



UPPSALA
UNIVERSITET

W 16014

Examensarbete 30 hp
Maj 2016

Accuracy in Swedish unsegmented and segmented rating curves

Accounting for measurement uncertainty and
heteroscedasticity

Mattias Sörengård

ABSTRACT

Accuracy in Swedish unsegmented and segmented rating curves

Mattias Sörengård

River discharge estimation is the basic hydrological information for most hydrological applications in various socioeconomic planning. Increasing the accuracy of the traditional rating curve in relation to river discharge estimation would be very valuable to hydrological applications. Suggestions have been made that the traditional power function rating curve should be divided into several segments because this is often motivated by the physical characteristics of the river. Each curve is commonly constructed by regression and each requires 3 estimated parameters. However stage-discharge data is often scarce, and this scarcity could lead to overparametrization and deterioration of accuracy.

By constructing many unsegmented and segmented rating curves accounting for measurement uncertainty, the models can be validated, it can be determined if segmented rating curves suffers from overparametrization. The results showed that two-segmented rating curves did not yield better fits to data, and nor did it generate larger errors than unsegmented rating curves in extrapolation. Segmentation only reduced errors in low flow interpolation, when there is a clear segmentation. It could also be concluded that unsegmented rating curves were slightly more robust when extrapolating. The biggest impact on rating curve errors was shown not to be determined by segmentation, but rather much more dependent on the amount of discharge measurement uncertainty or choice of regression method. With a mean discharge uncertainty of ± 5 %, the errors from in high flow was 60 % in interpolation and 35 % in extrapolation. For low flows, the interpolation errors were around 95 % and extrapolation error estimation was 250 %. Conclusions could also be made that the relative errors from rating curves increased with lower discharges.

Other important regression factors, such as heteroscedasticity, sometimes showed to have substantial impact on rating curve regressions, generally reduced from 59 % occurrence in unsegmented rating curves to 14-15 % in segmented rating curves.

Keywords: Rating curve, Non-linear-regression, Projection variable method, river discharge, gauging station, stage, uncertainty, validation, Overparametrization, overfit, hydrology

Department of Earth Sciences, Program for Air, Water and Landscape Sciences,
Uppsala University
Box 337, SE-751 05 Uppsala

REFERAT

Noggrannhet i svenska segmenterade och osegmenterade i avbördningskurvor

Mattias Sörengård

Uppskattning av vattenflöden i vattendrag är den grundläggande informationen för de flesta hydrologiska applikationer vid olika typer av socioekonomisk planering. Att förbättra noggrannheten i avbördningskurvor då vattenflödet uppskattas vid en mätstation skulle vara värdefullt för de flesta tillämpningar där vattenflöden används. Tidigare studier har föreslagit att avbördningskurvor borde delas upp i flera segment, eftersom vattendrag inte sällan har olika segment med olika fysikaliska karaktärer. Varje segment kräver dock att 2-3 regressionsparametrar bestäms, men flödesmätningar vid olika vattennivåer är ofta få, och knappheten kan göra att en utökad modell blir överparametriserad och än mer osäker.

Genom att konstruera många avbördningskurvor, segmenterade och osegmenterade, kan dessa valideras mot valideringsdata och var det möjligt se om segmenterade avbördningskurvor blev överparametriserade. Studien visade att segmenterade avbördningskurvor vid kalibrering, interpolation och extrapolation generellt inte gav bättre prediktion än osegmenterade avbördningskurvor. Vid låga flöden och tydligt motiverade segmenteringar gav segmenterade avbördningskurvor en bättre interpolation, men dock inte vid extrapolation, vilket är en indikation att segmenterade avbördningskurvor var något överparametriserade. Den största inverkan på att minska felen i avbördningskurvor var att minska mätosäkerheten i flödesmätningarna. Med en genomsnittlig mätosäkerhet i flödesmätningarna på ± 5 % kunde osäkerheten kvantifieras till kring 60 % för interpolerade osegmenterade avbördningskurvor vid höga flöden och kring 95 % vid låga flöden. Variansen var dock stor. Osäkerheten från modellvalideringen av extrapolation för osegmenterade avbördningskurvor vid höga flöden kvantifierades till kring 35 % vid höga flöden och kring 250 % vid låga flöden. Resultaten visade att de relativa felen från avbördningskurvor blev större för ju lägre flödet blir.

Heteroskedastitet, som kan generera osäkerheter i avbördningskurvor, visade sig vara vanligare (59 %) i osegmenterade avbördningskurvor jämfört med segmenterade (14-15 %). Även antalet flödesmätningar hade en betydelse för felen i avbördningskurvor.

Nyckelord: Avbördningskurva, icke-linjär regression, flöde, flödesmätning, osäkerhet, validering, hydrologi, överparametrisering, vattenstånd

Institutionen för geovetenskaper, Luft-, vatten- och landskapslära, Uppsala Universitet
Geocentrum, Villavägen 16, SE-752 36 Uppsala

PREFACE

This is a Master thesis corresponding to 30 ETCS and made as the final and independent work of my M.Sc. in Environmental and Water Engineering at Uppsala University. It started as a small project on the side of my other studies and later developed into a full scale master thesis.

The thesis was supervised by Giuliano Di Baldassarre, Professor in Hydrology at Uppsala University at the Department of Earth Sciences, to whom I would like to express my warmest gratitude. He has generously provided me with ideas to the outline of the project and continuously been supporting me with methodology and motivation. The subject reviewer was Thomas Grabs, senior lecturer in hydrology at Uppsala University at the Department of Earth Sciences, to whom I want to thank for his thoughtful reflections on the subject and on the report. And thank you Anna Sjöblom, senior lecturer in metrology at Uppsala University at the Department of Earth Sciences, as examiner, for the fantastic support on improving the quality of the report.

Early contact was made with the Swedish Meteorological and Hydrological Institute (SMHI) where Maud Goltsis Nilsson, hydrologist, kindly agreed to provide with data and with professional knowledge upon the subject. I want to express my gratitude to her and all other hydrologists at SMHI for their warm welcoming and valuable input. Also I want to thank Ida Enjebo for granting the permission to reproduce her figure in Enjebo (2014) as a part of Figure 1 in this report.

Finally I want to thank my family, partner and friends who have supported me through all my education and thesis.

Stockholm, Sweden, April 2016

Mattias Sörengård

Copyright © Mattias Sörengård and Department of Earth Sciences, Air, Water and Landscape Science, Uppsala University.

UPTEC W 16 014, ISSN 1401-5764

Published digitally at the Department for Earth Sciences, Uppsala University, Uppsala, 2016.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Att veta hur mycket vatten som flödar fram i Sveriges vattendrag är väldigt användbar information för vårt samhälle. Den används för att planera samhället med hänsyn till ovanligt stora regn, som kan orsaka farliga översvämning och stora materiella skador. Ibland kan det råda brist på vatten, och då strider blandade intressen över tillgången. Den begränsade mängden vatten är en förutsättning för viktiga ekonomiska- och samhällsfunktioner, så som dricksvatten, industrier, vattenkraft, odling, offentliga estetiska intressen och inte minst miljömässiga. För att samsas om, och ibland skydda sig mot, denna värdefulla resurs, ligger ofta politiska beslut bakom vattnets fördelning till olika aktörer. En grundläggande förutsättning för en korrekt resursfördelning, är att veta hur stor tillgången på vatten är.

Ett naturligt sätt att få reda på hur stor vattenresurs en instans förfogar över, är att mäta hur mycket vatten som rinner i det berörda vattendraget. Men Sveriges vattendrag är så många, långa och föränderliga att detta skulle vara omöjligt. Därför har man byggt datoriserade modeller som med hjälp av regn, temperatur och kända markförhållanden räknar ut hur mycket vatten som borde flöda i alla Sveriges vattendrag. Dessa beräkningar måste naturligtvis stämma överens med verkligheten, och då gärna så nära som möjligt. För att kunna jämföra modell med verklighet, så mäter man kontinuerligt hur mycket vatten som flödar på noga utvalda platser, från norr till söder, från bäckar till älvar.

Denna rapport handlar om själva mätningen av vattenflödet, som rent tekniskt är en utmaning. Att mäta vattenflödet så noggrant som möjligt kräver avancerad utrustning, och att en utbildad person är på plats vid mätningen. Det som behövs är kontinuerlig mätdata, d.v.s. mätningar varje minut eller timme. Lösningen på problemet är att man mäter vattendragets vattennivå. Den översätts sedan till vattenflöde enligt en matematisk formel, en avbördningskurva, som är unik för varje vattendrag och kan förändras över tid. Hur väl denna avbördningskurva fungerar avgör därför hur noggrann den kontinuerliga mätningen vattenflödet blir, och i realiteten hur noggrann och säker information samhällets vattenanvändare har på sina tillämnningar.

Tidigare studier bekräftar att avbördningskurvan genererar stora osäkerheter i uppskattningen av vattenflödet. En ny förbättrad modell av avbördningskurvan har föreslagits av en grupp forskare. Den föreslår att många vattendrag borde ha flera avbördningskurvor för olika vattennivåsegment, eftersom det bl. a. motsvarar vattendragets geometri bättre då den t.ex. förändras då vattendraget svämmas över. Problemet är att lite data finns att tillgå när man bygger dessa matematiska avbördningskurvor, vilket gör att den totala osäkerheten kan bli sämre, speciellt om man använder sig av avbördningskurvor utanför det inmätta området för väldigt stora respektive små flöden.

Det huvudsakliga syftet med den här rapporten är att undersöka om användningen av två segment i avbördningskurvor ger mer eller mindre osäkerhet inom det inmätta området, respektive utanför det inmätta området. Metoden i rapporten använder sig av svensk mätningar av flöde och vattennivå och bygger därefter upp avbördningskurvor som är antingen med ett segment eller med två segment. Den modell som passar bäst med kontrolldata, är den som ger minst osäkerhet och är då att föredra.

Resultaten visar att den nya modellen med två segment tycks ge mindre osäkerhet, speciellt för låga flöden, då översättningsformeln används kring vattennivåer där flödet tidigare mätts av en hydrolog. Men om avbördningskurvan används kring vattennivåer där ingen har varit mätt flödet, t.ex. vid översvämning och låga flöden, då fungerar den gamla modellen med ett segment av avbördningskurvan.

För att få en ungefärlig uppskattning av hur stora osäkerheterna i avbördningskurvorna så har osäkerheterna uppskattats till 60 % för höga flöden, med utförda flödesmätningar, och 95 % för låga flöden. Om inga flödesmätning har gjorts vid vattennivån av intresse, men man ändå vill använda avbördningskurvan, så ligger osäkerheterna på 35 % för höga flöden, och 250 % för låga flöden. Detta är värdefull information för de som arbetar med vattenmängder i naturliga vattendrag.

GLOSSARY

Calibration	Setting model parameters so that the fit a data set, e.g. a regression so the model can be used
Extrapolation	Using a model outside the range of data that was used in calibration
Heteroscedasticity	Variation of variance
Interpolation	Using a model inside the range of data that was used in calibration
MRMSE	Mean Root Mean Square Error
Overparametrization	Using too many model parameters in a model, that the model becomes more uncertain
Projection Variable Method	A Regression method
Rating curve	A curve that explains the relation between stage and discharge
Regression	Fitting a mathematical function to a data set
RMSE	Root Mean Square Error
Segmented rating curve	Rating curve constructed by two or more equations
SMHI	Swedish Metrological and Hydrological Institute
Stage	Relative water level
Unsegmented rating curve	Rating curve constructed by one equation
Validation (model)	Controlling the model with independent data

Table of Content

1 Introduction	1
2 Background.....	2
2.1 Importance of discharge data.....	3
2.2 Quantifying discharge.....	4
2.3 Accuracy of rating curves.....	7
2.4 Segmented rating curves.....	8
2.5 Evaluating goodness of fit	10
2.6 Heteroscedasticity.....	11
2.7 Current practice of setting stage curves at SMHI.....	12
3 Method.....	14
3.1 Set-up.....	14
3.2 Monte Carlo simulation.....	17
3.3 Heteroscedasticity.....	18
4 Results	19
4.1 Calibration, interpolation and extrapolation.....	20
4.2 Slope ratio.....	21
4.3 Error distributions.....	22
4.4 Parameter distribution	23
4.5 Visual heteroscedasticity	24
5 Discussion.....	26
5.1 Unsegmented or segmented rating curves.....	26
5.2 Heteroscedasticity.....	28
6 Conclusions	29
7 References	30

1 Introduction

Water resource related problems are increasing as when the demand of water is growing but the water resources are unevenly distributed, both locally and globally (WMO, 2008). WMO (2008) advises fair allocation of resources and quality control should be in focus when executing water related management and economic investments. This will ensure a sustainable use of water and secure the living conditions for people and the environment.

In order to get an objective opinion of the water supply when closing water agreements, either domestically or transboundary, it is crucial to reach consensus of the limited water availability and potential depletions (WMO, 2008). Monitoring and understanding hydrological conditions is the fundamental knowledge in order to assure that regulations are followed and agreements fulfilled. Water related problems can then objectively be assessed and properly prevented. Providing society with necessary and accurate information is therefore an important field under continuous development by scientists and engineers.

This report will focus on the technical implication of accuracy and its importance when quantifying stream flow when acquiring data from gauging stations. A gauging station is a construction which is placed in a river or a stream where it records the discharge. Because of technical limitations, gauging stations usually only record the water level, which by hydrologist is referred to as stage, and later translated into discharge. The stage-discharge relationship is therefore a crucial step in discharge estimation. By measuring the discharge at different stages a relationship is conventionally set up with a regression curve, called a rating curve.

Development in precision and accuracy of rating curves are of great interest for hydrologist since errors will propagate in all further hydrologic data. One novel approach is to separate the commonly used power law rating in more than one segment, because it is believed rivers often have natural segmentation relationship in cross section geometry e.g. the transition into floodplain. However the effect of using segmented rating curves has not been thoroughly researched and authors requests more applied studies of rating curve segmentation, which what this report intends to do.

One technical issue with this approach is that discharge data often is limited, and performing regression on two separate curves instead of one can generate even more errors, often referred to as overfit. This report aims to compare calibration and extrapolation properties between two-segmented rating curves and unsegmented rating curves, primarily in the perspective of overfitting, based upon Swedish stage-discharge measurements. Furthermore the study aims to conclude if there is a difference on rating curve accuracy if segmentation is done on rating curves that shows a clear sign of segmentation.

There is an added benefit of the segmentation method comparison, as the error estimation on Swedish rating curves are of great value to the Swedish Meteorological and Hydrological Institute (SMHI) since no previous and comprehensive error estimations has been made on rating curves based errors on this data set. Because of this reason the estimated values as such are valuable and worth being highlighted.

Another aspect that is believed to have an effect on rating curve regression is the variation of variance (heteroscedasticity) in discharge measurements. This report also aims to quantify the occurrence of heteroscedasticity in Swedish gauging stations, and to investigate how segmentation will effect this property compared to unsegmented rating curves. Finally the report aims to qualitatively rise awareness and suggest a solution to heteroscedasticity implications on the procedure of a weighted point of no flow. That is a common practice that is used at the Swedish Meteorological and Hydrological Institute (SMHI) when constructing rating curves.

2 Background

The world's water demand has been estimated to have increased a 10-fold from 1900 to 2000 and the demand is now accounting for half of the world's freshwater (World Meteorological Organization (WMO), 2008). Even though the allocation of the water resources have changed over time, the largest proportion is still agricultural irrigation and food production with 62.6 %, followed by industry with 24.7 % of the water usage and the third largest share with 8.5 % is used for domestic purposes (Kjellén et al., 1997). The need for freshwater has been predicted to increase even further with population growth and development that will improve living standards for many (WMO, 2008).

Reliable hydrological information is required in order to have the ability to assess the issues above (WMO, 2008). WMO, (2008) states that quantification and quality control of surface and ground water are the most essential information for understanding and assessing water problems. Stream flow gauging and weather station networks are the fundament for quantification, but also for physical, chemical and ecological water quality variables. Hydrological data collection, measurement quality control and open source data storage are important steps, which ultimately leads to valuable and available information for agriculture, power production, industries and environmental management.

The information from discharge gauging stations can further be used in assessing countries water resources, planning and operating water projects, predicting environmental and socio-economic impacts, achieving regulatory requirements, increase security from hydrological extremes and in detail allocating water recourses within and outside a country (WMO, 2008).

2.1 Importance of discharge data

The water cycle naturally incorporates many uncontrollable processes and hence dangerous complications can arise. The most direct and hazardous are the hydrological extremes: floods and droughts. A flood is defined as a brief rise of water level which recedes at a slower rate (SMHI, 2014a). Because of large precipitation or snow melt, a land that otherwise is not submerged with water is called flooded at times of peak flow. Flooding is a natural phenomenon that can be necessary for ecosystems as they replenish wetlands with water. It is common for humans to live within floodplain areas because of rich soils, abundant water supply and other favorable economic conditions (Di Baldassarre et al., 2013). However there are consequences of living in floodplain areas such as an increased risk of destruction of both lives and property. After a flood event excess load water can linger for a long time in the flooded areas causing water related diseases such as Malaria, Dengue fever and Bilharzia infections (Hendriks, 2010). With detailed knowledge of flood events effective warning systems can be constructed and also allow planners to build environments with consideration of risk of flooding (WMO, 2008). Constructions can be designed to handle peak flow occurring within a set probability and time interval, commonly referred to as design floods (Haan, 2002).

Droughts are the opposite extreme of flooding and represents an unnatural low availability of water (WMO, 2008). Unavailable irrigation can lead to disastrous reduced food production, which can lead to severe malnutrition for the surrounding population. Low flows can also have a critical effect on water quality since the concentration of pollutants, such as waste water and harmful chemicals, gets higher when there is less dilution within a smaller waterbody. Waste water treatment plants and other industries can then not reach the limits of exceedance which can influence the drinking water quality and the ecosystems. Aquatic ecosystems are also sensitive to low flows as morphological obstructions can limit species migration (Hadwen and Cooperative Research Centre for Sustainable Tourism, 2005). Droughts can therefore be a bottle neck for aquatic biodiversity which is considered as a loss of cultural, esthetical, scientific and educational values.

Many factors can alter the hydrologic regime. Deforestation, urbanization and draining wetlands can decrease the infiltration capacity of rain and increase the runoff and surface water (WMO, 2008). This can cause faster and higher peaks of discharge and which further can increase the risk of flooding and droughts. Climate change is predicted to have an impact on the frequency of hydrologic extremes and should be considered and accounted for in water management.

Erosion and sediment transport are other issues to be considered. Erosion can degrade otherwise useful land areas and worsen the quality of soils (WMO, 2008). The subsequent sediment transport can have an impact on the physical and chemical water properties causing ecological changes. Eroded particles can settle in undesired places such as clogging diversions or accumulating in dams, the later increasing potential dam failure. Deforestation

and bare soils are strong factors to erosion risks, but also construction of infrastructures and settlements at sensitive locations.

Direct water pollution from waste water plants and industries but also indirect leaching, are a potential threats to declining water quality (WMO, 2008). Thresholds from polluters must therefore be regulated and controlled. This also accounts for deposition of air born pollutants can be hazardous for water and transported long distances involving many countries. Morphological changes are yet another possible risk to a sustainable water usage. Dam building can change the rhythm of the hydrological regime, as the water storage will reduce the natural hydrologic variation that are important factors for both aquatic ecosystems and the population nurturing from the water downstream.

2.2 Quantifying discharge

Intensive discharge measurements are often rare, mainly because they rely on expensive equipment and rivers are mostly located far away from hydrological offices. Discharge measurements are even more rare because of: (1) high velocities (2) danger for staff and equipment (3) operational timing issues and (4) destruction of gauging station (Petersen-Overleir and Reitan, 2005). A lengthy presentation of discharge measurements can be found in *section 2.7*. A gauging station is a construction which is placed in a river or a stream, where it records discharge by measuring the stage. The stage is the height of the water surface level relative to a local benchmark, Figure 1.

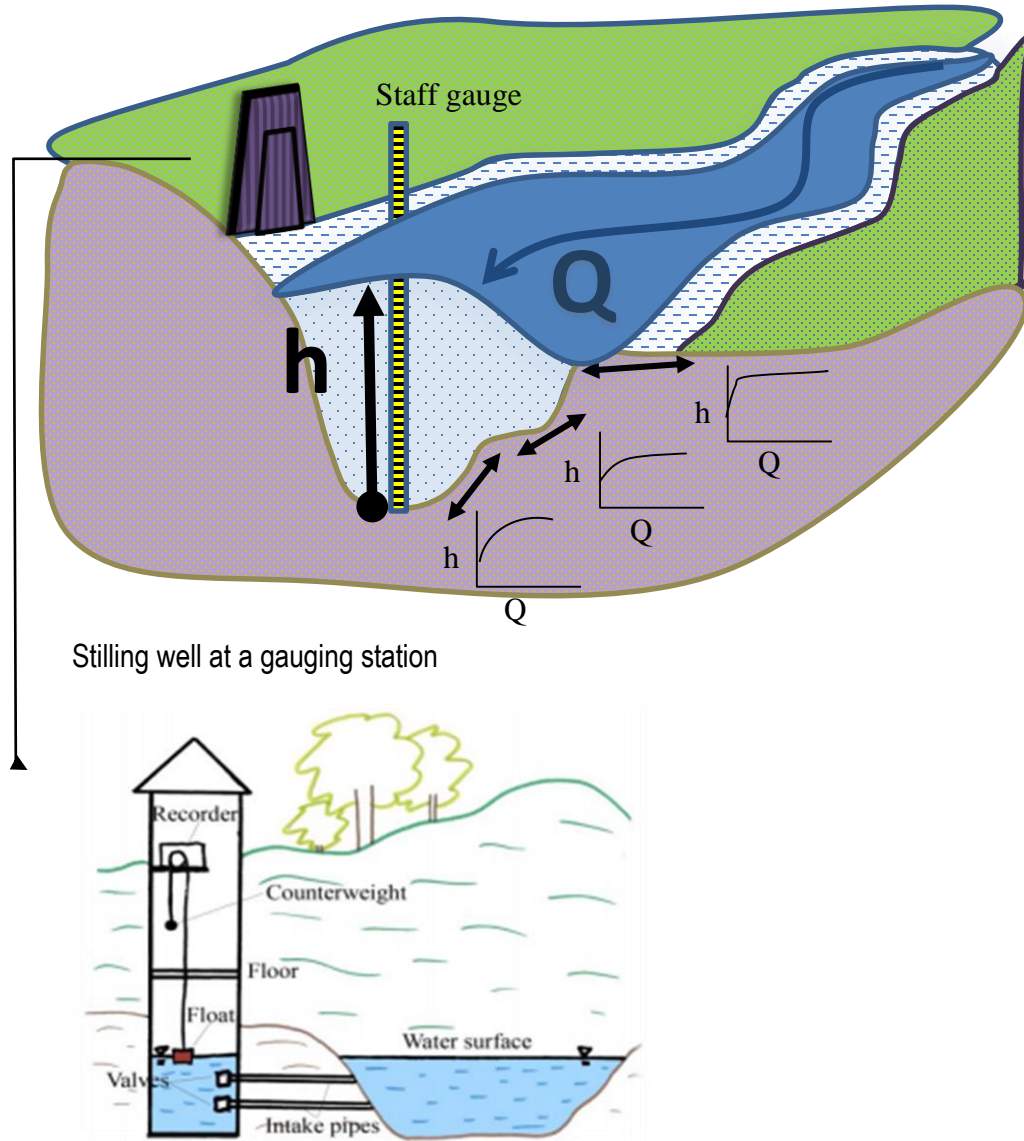


Figure 1. River and stream flow discharge recordings are based on information on discharge, Q , and stage, h , acquired at a gauging station (upper). Stage is continuously recorded at gauging station, often automatically within a stilling well (lower) but sometimes only with a manually recorded staff gauge (upper). The relationship can look differently at different parts of the cross section because of changes in geometry and friction, especially when entering a flood plain (upper right). Source (lower): (Enjebo (2014), reproduced with permission of the author).

Since the gauging stations generally only records the stage and not the discharge, it is crucial to know what the discharge is at a certain stage, the stage-discharge relationship. At controlled passages of streams, the stage-discharge relationship can sometimes be known from mathematical derivations at dams or weirs with known geometry, such as various weir formulas e.g. V-notch stage-discharge relationship, Eq. (1) (Hendriks, 2010). Q is the discharge, ϕ is the angle of the V-notch weir, h is the upstream water level above the V-notch and C is a constant accounting for friction losses.

$$Q = C \tan\left(\frac{\theta}{2}\right) h^{\frac{5}{2}} \quad (1)$$

At most other gauging stations, stage-discharge relationship have to be prepared manually, primarily by constructing rating curves (WMO, 2008). Rating curves are constructed with regression models from discrete stage and discharge measurements, Figure 2.

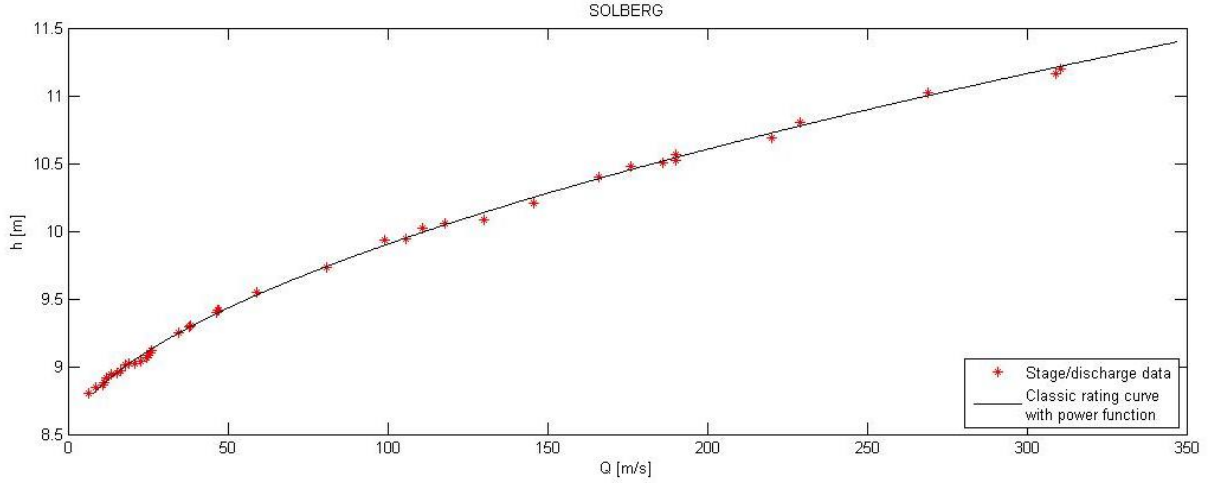


Figure 2. A rating curve constructed from stage and discharge measurements from a cross section at Solberg. The red stars are measurement data points and the black line is the rating curve.

Sweden and many other countries have rigorous ongoing field campaigns of continuous measurements in order to construct and maintain rating curves (WMO, 2008; SMHI, 2014b). When a rating curve has been constructed, it must be controlled and updated continuously because the river or stream section can change over time due of vegetation, constructions, debris, erosion and ice (Pappenberger et al., 2006).

There are several ways to record the stage, and the most basic one is to use a staff gauge, a graded scale placed in the water, which can be recorded manually (Hendriks, 2010). Developed countries often have continuous measurements of the stage at their gauging stations (WMO, 2008; e.g. SMHI, 2014b). The stage should be measured where downstream water, such as waves and dams does not influence the discharge or the stage across an assigned measurement section (SMHI, 2014b).

Well-functioning gauging stations often give most accurate discharge data available. That data is therefore used as a reference when calibrating models, e.g. HBV and S-HYPE used by SMHI to model stream flow and chemical transport for ungauged streams and rivers (Strömqvist et al., 2012). The modeled estimated flows are considered to have a significantly lower accuracy than the gauging stations. Therefore, wise positioning of gauging stations is important in order to cost effectively cover a hydrologic network comprehensively (WMO, 2008). WMO (2008) recommends the network to be evenly distributed and to include various climatic, topographic and geologic conditions in both small and large rivers and streams within the domain of interest.

2.3 Accuracy of rating curves

The basis for production of hydrological data is primarily computed from rating curves; the estimated one-to-one relationship between discharge and stage. Precision and accuracy of rating curves are therefore crucial, since errors will propagate in all further hydrologic analysis based on discharge information (Di Baldassarre et al., 2012; Petersen-Overleir and Reitan, 2005). Errors effecting discharge information originating from rating curves are often ignored, but many studies show that they should not be neglected, e.g. Pelletier (1988) and Di Baldassarre and Montanari (2009). With today's advancements in building computer based models to preform hydrological forecasting, the need for a better understanding of errors in hydrological data has been emphasized, in order to have a more correct model results that are used in decision making (Pappenberger et al., 2007).

Extreme flows particularly generate big errors, since hydrological applications require extrapolation of the rating curve far beyond the measurement range (Pappenberger et al., 2006). Despite the necessity of preforming such procedures, some authors recommend that extrapolation should not be done at all (Kuczera, 1996). In flood events, errors might generate practical uncertainties up to 30 % (Di Baldassarre and Montanari, 2009). Measurements in the high- and low-flow register are rare, e.g. only 12 % of gauging stations in France are calibrated for a 2-year flood, and 65 gauged stations in Australia are extrapolated on average 5.45 % of the time (Lang et al., 2010; Pena-Arancibia et al., 2015). The occurrence of extrapolation differs between countries depending on rating curve stability and measurement policies in gauging networks.

Many countries use the standardized power equation, Eq. (2), in natural rivers when constructing rating curves (ISO 1100-2, 1998; Lambie, 1978). The function is mathematically originating from Manning's formula, which is a physical explanation of open stream flow (Petersen-Overleir, 2009). An iterative non-linear least squares regression is commonly used to solve the power function of the rating curve (e.g. Petersen-Overleir and Reitan, 2005; Goltsis Nilsson, 2014). The parameter c is the stage of no flow while a and b are constants.

$$Q = a(h - c)^b \quad a_j, b_j, c_j > 0 \quad (2)$$

Eq. (2) is the standardized power function rating curve (ISO 1100-2, 1998). When using Eq. (2) as a rating curve some assumptions of idealized non-tidal steady conditions has to be made (Fenton, 2001; Petersen-Overleir and Reitan, 2005): (1) obey the Q-h relationship, (2) negligible control changes of cross section, (3) no hysteresis effect and (4) randomly distributed errors. When extrapolating rating curves with the power equation, predictions are either overestimated or underestimated, which means that errors increase with higher discharges (Di Baldassarre et al., 2012; Reitan and Petersen-Overleir, 2009).

2.4 Segmented rating curves

Novel approaches on rating curve construction have been developed in order to increase the accuracy of rating curves. One such approach is to separate the power law rating in more than one segment (Petersen-Overleir and Reitan, 2005; Reitan and Petersen-Overleir, 2009). More than two segments can sometimes be motivated as natural rivers can have sections where the Q-h relationship looks different, often due to changes in cross section geometry, Figure 1 (Petersen-Overleir and Reitan, 2005). A common and distinct change in the shape of the section is when the water is entering the flood plain (Hersch, 1999). It not only the geometry that changes, but also a possible change in friction due to e.g. vegetation (Lambie, 1978; Hersch, 1999).

The standardization document (ISO 1100-2, 1998) also acknowledges the extension of Eq. (2) to segmented rating curves where Eq. (3) is a two-segmented rating curve. More segments than two can be applied, but the scope of this study will restrict to one- and two-segmented rating curves, and will hereafter be referred to as unsegmented and segmented rating curves.

The additional h are the ranges in stage that the segmented curve applies for, h_{min} is the stage of no-flow, h_0 is the intersection of the segments and h_{max} is the upper limit of the rating curve.

$$Q = \begin{cases} a_1(h - c_1)^{b_1} & h_{min} \leq h \leq h_0 \\ a_2(h - c_2)^{b_2} & h_0 \leq h \leq h_{max} \end{cases}, \quad c_1 \leq h_{min}, \quad c_2 = h_0, \quad a_j, b_j, c_j > 0 \quad (3)$$

Eq. (3) is a two-segmented power law rating curve, and it is a more flexible option than an approximation with a single power curve. Two problems with the segmented rating curve set-up are the sudden repositioning of control at h_0 and that there are not idealized hydraulic conditions enough to motivate the power function at all (Petersen-Overleir and Reitan, 2005).

Sweden is a country that is currently using the concept of segmented rating curves in their network of gauging stations, and has applied the method since 1979 (Sjödén, 2009). In the Swedish gauging network, 22 % of the rating curves are unsegmented, 69 % are two-segmented and 9 % are three-segmented. SMHI is practicing the traditional way of constructing a segmented curve, which is to initially determine the number of segments and fit each segment separately (Goltsis Nilsson, 2014). The segment intersection h_0 is determined visually from a log-log-plot, Eq. (4), since each power curve function should be linear when transformed logarithmically, Figure 3 (Hersch, 1999).

$$\log(Q) = \log(a) + b \log(h + c) \quad (4)$$

Two segments are subsequently fitted to the two datasets within a limited segmentation interval, and then manually merged to get a continuous and as close to a smooth curve as possible.

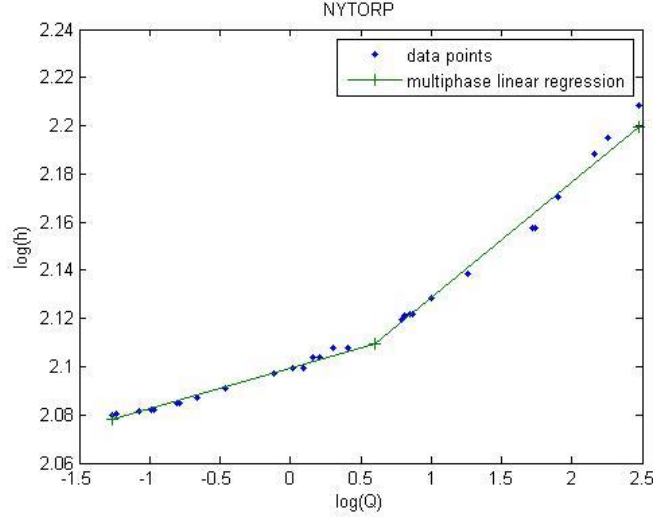


Figure 3. A segmented rating curve at Nytorp constructed from Eq. (3) on the log-log scale where each section (green line) is modelled with Eq. (4). The measurement data points (blue) are naturally aligned in two lines rather than one which is an indication of a real segmentation. The segmentation is done with multiphase linear regression and the middle (+) indicates the point of segmentation h_0 .

Petersen-Overleir and Reitan (2005) developed a non-linear regression method with the purpose of achieving an objective segmentation in order to minimize uncertainty in rating curves, hence avoiding the risk of errors in hydrology generated by human methodology as highlighted by Jónsson et al. (2002). Because of the challenges with multimodality in the likelihood surface when using non-linear regression method (Petersen-Overleir and Reitan, 2005), Petersen-Overleir and Reitan instead turned to a Bayesian approach, thereby also allowing the determination of the optimal number of segments (Reitan and Petersen-Overleir, 2009).

Regardless of what method that is used for constructing segmented rating curves, it nevertheless increase the number of parameters with three, for each additional segment, e.g. Eq. (3). Applying more parameters to a model that already uses sparse data amounts, puts the rating curve at risk of being over-parameterized and then subsequently overfitted. Overfitting occurs when a model starts to fit both the intended observations as well as the noise (Hawkins, 2004). In the case of rating curves, the noise is mainly the measurement uncertainty. Overfitting can generate even more uncertainty, especially when extrapolating rating curves (Di Baldassarre et al., 2012). Typically, an increased number of parameters leads to reduced model errors for the data used for calibration, but it can also deteriorate the predicting accuracy for other similar data (Jakeman and Hornberger, 1993; Beven, 1993). On the other hand, having too few model parameters can generate a prediction bias by not explaining the process enough. Setting the right number of model parameters can be regarded as a tradeoff function of bias and variance where the minimum represents the right number of parameters (Faber and Rajko, 2007). The general principle of the number of

model parameters is to not violate parsimony, which basically means the least number of necessary predictors should be used to explain the relationship (Hawkins, 2004).

Another advantage of few parameters is increased portability and practical use. Establishing a rating curve is very valuable for all fields in hydrology and should therefore be manageable by many. In the case of segmented rating curves, advanced numerical based computations or exclusive software are necessary (Petersen-Overleir and Reitan, 2005; Reitan and Petersen-Overleir, 2009) which might be a disadvantage.

The conventional procedure for model selection and simultaneously minimizing overfitting, supported by available data, is by using Aikakes Information Criterion (AIC) (Akaike, 1974) or Bayesian Information Criterion (BIC) (Schwartz, 1979). However, to significantly determining the existence of overfitting is generally done by model validation (Faber and Rajko, 2007), where the most common are statistical and residual validation analysis (Doherty and Hunt, 2009).

2.5 Evaluating goodness of fit

This practice is often done by comparing simulated values with observed values. Comparisons are often quantified in various ways by using the difference between simulated values with observed values e.g. residuals are the absolute differences between an estimated value and the data, Eq. (5). One commonly used estimate is the Root Mean Square Error (RMSE) Eq. (6). E is the residual error, \hat{Q} is the modelled discharge and Q_{val} is the measured discharge, the validation data. n is the number of validation measurements.

$$E = \hat{Q} - Q_{val} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{Q} - Q_{val})^2}{n}} \quad (6)$$

There are two different ways to validate a model, either externally by comparing model results with independent external data or internally where validation samples are taken from the same observations as used for model calibration. The latter approach is more resource effective if the data is scarce and needed for model calibration. Because of data dependent properties further approaches of extensive methods such as cross-validation, leverage correction, bootstrapping or Mallows' C_p are needed (Faber and Rajko, 2007). Although external validation is somewhat data wasteful, it is used in this study because it is a closer assessment of RMSE prediction and easier to apply.

The external validation idea is to divide the available data into one training set for calibrating the model, and into a validation set. How much of the available data that should be training data and validation data respectively has been debated. One suggestion to determine the required number of validation points is proposed to be dependent on the precision of the

measurement errors (Faber and Rajko, 2007). However, this varies greatly between applied external validation studies and therefore the widely used 30 %, discretely rounded upwards, will be assessed (recommended by Carlsson, (2014)). This ensures having enough validation data.

An important tool in avoiding overfitting is restricting the range or distribution of the parameters by prior information of realistic values of the parameter (Tikhonov, 1977). A study was done with the estimation of b (Eq. (2)-(4)), as the parameter mostly should be between 1.5-3.5, while $b > 4$ should be carefully used since an estimation error in the exponent can generate a very large over or under prediction in a flood event (Gawne and Simonovic, 1994). Lack of data or large measurement errors could alter the b parameter to be out of range, but it could also indicate that another number of rating curve segments is motivated. An analysis of the range of b can be used for evaluating bad rating curve performance (e.g. Petersen-Overleir and Reitan, 2005)

2.6 Heteroscedasticity

Another aspect of uncertainty in rating curves is heteroscedasticity, which refers to the heterogeneity in variance of residual errors. Routinely it is assumed that discharge measurement errors are normally distributed with the same variance along the rating curve (Petersen-Overleir and Reitan, 2005; Di Baldassarre et al., 2012). This has shown not to be the case as 29 out of 65 stream flow stations in southeast Australia shows signs of heteroscedasticity (Pena-Arancibia et al., 2015). The ISO 1100-2 (1998) standard of rating curves suggests non-linear least squares as the regression method to solve Eq. (1), but that method is particularly bad in accounting for heteroscedasticity (Petersen-Overleir, 2004). Generally it is well known by statisticians that heteroscedasticity can affect aspects in curve fitting and regressions. Solving Eq. (2) with non-linear least squares assumes that the uncertainty in discharge measurements are linearly proportional to the discharge, which has been shown by Petersen-Overleir, (2004) to generally be incorrect. The same article suggests that another rating curve model parameter can be added in order to account for the heteroscedasticity.

Heterogeneity in discharge measurements uncertainty has been shown to increase the risk of introducing large uncertainties in the extrapolation of rating curves (Westerberg et al., 2011). One substantial problem when solving the power function, Eq. (2), is that it is commonly done with non-linear regression, an optimization of a curve between residuals. As the residuals in rating curves naturally is much larger for high flow, different weight is put on the different measurements. It is suggested in the literature that there is a need to increase the knowledge of heteroscedasticity in rating curves in different regions of the world (Pena-Arancibia et al., 2015; Petersen-Overleir, 2004).

2.7 Current practice of setting stage curves at SMHI

SMHI has a long history of measuring flow in Swedish streams and rivers, with records available from the early 20th century (SMHI, 2014b).

In Sweden most gauging stations use float recorders inside of a stilling well, Figure 1. The water is let into the well by intake pipes, thereby measuring the water level with reduced oscillations and turbulence due to the stream flow and wind conditions (Hendriks, 2010). Hendriks (2010) further explains how the float with its counterweight is connected to a shaft encoder and how the data is stored with an automatic logger or on a paper chart that is manually changed and registered. Another common method to measure the stage is to mount one pressure sensor at the bottom of the stream and one at the surface. Accounting for the water density, the pressure difference of the total head (air and water) and the air pressure is the stage of the stream. Digital recordings can transmit the stage in real time through GSM or GPRS to faster predict discharges and hazards.

The gauging network is distributed throughout Sweden and consists of 230 stations continuously recording the stage. Regular measurement of the discharge is done approximately every second year in order to confirm or update the rating curve (Lennermark, 2015). All active Swedish stations are included in this study and the data comes from the measurements of stage and discharge that are currently used for constructing rating curves. Stage is manually recorded from reading the gauging staff compared to three on-land benchmarks while discharge has been measured with different methods depending on situation and what year it was done (Lennermark, 2015).

At SMHI the accuracy in stage measurement should not be greater than ± 3 mm accordingly to ISO 1100-2 (1998) but the practical threshold is ± 10 mm (Sjödin, 2009). McMillan et al., (2012) suggests that uncertainty is typically less than ± 10 mm but local water oscillations and clogging of stilling wells can add another ± 20 mm uncertainty (Vandermade, 1982; Dottori et al., 2009).

2.7.1 Discharge measurements with mechanical current meters

Mechanical current meters were predominantly used until the middle of the 1990s and those measurements are actively being used in today's rating curves and hence included this study (SMHI, 2014b; Goltsis Nilsson, 2014). They consists of a propeller and the measurement is done by counting the number of revolutions which is related to water velocity (Hendriks, 2010). A current meter measures velocity in a single point in the stream, and in order to have a good estimate of the total velocity profile in a stream many measurements must be done at various depths and along the whole cross section (Hendriks, 2010). The discharge is estimated by interpolating and extrapolating the velocity profile on a cross section area which is called the area-velocity method. A comprehensive literature review on uncertainty flow measurement with mechanical current meters indicates a discharge measurement uncertainty of 4-17 % (Pelletier, 1988; McMillan et al., 2012).

2.7.2 Discharge measurements with ADCP

The mechanical current meters have mostly been replaced with various Acoustic Doppler Current Profilers (ADCP) by SMHI during the last 20 years (SMHI, 2014b). ADCP uses sound waves and the Doppler effect to determine velocity through changes in wave frequency as the sound reflects in the water (Hendriks, 2010). Velocity profiles can either be measured with an ADCP from a boat or with a hand-held device called Acoustic Doppler Velocimeter (ADV) or FlowTracker (Lennermark, 2015). The Flow Tracker is used with the same methodology as a current meter and a study by McIntyre and Marshall (2008) shows that 93 % of the samples with a FlowTracker is within 20 % deviation from mechanical current meters.

Uncertainties in boat-held ADCP are often calculated from several measured transects and the difference from each transect is the relative error. Studies have shown errors of 3-5 % or 5-7 % when comparing several transects (Mueller, 2003). The measurement routine at SMHI ensures that no data points used for a rating curve measured with ADCP has a larger deviation than 10 %; mostly between the range 3-7 % (Lennermark, 2015). In order to decrease the variation of uncertainty in the routine, all field hydrologists follow the same homogenized measurement procedures (Lennermark and Nyman, 2014). The main effort when measuring discharge is according to Lennermark and Nyman (2014) to find a proper non-curved section with a smooth floor and laminar flow conditions. Lennermark (2015) means that the section should not be influenced by backwater effects. At least four transects should be measured and if the relative difference is larger than 5 % another four transects are made. If there is a varying water level during the measurement, several transects should be made for each water level. The moving bed test is made to ensure that the acoustic signals do not get affected by transportation of sediments in the riverbed. Adjustments are also made to adapt to the level of particles in the water, because too clear water can prevent effective backscattering and the signal does not reach the bottom floor. The ADCP is unable to reach all the way to the shore, so the river is extrapolated to the riverbank. Guidelines at SMHI recommend that only 5 %, maximum 10 %, of the total discharge should be extrapolated. Low velocities and representable conditions are preferred in shoreline extrapolation areas. Change in density and viscosity due to water temperature effects are double checked and the equipment is adjusted to the magnetic north compared to the real north. Other considered sources of uncertainty is the extrapolation of discharge above the ADCP-sensor, magnitude of velocity gradient and the velocity profile resolution.

The technology of ADCP is advancing rapidly on performance, accuracy and methodology meaning that all discharge measurements available for rating curve construction with ADCP are not sure to be entirely equivalent (Lennermark, 2015).

2.7.3 Discharge measurements with salt injection

A third, but less practiced method in Sweden, is salt injection (SMHI, 2014b; Hendriks, 2010). It can be applied in smaller and turbulent streams with the assumption of well mixed

conditions. A known volume and concentration of saltwater is mixed upstream and recorded with two conductivity meters. Time differences in the conductivity peak can be related to the stream velocity (Hendriks, 2010). Studies show uncertainties of 2.3 %-7.1 % in salt dilution discharge measurement in small turbulent streams (Hamilton and Moore, 2012).

2.7.4 Rating curve construction practice

At SMHI the measurements are stored in a database (WISKI) and from that rating curves are constructed, either with a weir equation e.g. Eq. (1) or most commonly with the power function, Eq. (2) (G. Nilsson, 2014). New measurements are subjectively compared to previous measurements and to the current rating curve in order to control if any anomalies such as backwater influences or changes in the section are visible. If several measurements from different measurement occasions indicates a new rating curve relationship, the rating curve will be updated with the new measurements included and older measurements are removed if needed. By visual log-log inspection on Figure 3, it is determined whether the rating curve needs more on segmentation Eq. (3) or more Eq. (3).

3 Method

The design of the method aims to show how the accuracy in rating curves is effected by the choice of segmentation, both in calibration, interpolation and extrapolation. In order to get an accurate understanding, the measurement uncertainties were accounted for. Many factors can effect non-linear least squares rating curve regression with Eq. (2) and Eq. (3). One such important factor that will be reviewed is heteroscedasticity. Also the generation on specific procedures at SMHI will be evaluated. The results aims to give a comprehensive and valuable picture of the uncertainty in Swedish rating curves.

3.1 Set-up

Unsegmented and segmented rating curves were constructed by using stage and discharge measurements from SMHI's gauging network. In six separate experiments, the upper segment, lower segment and unsegmented rating curves constructed with two different methods were evaluated in calibration, interpolation and extrapolation. Since the non-linear approach of solving the segmented segmentation curve was found unreliable (Petersen-Overleir and Reitan, 2005) the regression of each segmented section was preformed separately as unsegmented rating curves. This ignores the smoothness requirements but suits well for exploring extrapolation properties.

The method used non-backwater influenced stage and discharge measurements from SMHI. The measurements from each station were then divided into an upper segment and a lower segment. The segmentation was done by identifying the intersection, h_0 , from the optimal fit of two linear regressions on the logarithmic scale, Figure 3. The procedure is the same approach as the one used by SMHI, but instead of a visual estimation a multiphase linear

regression method was used (Hudson, 1966; Ganse, 2015). A difference in slope of the two sections on the logarithmic scale indicates of a segmentation in the stream.

Solving for unsegmented rating curves with non-linear regression from Eq. (2) can be numerically challenging, as non-linear regression solutions depend on iterative algorithms which can both be unstable or diverge to no or a wrong solution (Chapra, 2008). The choice of numerical solving methods could have affected the construction of rating curves, but since the objective of this study primarily is to evaluate segmentation and not numerical solving methods, only two numerical method were used.

3.1.1 Projection Variable Method

In order to increase the convergence of the regression, the Projection Variable Method was used to solve separable non-linear least squares problems such as Eq. (2) (Petersen-Overleir and Reitan, 2005). The linear parameters, in this case a , can be expressed as a function, g , of the non-linear parameters b and c , Eq. (7) (Golub and Pereyra, 2003). This reduces the number of parameters and increases convergence towards a global minimum, given proper initial guesses and boundaries of the non-linear parameters, Eq. (8). Parameter boundaries given in the iterative algorithms were $a > 0$, $b \in [1, 5]$ and $c > 0$. Due to unrealistic convergence of b to $b = 0$, $b = 1 \pm 0.001$ and $b = 5 \pm 0.001$ corresponding rating curves were removed.

$$a = g(b, c) \quad (7)$$

$$Q = g(b, c)(h - c)^b \quad a_j, b_j, c_j > 0 \quad (8)$$

3.1.2 Log Method

The Log Method is an intuitive and easy to use method. It uses the linear properties of log-log-transformed h - Q as in Figure 3. The intersection on the y-axis of a linear regression determined c , or the stage of no flow. An additional linear regression (Eq. (4)), was further used to determine a and b

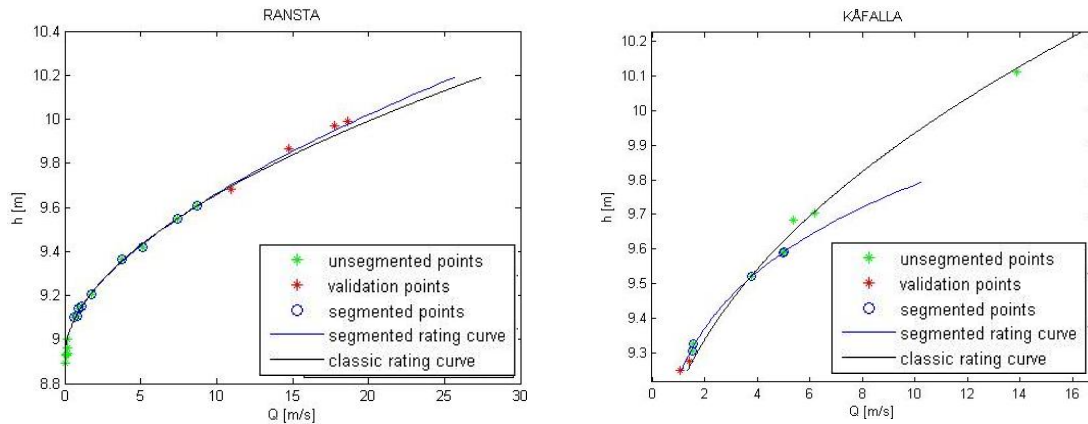


Figure 4. Extrapolation of segmented (blue lines) and unsegmented (black lines) rating curves in high flow (left) and low flow (right) register. The fit of rating curves are evaluated with validation measurement data (red stars) by MRMSE in the extrapolation register, Eq. (9). The rating curves was modeled with the Projection Variable Method, Eq. (8).

3.1.3 Error evaluation

Three forms of error evaluation was made: calibration, interpolations and extrapolation. From each gauging station one third (rounded upwards) of all available measurement data in the segmented rating curve was used as validation data. The rest of the data was used to construct the rating curve, Eq. (2). In interpolation, for upper and lower segments respectively, one third of the data was randomly extracted and used for validation data. For calibration, the same data as in interpolation was not extracted, but used for error measurements. For extrapolation, the most upper and lower third of the data was used as validation data.

All rivers and streams have a different range of discharge, and therefore the absolute flow errors was normalized by dividing with the mean flow of the validation measurements in order to be comparable. The evaluation was made by regression analysis, by calculating the discharge-normalized Root Mean Square Error (MRMSE), Eq. (9), an extension of RMSE, Eq. (6).

$$MRMSE = \sqrt{\frac{\sum_{t=1}^n (\bar{Q} - Q_{val})^2}{n * Q_{val}^2}} \quad (9)$$

The procedure allowed an estimation of the mean relative error for each rating curve, which can be viewed as average relative uncertainty in the range of validation data.

The combined evaluation from all gauging stations was visualized through boxplots. Boxplots are visual description of numerical data by quartiles (Alm and Britton, 2008). Top and bottom of a boxplot is the 1st and 3rd quartiles, meaning that 50 % of the numerical data is within the box. A band in the middle of the box represents the median. The vertical lines from the boxes are called whiskers and contains 1.5 of the interquartile range, which is the 3rd quartile subtracted with the 1st quartile.

Additional experiments were done by visually analyzing errors in relation to a set of factors (*below*), which possibly could have an impact on the accuracy of the rating curves. This was done in order to more deeply understand the behavior of extrapolation errors originating from either segmented or unsegmented rating curves, such as when there should be a segmentation or not.

The factors evaluated were:

- *Total number of measurements* used for constructing the rating curves. More measurements should decrease the impact of each measurement uncertainty and hence stabilize the rating curve.
- *Difference in slope* on the log-log scale between the upper and lower segment from Eq. (4). A large difference should indicate a more pronounced change in the section motivating a real segmentation.
- *Slope ratio* with the same argument as above but it is a better estimate because slope differences can neglect small slopes, Eq. (10).

$$\text{Slope ratio} = \frac{\text{lower slope}}{\text{upper slope}} \quad (10)$$

- *Magnitude of slopes*
- *Number of validation points*

3.2 Monte Carlo simulation

Many scientists within the field of hydrology have moved away from a deterministic approach of handling measurements and instead regard them with a probabilistic perspective. A discharge measurement should not be viewed as a rigid data point but as a manifestation of a distribution. Petersen-Øverleir (2004) suggests that a measurement is normal distributed, with a mean and a variance that depends on the measurement error, ε . Stage errors, ε_h , are commonly expressed in cm, Eq. (11) while discharge errors, ε_Q , are often expressed in percentage, Eq. (12). The magnitude of errors in the resampled data reflects realistic errors mentioned in the literature, *section 2.7.2-2.7.4*. The stage was simulated with 3 cm (the largest stage error mentioned) and the discharge with 1, 5 and 10 %.

The measurement uncertainties were assumed to be normal distributed with a 95 % confidence interval, therefore simulation with a certain error has half the original variance. The simulation equations are different, because errors in the stage is expressed in cm, while in the discharge it is expressed in percentage; Eq. (11) and Eq. (12), μ being the mean and σ the variance.

$$h = h + \varepsilon_h \quad \varepsilon_h \sim \mathcal{N}(\mu, \sigma) \quad (11)$$

$$Q = Q + Q \varepsilon_Q \quad \varepsilon_Q \sim \mathcal{N}(0, \sigma) \quad (12)$$

All rating curves was reconstructed 100 times with the original data resampled with induced errors from Eq. (11) and Eq. (12). The results was evaluated and validated, *previous section*. This way of resampling is called a Monte Carlo simulation and is a common method that accounts for the uncertainty in measurements.

By taking into account the uncertainty, one can evaluate how the measurement uncertainty, both in discharge and stage, propagates into extrapolation errors and the fit during calibration. By stepwise increasing the measurement uncertainty in Eq. (11) and Eq. (12), it was possible to estimate the uncertainty in rating curves and also to determine whether if segmented rating curves are overfitted or not.

If overfitting occurs, the extrapolation errors will grow faster for the segmented curve compared to the unsegmented rating curve, because an over-parametrized model starts fitting the noise rather than the real system, and is more likely to worsen the extrapolation properties (Faber and Rajko, 2007). A too simple model can instead generate bias by not explaining the reality thoroughly enough. Faber and Rajko (2007) argues that the number of parameters is a tradeoff between bias and overparametrization.

An important aspect of rating curve segmentation practice is to explore whether there is any difference of using unsegmented and segmented rating curves when there is a clear sign of segmentation. A sign of segmentation was quantified as slope ratios (defined in Eq. (10)) when applying multiphase regression on logarithmic data. Simulations were therefore done by extracting data with a clear sign of segmentation. The clear indication of segmentation value, the *slope ratio* was set to <0.20 , this because the *slope ratio* factor simulations in this study showed an even distribution of errors between 0 and 1, and that SMHI suggests 59 % of its rating curves should be segmented. This means that at least one third of the rating curves with the smallest *slope ratio* >0 ought to be segmented.

3.3 Heteroscedasticity

Heteroscedasticity residual analysis was performed visually on both segmented and unsegmented rating curves. Visual analysis was done by classifying the residuals in the same way as in Peña-Arancibia et al. (2015); where type A has a trumpet shape residuals indicating heteroscedasticity (also Fig. 1 in Petersen-Overleir, (2004)), type B has a non- to slight trumpet shape and type C means that it cannot be determined, Figure 5. Rating curves with six or less data points were excluded because the evaluation of heteroscedasticity were then visually difficult and would always be classed as C.

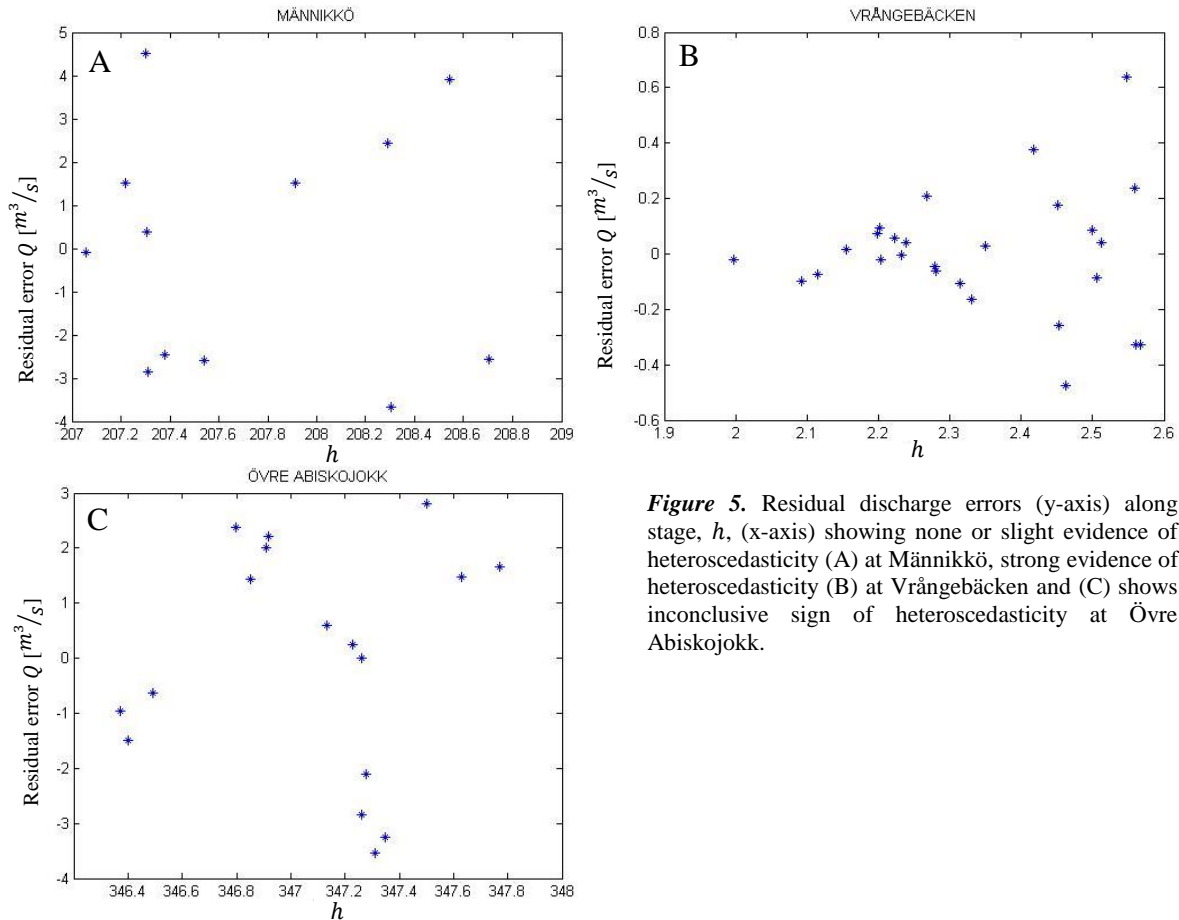


Figure 5. Residual discharge errors (y-axis) along stage, h , (x-axis) showing none or slight evidence of heteroscedasticity (A) at Männikkö, strong evidence of heteroscedasticity (B) at Vrångebäcken and (C) shows inconclusive sign of heteroscedasticity at Övre Abiskojokk.

At SMHI there were residual records of the relative uncertainty for each measurement in the rating curve.

$$E = \frac{\bar{Q} - Q}{Q} \quad (13)$$

There was a notable difference between the residuals from SMHI and the residuals from the generated curves with the Projection Variable Method. When heteroscedasticity occurs in rating curves, the variance was the largest at high flows, Figure 5(b), because residual are absolute errors and therefore residuals are larger for higher discharge values (Di Baldassarre et al., 2012). Large low flow residuals indicate a larger low flow measurement uncertainty, which is well known at SMHI (Lennermark, 2015).

The hypothesis of what is causing this difference was believed to arise from a weighted no-flow procedure that is used at SMHI. Sometimes it is possible to measure the cease to flow value in the field (Lennermark, 2015). This means it is known where the rating curve should intersect with the y-axis (stage), which is normally estimated from a regression. A forcing point, a 100-times weighted artificial measurement point, is then placed at the measured intersection. This could have a statistical impact on the regression, especially close to the forcing point. A bias is therefore induced into the rating curve and could have impacts on the accuracy on the whole curve. An investigation of phenomena and its effect on the residuals was done by adding a modelled forcing point with the similar weight of 100 measurements.

Here a suggestion is made of how to avoid the statistical interference of a forcing point, but still being able to set a cease to flow point. In Eq. (2), the parameter c is designed to be used as the point of no flow (only for unsegmented or lower segment rating curve) and that parameter was set to be the measured no flow point. Determining c beforehand is also beneficial in a numerical perspective since estimating a and b becomes much easier with a linear regression. This method also avoids complicated iterative non-linear regressions as conventional linear regressions can be performed with Eq. (4) with the benefit of easier handling.

The weighted no flow point practice was compared qualitatively by modelling rating curves 0.7 meters below estimated intercepts on previously constructed curves, both with the weighted forcing point and suggested method of pre-setting parameter c . Although construction of rating curves with the power law becomes a lot simpler, it can be misleading to determine a no-flow point at all, as this point can vary greatly with time and should therefore only to be used if the no-flow elevation is certain, such as at a fall or a weir.

3 Results

4.1 Calibration, interpolation and extrapolation

The most obvious result from analyzing the errors of rating curves, it is possible to observe how the biggest impact on the magnitude of errors is the uncertainty of discharge magnitude, Figure 6. Secondly, the error estimation is larger in for low flow compared, especially for low flow in extrapolation, Figure 6. Generally the Projection Variable Method constructs rating curves with less errors than the Log Method, with the exception to unsegmented rating curves in low flow, Figure 6 (bottom right). Generally segmented rating curves preforms with slightly less errors, compared to the unsegmented counterparts. This seems to not hold true for all high flows with the Projection Variable Method, or the low flow extrapolation with the Log Method. No obvious signs of overparametrization can be observed in the results, this can be observed by that the segmented rating curves were observed not to preform worse at high uncertainties, compared to unsegmented rating curves.

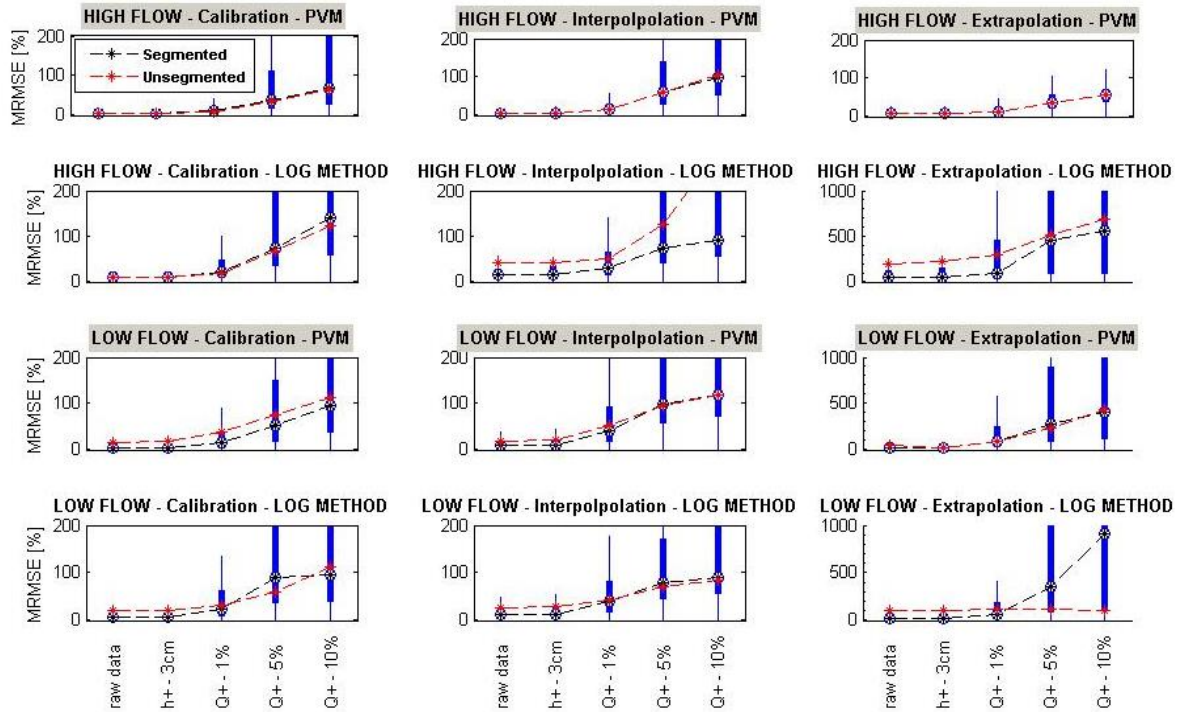


Figure 6. The median of the distributions (the distribution is in blue boxplots) of calibration, interpolation and extrapolation MRMSE in high and low flow, when constructing segmented (black) and unsegmented (red) rating curves with varying measurement uncertainty. The methods used were the Projection Variable Method Eq. (8) (denoted PVM and marked in gray) and linear regression on log-log-transformed h-Q data (denoted LOG METHOD). Y-axis shows the MRMSE, which indicates the amount of errors in the rating curves, and can be interpreted as percentual error. In order to thoroughly account for the impact of the measurement uncertainty a resampling was done 100 times with Eq. (11) and (12), a normal distribution with a stage measurement uncertainty of 3cm (denoted h + - 3cm) and a discharge measurement uncertainty of 1, 5 and 10 % (Q + - 1 %, Q + - 5 % and Q + - 10 %). No resampling without uncertainty is denoted with raw data.

Comparing the two different regression models, it is possible to observe that the Projection Variable Method generally generated less errors for high flow compared to the LOG Method, Figure 7. The Log Method generated less errors in the low flow, with exception of segmented rating curves in extrapolation.

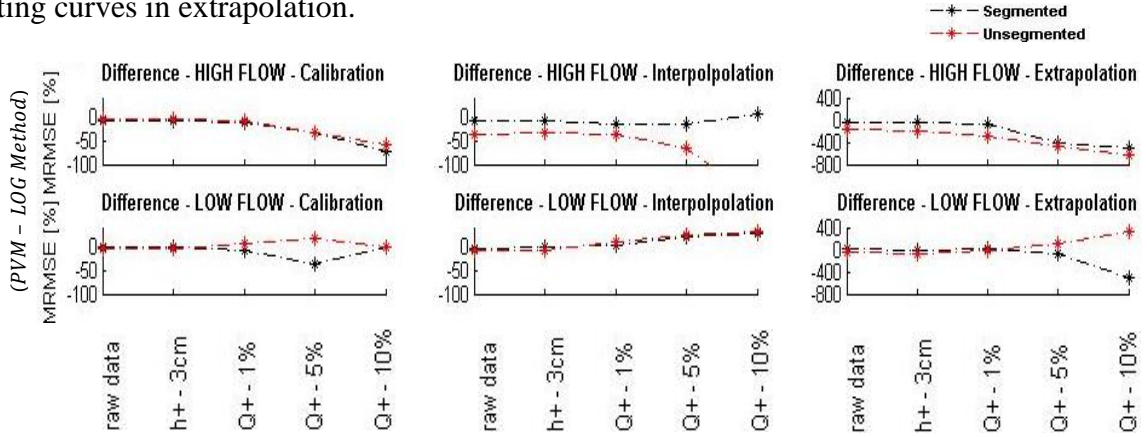


Figure 7. The error (MRMSE) difference between the Projection Variable Method and the LOG Method for calibration, interpolation and extrapolation for high and low flow (*PVM - LOG Method*). The procedure is done for both unsegmented (red) and segmented rating curves (black), meaning that negative values on y-axis indicates that *PVM* generates less error than the LOG Method.

4.2 Slope ratio

A small slope ratio, defined in Eq. (10), of h-Q data on the log-log-scale, Figure 3, indicates of clear visual segmentation and therefore motivated to use segmented rating curves. When the same analysis as in previous section was applied on rating curves with slope ratios < 0.2 , it is possible to observe that the impact was small in the high flow, Figure 8. Although rating curves with slope ratios < 0.2 generates more errors compared to the general result (Figure 7), segmentation generated less errors in calibration and for interpolation with small errors. However, when the errors increase in low flow interpolation and extrapolation the unsegmented rating curves generated less errors instead, Figure 8.

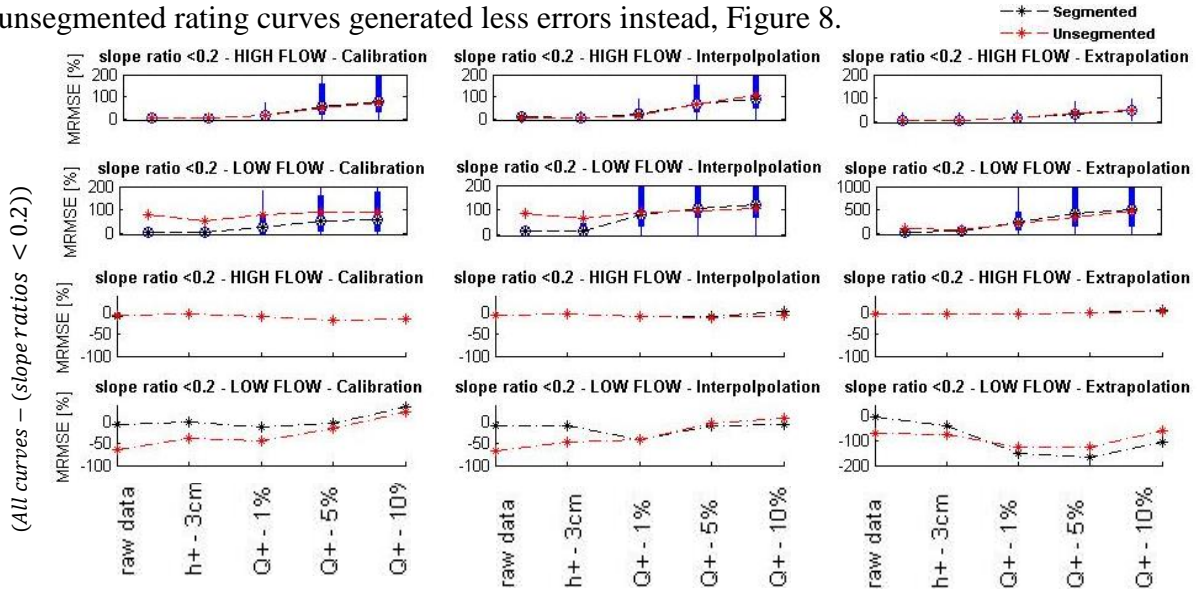


Figure 8. The median of the distributions (the distribution is in blue boxplots) of calibration, interpolation and extrapolation MRMSE in high and low flow, when constructing segmented (black) and unsegmented (red) rating curves with varying measurement uncertainty and slope ratios < 0.2 . The error (MRMSE) difference between rating curves with a slope ratios < 0.2 (a visually clear sign of segmentation) and the total set of rating curves.

4.3 Error distributions

No visual differences in the distribution of errors was found when comparing segmentation procedures. Parameters such as: *number of measurements*, *slope difference*, *slope magnitude*, *slope ratio* and *number of validation points* visually had the same effect on both sementation methods. However, the number of measurements had a noteworthy effect on the distribution of both rating curve methods, Figure 9. More measurements did not generate the smallest error, but seemingly less outliers and smaller variance.

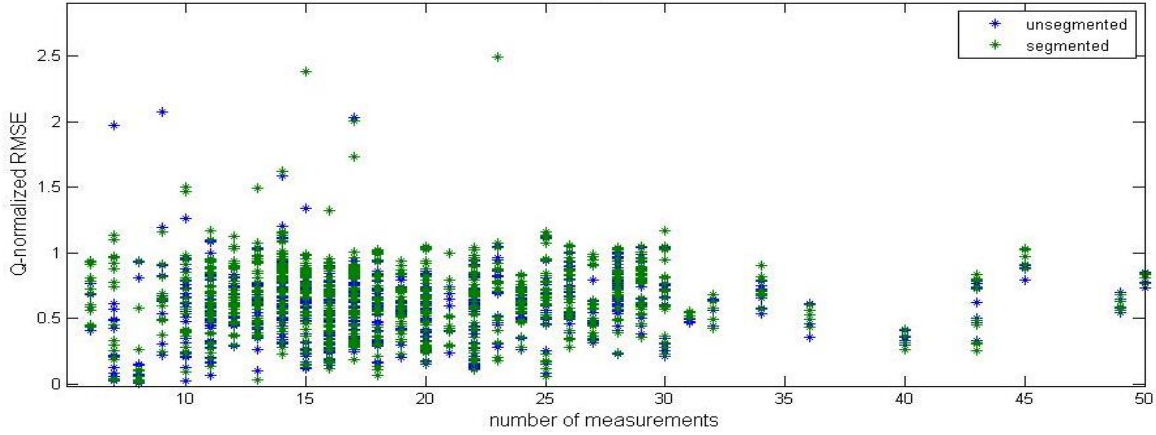


Figure 9. The distribution of the extrapolation MRMSE in high flow register depending on the number of measurements available when constructing segmented (green) and unsegmented (blue) rating curves with Eq. (8). A large RMSE indicates of large errors when predicting discharges outside of the measurement range of high flows. In order to clearer visualize the impact of the large number of parameters a resampling was done 5 times with Eq. (11) and (12) with a stage measurement uncertainty of 1cm and a discharge measurement uncertainty of 5 %. If the MRMSE is multiplied with 100, it can be interpreted as percent.

The *slope ratio* distribution showed that most of the slope ratios are in the range between 0 and 1, Figure 10. This means that the segmentation is mostly done so that the first segment has a smaller exponential increase of discharge with stage, such as in Figure 3. This scenario can be compared with the stream entering a flatter surface, e.g. a flood plain. When the *slope ratio* is larger than 1 it means that there is an exponential decrease in the upper segment, which can be interpreted as a contraction or a sudden increase of friction in the cross section. A negative value means that one of the slopes has a negative stage discharge relation. There could possibly be a trend of larger *slope ratio* up from 0 to 1 where a smaller slope ratio would give a smaller error.

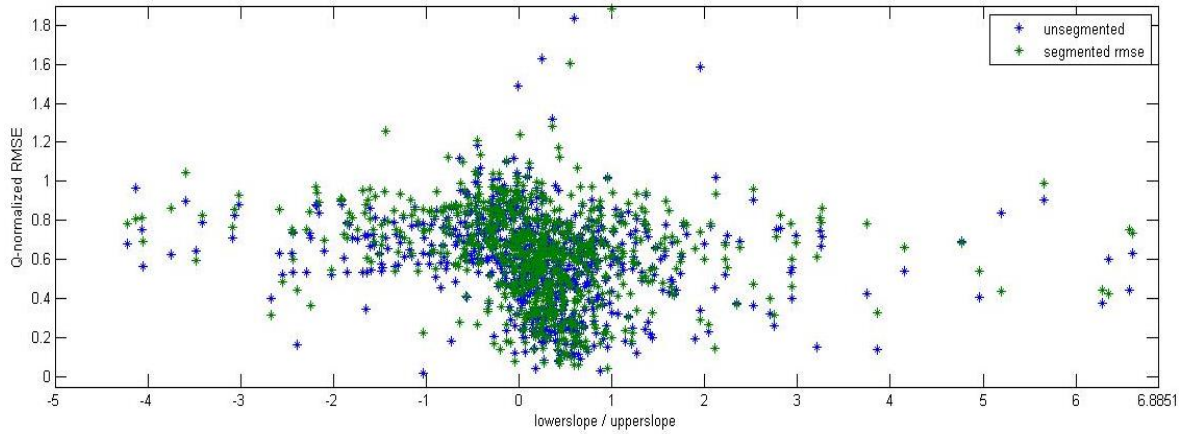


Figure 10. The distribution of the extrapolation MRMSE in high flow register depending on the slope ratio, Eq. (10), when constructing segmented (green) and unsegmented (blue) rating curves with Eq. (8). A large RMSE indicates of large errors when predicting discharges outside of the measurement range of high flows. In order to clearer visualize the impact of the large number of parameters a resampling was done 5 times with Eq. (11) and (12) with a stage measurement uncertainty of 1cm and a discharge measurement uncertainty error of 5 %. If the MRMSE is multiplied with 100, it can be interpreted as percent.

4.4 Parameter distribution

The model parameter control can be an important tool for evaluating the validity of a model, which in the case of power law rating curves can be done on exponent b in Eq. (2). The variance within the boxes of the boxplots contains the critical parameter range between 1 and 4 for low flow and high flow. This hold true for both the segmentation and the unsegmented rating curves and for calibration and extrapolation simulations. Segmented rating curves consistently have a smaller exponent than unsegmented rating curves and that the exponent increases with uncertainty, Figure 11. That feature is largest when extrapolating. The distributions are often skewed to a smaller parameter value.

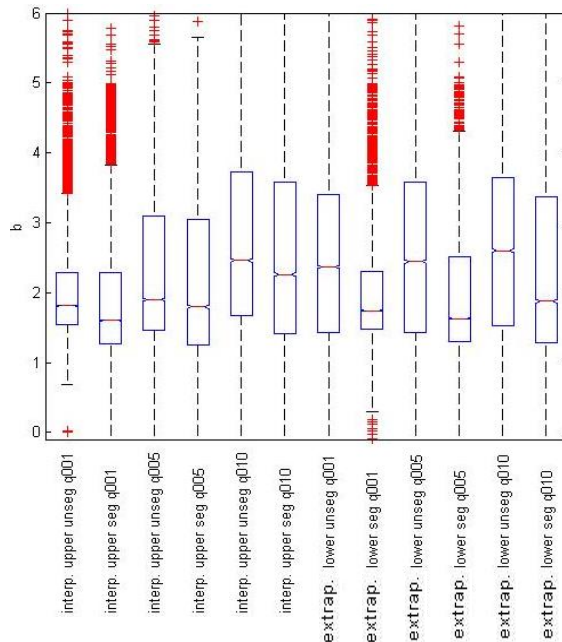


Figure 11. Distribution of the exponent parameter b in Eq. (2), (3) and (8) when modeling unsegmented and segmented rating curves in high flow register with Eq. (8). $b > 4$ is considered large and $b < 1$ unrealistic.. Trunkated at $b = 6$.

4.5 Visual heteroscedasticity

Evaluation of SMHI relative error residual plots shows visual evidence of heteroscedasticity on measurement uncertainty. The occurrence of heteroscedasticity is estimated to be 57 % at SMHI, Figure 11. Heteroscedasticity occurring due to discharge residual differences occurred in 59 % of the unsegmented rating curves constructed with the Projection Variable Method. By using segmented rating curves, the occurrence of heteroscedasticity shrinks to around 15 %, while the occurrence of ‘none or slight evidence’ of heteroscedasticity increases, Figure 12.

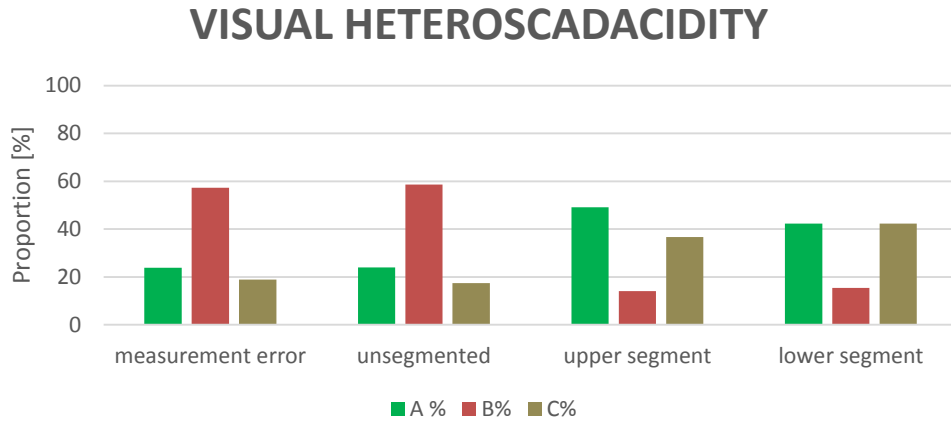


Figure 12. The impact of visual classification of procetual occurance of heteroscedasticity according to Figure 5. A is none or slight evidence of heteroscedasticity, B is strong evidence of heteroscedasticity and C shows inconclusive sign of heteroscedasticity. Those denoted Measurement errors are evaluated from measurement error plots relative to discharge, a heteroscedasticity arising from absolute uncertainties have larger relative uncertainties. Unsegment, upper segment and lower segment corresponds to the heteroscedasticity occurring after the non-linear regression by unsegmented and segmented rating curves been modeled with Eq. (8).

A representative example of consequences when using forcing points can be observed in the unsegmented rating curves at Brusafors gauging station, Figure 13. When simulating a fictional no-flow forcing point, Figure 13, the five subsequent low flow discharge measurements are underestimated in the regression. However when applying the pre-set no flow parameter c in Eq. (4), then it reduces to two underestimated discharge measurements with a smaller residual magnitude. It is worth noting that both procedures of changing in the low flow, affect how the rating curve looks in the high flow.

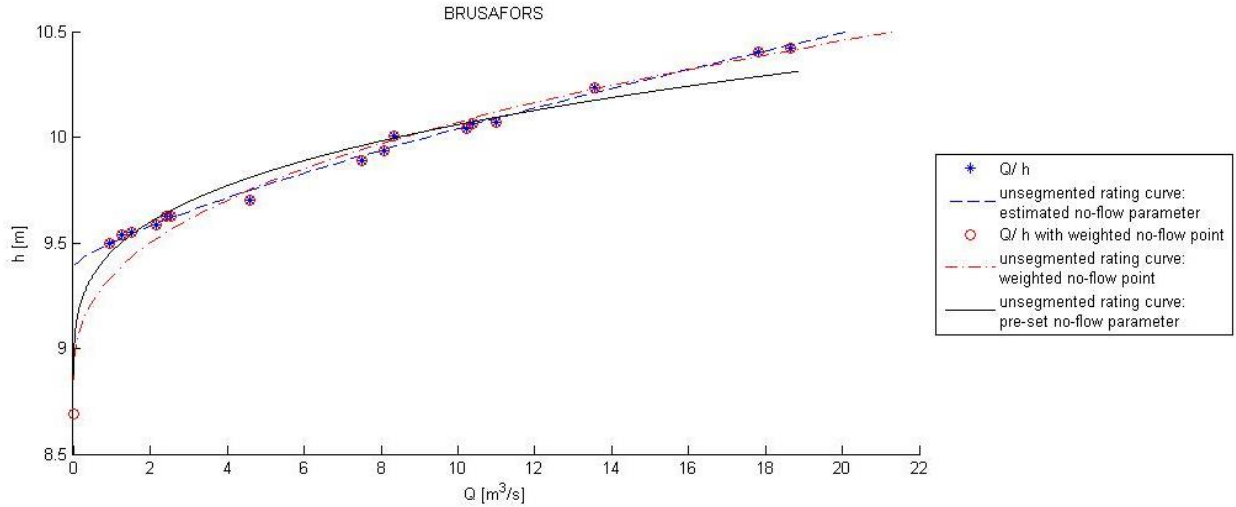


Figure 13. 3 rating curve constructed with three different methods by data from Brusafors (blue stars) with a “known” no-flow point at a stage of 8.7m. Q/h stand for discharge and stage measurement. The dashed curve (blue) is modelled with Eq. (8). The double dashed curve (red) is also modelled with Eq. (8) from Brusafors data but also with a weighted no-flow point at 8.7 (red circle) forcing the rating curve down to the no flow point. The whole line (black) rating curve is modelled with original Brusafors data (blue stars) with Eq. (12) and c is pre-set to the no-flow point 8.7.

Another no-flow consequence sample could clearly be visualized in a residual plot at Snapparp gauging station, Figure 14. The forcing point is changed the configuration of residuals by both increasing the residual variance, especially in the lower section of the rating curve. Not only did it increase the uncertainty, but it also generated a bias which was not there before.

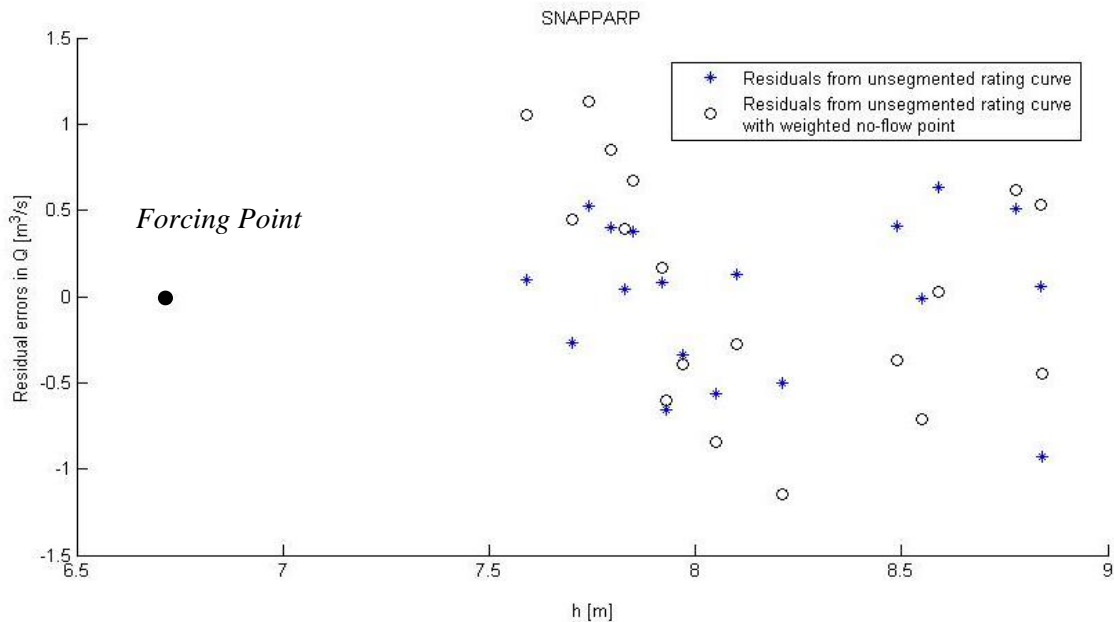


Figure 14. An example of rating curve residuals in Snapparp visualizing the impact when using the weighted point of no-flow (red circles) in contrast to the pre-set parameter c and modelled with Eq. (12) (blue stars).

5 Discussion

There were big differences of errors within the rating relationships, when the measurement uncertainties were included in rating curve accuracy, Figure 6. Unsegmented rating curve median error in calibration error of high flow was estimated to around 4 %, but when including an average discharge uncertainty of ± 5 % the median error from the rating curves were estimated to be close to 60 %, Figure 6. If the same measurement uncertainty was considered when extrapolating the error, it was smaller with errors of 36 %. The latter is a result with the same magnitude as previous studies of flooding uncertainties of 30 % (Di Baldassarre and Montanari, 2009). It is a surprising result, that extrapolation in high flow shows less errors than interpolation. The reason for this is probably that extrapolation is measured in at a higher discharge, and the larger the discharge, the smaller the relative errors becomes.

When considering low flow, the accuracy became even lower. In unsegmented calibrated low flow the median error was 41 %, Figure 7, but regarding the 5 % measurement uncertainty the median estimated error was 95% in interpolation and 250 % when extrapolating, Figure 9. These are considerably large errors, and therefore crucial that both hydrologist constructing rating curves and professionals using hydrological products (hydrographs and models) are well aware of this. Especially since a rating curve with a great fit, but unaccounted measurement uncertainty, can give an impression of low errors in the regression, when in reality it is not possible to know if the curve has been fitted to the true discharges.

5.1 Unsegmented or segmented rating curves

The suggested solution of reducing errors by segmenting the power law rating curve (Eq. (2)) into two parts (Petersen-Overleir and Reitan, 2005, Reitan and Petersen-Overleir, 2009), was evaluated on over 200 rating curves in Sweden. Each segment was modeled and evaluated separately, both with the Projection Variable Method and the Log Method. Determination of the point of segmentation was done with multiphase linear regression when the data are log-log-transformed. This approach makes the study of segmentation possible, but it ignores the requirement that the segments should be merged into a smooth curve. This requirement could possibly have an impact on the curves error properties, but was neglected.

The results showed that the Projection Variable Method generally constructs rating curves with the least amount of errors, for both segmented and unsegmented rating curves, Figure 7. The indication was strongest for high flow, and the Projection Variable Method preforms equally with the Log Method for low flows. The exception was that the Log Method generates the smallest errors in extrapolation for low flow unsegmented rating curves.

When it comes to the performance of segmentation with the Projection Variable Method, little difference could be observed in either calibration, interpolation or extrapolation, especially when comparing with the impact of discharge measurement uncertainty, Figure 6.

Looking at the small differences, segmented rating curves in the low flow seemed to be beneficial with the Projection Variable Method. For the Log Method, segmented rating curves showed clearly, that it performed with less errors compared to unsegmented.

When simulating cross sections at gauging stations that have a clear sign of natural segmentation (*slope ratio* < 0.2), rating curves were generally generating more errors than if they did not have a small slope ratio, Figure 8. In this case, segmentation had little impact on high flow, but had a notable impact on low flow. This were clearest at calibration and interpolation at small discharge uncertainties. In extrapolation and high discharge uncertainties ($< Q \pm 5$), unsegmented rating curves preformed with less error. This could be a sign of overparametrization and overfit, because the effect of overparametrization would increase the errors faster for segmented rating curves when increasing the measurement uncertainty. No sign of overparametrization could be observed in the results including all rating curves, Figure 6, but slightly for rating curves with slope ratios < 0.2 , and the conclusion is that segmented rating curves slightly suffer from overfit compared to unsegmented rating curves.

The second conclusion is that segmentation has a minor impact on errors segmented rating curves, and the biggest impact are primarily discharge measurement uncertainty followed by choice of regression method. The biggest, and maybe only, benefit with segmentation was for low flow interpolation with low discharge uncertainties. On the other hand, segmentation does not seem to generate more errors. It can be argued that the difference between the two segmentation methods are too small to make the extra and complicated work of constructing segmentations in rating curves.

Another factor that reduced the uncertainty, was found in the study to be the number of measurements at a gauging station, Figure 9. In more than 17 measurements, large error outliers seemed to be reduced, which means that the curve becomes more stable. In more than 30 measurements the error variance was reduced clearly, possibly due to the small sample size. The number of measurement that can be made is a budget question and this report can work as a foundation arguing for a more extensive budget for field hydrologists.

Within the simulation of the rating curves some unrealistic parameter estimations have occurred. An example when the b exponent in Eq. (2) sometimes happens to become < 1 and sometimes even < 0 , Figure 11. If b is < 1 it would suggest that the river was hyperbolic rather than parabolic, which is unlikely. When parameter $b < 0$ it means that the stage-discharge relationship is reversed, with less discharge at larger stages, which is physically impossible. Beacaus of the large sample size and multiple simulations, this should not have caused too many problems for median values.

5.2 Heteroscedasticity

The effect of heteroscedasticity when performing regression is problematic, as different weights are put on different measurement. Petersen-Overleir (2004) extension of the power function accounting for heteroscedasticity would be an interesting resumption in the future. This because of signs of heteroscedasticity occurred on as many as 59 % of the unsegmented rating curves, Figure 12, which can be compared to the occurrence of 45 % in a study in Australia (Pena-Arancibia et al., 2015). Heteroscedasticity was more uncommon in segmented rating curves with an occurrence of 14 % for upper segments and 15 % for lower segments. The lower occurrence of heteroscedasticity in the lower segment could be one explanation why segmented rating curves for low flow generates much less errors than the unsegmented.

Another source of heteroscedasticity is originating from the measurement uncertainty. The relative measurement uncertainty was as expected to be larger in low flow discharge measurement and the discharge data showed signs of heteroscedasticity in 57 % of the rating curves. The increased variance of low flow could be a counter weight to the high flow heteroscedasticity, or it could be needed to handle this variance separately.

What was peculiar in the SMHI rating curve discharge residual documents was the reoccurring one sided residual heteroscedasticity. In a well-functioning regression the data should be distributed equally on both sides of the regression curve. But sometimes in the low flow in SMHI residual plots of Eq. (13) the discharge data was often clustered above the rating curve. The hypothesis was that forcing points with the weight of 100 data points and no variance was added at a known point of no flow. This study shows that such procedure adds heteroscedasticity to the regression and heavily influences all parts of the regression curve, even in the high flow register, Figure 13. The residuals looked quite different adding heteroscedasticity with the forcing point, Figure 14. The suggested solution to add a point of no flow instead of adding a forcing point was to put the point of no flow directly in the c parameter which is the hydrological interpretation of that parameter. Although it is not recommended to directly estimate c unless the river floor is stable, such as with a concrete or stone floor, there are advantages. There is a reduced risk of overparametrization and simple linear regression can be applied when solving Eq. (2) by Eq. (4). The results showed that the method can reduce the problem of underestimation when adding a point of no flow. On the other hand, that method can generate a large change in the high flow register which this study did not evaluate.

6 Conclusions

- Errors from rating curves were large when accounting for measurement uncertainties, compared to the fit of the rating curve.
- With a mean discharge uncertainty of $\pm 5\%$, the errors from in high flow were around 60 % in interpolation and 36 % in extrapolation. For low flows, the interpolation errors were around 95 % and extrapolation error estimation is 250 %.
- Rating curve relative errors reduced with increased discharge
- Discharge measurement uncertainty had the largest impact on errors generated from rating curves.
- Stage measurement uncertainty was almost negligible.
- Choice of regression method had a considerable impact on rating curve performance.
- Segmentation of rating curves had little or no positive impact on rating curve performance in calibration, interpolation and extrapolation in high flow.
- Segmentation of rating curves had a slightly more positive impact on rating curve performance in calibration and interpolation, especially with a clear segmentation in a log-log-distribution or a small slope ratio.
- Segmentation of rating curves did not generate more errors.
- Segmented rating curves showed slight signs of overparametrization.
- Rating curves with a clear segmentation (small slope ratio) generated on average more errors, regardless segmentation method or regression method and should preferably not be chosen as a cross section.
- Heteroscedasticity occurred in 59 % of the unsegmented rating curves, while only 14-15 % in the segmented rating curves.
- The weighted point of no flow procedure at SMHI was shown to have, sometimes drastically, heteroscedasticity implications.

7 References

- Akaike, H., 1974. A new look at the statistical model identification, *IEEE Trans. Autom. Control* AC-19.
- Alm, S.E., Britton, T., 2008. *Stokastik: sannolikhetsteori och statistikteori med tillämpningar*. Liber, Stockholm.
- Beven, K., 1993. Prophecy, Reality and Uncertainty in Distributed Hydrological Modeling. *Adv. Water Resour.* 16, 41–51. doi:10.1016/0309-1708(93)90028-E
- Carlsson, B., 2014. Training data vs. Validation data.
- Chapra, S.C., 2008. *Applied numerical methods with MATLAB for engineers and scientists*. McGraw-Hill Higher Education, Boston.
- Di Baldassarre, G., Laio, F., Montanari, A., 2012. Effect of observation errors on the uncertainty of design floods. *Phys. Chem. Earth* 42-44, 85–90. doi:10.1016/j.pce.2011.05.001
- Di Baldassarre, G., Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. *Hydrol. Earth Syst. Sci.* 13, 913–921. doi:10.5194/hess-13-913-2009
- Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Salinas, J.L., Bloeschl, G., 2013. Socio-hydrology: conceptualising human-flood interactions. *Hydrol. Earth Syst. Sci.* 17, 3295–3303. doi:10.5194/hess-17-3295-2013
- Doherty, J., Hunt, R.J., 2009. Two statistics for evaluating parameter identifiability and error reduction. *J. Hydrol.* 366, 119–127. doi:10.1016/j.jhydrol.2008.12.018
- Dottori, F., Martina, M.L.V., Todini, E., 2009. A dynamic rating curve approach to indirect discharge measurement. *Hydrol. Earth Syst. Sci.* 13, 847–863.
- Enjebo, I., 2014. Inventory of hydrological measurements in Sweden (Master Thesis No. W14043), UPTec W, ISSN 1401-5765 ; 14043. Department of Earth Sciences, Uppsala University, Uppsala.
- Faber, N.M., Rajko, R., 2007. How to avoid over-fitting in multivariate calibration - The conventional validation approach and an alternative. *Anal. Chim. Acta* 595, 98–106. doi:10.1016/j.aca.2007.05.030
- Fenton, J.D., 2001. Rating curves: part 2 – representation and approximation., *Proceedings of the Conference on Hydraulics in Civil Engineering*. Institute of Engineer Australia, Hobart.
- Ganse, A.A., 2015. Multi-phase linear regr. [WWW Document]. URL <http://staff.washington.edu/aganse/mpregression/mpregression.html> (accessed 4.12.15).
- Gawne, K.D., Simonovic, S.P., 1994. A computer-based system for modelling the stage-discharge relationships in steady state conditions. *Hydrol. Sci. J.* 39, 487–506. doi:10.1080/02626669409492770
- Goltsis Nilsson, M., 2014.
- Golub, G., Pereyra, V., 2003. Separable nonlinear least squares: the variable projection method and its applications. *Inverse Probl.* 19, R1–R26. doi:10.1088/0266-5611/19/2/201
- Haan, C.T., 2002. *Statistical methods in hydrology*, 2nd ed. ed. Iowa State Press, Ames, Iowa.

- Hadwen, W.L., Cooperative Research Centre for Sustainable Tourism, 2005. Effects of tourism on Fraser Island's dune lakes. CRC for Sustainable Tourism, Gold Coast, Qld.
- Hamilton, A.S., Moore, R.D., 2012. Quantifying Uncertainty in Streamflow Records. *Can. Water Resour. J.* 37, 3–21. doi:10.4296/cwrj3701865
- Hawkins, D.M., 2004. The Problem of Overfitting. *J. Chem. Inf. Model.* 44, 1–12. doi:10.1021/ci0342472
- Hendriks, M.R., 2010. Introduction to physical hydrology. Oxford University Press, Oxford ; New York.
- Herschy, R.W. (Ed.), 1999. Hydrometry: principles and practices, 2nd ed. ed. John Wiley, New York.
- Hudson, D., 1966. Fitting Segmented Curves Whose Join Points Have to Be Estimated. *J. Am. Stat. Assoc.* 61, 1097–&. doi:10.2307/2283203
- ISO 1100-2, 1998. Measurement of liquid flow in open channels - Part 2: Determination of the stage-discharge relation.
- Jakeman, A., Hornberger, G., 1993. How Much Complexity Is Warranted in a Rainfall-Runoff Model. *Water Resour. Res.* 29, 2637–2649. doi:10.1029/93WR00877
- Jónsson, P., Petersen-Øverleir, A., Nilson, E., Edstöm, M., Iversen, H.L., Sirvio, H., 2002. Methodological and personal uncertainties in the establishment of rating curves, NHP Report No. 47. XXII Nordic Hydrological Conference, vol. 1. Killingtveit, Å.
- Kjellén, M., McGranahan, G., Stockholm Environment Institute, 1997. Urban water: towards health and sustainability. Stockholm Environment Institute, Stockholm, Sweden.
- Kuczera, G., 1996. Correlated rating curve error in flood frequency inference. *Water Resour. Res.* 32, 2119–2127. doi:10.1029/96WR00804
- Lambie, J.C., 1978. Measurement of flow–velocity–area methods, first ed. ed. Wiley, England.
- Lang, M., Pobanz, K., Renard, B., Renouf, E., Sauquet, E., 2010. Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis. *Hydrol. Sci. J.-J. Sci. Hydrol.* 55, 883–898. doi:10.1080/02626667.2010.504186
- Lennermark, M., 2015.
- Lennermark, M., Nyman, S., 2014. So Instruktion - Vattenföringsmätning med ADCP (No. 16617). Swedish Meteorological and Hydrological Institute.
- McIntyre, N., Marshall, M., 2008. Field verification of bed-mounted ADV meters. *Proc. Inst. Civ. Eng.-Water Manag.* 161, 199–206. doi:10.1680/wama.2008.161.4.199
- McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrol. Process.* 26, 4078–4111. doi:10.1002/hyp.9384
- Mueller, D.S., 2003. Field evaluation of boat-mounted acoustic doppler instruments used to measure streamflow. *IEEE*, pp. 30–34. doi:10.1109/CCM.2003.1194278
- Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R., Matgen, P., 2007. Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach. *J. Hydrol.* 333, 275–287. doi:10.1016/j.jhydrol.2006.08.017
- Pappenberger, F., Matgen, P., Beven, K.J., Henry, J.-B., Pfister, L., Fraipont de, P., 2006. Influence of uncertain boundary conditions and model structure on flood inundation

- predictions. *Adv. Water Resour.* 29, 1430–1449.
 Žšdoi:10.1016/j.advwatres.2005.11.012
- Pelletier, P., 1988. Uncertainties in the Single Determination of River Discharge - a Literature-Review. *Can. J. Civ. Eng.* 15, 834–850.
- Pena-Arancibia, J.L., Zhang, Y., Pagendam, D.E., Viney, N.R., Lerat, J., van Dijk, A.I.J.M., Vaze, J., Frost, A.J., 2015. Streamflow rating uncertainty: Characterisation and impacts on model calibration and performance. *Environ. Model. Softw.* 63, 32–44. doi:10.1016/j.envsoft.2014.09.011
- Petersen-Overleir, A., 2009. Avbördningskurvan ur ett hydrauliskt perspektiv.
- Petersen-Overleir, A., 2004. Accounting for heteroscedasticity in rating curve estimates. *J. Hydrol.* 292, 173–181. doi:10.1016/j.jhydrol.2003.12.024
- Petersen-Overleir, A., Reitan, T., 2005. Objective segmentation in compound rating curves. *J. Hydrol.* 311, 188–201. doi:10.1016/j.jhydrol.2005.01.016
- Petersen-Overleir, A., Reitan, T., 2005. Objective segmentation in compound rating curves. *J. Hydrol.* 311, 188–201. doi:10.1016/j.jhydrol.2005.01.016
- Reitan, T., Petersen-Overleir, A., 2009. Bayesian methods for estimating multi-segment discharge rating curves. *Stoch. Environ. Res. Risk Assess.* 23, 627–642. doi:10.1007/s00477-008-0248-0
- Schwartz, G., 1979. Estimating the dimension of a model (No. 6), *Ann. Statist.*
- Sjödin, N., 2009. Dagens ISO-standard för fältmätning av vattenstånd och vattenföring. SMHI, 2014a. Översvämningar [WWW Document]. SMHI. URL <http://www.smhi.se/kunskapsbanken/hydrologi/oversvamningar-1.5949> (accessed 4.10.15).
- SMHI, 2014b. SMHIs vattenföringsmätningar [WWW Document]. SMHI. URL <http://www.smhi.se/kunskapsbanken/hydrologi/smh-is-vattenforingsmatningar-1.80833> (accessed 4.10.15).
- Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., 2012. Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. *Hydrol. Sci. J.* 57, 229–247. doi:10.1080/02626667.2011.637497
- Tikhonov, A.N., 1977. Solutions of ill-posed problems, Scripta series in mathematics. Winston ; distributed solely by Halsted Press, Washington : New York.
- Vandermade, J., 1982. Determination of the Accuracy of Water Level Observations. *Hydrol. Sci. J.-J. Sci. Hydrol.* 27, 183–183.
- Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K.J., Halldin, S., 2011. Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrol. Process.* 25, 603–613. doi:10.1002/hyp.7848
- WMO (Ed.), 2008. Guide to hydrological practices, 6th ed. ed, WMO. WMO, Geneva, Switzerland.