figure 3 Schematic structure of the HBV model, its routines and parameters.

Table 2 HBV model parameters and used parameter ranges (for the cluster group) in the calibration procedure (modified from Seibert (1999)).

Parameter	Explanation	Unit	Lower bound	Upper bound
Snow routine				
TT	Threshold temperature	°C	-2.6	2
CFMAX	Degree-day factor	mm °C^{-1} d^{-1}	0	7
SFCF	Snowfall correction factor	-	0.4	1.5
CWH	Water holding capacity	-	0	0.2
CFR	Refreezing coefficient	-	0	0.25
Soil routine				
FC	Maximum of SM (storage in soil box)	mm	45	550
LP	Threshold for reduction of evaporation (SM/FC)	-	0.3	1
BETA	Shape coefficient	-	0.2	7
CET	Correction factor for potenial evaporation	°C^{-1}		
Response routine				
K_0	Recession coefficent	d^{-1}	0.01	0.9
K_1	Recession coefficent	d^{-1}	0.005	0.6
K_2	Recession coefficent	d^{-1}	5.00E-07	0.15
UZL	Treshold parameter	mm	0	140
PERC	Maximal flow from upper to lower box	mm d^{-1}	0	12
MAXBAS	Routing, length of weighting function	d	1	13

2.2.2 Calibration

To make the automatic calibration of the HBV model in an efficient way a built-in genetic algorithm was used (Seibert, 2000). This is a faster and more computationally efficient calibration method compared to the Monte Carlo procedure that is also available as a built-in function in the HBV-light software. The genetic algorithm is a method of optimization that is based on the natural occurring biological evolution. That means that the parameter set is treated as a set of genes that then are evolved by melding, mutations and off-spring processes through the calibration process (Beven, 2009).

The algorithm initially starts with a set of *n* (set to 50 in this study) random sets of parameter values, whose performance is evaluated based on the value of the objective function. The next generation of parameters is thereafter created by *n* times fusing two 'parent' parameter sets randomly, but with good performing sets having a higher probability of being chosen. The new generation of parameter sets is then again evaluated and if it gives a better fitness from the objective function that set replaces the old one. These evolutionary steps are then repeated many times (set to 5000) until an optimal set is obtained. The total procedure was repeated several times (25 for all catchments and 100 for the Cluster groups) giving multiple parameter sets for every catchment (Seibert, 2000). In HBV-light there is also a last step in the genetic calibration called Powell optimization, where the parameters are fine-tuned using Powell's quadratically convergent method (Press, 2002).

The calibration includes a set of ranges of possible values for the parameters (Table 2). The initial ranges used in this study were based on values provided by Seibert (2000). If many catchments had parameters that hit the limits of the set range the original range was widened and a new calibration was done.

2.2.3 Weighted objective function

During calibration the parameter set has to be evaluated for its performance. For this, an objective function is needed to estimate how good the model can simulate the streamflow.

Typically, one measuring function is not able to evaluate all aspects and to present the dynamics of the runoff (Solomatine and Wagener, 2011). For example one measure could catch all the dynamics of the peak flows well, but could not be able to evaluate the total volume error. Because of this problem, multiple weighted objective functions were used in this study to evaluate the parameter sets as proposed by Seibert (1997): The Nash Sutcliff efficiency (R_{eff}) (Equation 2) is a standard measure for model performance, but it puts an emphasis on high flows. Therefore, R_{eff} was given only a weight of 0.8 and $\log R_{eff}$ (Equation 3), which is a better measure for low flows, and Volume Error (Equation 4) were also included with a weight of 0.1 for each of them.

Nash-Sutcliffe (R_{eff})

$$R_{eff} = 1 - \frac{\sum(Q_{obs} - Q_{sim})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2} \tag{2}$$

$\mathrm{Log}R_{eff}$

$$logR_{eff} = 1 - \frac{\sum(lnQ_{obs} - lnQ_{sim})^2}{\sum(lnQ_{obs} - \overline{lnQ_{sim}})^2} \tag{3}$$

Volume Error

$$Volume\ error = 1 - \frac{\sum|(Q_{obs} - Q_{sim})|}{\sum(Q_{obs})} \tag{4}$$

Joining these three weighted objective functions results in a fuzzy measure, which is defined as one for a perfect fitted model and zero for a poor model. In this study, a value above 0.7 was defined as 'good', while values below 0.7 were not satisfying. All the catchments that resulted in fuzzy measure below 0.7 were excluded from the further work.

Definition of model efficiency based on the fuzzy measure:

> 0.85 Excellent
0.7-0.85 Good
< 0.7 Not satisfying

2.2.4 Automatization

In this project, codes were developed to run the calibration automatically for many catchments. This was done with a Matlab script that constructed a database of folders for all catchments with all the input files needed to run the calibration. All data were prepared or gathered from other databases and compiled to the correct input file format. The regular way of running the HBV-light software is to run the Graphic User Interface (GUI). But in this project 324 catchments were calibrated, and therefore the Command-line interface (CLI) function was used instead. The CLI has the same functionality as the GUI, but it is running the software through the Microsoft Windows CLI (DOS) instead. By running a batch file in the Windows command line window, the calibration could easily be run many times, and it was also simple to run the calibration on multiple computers, which is desirable since the calibration could be very time consuming on a single computer. A short summary and an example of the code for this approach are shown in the appendix A.

2.3 Data

The needed input data for the HBV model are daily values of temperature, precipitation and streamflow but long-term monthly means of temperature and potential evaporation are also mandatory. A summary of all the used data are presented in Table 3 and for a more detailed information see the following sections.

Table 3 Summary of used data.

	units, resolution	**Avilable** time period	**Calibration in all Sweden** time period	**Calibration in Balsjö** time period
Precipitation	mm, daily	~1961 - 2010	1991-1998	2004-2010
Temperature	°C, daily	~1961 - 2010	1991-1998	2004-2010
Discharge	mm, daily	~1961 - 2010	1991-1998	2004-2010
Potential ET	mm, monthly	1961-1978	1961-1978	1961-1978
Landuse	6 classes, 100 m	2000, 2006	2000	2006

2.3.1 Precipitation and temperature

The input data for temperature and precipitation was retrieved from SMHI's service called PTHBV which is a climate database commonly used in research studies. It provides daily values of temperature and precipitation since 1961 and presently covers all of Sweden with a grid of 4x4 km. The gridded values were interpolated based on measurements from all of SMHI's Swedish weather stations and also some stations in Norway. Measured weather data of precipitation were corrected for measuring losses due to wind turbulence around the station that can give losses because snow and rain blowing past the rain gauge (Johansson, 2002).

2.3.2 Discharge

Measured discharge data were obtained for 324 stations from SMHI's service 'Vattenwebb' (English: 'water web'), which provides information on several hydrological variables for download. About 220 of the available stations are managed by SMHI and the rest are owned and run by the hydropower companies. The majority of the stations are measuring the level and then determine the discharge with a rating curve (SMHI, 2015). For each streamflow station, the corresponding catchments boundary was downloaded as shape file from SMHI's water archive (SVAR, Svenskt VattenARkiv) (Henestål et al., 2015) and was imported into ArcGIS – a geographical information system (GIS) software for further work.

2.3.3 Land use and catchment characteristics

To be able to relate calibrated parameters to forest cover in each catchment, data were needed for land use and forest cover in particular. A land use database provided by the European Environmental Agency (EEA), called Corine Land Cover (CLC) was selected. It is derived from satellite data that were collected during the years of 1990, 2000 and 2006 plus/minus one year and the data are available with resolutions of both 100 and 250 meter. This data is freely accessible for everyone and the original data consist of 44 different land cover classes (EEA, 2007).

In this study the data from year 2000 (closest to the used time interval) with the higher resolution of 100 meter is used. The 44 land cover classes were aggregated to 6 different main classes: agricultural land, urban areas, forests, wetlands, water bodies and open areas. The forest class contains three original classes called broad-leaved forest, coniferous forest and mixed forest (Appendix B). The set of used catchments included a wide range of forest cover percentage and covered the full spectra from 0 to 100 percent (figure 4), which also is desired when the main focus is on the forest cover impact.

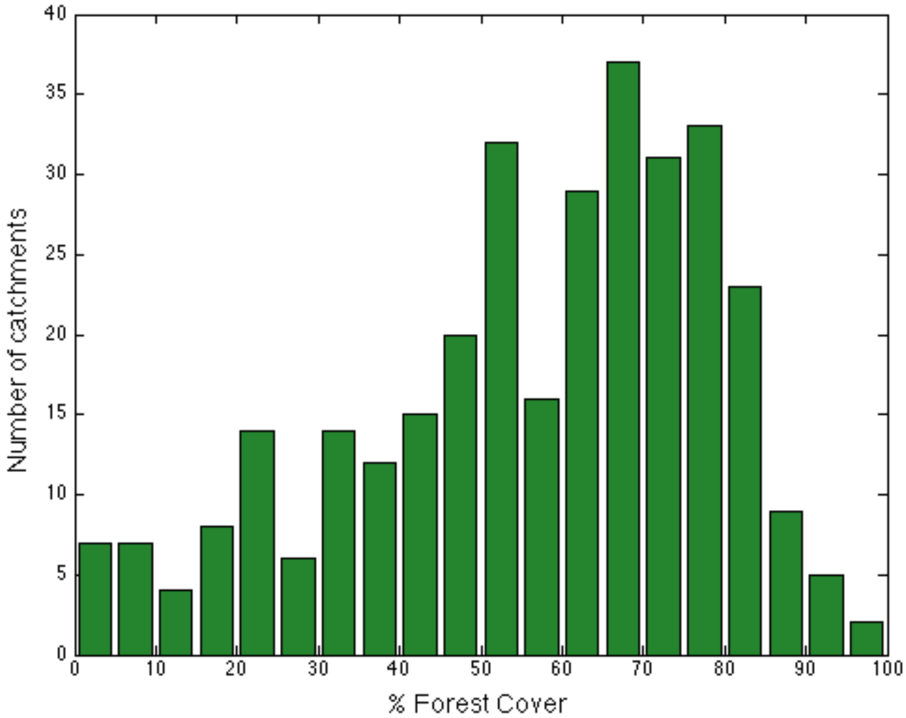

Figure 4 Histogram showing the distribution of forest cover percentage for all included catchments.

For the three small Balsjö catchments, the Corine data set has too low resolution to give good approximations of land cover. Therefore, data provided by Schelker et al. (2013) were used instead.

To be able to group the catchment according to their similarities, some catchment characteristic data were needed. For each catchment, data for latitude, longitude, catchment area (Figure 5) and mean elevation (Figure 6) could be downloaded from the SVAR database (Henestål et al., 2015). The relation between forest cover and other catchment characteristics such as mean elevation, lake percentage and catchment area was investigated by calculating the Spearman’s correlation between them.

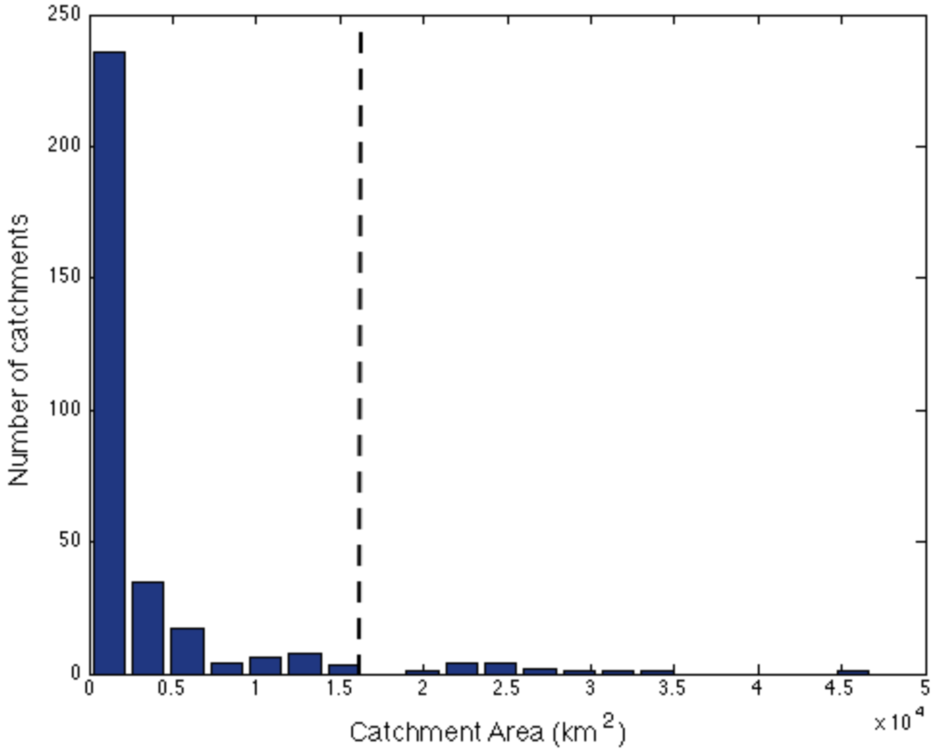

Figure 5 Histogram displaying the frequency of catchment sizes. The dashed line shows the limit of excluded catchments.

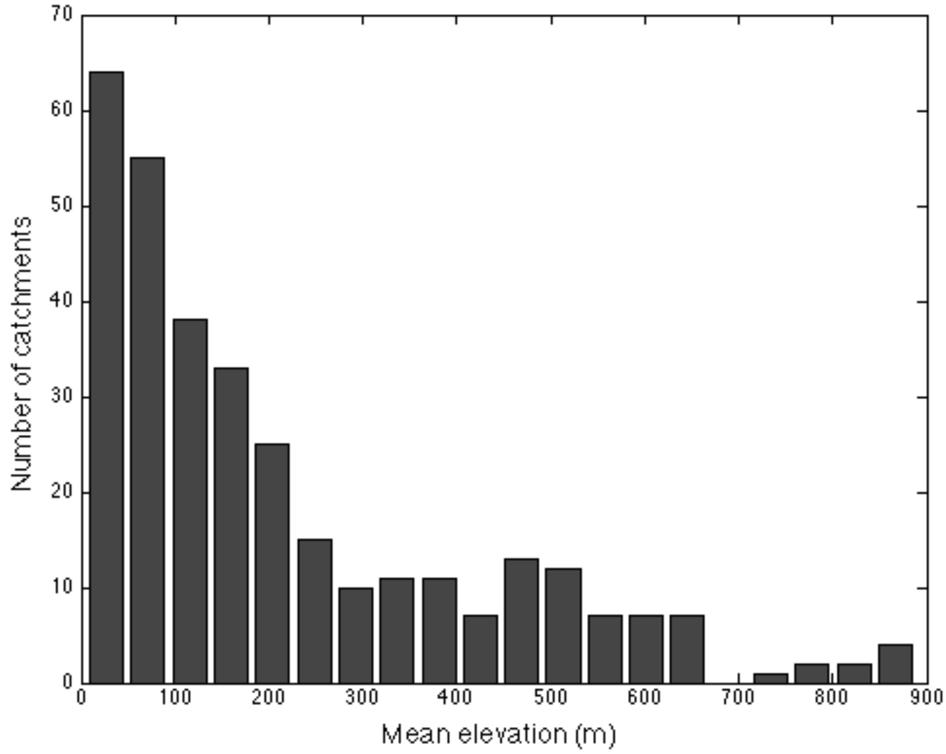

Figure 6 Histogram showing the distribution of mean elevation for the included catchments.

2.3.4 Potential evaporation

For the needed input of long-term monthly mean values of potential evaporation, data from Eriksson (1981) were used. Eriksson (1981) calculated potential evaporation using the Penman equation (Penman, 1948) based on meteorological measurements from SMHI at 152 stations across all Sweden during the period 1961-78. The equation is based on measurements of global radiation, albedo, air temperature, cloud amount and wind speed. The data are presented as monthly mean of potential evapotranspiration for all stations (Eriksson, 1981).

After removing stations with uncertain values due to missing data, potential evaporation data were available for 144 stations. The data from Eriksson (1981) were digitalized and georeferenced in ArcGIS and a Kriging interpolation was performed for all stations, which resulted in 12 maps of long-term mean potential evapotranspiration for every month of the year and one map with annual mean values (Figure 7). From these maps, interpolated values for each studied catchment were used as input data for the HBV model.

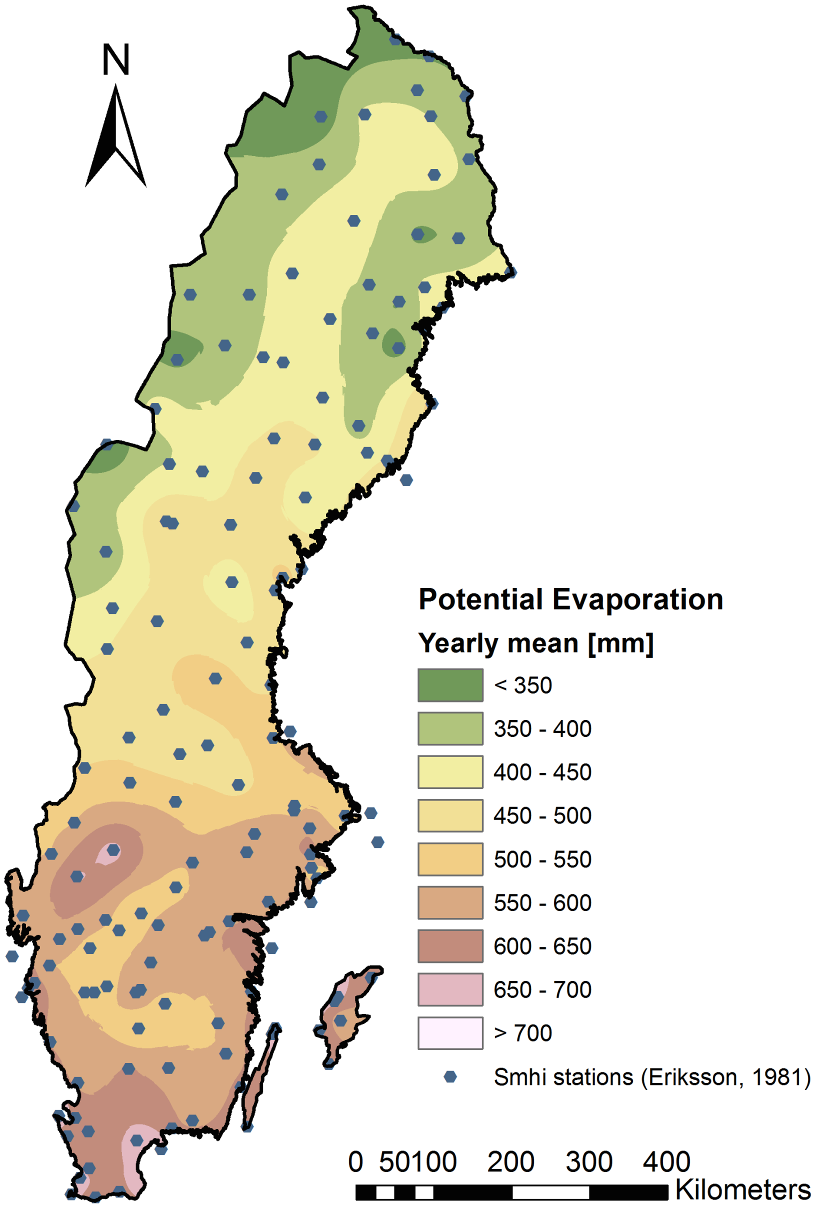

Figure 7 Yearly mean of potential evaporation interpolated from SMHI stations (Eriksson, 1981).

Even though the original data provided by Eriksson (1981) is from a different period than the remaining data, HBV only requires long-term mean estimates of the potential evaporation that are adjusted according to measured temperature values within the model. Since these long-term values are likely to not have changed significantly, it is, thus, still a standard practice to use them for the HBV model in Sweden (Teutschbein and Seibert, 2012; Wrede et al., 2013).

2.4 Spearman's rank correlation

To examine correlations between calibrated HBV model parameter values and forest cover they were plotted against each other and the Spearman rank correlation was calculated. It was chosen because it is a robust measure and no shape or functional relationship had to be assumed (Seibert, 1999).

Spearman's rank correlation only evaluates the correlation for monotonic relationships between two variables and other statistics have to be used if there are non-monotonic relationships in the data. The correlation from this test is described with Spearman's coefficient *rho* (r_s) and it can be explained as the linear correlation coefficient calculated from the rank of every parameter. Mathematically it can be described as an equation whit *n* observations and parameters x and y and the rank for x and y is Rx and Ry (equation 5) (Helsel and Hirsch, 2002).

$$rho = \frac{\sum_{i=1}^{n}(Rx_i Ry_i) - n(\frac{n+1}{2})^2}{n(n^2-1)/12} \qquad (5)$$

2.5 Cluster analysis

As a part of the method for this project there was a need to find a smaller group of catchments 'similar' to the Balsjö catchments. Therefore, a cluster analysis (Cattell, 1943) was performed, which made it possible to identify non-overlapping groups (so-called clusters in the large multivariate data set of 324 catchments. This was done in such a way that catchments within the same cluster are as similar as possible, while catchments in different groups are as dissimilar as possible (Kaufman and Rousseeuw, 2009). Here, K-means clustering was used, which is a popular and simple way to perform a cluster analysis and it has been used for more than 50 years now (Jain, 2010). The K-mean algorithm groups a set of *n d*-dimensional points into a set of K clusters with a method of minimizing the squared error between the empirical mean of the cluster and the points in the data (Jain, 2010).

In this project, 6 cluster groups were assumed to be a suitable amount of clusters to separate the catchments in groups of desired sizes. The catchment characteristics for this analysis were chosen to not be strongly correlated with forest cover percentage, since a wide range of forest cover percentage within each group was still needed for the correlation analysis. Because of this, the land cover and hydrological statistics were not included. Instead the catchments size, latitude, longitude and mean elevation were used.

3 RESULTS

3.1 General performance

The calibration of all 324 catchments resulted in a wide range of model efficiency measures calculated by the weighted objective function. Only 218 out of 324 catchments could be modelled properly with acceptable values of weighted objective function above 0.7, had enough data values and were not too large (< 16 000 km^2). The spatial distribution for the model efficiency (Figure 8) showed no clear overall pattern.

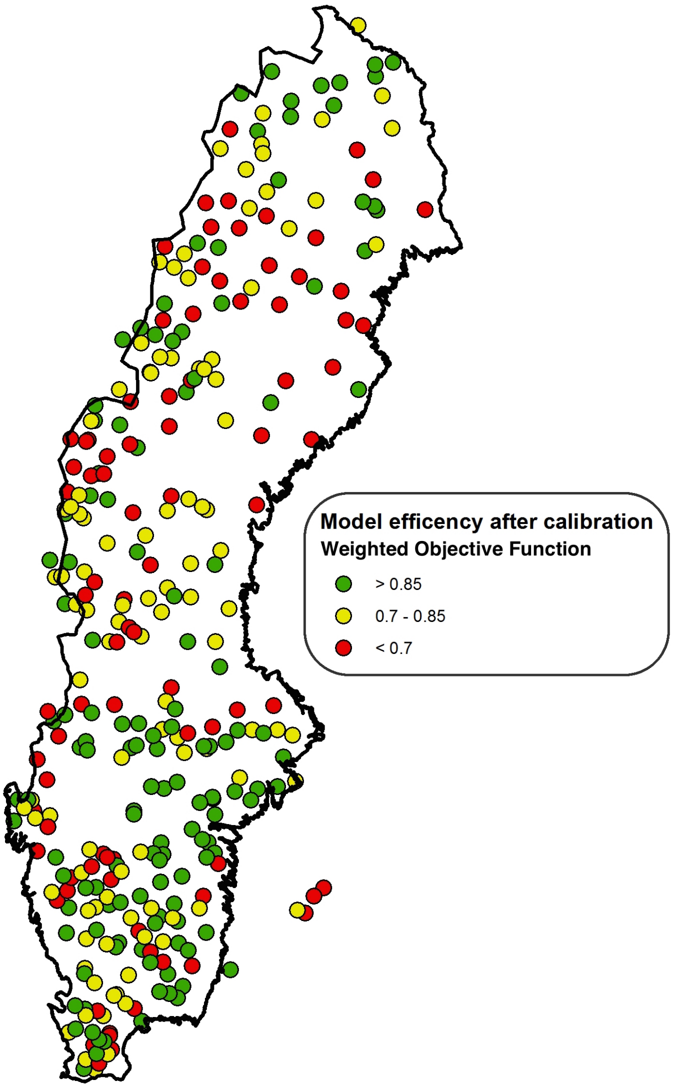

Figure 8 The model efficency (based on the weighted objective function) for all catchments represented by a point at the catchment centroid**.**

The general water balance was checked for all catchments by comparing the yearly discharge with the yearly sum of precipitation. It became clear that there are severe problems with the data in the mountainous region in northwest Sweden (Figure 9). These catchments were excluded from the following analysis.

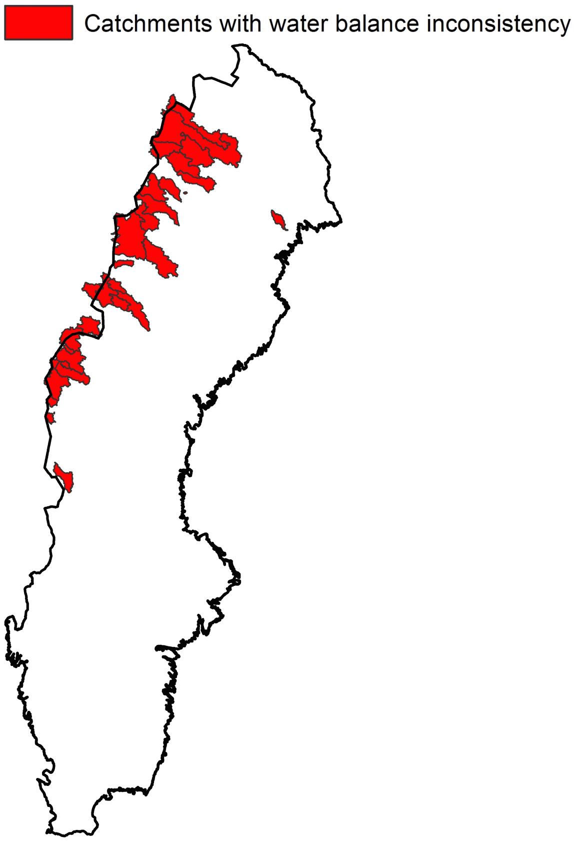

Figure 9 Catchments with an inconsistent water balance. Data were summarized over the hydrological year (Oct-Sept).

3.2 Relationship between model parameters and forest cover

A calibration was conducted for all of the well performing catchments. The calibrated parameter sets were related to the forest cover, which gave significant correlations for 8 of the 15 HBV model parameters (Table 4).

Table 4 Correlation between parameter median from the calibration and the forest cover for all the catchments showing the Spearman's Rho coefficient and the p-value. Bold values are significant at a 5% confidence level.

Parameters	r_s	p-value
Perc	-0.05	0.453
UZL	-0.06	0.377
K_0	**-0.17**	**0.014**
K_1	0.06	0.353
K_2	**0.41**	**0.000**
MAXBAS	**-0.16**	**0.017**
CET	-0.02	0.738
TT	**-0.14**	**0.045**
CFMAX	**-0.48**	**0.000**
SFCF	**-0.52**	**0.000**
CFR	**0.23**	**0.001**
CWH	0.07	0.284
FC	-0.06	0.393
LP	-0.02	0.803
BETA	**0.14**	**0.035**

For further examination, each parameter has been plotted against forest cover. Here, only the example of the snowfall correction factor SFCF (Figure 10) and CET (Figure 11) are shown. The snowfall correction factor SFCF shows a significant negative trend against forest cover, i.e., the SFCF is decreasing with higher forest cover. It can clearly be seen that the boxes for some catchments are very wide and have long whiskers, which shows the uncertainty of the parameter and the calibration.

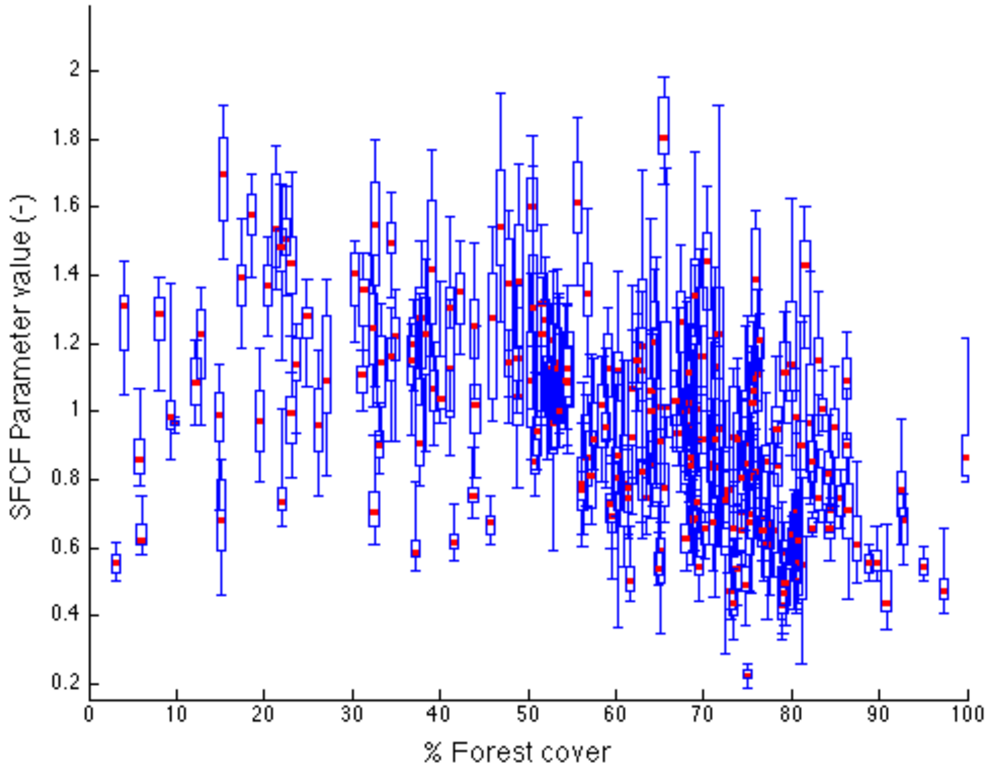

Figure 10 Relationship between the model parameter SFCF and forest cover. Each box represents 25 SFCF values of the 25 best calibrated model parameter sets for all catchments each with a specific forest cover percentage. The red line shows the median, the box marks 25 and 75 percentiles ant the whiskers are 95 percentiles.

For the CET parameter (Figure 11), there is no significant trend and one can see that they have very wide ranges of parameter values and many values are stretching almost from the floor to the roof, which is zero to one. Catchments with high percentages of forest cover seem to have more narrow ranges of good parameter values while the lower percentages have very wide ranges and high uncertainty.

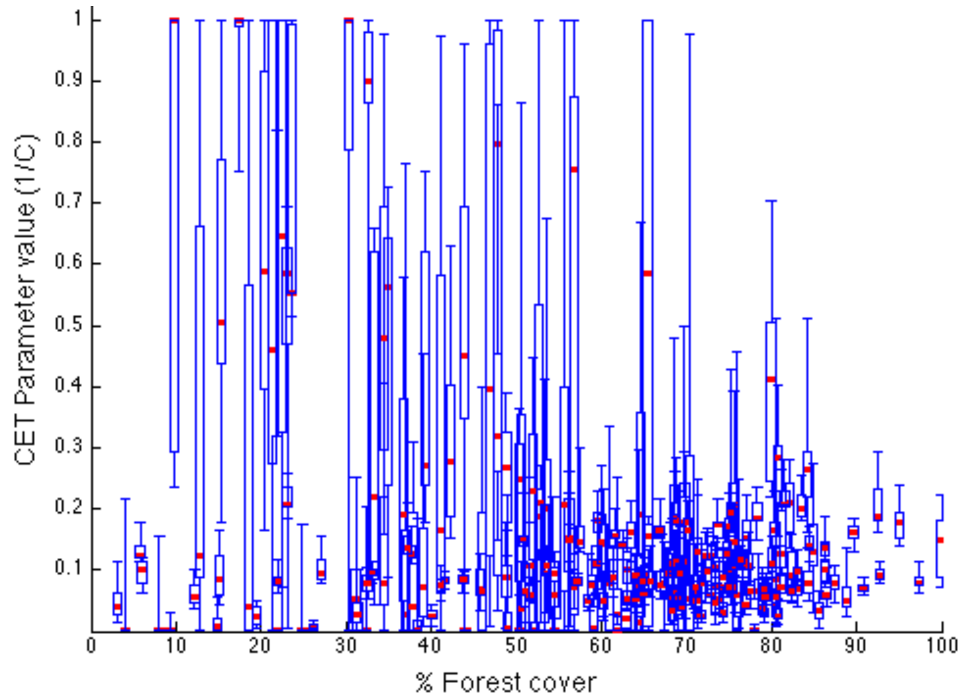

Figure 11 Relationship of the model parameter CET against forest cover. Each box represents 25 CET values of the 25 best calibrated model parameter sets for all catchments each with a specific forest cover percentage. The red line shows the median, the box marks 25 and 75 percentiles ant the whiskers are 95 percentiles.

An easier way to visualize the correlation is to just plot the medians (without boxes) against the forest cover, where every dot represents the median of the good parameter values for one catchment (cf., Figure 12). In the case of SFCF (Figure 12), the trend is clearer even if the values are quite spread out. A clear trend can also be seen for model parameter K_2 (Figure 13), but its relationship to forest cover seems to be nonlinear, and rather like a polynomial of second degree.

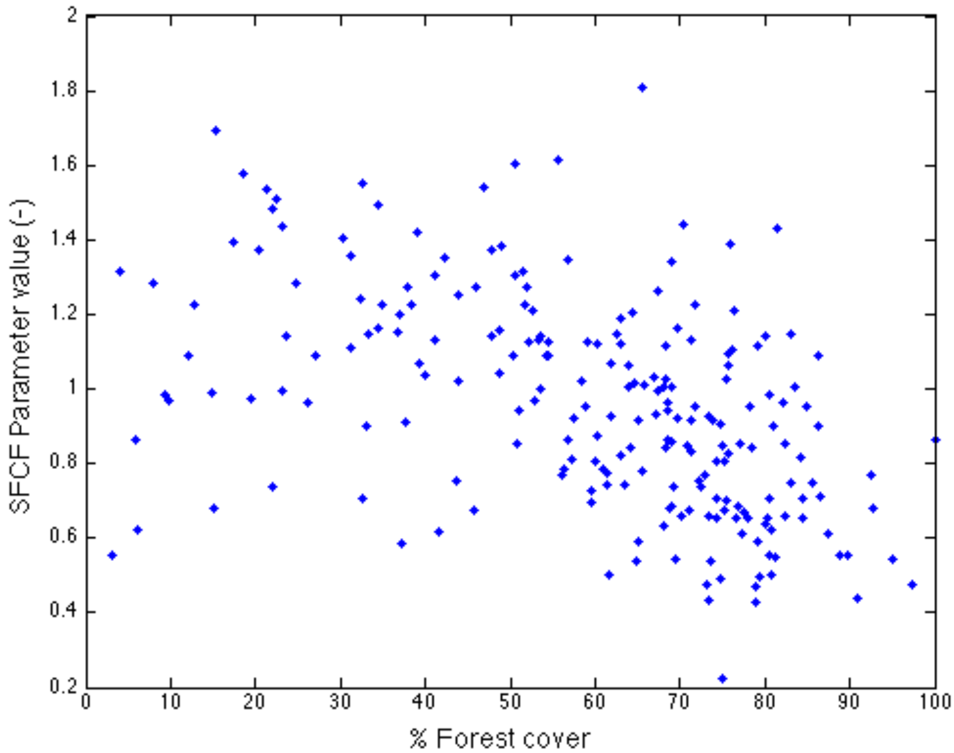

Figure 12 Medians of the parameter SFCF plotted against forest cover.

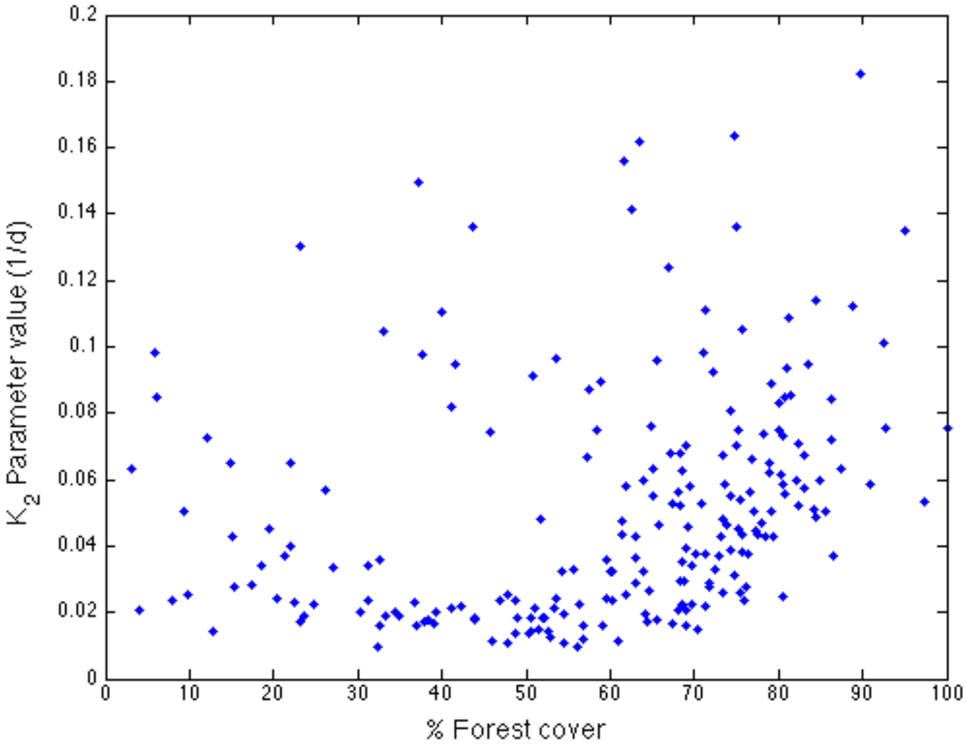

Figure 13 Medians of the parameter K_2 plotted against forest cover.

The results from the calibration for all catchments gave correlation for some parameters but not as strong and clear relationship as expected and desired. Therefore a further analysis was done to try the same approach but on a smaller scale and fewer catchments to see if the correlations were stronger in that case and then a general model for all Sweden might not be supported.

3.3 Cluster analysis

The cluster analysis resulted in six different groups (Figure 14), with the Balsjö catchments belonging to cluster group number six together with 25 other catchments. The correlations for the other cluster groups were calculated but not analysed and only group number six was included in the further analysis.

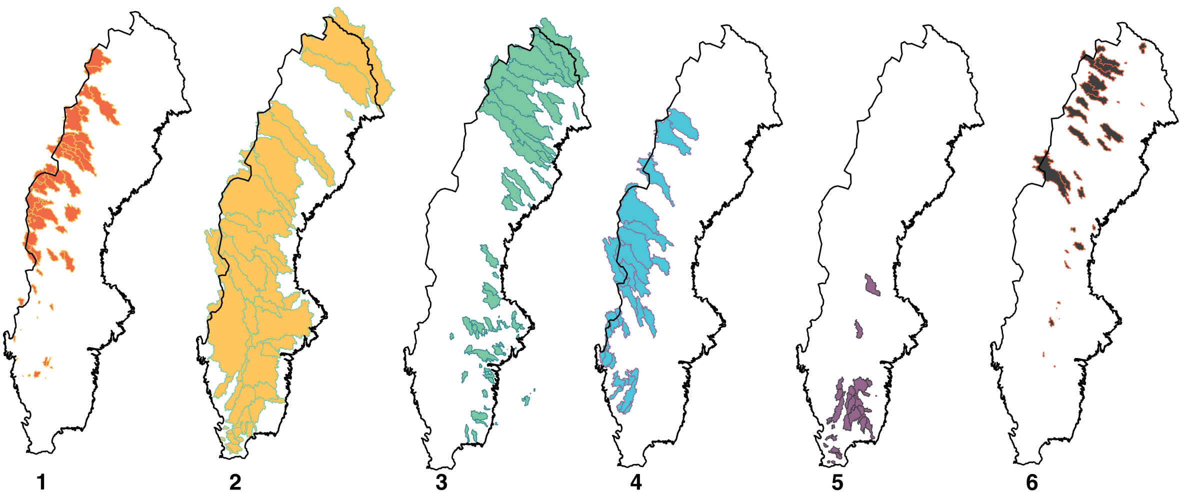

Figure 14 The six catchment groups from the cluster analysis.

3.4 Relationship between model parameters and forest cover (Balsjö cluster group six)

A new calibration was performed for all catchments belonging to cluster group number six, and for this smaller group of catchments the calibration could be done 100 times for every catchment (instead of 25 as done before) within a reasonable time. When correlating the median of the best 100 parameter values against forest cover, significant correlations were found for 6 parameters (Table 5), which are two parameters less than for the calibration of all the catchments (Table 4). The analysis of this smaller group did not show a correlation for the parameters $K_{0,}$ MAXBAS or CFMAX but instead had a significant correlation for LP (P=0.034) and almost for FC (P=0.056). It should also be noted that the r_s for TT had switched from minus to plus: while there was only a weak negative correlation when analysing all catchments, there was a very strong positive correlation with forest within cluster group six.

Table 5 Spearman rank correlation for the Balsjö-like catchment group (6). Bold values are significant at 5% shows correlation within 5% significance level.

Parameters	r_s	p-value
Perc	-0.24	0.251
UZL	-0.30	0.151
K_0	0.36	0.080
K_1	-0.03	0.905
K_2	**0.40**	**0.051**
MAXBAS	0.21	0.310
CET	-0.10	0.641
TT	**0.63**	**0.001**
CFMAX	0.09	0.657
SFCF	**-0.40**	**0.047**
CFR	**0.48**	**0.015**
CWH	0.32	0.124
FC	0.39	0.056
LP	**-0.43**	**0.034**
BETA	**0.59**	**0.002**

3.5 Comparison of Balsjö pre- and post-clearcutting parameters

For some parameters the obtained values for pre- and post-clearcutting conditions in the Balsjö catchments fit nicely with the trends demonstrated earlier (Figures 15, 16), while the exact opposite trend could be seen for others (Figure 17).

The most clear trend and best validation with the Balsjö catchments is seen with the parameter TT (Figure 15). The pre- and post-clearcutting parameter values of the three Balsjö catchments nicely match the clear positive trend of the other catchments belonging to the same cluster group. All three Balsjö catchments were calibrated with three different lengths of datasets after the clear cut (1, 3 and 5.5 years). The different lengths are plotted as different dots in the figures below, which are shown by the three colored dots for the post-clearcutting catchments. The three different data length calibrations are not showing a very wide range of values and they are all showing similar values as the cluster group did.

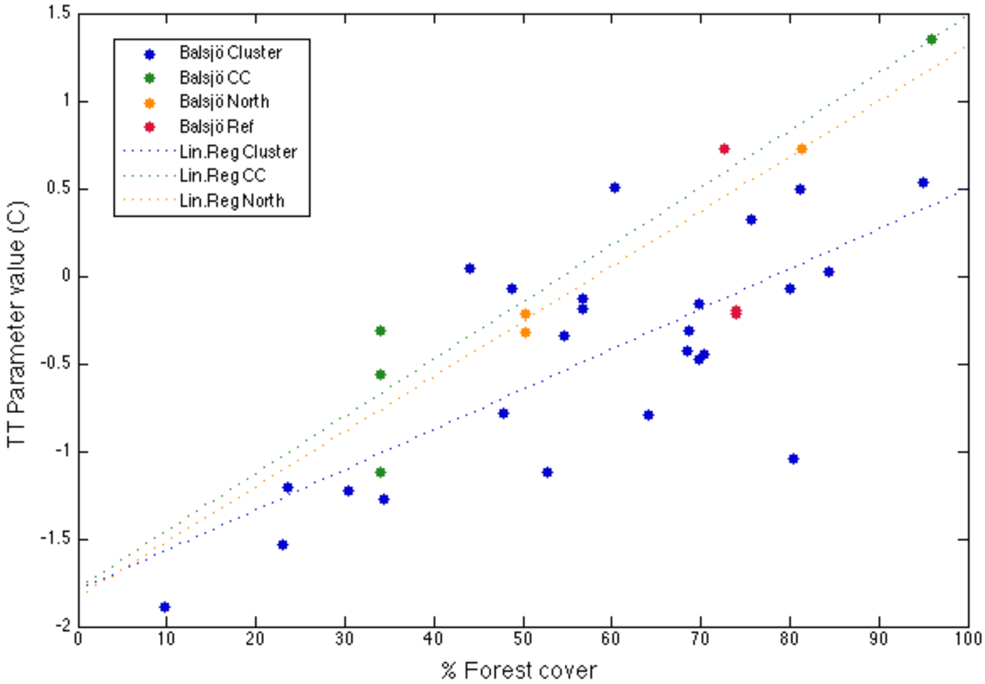

Figure 15 Plot of the median values of parameter TT for the cluster group 6 catchments (blue dots) and the three Balsjö catchments (green, yellow and red dots) are both before and after CC. The linear regression lines are also shown for the cluster group (blue dotted line) and for the two harvested catchments (green and yellow dotted lines).

Similarly, also the correlation for the parameter BETA could be validated with the Balsjö catchments (Figure 16). A positive correlation between BETA and forest cover can be seen both for the entire cluster group and the Balsjö catchments, but the latter shows a much weaker correlation.

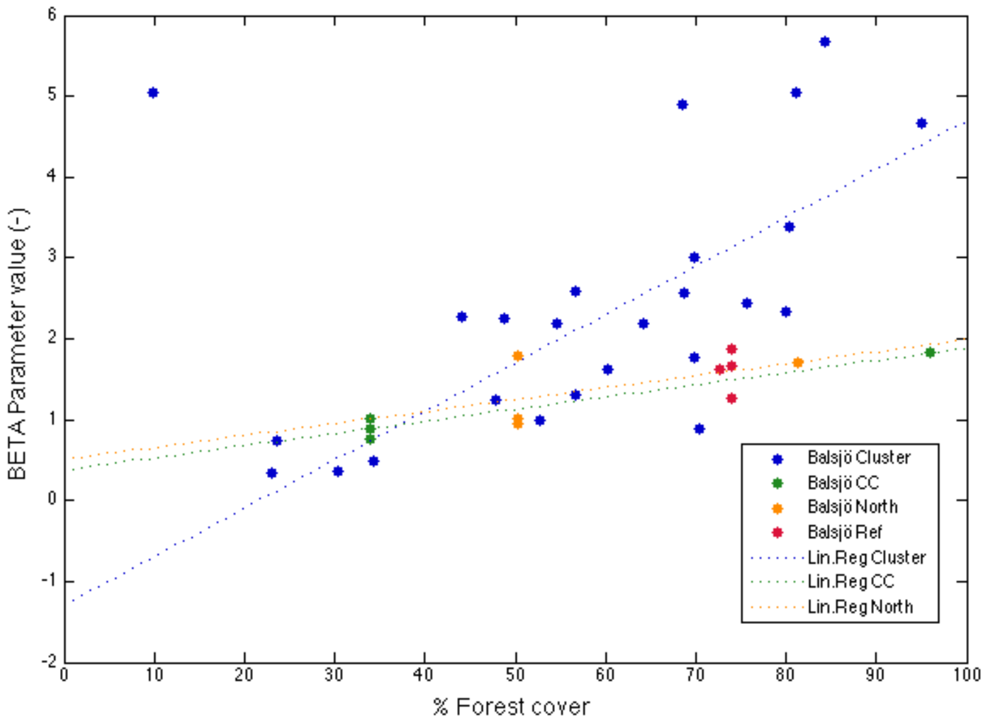

Figure 16 Plot of the median values of parameter BETA for the cluster group 6 catchments (blue dots) and the three Balsjö catchments (green, yellow and red dots) are both before and after CC. The linear regression lines are also shown for the cluster group (blue dotted line) and for the two harvested catchments (green and yellow dotted lines).

For other parameters, such as CFR, the correlation in the cluster group could not be verified with the Balsjö catchments (Figure 17).

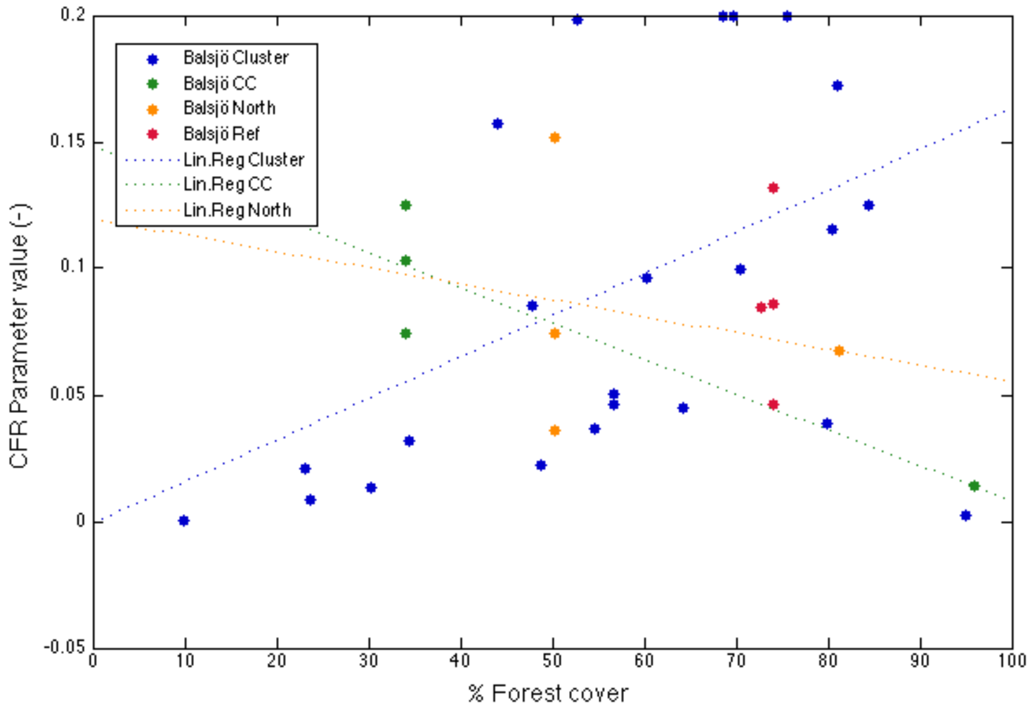

Figure 17 Plot of the median values of parameter CFR for the cluster group 6 catchments and the Balsjö catchments.

The parameter FC did not result in a significant correlation with forest cover at a 5% significance level, but with a p-value of 0.056 it is significant at a 10% significance level and is therefore considered good enough. The Balsjö catchments could actually confirm this correlation quite well (Figure 18, 19). The box plot shows an example of the uncertainty of the value with wide ranges in the box plot. Both the cluster group catchments and the three Balsjö catchments have some catchments with wide box ranges but catchments over 70% of forest cover have narrower ranges.

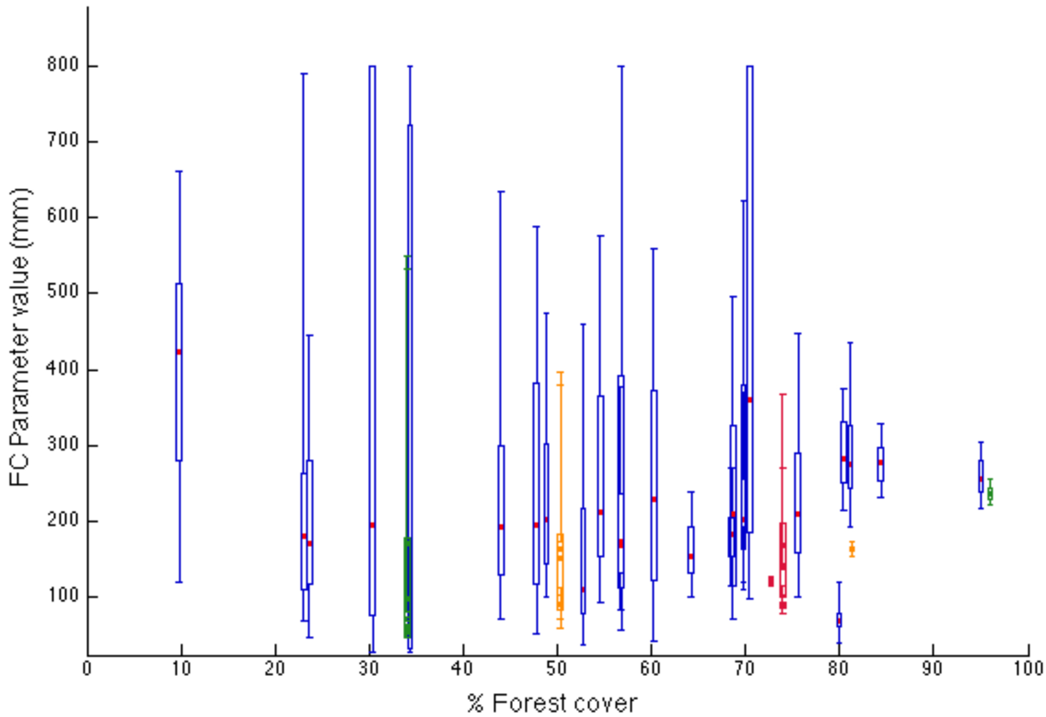

Figure 18 Box plot for the parameter values of FC based on 100 calibrations. Showing the cluster group six catchments (blue), CC (green), North (yellow) and Ref (red). The line in the middle of the boxes shows the median, the box is 25 percentiles ant the whiskers are 95 percentiles.

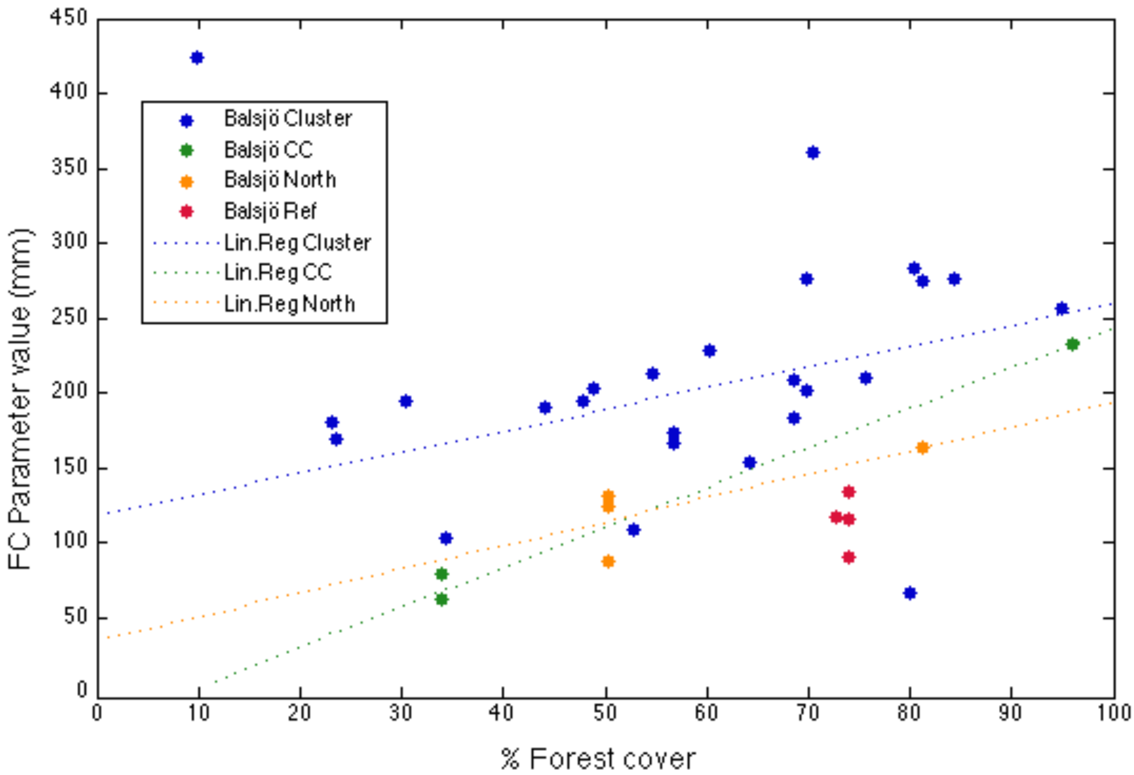

Figure 19 Plot of the median values of parameter FC for the cluster group 6 catchments and the Balsjö catchments.

4 DISCUSSION

4.1 General Performance

The majority of the catchments could be calibrated with satisfying model efficiency but still, there were a large number of catchments that could not. The poor model performance could have different causes but some possible reasons could be disturbance from human impact, regulations of streams or water reservoirs. A quick analysis was done to compare the weighted objective function against 'Degree of regulation' and 'Catchment area' but no clear pattern were found. For the large catchments the problem could be something similar as for the mountainous region, because the interpolated value of precipitation for the whole catchment could not describe the precipitation for that catchment due to the big variation of precipitation in the catchment area. Similar problem can occur for very small catchments if the precipitation station is placed far away from the catchment, then the interpolated value can be very different from the "true" value.

A problem with inconsistent water balances was found, especially in the mountainous parts in northwestern Sweden. This is possibly caused by large spatial variations in precipitation patterns due to the elevation differences and due to a low density of measuring stations. This problem in the PTHBV data set was also described in van der Velde et al. (2013). They solved it by using an additional data set with E-Obs data from Haylock et al. (2008), which have higher precipitation values in the higher regions, because this data set takes more measurement stations from the Norwegian side of the mountains into account. This was not done in this study, because the resolution of the E-Obs data set is too coarse for the relatively small catchments analysed in this study. Instead, catchments with higher discharge than total precipitation were excluded. But there were still a couple of catchments with unreasonably high runoff coefficients (runoff to precipitation ratio > 0.8), which probably contribute to the uncertainties in the data and in the interpretation of the results.

4.2 Relationship between model parameters and forest cover

The model parameters are partly based on physical concepts and some correlation with forest cover was expected. At the same time it is not reasonable to have good correlation for all of the parameters since all of them are not expected to be dependent on forest cover percentage. But the present study also revealed some relationships between parameters and forest cover that were not expected based on earlier findings by Seibert (1999). Hereafter follows an attempt to explain the relationship with forest cover for the significantly correlated parameters found in this study:

- The recession coefficient K_0 influences the peak flow recessions and lower K_0 values leads to lower flood peaks (Seibert, 2005). Forested catchments typically show lower flood peaks compared to catchments with more open, urban or agricultural areas (e.g., Hundecha and Bárdossy, 2004; Fitzpatrick et al., 1999). Thus, the negative correlation found in this study (i.e., lower K_0 values with higher forest percentage) would have been expected. However, even though a significant correlation was found, the correlation is only weak with r_s=-0.17.
- The K_2 recession coefficient is responsible to change the slope of the baseflow (Seibert, 2005). Higher K_2 values cause lower simulated baseflow, which is typically associated with higher forest cover (Price, 2011) and, thus, confirms the found positive correlation in this study.

- MAXBAS is a routing parameter that transforms the runoff in the final step of the model by a triangular weighting function (Seibert, 1999). Higher values would cause more delay in the runoff, giving lower peaks and longer and flatter recession curves, which is often associated with higher percentages of lakes or with larger catchments areas (Seibert, 1999). A weak but significant negative correlation was found between forest cover percentage and catchment area (r_s=-0.28) as well as lake percentage (r_s=-0.30) (Appendix C), which could be a possible reason for the weak negative correlation of MAXBAS and forest cover.
- TT, the threshold temperature, is related to snow melt and the refreezing of melt water. According to HBV's melting and refreezing equations in (Seibert, 1999), a higher value of TT causes less snow to melt and to melt the snow later at higher temperatures, which would be expected in a forested area with more shadow and less available snow pack. This relationship could, however, not be confirmed in this study, as the found correlation was slightly negative with r_s=-0.14.
- CFMAX is the degree-day factor, which was expected and confirmed to decrease with higher percentage of forest (Seibert, 1999) due to the fact that more forest with more shadow and less snow on the ground will give less snowmelt. The wind is also an important factor in the snow melting process and is also affected by a clear cut.
- This is also connected to the refreezing coefficient CFR that is increasing with increasing forest cover and this could be explained as a compensation for the opposite trend of CFMAX (Seibert, 1999).
- The snowfall correction factor SFCF compensates for the fact that evaporation and sublimation of snow, which may be significant in forested areas, is not simulated by HBV (Seibert, 1999). Lower values of SFCF reduce the water equivalent of the snow pack, which is usually the case in forested areas and this is also confirmed with the found correlation.
- Finally, the parameter BETA determines the relative contribution from rain and snowmelt to runoff at a given soil moisture deficit. This is connected to the soil characteristics and land cover and was therefore expected to be correlated to forest cover (Seibert, 1999) because a clear cut is highly affecting the root zone and ability of the soil to keep the moisture in the soil. The correlation was however rather weak with r_s=0.14.

The correlations found in this study for CFMAX, SFCF, CFR and K_2 were also demonstrated by Seibert (1999), but the additional correlation for K_0, TT, BETA and MAXBAS were new findings. Especially SFCF and CFMAX show correlations in a similar order of magnitude compared to Seibert (1999), but the K_2 correlation is showing the opposite sign of correlation.

The problem with the presented correlation analysis is the large amount of variable factors included in the large catchment data set stretching from far North to far South. Other factors such as wetland percentages, catchment size, local climate, elevation or slope might dilute the possible clear relationships to forest cover.

One pattern that was observed in many of the box plots of the parameters was that the uncertainty (length of the box) was in general higher for the catchments with lower percentage of forest cover. Especially over 70% of forest cover gave a distinct decrease in uncertainty. This observation could be connected to the problem with poor data in higher regions, because there is a strong negative correlation between forest cover and elevation (Appendix C). The phenomena of poor data in the mountainous region could therefore be a reason for this observation.

4.3 Cluster analysis

The cluster analysis gave a clear partition of the catchments that could be more useful than for example clustering them regionally in six geographical parts. Some patterns are obvious in the results and it is clear that group one and five consist mainly of small catchments in west and south respectively (Figure 14). Group two mainly includes very large catchments. Group three and four corresponds mainly to mid-sized catchments in the northeast and west parts respectively. The used group number six is made up by very small catchments and larger catchments in the north closer to the Balsjö catchments.

4.4 Relationship between model parameter (Balsjö cluster group six)

For the Balsjö catchment group, similar correlations as for all Swedish catchments (see sections 3.2 and 4.2) were found for only two parameters: K_2 and SFCF. Additionally, CFR and BETA showed correlations in the same direction as before, but correlations were much stronger for the cluster group. The previously found weak correlations for K_0, MAXBAS and CFMAX could not be confirmed. TT, which had only a weak negative correlation before, now has a strong positive correlation as would been expected (see explanation in section 4.2). LP, which is included in the soil moisture routine and the soil moisture, was expected to correlate with the forest cover, especially due to the relationship between land cover soil types (Seibert, 1999). The similarities to earlier findings are hopeful results because the study by Seibert (1999) only used nine smaller catchments but they were catchments with good measurements in the south central Sweden compared to this study that uses large catchments with some problematic data.

The results shown here are based on median values of the 100 best parameter values. This is not optimal because there could be several ranges of values that are included and other local optimal solutions are possible. It is at least more reasonable to use the median than the mean value, which in the case of two local optimal ranges could show a value somewhere in between. This value, unlike the median, has not been a part of a well performing parameter set and would have had even more uncertainty. The box plots can be used to show this uncertainty. All models do have some parameter uncertainty and this is also a fact for the HBV model since it is a strong simplification of a complex natural system. Parameter uncertainty for the HBV model have been studied by Seibert (1997) and Uhlenbrook et al. (1999) and they found the parameter uncertainty to be considerably high for the major part of the parameters. This has to be considered in the analysis and a good way is to show the ranges or probability distributions and not a single value. Box plots are a good way to do this, but in this study mostly median values were chosen for the sake of simplicity and simple visualization.

The results of the general calibration and for the calibrations of the Balsjö-like catchments in cluster group six differ considerably and many of the correlations show different relationships. This could be explained in two ways. Either the calibration method in this study is poor and does not give proper results or that a general correlation for all Sweden and all kind of catchments is not possible and smaller groups and regionalization of the catchments have to be used to model the impact with this method. It is a big country and the climate, geology and land use are very different in different parts, which would make the last cause reasonable.

4.5 Comparison of Balsjö pre- and post-clearcutting parameters

The validation of the method with the calibrations of the Balsjö catchments is interesting and could show if the chosen approach really is reasonable. Unfortunately the lack of data is making this analysis not really reliable and could be the cause of some strange results. The calibration with this data could give very different results depending on if that one year is very wet or dry and when the spring flood appears and so on. A quick analysis shows that the used year 2006 seems to be close to average year with no special rain events and a normal spring flood compared to the years after the clearcutting. Just one year of data is not considered enough to get a trustful calibration but at least the uncertainty is made visible in the results by showing the calibration results for the different lengths of data series. However, this data set was the best available for this purpose and is still useful but the results should be used carefully. The FC, BETA and TT parameters are showing very promising results, because they are closely following the trend of the cluster group, although confirmation of just 3 out of 7 correlations is not that satisfying. If data from even more clearcutting sites were available, it would be very interesting to try this method there as well.

4.6 Equifinality

One major problem with the used modelling approach is equifinality (Beven and Freer, 2001), which implies that the parameter calibration for a catchment does not result in one true parameter set. Instead the calibration results in many possible and well performing parameter sets. Some methods to address this problem have been investigated. Hundecha and Bárdossy (2004) used different catchment characteristics as base for the regionalization and parameter values were initially associated with the characteristics and could then be transferred to the catchment scale using a transfer function. An investigation of the multilinear regression and the entropy could be used to find an internal structure of the parameters and then the structure could be treated as parameter vector (Das et al., 2006). Bárdossy (2007) concluded that the parameters of a rainfall-runoff model could not be identified as individual values and parameter vectors could be used instead. The reason for this is mainly the fact that model parameters have some interdependencies, which is when the value of one parameter can be compensated by a change in the value of some of the other parameters.

4.7 Recommendations and further research

To handle the problem with interdependency between parameters, the concept of copulas (Nelsen, 2006) could be studied. This is useful when there are strong dependencies or correlation between parameters in the model space (Dorota and Kurowicka, 2010). The principle of copulas is a modern statistic tool that can be described as functions that join multivariate distribution functions to their one-dimensional marginal distribution and it can be used to model the interdependence. A good introduction in this complex area is given by Nelsen (2006) and an example of a hydrological study using copula for regionalization is provided by Samaniego et al. (2010).

Another interesting method is demonstrated by Yadav et al. (2007), who identified the relationship between physical characteristics (forest cover) and dynamic streamflow responses of watersheds, which seemed to work better than comparing it to model parameters.

There are many different ways to do the calibration of the HBV model: Different algorithms have many settings that can be changed and that would result in different parameter sets. Further research with a more stable and reliable calibration method is suggested or one could try out different settings of parameter ranges and in the GAP calibration, than just the ones used in this study, where no deeper evaluation was done.

5 CONCLUSION

A parameter regionalization was done for 218 catchments in Sweden and relationships between forest cover and model parameter were assessed. Significant correlations were found for about half of the 15 HBV parameters and most of them could also be explained with hydrological reasoning and experiences from paired catchment studies. Validation with a paired catchment clearcutting experiment could partially confirm the previously found correlations for three parameters by showing the same parameter response caused by a change in forest cover. The results revealed some promising correlations but the results are showing a lot of uncertainties and some parameter correlations could not be seen in the validation.

This lead to the conclusion that the applied approach is likely too simple to be able to model the impacts from forest harvesting on streamflow. The main reasons for this seem to be equifinality and model parameter interdependencies. Despite this, the results provide a promising first step towards simulating "virtual forest clearcutting", but further research and a considerable expansion of the method is necessary to make it suitable for impact assessment of forest clearcutting.

6 REFERENCES

Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. Hydrology and Earth System Sciences 11, 703–710.

Bergström, S., 1995. The HBV model. SMHI 443–476.

Bergström, S., 1992. HBV model: its structure and applications. SMHI RH No 4.

Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments, Report RHO 7. SMHI, Norrköping, Sweden.

Beven, K., 2009. Environmental modelling: an uncertain future? : an introduction to techniques for uncertainty estimation in environmental prediction. Routledge, London, UK.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249, 11–29. doi:10.1016/S0022-1694(01)00421-8

Booij, M.J., 2005. Impact of climate change on river flooding assessed with different spatial model resolutions. Journal of Hydrology 303, 176–198. doi:10.1016/j.jhydrol.2004.07.013

Brandt, M., Bergström, S., Gardelin, M., 1988. Modelling the Effects of Clearcutting on Runoff: Examples from Central Sweden. Ambio 17, 307–313.

Brath, A., Montanari, A., Moretti, G., 2006. Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty). Journal of Hydrology 324, 141–153.

Calder, I.R., 1993. Hydrologic effects of land-use change. Chapter 13 In Maidment D.R Handbook of hydrology. McGraw-Hill, New-York. pp.50.

Cattell, R.B., 1943. The description of personality: basic traits resolved into clusters. The Journal of Abnormal and Social Psychology 38, 476–506. doi:10.1037/h0054116

Christiansen, L., Karlsson, S., Pettersson, J., 2014. Swedish Statistical Yearbook of Forestry, Official Statistics of Sweden, Chapter 14 - Economy. Swedish Forest Agency, Jönköping, Sweden.

Das, T., Bárdossy, A., Zehe, E., 2006. Assessment of Model Parameters Interdependency of a Conceptual Rainfall-Runoff Model. AGU Fall Meeting Abstracts 43, 0502.

De Roo, A., Odijk, M., Schmuck, G., Koster, E., Lucieer, A., 2001. Assessing the effects of land use changes on floods in the meuse and oder catchment. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere 26, 593–599. doi:10.1016/S1464-1909(01)00054-5

Dorota, K., Kurowicka, D., 2010. Dependence Modeling : Vine Copula Handbook. World Scientific Publishing Co., River Edge, NJ, USA.

EEA, 2007. CLC2006 technical guidelines. European Environmental Agency, Copenhagen, Technical report 2007.

Eklöf, K., Schelker, J., Sørensen, R., Meili, M., Laudon, H., von Brömssen, C., Bishop, K., 2014. Impact of Forestry on Total and Methyl-Mercury in Surface Waters: Distinguishing Effects of Logging and Site Preparation. Environ. Sci. Technol. 48, 4690–4698. doi:10.1021/es404879p

Ellison, D., N. Futter, M., Bishop, K., 2012. On the forest cover–water yield debate: from demand- to supply-side thinking. Glob Change Biol 18, 806–820. doi:10.1111/j.1365-2486.2011.02589.x

Eriksson, B., 1981. Potentiella evapotranspirationen i Sverige, 28/27 ed, RMK/RHO. SMHI, Norrköping, Sweden.

Fitzpatrick, F.A., Knox, J.C., Whitman, H.E., 1999. Effects of Historical Land-Cover Changes on Flooding and Sedimentation, North Fish Creek, Wisconsin. (Water-Resources Investigations Report No. 99–4083). U.S. Geological Survey, USGS.
Harlin, J., 1992. Modelling the hydrological response of extreme floods in Sweden. Nordic Hydrology 23, 227–244.
Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res. 113, D20119. doi:10.1029/2008JD010201
Helsel, D.R., Hirsch, R.M., 2002. Chapter A3: Statistical Methods in Methods in Water Resources. Book 4: Hydrological Analysis and Interpretation Techniques of Water-Resources. United States Geological Survey.
Henestål, J., Ranung, J., Gyllander, A., Johnsen, Å., Olsson, H., Pettersson, O., Westman, Y., Wingqvist, E.-M., 2015. Arbete med SVAR version 2012_1 och 2012_2, Svenskt Vattenarkiv, en databas vid SMHI. SMHI, Norrköping, Sweden.
Hundecha, Y., Bárdossy, A., 2004. Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. Journal of Hydrology 292, 281–295. doi:10.1016/j.jhydrol.2004.01.002
Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognition Letters, Award winning papers from the 19th International Conference on Pattern Recognition (ICPR) 19th International Conference in Pattern Recognition (ICPR) 31, 651–666. doi:10.1016/j.patrec.2009.09.011
Jia, Q.Y., Sun, F.H., 2012. Modeling and forecasting process using the HBV model in Liao river delta. Procedia Environmental Sciences, 18th Biennial ISEM Conference on Ecological Modelling for Global Change and Coupled Human and Natural System 13, 122–128. doi:10.1016/j.proenv.2012.01.012
Johansson, B., 2002. Estimation of areal precipitation for hydrological modelling in Sweden. Göteborg.
Jones, J.A., Post, D.A., 2004. Seasonal and successional streamflow response to forest cutting and regrowth in the northwest and eastern United States. Water Resour. Res. 40, W05203. doi:10.1029/2003WR002952
Kaufman, L., Rousseeuw, P.J., 2009. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons.
Lindstrom, G., Johansson, B., Persson, M., Gardelin, M., Bergstrom, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201, 272–288. doi:10.1016/S0022-1694(97)00041-3
Liu, J., Dietz, T., Carpenter, S.R., Folke, C., Alberti, M., Redman, C.L., Schneider, S.H., Ostrom, E., Pell, A.N., Lubchenco, J., Taylor, W.W., Ouyang, Z., Deadman, P., Kratz, T., Provencher, W., 2007. Coupled Human and Natural Systems. Ambio 36, 639–649.
Nelsen, R.B., 2006. An introduction to copulas, 2. ed. ed, Springer series in statistics. Springer, New York.
Penman, H.L., 1948. Natural Evaporation from Open Water, Bare Soil and Grass. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 193, 120–145. doi:10.1098/rspa.1948.0037
Press, W.H., 2002. Numerical recipes in C++: the art of scientific computing. Cambridge Univ. Press, Cambridge.

Price, K., 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. Progress in Physical Geography 35, 465–492. doi:10.1177/0309133311402714
Primožič, M., Kobold, M., Brilly, M., 2008. The implementation of the HBV model on the Sava River Basin. IOP Conf. Ser.: Earth Environ. Sci. 4, 012004. doi:10.1088/1755-1307/4/1/012004
Samaniego, L., Bárdossy, A., Kumar, R., 2010. Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. Water Resour. Res. 46, W02506. doi:10.1029/2008WR007695
Sanderson, E.W., Jaiteh, M., Levy, M.A., Redford, K.H., Wannebo, A.V., Woolmer, G., 2002. The Human Footprint and the Last of the Wild. BioScience 52, 891–904. doi:10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2
Schelker, J., Kuglerová, L., Eklöf, K., Bishop, K., Laudon, H., 2013. Hydrological effects of clear-cutting in a boreal forest – Snowpack dynamics, snowmelt and streamflow responses. Journal of Hydrology 484, 105 – 114.
Seibert, J., 2005. HBV-light, User´s Manual Version 2. Stockholm.
Seibert, J., 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. Hydrology and Earth System Sciences 4, 215–224. doi:10.5194/hess-4-215-2000
Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. Agricultural and Forest Meteorology 98–99, 279–293. doi:10.1016/S0168-1923(99)00105-7
Seibert, J., 1997. Estimation of parameter uncertainty in the HBV model. Nordic Hydrology 28, 247–262.
Seibert, J., McDonnell, J.J., 2010. Land-cover impacts on streamflow: a change-detection modelling approach that incorporates parameter uncertainty. Hydrological Sciences Journal–Journal des Sciences Hydrologiques 55, 316–332.
Seibert, J., Vis, M.J.P., 2012. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. Hydrology and Earth System Sciences 16, 3315–3325. doi:10.5194/hess-16-3315-2012
SMHI, 2015. Vattenföring. URL http://www.smhi.se/klimatdata/hydrologi/vattenforing (accessed 6.11.15).
SMHI, 2014. HBV. URL http://www.smhi.se/forskning/forskningsomraden/hydrologi/hbv-1.1566 (accessed 6.2.15).
Solomatine, D.P., Wagener, T., 2011. 2.16 - Hydrological Modeling, in: Wilderer, P. (Ed.), Treatise on Water Science. Elsevier, Oxford, pp. 435–457.
Sørensen, R., Ring, E., Meili, M., Högbom, L., Seibert, J., Grabs, T., Laudon, H., Bishop, K., 2009. Forest Harvest Increases Runoff Most during Low Flows in Two Boreal Streams. Ambio 38, 357–363. doi:10.1579/0044-7447-38.7.357
Stednick, J.D., 1996. Monitoring the effects of timber harvest on annual water yield. Journal of Hydrology 176, 79–95. doi:10.1016/0022-1694(95)02780-7
Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology 456–457, 12–29. doi:10.1016/j.jhydrol.2012.05.052
Uhlenbrook, S., Seibert, J., Leibundgut, C., Rodhe, A., 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. Hydrological Sciences Journal 44, 779–797.

Van der Velde, Y., Lyon, S.W., Destouni, G., 2013. Data-driven regionalization of river discharges and emergent land cover–evapotranspiration relationships across Sweden. J. Geophys. Res. Atmos. 118, 2576–2587. doi:10.1002/jgrd.50224
Vitousek, P.M., 1994. Beyond Global Warming: Ecology and Global Change. Ecology 75, 1862–1876. doi:10.2307/1941591
Wooldridge, S., Kalma, J., Kuczera, G., 2001. Parameterisation of a simple semi-distributed model for assessing the impact of land-use on hydrologic response. Journal of Hydrology 254, 16–32. doi:10.1016/S0022-1694(01)00489-9
Wrede, S., Seibert, J., Uhlenbrook, S., 2013. Distributed conceptual modelling in a Swedish lowland catchment: a multi-criteria model assessment. Hydrology Research 44, 318. doi:10.2166/nh.2012.056
Yadav, M., Wagener, T., Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources 30, 1756–1774. doi:10.1016/j.advwatres.2007.01.005
Zegre, N., 2011. Evaluating the hydrologic effects of forest harvesting and regrowth using a simple rainfall-runoff model. Journal of Environmental Hydrology 19.
Zhang, M., Wei, X., Sun, P., Liu, S., 2012. The effect of forest harvesting and climatic variability on runoff in a large watershed: The case study in the Upper Minjiang River of Yangtze River basin. Journal of Hydrology 464–465, 1–11. doi:10.1016/j.jhydrol.2012.05.050

Appendix A – HBV CLI function

The available information and instruction for how to use the HBV-light-CLI function is very limited. Therefore I want to publish a short guide here to clarify the usage of this function.

The HBVsoftware can run through the Windows CMD with the same function as the GUI but it can easily be programmed for many catchments. It is important to have capital letters where it supposes to be. Before this can be done, all the needed input data have to be in correct files and structure in the folder. Just as they are when the GUI is used (Seibert, 2005).

The first thing to do is to enter the path to the folder where HBV-light is installed and make sure there is a file called HBV-light-CLI.
Navigate to correct folder with the commando CD [*folder]*
C:
CD program files

Then the HBV-light software can be started with the commando:
hbv-light-cli

Then the different functions and settings are displayed.
But to run the calibration used in this study the following code is used:
hbv-light-cli Run "path to input data folder" GAPRun "path to results folder"

Examle:
hbv-light-cli Run P:\HBVdata\1 GAPRun P:\HBVdata\1\results

To calibrate many catchments automatically a batchfile can be written like this:

Write a .txt file and save as Batchfile (.bat)

hbv-light-cli Run P:\HBVdata\1 GAPRun P:\HBVdata\1\results
hbv-light-cli Run P:\HBVdata\2 GAPRun P:\HBVdata\2\results
hbv-light-cli Run P:\HBVdata\3 GAPRun P:\HBVdata\3\result

Appendix B – Land use reclassification

Table showing the reclassification of land classes in different colors.

GRID_CODE	CLC_CODE	LABEL1	LABEL2	LABEL3	RGB
1	111	Artificial surfaces	Urban fabric	Continuous urban fabric	230-000-077
2	112	Artificial surfaces	Urban fabric	Discontinuous urban fabric	255-000-000
3	121	Artificial surfaces	Industrial, commercial and transport units	Industrial or commercial units	204-077-242
4	122	Artificial surfaces	Industrial, commercial and transport units	Road and rail networks and associated land	204-000-000
5	123	Artificial surfaces	Industrial, commercial and transport units	Port areas	230-204-204
6	124	Artificial surfaces	Industrial, commercial and transport units	Airports	230-204-230
7	131	Artificial surfaces	Mine, dump and construction sites	Mineral extraction sites	166-000-204
8	132	Artificial surfaces	Mine, dump and construction sites	Dump sites	166-077-000
9	133	Artificial surfaces	Mine, dump and construction sites	Construction sites	255-077-255
10	141	Artificial surfaces	Artificial, non-agricultural vegetated areas	Green urban areas	255-166-255
11	142	Artificial surfaces	Artificial, non-agricultural vegetated areas	Sport and leisure facilities	255-230-255
12	211	Agricultural areas	Arable land	Non-irrigated arable land	255-255-168
13	212	Agricultural areas	Arable land	Permanently irrigated land	255-255-000
14	213	Agricultural areas	Arable land	Rice fields	230-230-000
15	221	Agricultural areas	Permanent crops	Vineyards	230-128-000
16	222	Agricultural areas	Permanent crops	Fruit trees and berry plantations	242-166-077
17	223	Agricultural areas	Permanent crops	Olive groves	230-166-000
18	231	Agricultural areas	Pastures	Pastures	230-230-077
19	241	Agricultural areas	Heterogeneous agricultural areas	Annual crops associated with permanent crops	255-230-166
20	242	Agricultural areas	Heterogeneous agricultural areas	Complex cultivation patterns	255-230-077
21	243	Agricultural areas	Heterogeneous agricultural areas	Land principally occupied by agriculture	230-204-077
22	244	Agricultural areas	Heterogeneous agricultural areas	Agro-forestry areas	242-204-166
23	311	Forest and semi natural areas	Forests	Broad-leaved forest	128-255-000
24	312	Forest and semi natural areas	Forests	Coniferous forest	000-166-000
25	313	Forest and semi natural areas	Forests	Mixed forest	077-255-000
26	321	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Natural grasslands	204-242-077
27	322	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Moors and heathland	166-255-128
28	323	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Sclerophyllous vegetation	166-230-077
29	324	Forest and semi natural areas	Scrub and/or herbaceous vegetation associations	Transitional woodland-shrub	166-242-000
30	331	Forest and semi natural areas	Open spaces with little or no vegetation	Beaches, dunes, sands	230-230-230
31	332	Forest and semi natural areas	Open spaces with little or no vegetation	Bare rocks	204-204-204
32	333	Forest and semi natural areas	Open spaces with little or no vegetation	Sparsely vegetated areas	204-255-204
33	334	Forest and semi natural areas	Open spaces with little or no vegetation	Burnt areas	000-000-000
34	335	Forest and semi natural areas	Open spaces with little or no vegetation	Glaciers and perpetual snow	166-230-204
35	411	Wetlands	Inland wetlands	Inland marshes	166-166-255
36	412	Wetlands	Inland wetlands	Peat bogs	077-077-255
37	421	Wetlands	Maritime wetlands	Salt marshes	204-204-255
38	422	Wetlands	Maritime wetlands	Salines	230-230-255
39	423	Wetlands	Maritime wetlands	Intertidal flats	166-166-230
40	511	Water bodies	Inland waters	Water courses	000-204-242
41	512	Water bodies	Inland waters	Water bodies	128-242-230
42	521	Water bodies	Marine waters	Coastal lagoons	000-255-166
43	522	Water bodies	Marine waters	Estuaries	166-255-230
44	523	Water bodies	Marine waters	Sea and ocean	230-242-255
48	999	NODATA	NODATA	NODATA	
49	990	UNCLASSIFIED	UNCLASSIFIED LAND SURFACE	UNCLASSIFIED LAND SURFACE	
50	995	UNCLASSIFIED	UNCLASSIFIED WATER BODIES	UNCLASSIFIED WATER BODIES	230-242-255
255	990	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	

Appendix C – Correlation of catchment characteristics

Correlations were also found between forest cover percentage in all the catchments and the following parameter in the table, which also shows the Spearman's correlation coefficient r_s and the p-value.

	r_s	p-value
Catchment Area	-0.28	0
Lake percentage	-0.3	0.0369
Mean elevation	-0.39	0