

MASTER THESIS

Water balance in the Dutch river Rhine and
uncertainty of rating curves

Jetze J. Twijnstra
March 2020

UNIVERSITY OF TWENTE.

MSc. Thesis:

Water balance in the Dutch river Rhine and uncertainty of rating curves

by

Jetze J. Twijnstra

March 2020,

Lelystad

Contact:

jetze.twijnstra@gmail.com

Graduation Committee:

prof. dr. S.J.M.H. Hulscher

University of Twente

dr. J.J. Warmink

University of Twente

ir. M.R.A. Gensen

University of Twente

dr. F. Huthoff

University of Twente & HKV lijn in water

ir. G.L Horn

HKV lijn in water



UNIVERSITY OF TWENTE.

Preface

This report presents the thesis of my Master Civil Engineering and Management at the University of Twente. The thesis was carried out at HKV Lelystad and at the University of Twente, in collaboration with Rijkswaterstaat Oost-Nederland.

First, I would like to thank all members of the graduation committee for taking time and effort for supervision and support throughout this thesis. Second, I would like to personally thank Matthijs for always giving near immediate replies on questions via email and for the energetic feedback sessions in which usually hours flew by. Third, I would like to personally thank Geerten for giving clever insights from his background as hydrologist and meteorologist, and for providing datasets, information and QGIS skills. Fourth, I would like to personally thank Freek for keeping sight on the bigger picture of this thesis and for the hard work on the paper for the RiverFlow2020 conference. Fifth, I would like to personally thank Koen for getting me started with python and for the great speed course on Bayesian statistics. Sixth, I would like to personally thank Jord for giving very specific and detailed feedback. Seventh, I would like to personally thank Suzanne for cowriting the RiverFlow2020 and NCR days conference papers. Finally, I would like to thank all my colleagues at HKV for the great time at the office and for the fun group activities.

I hope you find this report informative and that you enjoy reading it.

Jetze J. Twijnstra

Lelystad, March 2020

Summary

Accurate rating curves are essential for a wide range of river management purposes, particularly as a basis for flood risk management. It is unknown how the Dutch river Rhine system behaves at extremely high discharges. To predict this behavior, hydrodynamic models are used which are calibrated and validated by rating curves. Rating curve accuracy is therefore important for the reliability of hydrodynamic model results. For this thesis, the three largest branches of the Dutch river Rhine and the bifurcation point Pannerdense Kop are considered. The rating curve locations Bovenrijn – Lobith (BR), Pannerdense Kop - Waal (WL) and Pannerdense Kop - Pannerdensch Kanaal (PK), are only 5 km apart without intermediate tributaries or significant water storage areas. Therefore, between the locations a nearly perfect water balance would be expected. Comparing the operational rating curves of 2018 for these locations shows that the water balance is not closing, up to 5% error. This gives a direct indication of the uncertainty in the rating curves. The objective of this thesis is to determine the uncertainty of single rating curves as related to flow measurement errors and to explore how water balance considerations can influence the uncertainty of rating curves.

Firstly, the uncertainty that is associated with single rating curves is quantified, using Bayesian inference and Markov chain Monte Carlo simulations, as based on homogenized measurement datasets. An indication of the obtained total rating curve uncertainty bandwidths is as follows: for local water levels of 12 m +NAP the total rating curve uncertainty bandwidths are $430.1 \text{ m}^3\text{s}^{-1}$, $364.5 \text{ m}^3\text{s}^{-1}$ and $199.8 \text{ m}^3\text{s}^{-1}$, for BR, WL and PK, respectively, and for local water levels of 16 m +NAP the total rating curve uncertainty bandwidths are $1018.6 \text{ m}^3\text{s}^{-1}$, $1113.7 \text{ m}^3\text{s}^{-1}$ and $727.9 \text{ m}^3\text{s}^{-1}$, for BR, WL and PK, respectively.

Secondly, the uncertainty that is associated with single rating curves, as related to flow measurement errors, have been quantified. Therefore, standard discharge measurements error values (3 and 5%) obtained from ISO reports and scientific literature were imposed in Bayesian inference and MCMC simulations. It is found that from all three branches, for equal measurement errors, the measurement uncertainty is most dominant in BR. However, there is a large difference in variation explained by measurement errors when assuming discharge measurement errors of 3% versus 5%, especially for BR. It remains unknown what the exact measurement error is per location. The measurement uncertainty is likely highest for WL, since more measurement uncertainty is introduced due to assumed floodplain flow in the relatively large and complex cross-section of the floodplain of WL.

Finally, this thesis proposes a method for establishing rating curves based on water balance closure. Currently, in scientific literature, no method can be found that considers the water balance for the establishment of rating curves. The method of today's practice only considers locally measured discharges for rating curve construction. The proposed method also includes discharge measurements from other locations to incorporate a closing water balance in the separate rating curves. Compared to the method used in today's practice, the proposed method clearly reduces the systematic error in water balance and thereby provides more consistent rating curves for the river network of the Dutch Rhine. Next, reducing the water balance error led to a decrease of the rating curve uncertainty bandwidth, for water levels above 16 m +NAP at WL, and to a slight increase of the rating curve uncertainty bandwidth for PK.

Since rating curves are essential in the construction of discharge time-series from water levels and in the calibration of river models, it is important that the systematic error in rating curves are removed as much as possible. Especially if these discharge time-series and calibrated models are used to define and hydraulically model design flood events. Hydrodynamic models treat the river system as a closed water balance. To prevent systematic under- or overestimation of discharges or water levels on entire river branches, there must be no systematic error in the water balance and water levels of rating curves. In the Netherlands, the design discharge for the Rhine river network is far beyond any event that has ever been observed. Therefore, it is of utmost importance that models used for development of flood management norms and regulations do not contain systematic effects that distort realistic system behavior.

Contents

Preface	iii
Summary	iv
1. Introduction	3
1.1. Background.....	3
1.2. Problem context.....	4
1.3. Objective and research questions.....	7
1.4. Outline	7
2. Methods	8
2.1. Description of available data.....	8
2.2. Data homogenization.....	8
2.2.1. Selecting data without weir effects	9
2.2.2. Detrending water levels for river bed subsidence	10
2.3. RQ1: Rating curve uncertainty.....	11
2.3.1. Rating curve model.....	11
2.3.2. Deterministic optimized rating curve	12
2.3.3. Bayesian rating curve	13
2.4. RQ2: Influence of measurement errors	15
2.4.1. Discharge measurement errors of helical Ott-mill instruments.....	16
2.4.2. Discharge measurement errors of ADCP instruments.....	16
2.4.3. Bayesian inference.....	17
2.5. RQ3: Water balance in rating curves	18
3. Results	20
3.1. Homogenized data	20
3.1.1. Selecting data without weir effects	20
3.1.2. Detrending water levels for river bed subsidence	20
3.2. RQ1: Rating curve uncertainty.....	24
3.2.1. Deterministic optimized rating curves.....	24
3.2.2. Bayesian rating curves	25
3.3. RQ2: Influence of measurement errors	26
3.4. RQ3: Water balance in rating curves	28
4. Discussion	31
4.1. Bayesian approach	31
4.2. Data homogenization.....	31
4.3. Measurement instrument	31
4.4. Assumed floodplain flow.....	31

4.5. Discharge distribution	32
4.6. Flood management norms	33
5. Conclusion and recommendations	34
5.1. Conclusion	34
5.1.1. RQ1: Rating curve uncertainty	34
5.1.2. RQ2: Influence of measurement errors	34
5.1.3. RQ3: Water balance in rating curves	34
5.2. Recommendations	35
6. Acknowledgements	36
References	37
Appendix A: Physical background rating curve model	39
Appendix B: Rating curve parameter uncertainty	40
Appendix C: Remaining predictive uncertainty	43

1. Introduction

1.1. Background

Discharge time series are essential for river water management. In the Netherlands, discharge time series are mainly derived from measured water level time series using rating curves (Quartel, et al., 2011). Measuring techniques for discharge time series are evolving, but the technique for measuring water level time series is more straight forward and therefore more frequently applied (Quartel, et al., 2011). For several locations along the Dutch Rhine, one rating curve is established. A rating curve describes the relationship between water level and discharge. Rating curves generally serve one purpose: translating continually measured water level time series into a discharge time series estimate, using the stage-discharge relationship. The discharge time series on the other hand, serve many purposes such as hydrodynamic model calibration, navigation, dredging maintenance, solving water quality issues, flood risk management and drought management (Quartel et al., 2011; Le Coz, 2012). In this thesis the focus lies on flood risk management.

A traditional rating curve is a single relation between water level and discharge for a certain location (Quartel, et al., 2011), and it provides the most common and most simple method for the determination of discharge time series in open channels (Rantz, 1982; Schmidt, 2002; Le Coz, 2012). The traditional method for obtaining single rating curves, is fitting a curve through paired measurements of stage and discharge (Kean & Smith, 2010). This traditional method has been used worldwide for over a century and is mainly developed by the US Geological Survey (Rantz, 1982) and based on ISO reports. The traditional rating curve implies that there are no temporal variations, which is a major shortcoming. Also, the traditional rating curve implies a steady open channel flow and a permanent control (reference hydraulic situation) (Le Coz, 2012).

In natural rivers like the Dutch river Rhine temporal variations occur, both in the channel flow and in the hydraulic control (Quartel, et al., 2011). The temporal variation can be:

- continuous variation (scour and fill sand-bed channel, variable backwater and changing discharge/hysteresis);
- seasonal variation (aquatic vegetation, ice cover and overflow and ponding); and
- permanent change (change to channel cross-section, human interventions and riverbed subsidence) (Hersch, 2009).

A Dutch method to incorporate temporal variations is the Qf-relationship. The method is developed in 1998 by Royal HaskoningDHV in collaboration with Rijkswaterstaat-RIZA and has been operational since 2004. This method corrects the derived discharge for the following four temporal variations: 1) hysteresis, 2) autonomous subsidence, 3) weir effects and 4) human interventions to the river. This solves some of the major shortcomings of a traditional rating curve.

Rating curves for the Dutch Rhine branches are available for different historical periods and have throughout the years been obtained due to various field measuring campaigns. A crucial issue for establishing rating curves is to accurately measure the cross-sectional flow velocity and surface area, to be able to calculate the discharge for a certain water level (Le Coz et al., 2014). Discharge measurements are therefore not directly measured but calculated, using cross-sectional flow velocity- and surface area measurements. However, they are referred to as 'discharge measurements'. Note that there is a strict distinction between 'measured' discharge and 'derived' discharge. The latter is derived from a rating curve, based on a measured water level for a specific location. According to Le Coz (2012) at least three types of errors need be considered when discharge measurements are to be used for the establishment and the evaluation of single rating curves:

1. errors in discharge measurements which are related to the measurement process itself. Combining instrumental errors, assumptions for floodplain flow, environmental errors, human errors and spatial integration errors;
2. errors in time integration to flow non-stationarity during the measurement (e.g., hysteresis); and
3. systematic errors due to deviation of hydraulic conditions from the reference flow regime.

Because of the demanding efforts for data collection and processing in setting up rating curves, they are often created only at relatively few key river locations. Along a single river reach, consecutive rating curve locations are commonly far apart, because it is assumed that in a single channel the intermediate discharge-variations are small. An exception to this is when tributaries or bifurcations are present in the river. Rating curves for different branches are then required to fully describe the distribution of water over the branches. Together, these rating curves in a network are expected to give a closing water balance, where water input from upstream is traceable to downstream, apart from intermediate minor additional runoff, inflows, or stored or extracted water quantities.

The discharge distributions at the major bifurcation points of the Dutch Rhine system, and thereby the amount of discharge per branch, have a dominant influence on the water levels along the downstream branches. Therefore, an accurate estimate of the discharge distribution and discharges per branch are essential for flood risk management. To obtain accurate discharge estimates from rating curves, a good rating curve accuracy and mass balance accuracy are needed (e.g., Vervoorn, 1998; HKV, 2009). However, in practice it is not possible to maximize the rating curve accuracy and the water balance accuracy simultaneously (Vervoorn, 1998), see Figure 1. This is due to errors in the measurements, for example resulting from assumption for floodplain flow. A trade-off has to be made between rating curve accuracy and water balance accuracy (Vervoorn, 1998). For example, HKV (2009) only considers water balance accuracy in the validation of the rating curve models, to determine whether they are reliable and accurate. Furthermore, the best method for assessing rating curve uncertainty is an important and open scientific issue (Le Coz et al., 2014).

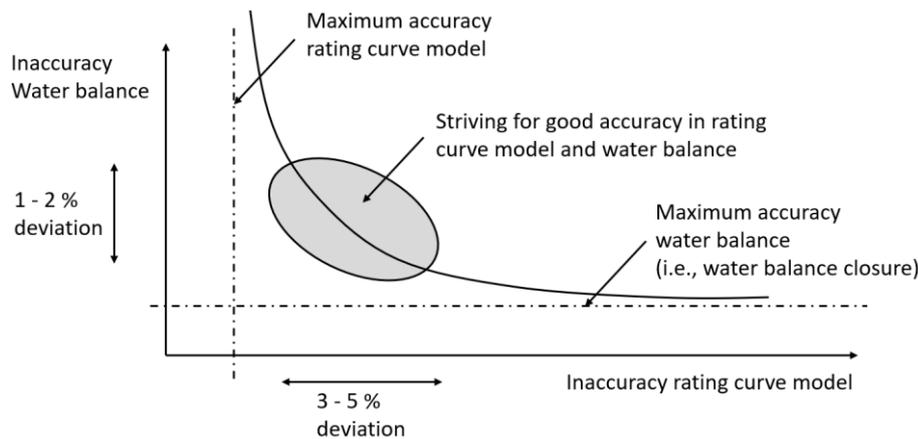


Figure 1. Visual indication of the relationship between rating curve accuracy and mass balance accuracy. Adapted from Vervoorn (1998).

1.2. Problem context

Under the new Dutch flood risk framework, quantifications of all uncertainties are required and should be calculated if possible. Currently, measurement errors (errors related to the measurement process itself) have not yet been quantified as a source of rating curve uncertainty. Errors in discharge measurements translate to errors in single rating curves, and in turn to errors in discharge time series estimates, see Figure 2. Hydrodynamic models are calibrated and validated by discharge time series estimates. Since it is unknown how the river system behaves at extreme upstream discharges, branch discharge estimates at extreme discharges are derived from hydrodynamic model studies and rating curve extrapolations. In this context, an accurate set up of rating curves becomes of high importance for the reliability of hydrodynamic model results (Pappenberger et al., 2006; Herschy, 2009), and thus flood risk management.

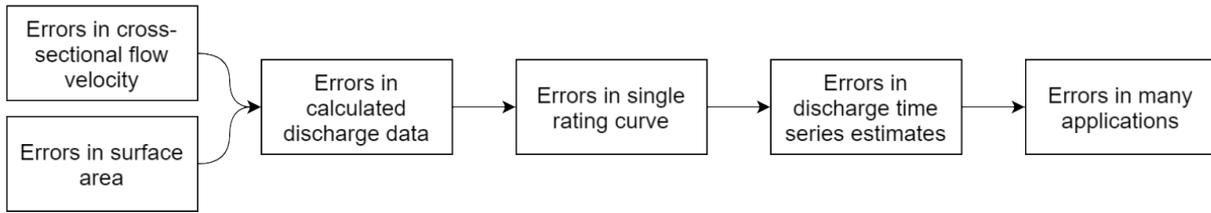


Figure 2. Schematization of the effect of measurement errors.

Next, in the Netherlands, a great opportunity presents itself to check the consistency and accuracy of rating curves at river stations around the bifurcation near the Pannerdense Kop, see Figure 3. This location in the river Rhine is of great significance for water management in the Netherlands. Amongst others, the discharge split at the Pannerdense Kop affects downstream flood risk, and it is actively controlled to aid river navigation. Therefore, the discharge split at the Pannerdense Kop is frequently measured at the following locations:

- Bovenrijn – Lobith (BR);
- Pannerdense Kop - Waal (WL); and
- Pannerdense Kop - Pannerdensch Kanaal (PK).

BR is located just upstream of the bifurcation point and WL and PK are located downstream the Pannerdense Kop bifurcation, see inset in Figure 3. The locations are only 5 km apart without intermediate tributaries or significant water storage areas. Therefore, between the stations a nearly perfect water balance would be expected. A small lake named 'De Bijland' is connected with the Bovenrijn between the locations. It is assumed that this lake has a negligible effect on the water balance. Water balance closure of the operational rating curves is important for flood management. Discharge time series are derived from operational rating curves, using observed water levels. Hydrodynamic models are calibrated and validated based on these discharge time series estimates and observed water levels. However, hydrodynamic models treat the whole river system as a closed water balance, while the discharge time series used in their calibration and validation, are not according to a closed water balance. This causes discrepancies between actual discharge time series and derived discharge time series from the rating curves.

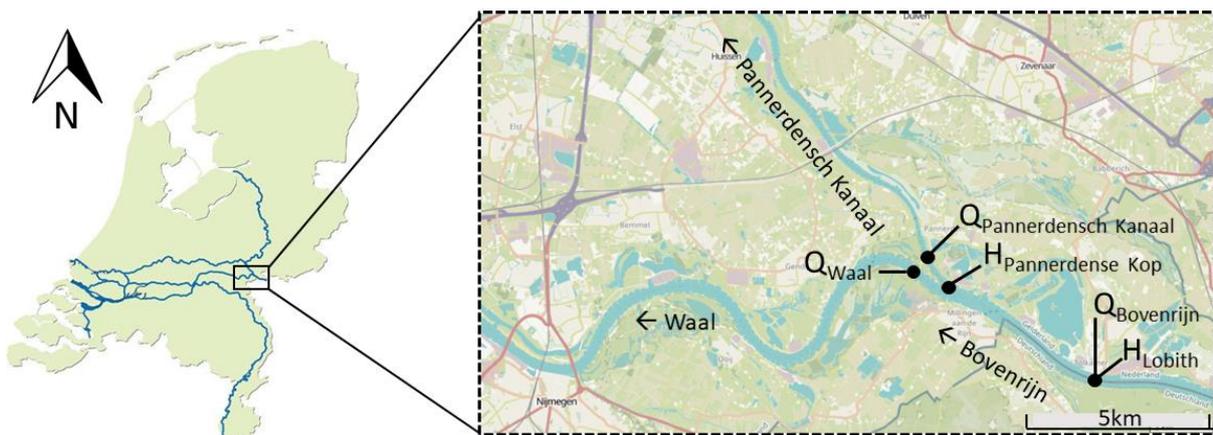


Figure 3. The river Rhine in the Netherlands near the bifurcation point at Pannerdense Kop. Also shown are the three rating curve locations. Adapted from Steenblik et al. (2020).

Figure 4 shows the calculated error in the water balance at this bifurcation point, as derived from operational rating curves at three locations, and from actual ADCP measurements that were carried out in 2018. This error in the water balance, which goes up to 5%, is a direct indication of the uncertainty of the respective rating curves. Only data for water levels above 10 m +NAP at Lobith

were considered, which is the regime of a freely flowing river where no active weir operating is taking place (Reeze et al., 2017). This threshold of 10 m +NAP at Lobith roughly corresponds with an upstream discharge of 2500 m³s⁻¹. The discharge data derived from the operational rating curves has a frequency of 10 minutes is open source (<https://waterinfo.rws.nl/>). For the ADCP measurement data, only same-day measurements (measurements with at least one measurement in each branch occurring at the same day) have been used. The positive systematic error reflects a higher upstream water volume, implying that in the bifurcating system water is lost.

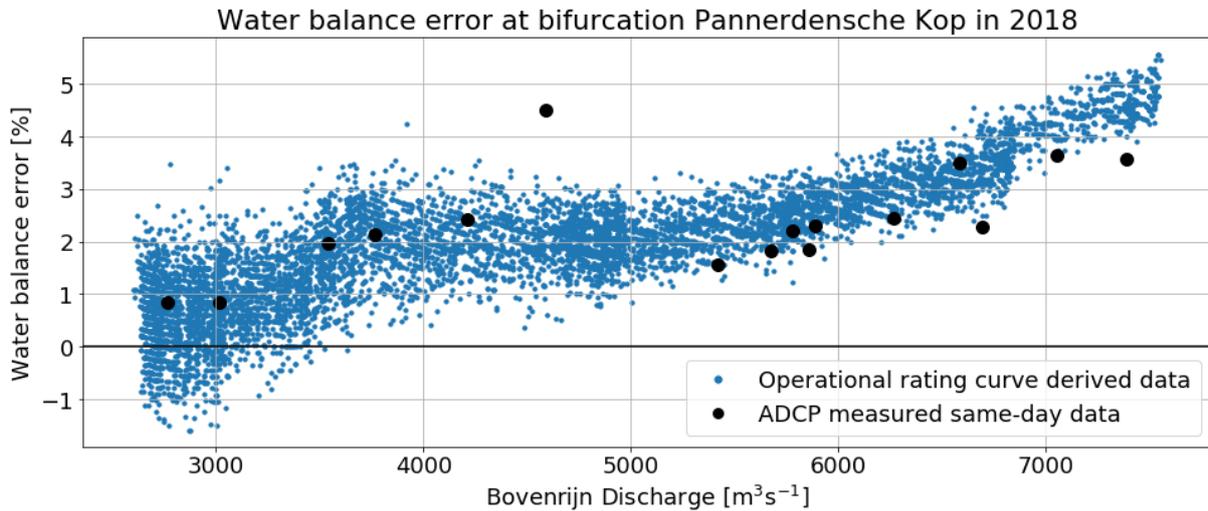


Figure 4. Water balance error at bifurcation Pannerdense Kop for discharge data derived from operational rating curves (source: <https://waterinfo.rws.nl/>) and for same-day ADCP flow measurements. The error is calculated as: $(Q_{\text{upstream}} - Q_{\text{downstream}}) / Q_{\text{downstream}}$. The figure presents the discharge domain without weir effects for year 2018.

The cross-sectional geometries of the three locations, obtained from Baseline 5.3.3. (2018), are shown in Figure 5. The wetted areas are representative for the 1995 flood, which is the highest water level in the data set. In Figure 5 it can be seen that at high discharges the floodplains contain flow. The floodplains are not accessible by measuring boats and therefore assumptions are made for floodplain flow. An assumed floodplain flow can result in errors in the cross-sectional averaged discharge measurement. This can lead to a water balance error of the rating curves and the measured discharges (see water balance error in ADCP data in Figure 4). In Figure 5 it can be seen that the floodplain of the Waal is large compared to the Bovenrijn and the Pannerdensch Kanaal. Making an assumption for floodplain flow is therefore most difficult in the Waal. The Waal could therefore be (most) responsible for the water balance error (Figure 4).

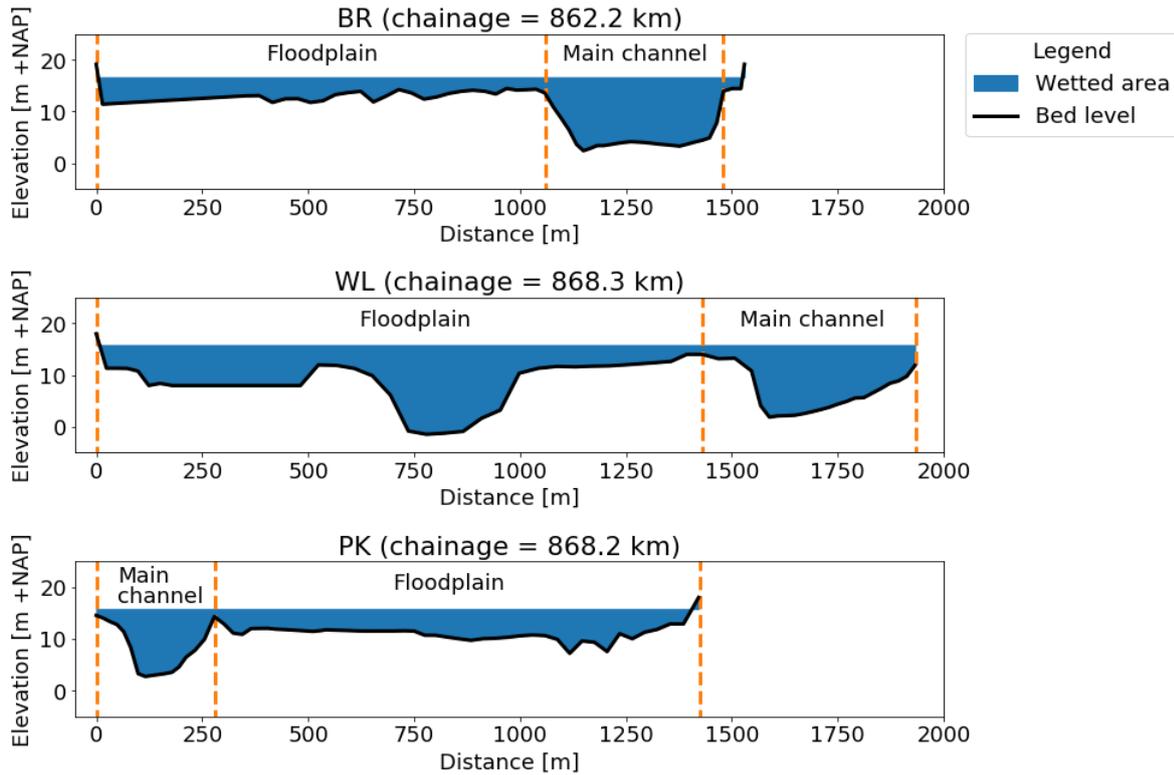


Figure 5. The wetted areas are representative for the 1995 flood, which is the highest water level in the data set (source: Baseline 5.3.3., 2018).

1.3. Objective and research questions

The objective of this thesis is to determine the uncertainty of single rating curves as related to flow measurement errors and to explore how considering the water balance in rating curve construction influences the rating curves uncertainty. The three largest Dutch Rhine branches (Bovenrijn, Waal and Pannerdensch Kanaal) and their bifurcation point (Pannerdense Kop) are studied. To meet the objective, the following research questions are addressed:

1. How large are the uncertainty bands of the single rating curves for the three branches?
2. To what extent is the uncertainty of single rating curves for the three branches induced by discharge measurement errors?
3. Can single rating curves be established based on water balance closure, and if so, how does this influence the uncertainty bands of the single rating curves for the three branches?

1.4. Outline

Following the outline of the research questions, Chapter 2 describes the methods and Chapter 3 describes the results. The methods and results are discussed Chapter 4 and finally the conclusions and recommendations presented in Chapter 5.

2. Methods

2.1. Description of available data

We use a dataset containing all available discharge measurements over a period from 1988 to 2018. The measurements have been collected at Lobith (Bovenrijn) and at the bifurcation Pannerdense Kop (Waal and Pannerdensch Kanaal), at each location containing 1303, 1202 and 1520 measurements, respectively. Each datapoint consists of three components: timestamp, measured stage and measured discharge, see Table 1. The stage has been measured continually with respect to the Amsterdam Ordnance Datum (NAP), using automatic float driven shaft encoders (Buschman et al., 2017). The discharge has been measured infrequently, using helical Ott-mills (mechanical hydrometric current meters) between 1988-2004 and Acoustic Doppler Current Profilers (ADCP) between 2001-2018. The data has been validated by data-owner Rijkswaterstaat, by which several unrealistic outliers were removed from the data.

Table 1. Overview of branches and measurement data (HKV, 2009).

Location	Timestamp	Discharge measurement	Water level measurement
	$[dd - mm - yyyy\ hh:mm]$	$[m^3 s^{-1}]$	$[cm + NAP]$
BR	t_{Lobith}	$Q_{Bovenrijn}$	H_{Lobith}
WL	$t_{Pann. Kop}$	Q_{Waal}	$H_{Pann. Kop}$
PK	$t_{Pann. Kop}$	$Q_{Pann. Kanaal}$	$H_{Pann. Kop}$

2.2. Data homogenization

A single rating curve implies a steady open channel flow and a permanent control (the reference hydraulic situation) (Le Coz, 2012). However, the measured stage-discharge data contains temporal variation and weir effects. Therefore, the data was homogenized. The data is homogeneous when there are no effects from:

- non-permanent flow (hysteresis);
- weirs (backwater);
- river bed subsidence; and
- river interventions.

In this thesis, effects of hysteresis and river interventions were not considered in the homogenization process. Hysteresis is not considered due to time limitation. Next, based on the work of Berends (2019), river interventions were not considered. Using the same data set, Berends (2019) found that there is no visible effect of river interventions on water levels and argued that the effect of historical river interventions may be hidden in rating curve uncertainty. By not considering effects of hysteresis and river interventions in the homogenization process, their uncertainty is implicitly included in the rating curve uncertainty. This implication will be discussed in the discussion (Chapter 4).

The raw dataset (Section 2.1) is homogenized, by:

- selecting data without weir effects (Section 2.2.1); and
- detrending water levels for river bed subsidence (Section 2.2.2).

2.2.1. Selecting data without weir effects

Backwater effects occur when the weir at Driel is active, which is located 25km downstream of the bifurcation point Pannerdense Kop in river Nederrijn. These backwater effects, influence water levels at the measurement locations (Lobith and Pannerdense Kop) and the discharge distribution at the Pannerdense Kop, see Figure 6. In the entire research period (1988-2018), the weir program of Driel has been as follows: as soon as the water level at Lobith dropped below 10.00 m +NAP the weir was put into operation (HKV, 2016; Reeze et al., 2017). However, in Figure 6 weir effects are visible for water levels at Lobith between 10 and 11 m +NAP. These weir effects are considered small and otherwise, if a threshold water level at Lobith higher than 10.00 m +NAP is used, very few data points remain. A water level of 10.00 m +NAP at Lobith is therefore the threshold for selecting data without weir effects, which corresponds with an upstream discharge of roughly $2500 \text{ m}^3\text{s}^{-1}$.

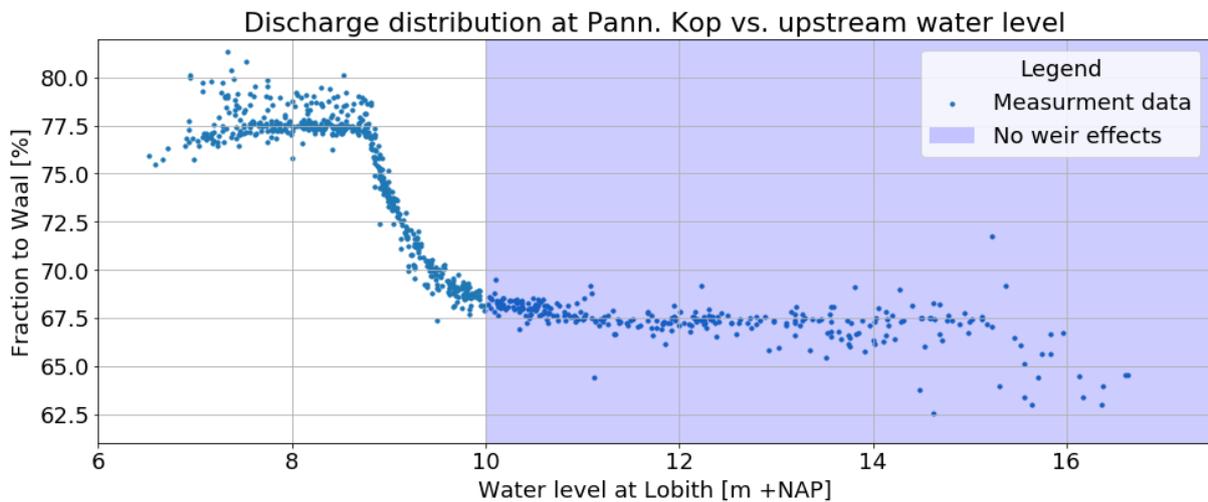


Figure 6. The influence of weir effects on discharge distribution.

The threshold at Lobith can directly be applied on the water level measurements at Lobith. To apply the threshold at Lobith on the water level measurements at the Pannerdense Kop, coinciding water levels between the two locations were used. HKV (2009) assumed that a water level at Lobith at 10.00 m +NAP coincides with a water level of ± 9.68 m +NAP at Pannerdense Kop and Rijkswaterstaat (2015) assumed that water level at Lobith at 10.00 m +NAP coincides with a water level of ± 9.65 m +NAP at Pannerdense Kop. The water levels Pannerdense Kop that coincide with a 10.00 m +NAP water level at Lobith, change over time, see Figure 7. Therefore, a single value for the coinciding water level at Pannerdense Kop should not be used as a threshold. In this thesis, the stage-discharge data of the WL and PK were merged with continually measured water levels at Lobith (frequency of 10 min) based on the exact same timestamps. Next, the threshold at Lobith was applied on the water level measurements at the Pannerdense Kop, using the coinciding water levels.

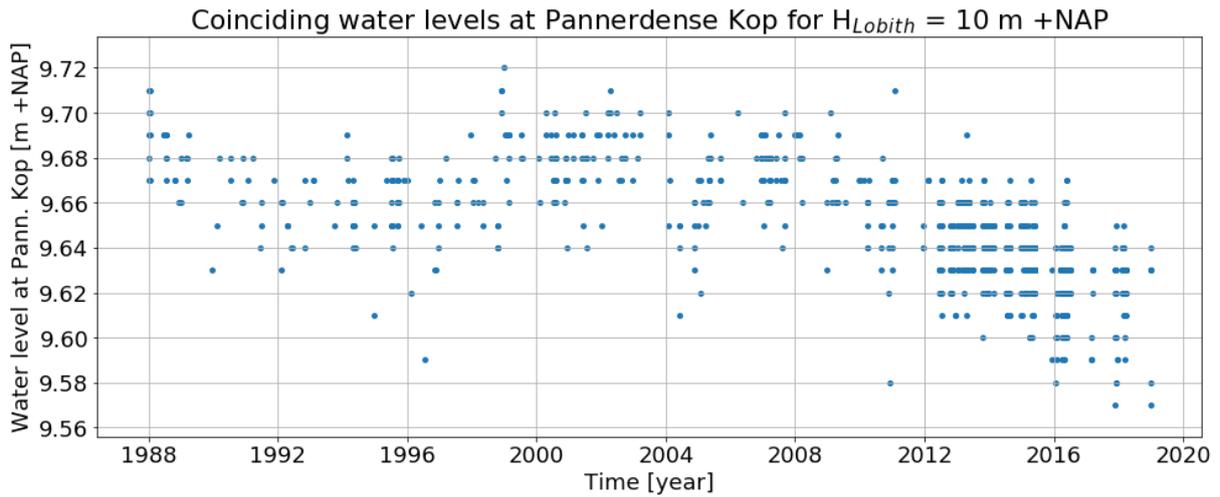


Figure 7. Coinciding water level at Pannerdense Kop given a water level of 10 m +NAP at Lobith (data source: <https://waterinfo.rws.nl>).

2.2.2. Detrending water levels for river bed subsidence

In Figure 8, river bed subsidence is clearly visible in the long term temporal change of the stage-discharge relationship. For all three branches, water levels have lowered at equal discharges due to river bed subsidence (up to roughly 0.6 m within the entire study period). To homogenize the stage-discharge data for river bed subsidence, measured water levels are detrended. The measured discharges remain unchanged.

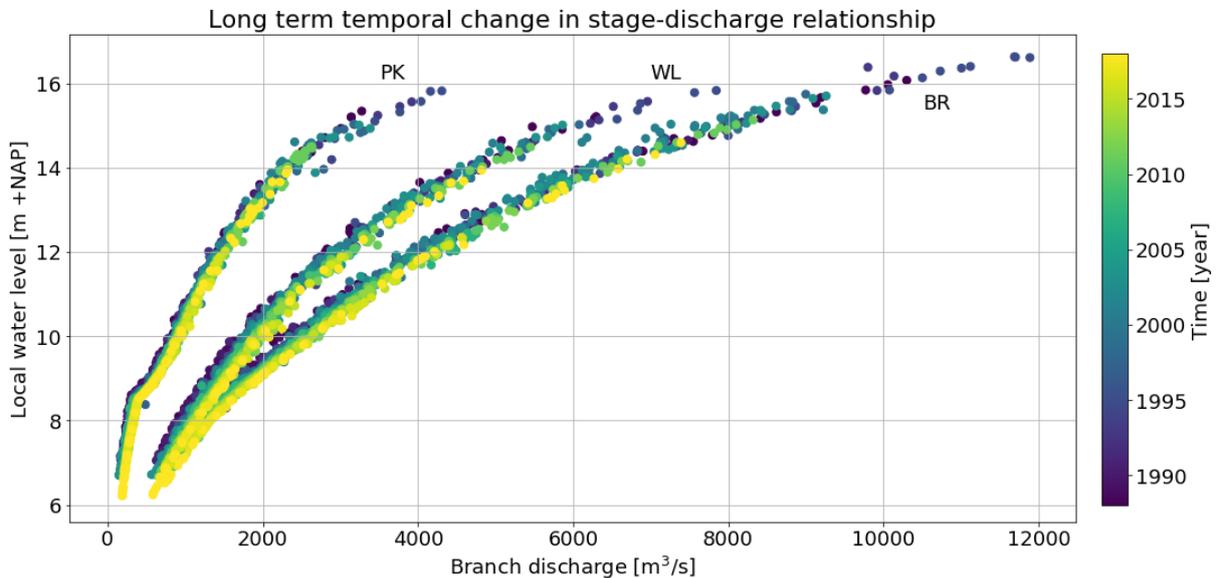


Figure 8. Long term temporal change in stage-discharge relationship of the three branches. Note, the figure contains data points including weir effects ($H_{Lobith} < 10.00$ m +NAP). Before river bed subsidence homogenization the data points including weir effects are excluded.

The first step in detrending the measured water levels, was to divide the dataset without weir effects (Section 2.2.1.) into subperiods. Per branch, the dataset was divided into 4 adjacent subperiods: 1988-1995; 1996-2003; 2004-2011; 2012-2018. Using more subperiods will give too

little data points for rating curve construction and using less subperiods will hide the temporal trend. Furthermore, the entire period of the dataset (31 years) is too small to use a moving average. Therefore, separate subsets are used.

The second step was to construct a rating curve through the subsets, representing the stage-discharge relationship. Polynomial functions were used as rating curve models, since there is too little data per subset to properly fit the rating curve model of Equation 1, which will be introduced in Section 2.3.1. For the polynomial functions, third order polynomials were used, since second order polynomials will under-fit the data points and fourth order polynomials will overfit the data points. The disadvantage of using polynomial functions as rating curves is that they are not suitable for extrapolation. Therefore, third order polynomial rating curves were only constructed for the domain for which all subperiods hold data points.

The third step was to linearly space 50 discharge values per branch in the domain for which all subperiods hold data points. Per subperiod, these 50 discharge values were translated into 50 water level values, using the subperiod specific rating curve. Per discharge value, 4 water level values are obtained for the 4 subperiods. Per discharge, the trend over time of the water levels was determined. This trend is assumed to be linear, based on the research of Berends (2019) who found there was a linear decrease over time of water levels for single discharge values. Berends (2019) have the same study area and period. The results of Berends (2019) are not used in this thesis, because they only obtained trends for 4 discharge values which is not sufficient for proper detrending. Also, even though Berends (2019) have the same study area, they only obtained results for the Pannerdense Kop and not for Lobith.

The final step was to determine the relationship between the trend in the water levels and the branch discharge. The water levels could then be detrended using the following information:

- the measured branch discharge, which is paired to the measured water level;
- the relation between the trend in the water levels and the measured branch discharge; and
- the time difference between the date of measurement and the reference date '2018-12-31 23:59:59'.

2.3. RQ1: Rating curve uncertainty

This research question aims to quantify the uncertainty bands of the single rating curves for the three branches, that stem from all sources of uncertainty present in the homogenous measurement dataset.

2.3.1. Rating curve model

The rating curve model used in this thesis is a summation of the one-dimensional power function adapted from WMO (2010) and ISO (2010). The power function is derived from the Manning-Strickler equation, and assumes a steady uniform flow in a wide rectangular channel. The physical background for the rating curve model used of this thesis is shown in appendix A. The model is given as:

$$Q = \sum_{i=0}^N a_i (h - b_i)^{+p_i}, \quad h > b_i \quad (1)$$

where the measured water level h [m + NAP] serves as input to derive output: the total discharge Q [$m^3 s^{-1}$]. The model parameters are the bed level b [m + NAP] (sometimes written h_0), the parameter related to the channel characteristics $a = Wn^{-1}S_b^{1/2}$ [$m^{4/3} s^{-1}$] and the hydraulic exponent parameter related to the hydraulic control p [-]. The parameters are determined based on available stage discharge data. Furthermore, the discharge is computed from a summation following the division of the channel cross-section in a number of N [-] subsections i [-]. Following HKV (2009), three subsections are assumed, representing the equivalent of the main channel, groin fields and flood plains.

2.3.2. Deterministic optimized rating curve

Per branch, a deterministic optimized rating curve is needed to derive the prior distributions of a_i for the Bayesian rating curves (Section 2.3.3). Furthermore, deterministic optimized rating curves are compared with the probabilistic Bayesian rating curves. After the homogenous (i.e., no weir effects & no rived bed subsidence) stage-discharge datasets are obtained, it is possible to fit the general rating curve form of Equation 1. Equation 1 is fitted through the measurement data points by using a standard regression analysis approach, namely 'least squares'. Least squares problems are either linear or nonlinear, depending on whether curve model is linear. Since Equation 1 is nonlinear, the nonlinear least squares method is used. The goal is to find the vector of parameters θ [a_i, b_i, p_i] that gives the best fit of the curve through the measurement data, also known as parameter optimization. In the least squares sense this means that the sum of the squares is minimized:

$$S = \sum_{i=1}^m r_i^2 \quad (2)$$

$$r_i = Q_i - f(h_i, \theta) \quad (3)$$

where S is the sum of the squares, m is the number of measurements i and r_i is the residual (in-sample prediction error). By optimization the vector of parameters θ is found that yields in a minimum of squares S . Other optimization criteria could have been used, for example RMSE, which puts a bit more weight on higher values. However, the results (Bayesian rating curves) are not sensitive for the optimization criteria, since the deterministic optimized rating curve is only needed to determine prior distributions of a_i for the Bayesian rating curves, which are moderately informative (Section 2.3.3).

For the optimization process, some parameters must have preset fixed values, since it is not possible to compute a deterministic optimized solution if there are too many unknown parameter values. Therefore, a decision must be made on which parameters are fixed and which are optimized. Furthermore, in the optimization process, the vector of parameters θ is subject to physical boundary conditions. The optimization approach is as follows (see Table 2 for an overview):

- a_i are optimized since they are needed to derive the prior distributions of a_i . It is too difficult to estimate fixed values for parameters a_i using the underlying physical parameters [W, n, S_b]. Furthermore, since the branches only flow in the positive direction, a_i are optimized for values larger than zero.
- b_0 is fixed, since there is limited measurement data for the low stage-discharge range due to exclusion of weir effects. Optimizing would therefore lead to uncertain results and unrealistic values. For all branches, an accurate estimation of the bed level is available in Baseline, namely annual- and hectare averaged multibeam measured main channel bed levels per section (location based on chainage). Furthermore, HKV (2009) determined b_i -values for an optimized stationary rating curve for period 2008-2009 for the same branches. Bed level values from Baseline were chosen as b_0 -values, since they are a more accurate approximation of the main channel bed level for 2018 (reference year). In Table 3 it can be seen that the bed level values from Baseline are slightly below the b_0 -values derived by HKV (2009) for a stationary rating curve for period 2008-2009. This could be the result of main channel erosion between 2009 and 2018.
- b_1 and b_2 are optimized, since they are too difficult to estimate. They represent schematized levels related to groin crests and floodplains, which contain large variation in space. To prevent interchangeability, parameters b_1 and b_2 have non-overlapping boundary conditions.
- p_i are fixed, because optimizing these parameters will lead to unrealistic values. For example, HKV (2009) who had the same study area obtained hydraulic exponents larger than 2, because in their optimization approach they used fixed values for b_i and optimized p_i . Based on the Manning equation $p_i \approx 1.667$, see Appendix A. Due to complex cross-sectional geometry and overbank flow, the hydraulic exponent p_i can show variability (± 0.1), but it is crucial to use realistic values for p_i (close to 1.667) for physical derivation of a and for extrapolation (Le Coz et al., 2014). Therefore, the hydraulic exponent p can deviate roughly ± 0.1 from the typical value of $5/3$ (≈ 1.667) (Le Coz et al., 2014). Furthermore, since b_1 and

b_2 are too difficult to estimate, p_i needs to be fixed to control the optimizing process. Otherwise there will be too many unknown parameters.

Table 2. Optimization approach for the vector of parameters θ .

	$i = 0$	$i = 1$	$i = 2$
a_i	$a_0 > 0$	$a_1 > 0$	$a_2 > 0$
b_i	$b_0 = \text{fixed}$	$b_1 > b_0$	$b_2 > b_1$
p_i	$p_0 = 1.667$	$p_1 = 1.667$	$p_2 = 1.667$

Table 3. Cease-to-flow levels.

Location	Chainage (km)	Section	Baseline; 2018	HKV; 2009		
			b_0	b_0	b_1	b_2
BR	862.2	8622	3.57	3.58	11	14
WL	868.3	8683	3.44	3.51	11	13.5
PK	868.2	8682	3.58	3.86	10.5	14.2

2.3.3. Bayesian rating curve

The best method for assessing rating curve uncertainty is an important and open scientific issue (Le Coz et al., 2014). Following other authors (Reitan and Petersen-Øverleir, 2011; Le Coz et al., 2014; Mansanarez et al., 2019; Berends (2019) Bayesian inference and Markov Chain Monte-Carlo (MCMC) is used to construct stochastic rating curves. The main advantage of Bayesian inference is that ‘hydraulic knowledge of the stage-discharge relation can be explicitly translated into prior distributions of the parameters of the assumed rating curve equation’ (Le Coz et al., 2014), which prevents physically unrealistic outcomes (Mansanarez, 2016). Furthermore, unlike deterministic optimization, where some parameters must have preset fixed values due to too many unknown parameter values, all rating curve parameters are described by probability distributions. This means that an uncertainty analysis can simply be done by using percentiles from the posterior distribution. Furthermore, the main advantage of MCMC is the computational efficiency in which samples are taken from the posteriors.

Following ISO (2007) for Ott-mills and ISO (2012) for ADCP, the assumption is made that the distribution of the discharge measurement errors are of normal nature. Therefore, a Normal error model that is proportional to the discharge is used to formalize the relationship between stage and discharge. Assuming that errors in stage measurement are negligible, the error model is given as:

$$Q(h|\theta, \sigma) = f(h|\theta) + \epsilon \tag{4}$$

$$\epsilon \sim N(0, \sigma f(h|\theta)) \tag{5}$$

where $f(h|\theta)$ is the rating curve model (Equation 1) and ϵ is the Normal error term. The normal error term ϵ is a Normal distribution around the rating curve model $f(h|\theta)$ (no bias), in which $\sigma f(h|\theta)$ is the standard deviation and σ is the relative standard deviation. Furthermore, as recommended by Le Coz et al. (2014), the standard deviation can also be written as $\gamma_1 + \gamma_2 \cdot f(h|\theta)$. “This affine model assumes that this standard deviation is dominated at low flows by a constant term c_1 whereas at high flows the error is proportional to the discharge estimation” (Mansanarez, 2016). This thesis is only focused on high flows and $\gamma_1 + \gamma_2 \cdot f(h|\theta)$ leads to values of γ_1 roughly between zero and four, which is very small compared to high flows (order size 10^5). Therefore, the standard deviation will not be written as $\gamma_1 + \gamma_2 \cdot f(h|\theta)$, but as given in Equation 5.

The error model of Equation 5 contains several unknown values, namely the rating curve parameters θ and the relative standard deviation σ . These unknown values are stochastic variables (described by probability distributions). "Prior to inference, the initial distributions should be chosen such that they cover the region where we expect the model to be accurate" (Berends, 2019). So, optimizing the parameters using Bayesian inference provides an understanding of what parameter values lead to an accurate model. This is formalized in Bayes' theorem:

$$p(P|O) = p(O|P) p(P) C^{-1} \quad (6)$$

where the posterior distribution $p(P|O)$ of the parameters P , given observations O , is the product of the prior distribution $p(P)$ and scaling term C to ensure unity. For this research question the theorem can be rewritten as:

$$\underbrace{p(\theta, \sigma | h_{obs}, Q_{obs})}_{\text{posterior}} = \underbrace{p(Q_{obs} | \theta, \sigma, h_{obs})}_{\text{likelihood}} \underbrace{p(\theta, \sigma)}_{\text{prior}} C^{-1} \quad (7)$$

The likelihood of the observed discharge measurements Q_{obs} is evaluated, given the pdf of a Normal distribution with mean $f(h_{obs}|\theta)$ and standard deviation $\sigma f(h_{obs}|\theta)$:

$$p(Q_{obs} | \theta, \sigma, h_{obs}) = p(Q_{obs} | N(f(h_{obs}|\theta), \sigma f(h_{obs}|\theta))) \quad (8)$$

The priors (prior distributions) of θ and σ are follows (see Table 4 and Table 5 for an overview):

- For a_i , moderately informative priors with Normal distributions centered at the deterministic optimized rating curve of the considered branch are adopted.
- For b_0 is a fixed prior is adopted, to control the inference because there is limited measurement data for the low stage-discharge range due to exclusion of weir effects. Similar to the deterministic rating curve optimization, the bed level values from Baseline are used as b_0 -values, since they are the most accurate approximation of the main channel bed level for 2018 (reference year).
- For b_1 and b_2 , priors with non-overlapping Uniform distributions to prevent interchangeability are adopted.
- For p_i , informative priors with Normal distributions centered at 1.667 (based on the Manning equation) are adopted.
- For σ , a non-informative prior with a Half-Normal distribution is adopted, following Gelman (2006).

The prior distributions of the underlying parameters (parameter uncertainty) of the rating curves are shown in Appendix B.

Table 4. Overview of prior choice.

Parameter	Level	Distribution	Notation
a_i	Moderately informative	Normal	$N(\mu, \sigma)$
b_0	Fixed	Fixed	x
b_1, b_2	Non-informative	Uniform	$U(x_l, x_u)$
p_i	Informative	Normal	$N(\mu, \sigma)$
σ	Non-informative	Half-Normal	$ N(\mu, \sigma) $

Table 5. Prior distributions.

Parameter	BR	WL	PK
a_0	$N(121,25)$	$N(87,25)$	$N(42,25)$
a_1	$N(78,25)$	$N(126,25)$	$N(32,25)$
a_2	$N(264,25)$	$N(280,25)$	$N(410,25)$
b_0	3.57	3.44	3.58
b_1	$U(3.57,12)$	$U(3.44,12)$	$U(3.58,12)$
b_2	$U(12,15)$	$U(12,15)$	$U(12,15)$
p_i	$N(1.667,0.25)$	$N(1.667,0.25)$	$N(1.667,0.25)$
σ	$ N(0,2) $	$ N(0,2) $	$ N(0,2) $

Finally, a Markov Chain Monte Carlo (MCMC) algorithm is used to sample from the posterior (Equation 7), using the Hamiltonian No-U-Turn sampler (NUTS) (Hoffman & Gelman, 2014). Typically, a sample size of 10^4 - 10^5 is used, according to Le Coz et al. (2014). Due to the large computational effort of taking samples, a sample size of 2000 is used. This fits the scope of the thesis. Furthermore, testing the sample size of 2000 resulted in proper chains, matching posteriors of the different chains and few divergences in the parameter space. The result is an ensemble of 2000 single rating curves samples, from which total uncertainty is derived. The total uncertainty consists out of model uncertainty and predictive uncertainty. The model uncertainty (or parametric uncertainty) results from the rating curve parameters θ . "Model uncertainty arises from fitting the rating curve model through a limited number of measurements" (Berends et al., 2020). The predictive uncertainty is present in the relative standard deviation σ of the Normal error term (Equation 5). Causes for predictive uncertainty include seasonal changes, hysteresis effects, and measurement error (Berends et al., 2020). To summarize, the model uncertainty depends on the number of measurements and the predictive uncertainty depends on the spread of these measurements.

2.4. RQ2: Influence of measurement errors

The aim of this research question is to quantify the uncertainty bands of the single rating curves for the three branches, that only stem from the measurement errors. In other words, which part of the total variation present in the measurement data, can be explained by measurement errors.

The variation (spread) of the measurement data is reflected by the predictive uncertainty, which includes variation of measurement errors, hysteresis effects and seasonal changes. The measurement errors cannot be derived from the available dataset, because the dataset does not contain the required metadata (e.g., number of verticals, duration of flow measurements). Therefore, similar to Le Coz et al. (2014), ISO reports and scientific literature were used to determine the measurement errors. These measurement errors are then be imposed in Bayesian inference and MCMC simulations. From the simulations the distribution of the remaining predictive uncertainty (hysteresis effects and seasonal changes) is obtained, which is needed to analyze to what extent the measurement errors define the predictive uncertainty.

Furthermore, two different instruments (Ott and ADCP) have been used by the data owner Rijkswaterstaat. Therefore, it was tested whether there is a difference in predictive uncertainty between the instruments, and if so, whether this difference should be considered for this research question. The Ott-mill measurements were split from the ADCP measurements, which then served as separate input for runs of Bayesian inference and MCMC simulations. This allowed for the possibility to spot differences in predictive uncertainty between the Ott-mill- and ADCP measurements.

2.4.1. Discharge measurement errors of helical Ott-mill instruments

Assuming that the uncertainties of the individual components are independent, the relative combined uncertainty for Ott measurements can be determined as follows (ISO, 2007):

$$r_Q = \sqrt{r_m^2 + r_s^2 + \left(\frac{1}{m}\right)\left(r_b^2 + r_d^2 + r_p^2 + \left(\frac{1}{n}\right)(r_c^2 + r_e^2)\right)} \quad (9)$$

where:

- r_Q : relative combined uncertainty in the discharge measurement;
- r_m : relative uncertainty due to limited number of verticals;
- r_s : relative uncertainty due to calibration errors in the current-meter, width measurement instrument, and depth sounding instrument;
- r_b : relative uncertainty in the width measurement;
- r_d : relative uncertainty in the depth measurement;
- r_p : relative uncertainty in mean velocity, due to limited number of depths per vertical;
- r_c : relative uncertainty in point velocity due to variable responsiveness of current-meter;
- r_e : relative uncertainty in point velocity due to velocity fluctuations (pulsations) in the stream;
- m : number of verticals; and
- n : number of depths per vertical.

Equation 9 gives a qualitative overview of how the individual components contribute to the combined uncertainty in discharge measurements. It is remarkable that Equation 9 does not specify how assumptions for floodplain flow are accounted for. The ISO (2007) does not consider this source of uncertainty for Ott measurements. Standard uncertainties can be found in look up tables in ISO (2007):

- $r_m = 3\%$ see Table G.6 ISO (2007), assuming: 15 verticals;
- $r_s = 1\%$ see Equation H.2 ISO (2007);
- $r_b = 0.15\%$ see Table G.1 ISO (2007), assuming: widths between 0 to 100 m;
- $r_d = 0.25\%$ see Table G.2 ISO (2007), assuming: depths between 6 and 14 m;
- $r_p = 2.5\%$ see Table G.4 ISO (2007), assuming: 5 points per vertical;
- $r_c = 0.5\%$ see Table G.5 ISO (2007), assuming: individual rating with velocity > 0.5 m/s;
- $r_e = (\sqrt{5 \cdot 3^2})\%$ see Table G.3 ISO (2007), assuming: exposure time = 1 min, velocity > 1 m/s;
- $m = 15$ see STOWA (2009); and
- $n = 5$ estimation.

Filling these standard uncertainties into the Equation (9) gives a value of $r_Q = 3.3\%$, which indicates a high accuracy. According to the handbook for measuring discharge in open channels of STOWA (2009), Ott measurements from a boat result in $r_Q = 3 - 8\%$. However, according to Table 9-2 in STOWA (2009) $r_Q \leq 5\%$ for more than 5 verticals. Rijkswaterstaat uses a minimum of 15 verticals (STOWA, 2009). So, the combined uncertainty in the discharge measurement is assumed to be between 3 and 5% for Ott-mills.

2.4.2. Discharge measurement errors of ADCP instruments

ISO (2012) on ADCP measurements, does not specify a methodology for computing the (combined) uncertainty of discharge measurements, but state that “the overall uncertainty will be a combination of the measured parameters, the computation methodology and the assumptions regarding the unmeasured portions of the channel” (ISO, 2012). According to STOWA (2009), the relative combined uncertainty for ADCP measurements can be determined as follows, assuming independency of the components:

$$r_Q = \sqrt{r_{Q_m}^2 + r_{Q_b}^2 + r_{Q_w}^2 + r_{Q_o}^2 + r_i^2} \quad (10)$$

Where:

- r_Q : relative combined uncertainty in the discharge measurement
- r_{Q_m} : relative uncertainty in the direct measurable zone of the cross-section
- r_{Q_b} : relative uncertainty in the bottom layer, calculated based on an assumed velocity profile
- r_{Q_w} : relative uncertainty in the surface layer, calculated based on an assumed velocity profile
- r_{Q_o} : relative uncertainty in the near bank- and floodplain area, estimated or calculated based on an assumed velocity profile
- r_i : relative uncertainty of instrumental errors.

The relative uncertainty in the near bank- and floodplain area r_{Q_o} , is not equal for all locations due to the different geometries. However, it is unknown how r_{Q_o} differs per location and therefore treated equally. Implications of this will be discussed in the discussion, see Chapter 4. STOWA (2019) do not give example values to fill in Equation 10, but state that ADCP measurements using the moving boat method result in $r_Q = 3 - 5\%$. In scientific literature, usually $r_Q = 5\%$ is used for ADCP measurements (e.g., Le Coz et al., 2014; Mansanarez, 2016). Finally, ADCP measurements are commonly considered more accurate than Ott measurements (e.g., Le Coz et al., 2014; Mansanarez, 2016).

2.4.3. Bayesian inference

From Section 2.4.1. and Section 2.4.2. it is concluded that $r_Q = 3 - 5\%$ for both Ott-mill and ADCP measurements. Following common hydrometry practice (e.g., ISO, 2007; ISO, 2012), uncertainty intervals of r_Q are defined as 95% confidence intervals. Now, with $\sigma_m = 3 - 5\%$ at 95% confidence, the remaining relative standard deviation σ_r has to be determined. The normal error model of Equation 4 is adapted to include measurement errors. Assuming that the measurement- and remaining error are independent and without bias, the adapted error model is given as:

$$Q(h|\theta, \sigma_m, \sigma_r) = f(h|\theta) + (\epsilon_m + \epsilon_r) \quad (11)$$

$$\epsilon_m + \epsilon_r \sim N\left(0, f(h|\theta)\sqrt{\sigma_m^2 + \sigma_r^2}\right) \quad (12)$$

where ϵ_m is the measurement error, ϵ_r is the remaining error, $f(h|\theta)\sqrt{\sigma_m^2 + \sigma_r^2}$ is the standard deviation, $\sqrt{\sigma_m^2 + \sigma_r^2}$ is the relative standard deviation, σ_m is the relative standard deviation of measurement uncertainty and σ_r is the relative standard deviation of remaining uncertainty. For this research question Bayes' theorem can be written as:

$$\underbrace{p(\theta, \sigma_m, \sigma_r | h_{obs}, Q_{obs})}_{\text{posterior}} = \underbrace{p(Q_{obs} | \theta, \sigma_m, \sigma_r, h_{obs})}_{\text{likelihood}} \underbrace{p(\theta, \sigma_m, \sigma_r)}_{\text{prior}} C^{-1} \quad (13)$$

The likelihood of the observed discharge measurements Q_{obs} is evaluated, given a Normal distribution with mean $f(h_{obs}|\theta)$ and standard deviation $f(h|\theta)\sqrt{\sigma_m^2 + \sigma_r^2}$:

$$p(Q_{obs} | \theta, \sigma_m, \sigma_r, h_{obs}) = p(Q_{obs} | N\left(f(h_{obs}|\theta), f(h_{obs}|\theta)\sqrt{\sigma_m^2 + \sigma_r^2}\right)) \quad (14)$$

The priors (prior distributions) of θ , σ_m and σ_r are as follows:

- For θ , containing $[a_i, b_i, p_i]$, priors equal to the priors of research question 1 are adopted, see Table 4 and Table 5.
- For σ_m , a fixed prior is adopted, with values obtained from ISO reports, which have been used by Rijkswaterstaat. A value of 3% and 5% are used, which are the most outer values for both Ott and ADCP.
- For σ_r , a non-informative prior with a Half-Normal distribution is adopted, following Gelman (2006).

An overview of the prior choice and distributions of σ_m and σ_r is given in Table 6 and Table 7. Finally, like research question 1, again a Markov Chain Monte Carlo (MCMC) algorithm is used to sample from the posterior, using the Hamiltonian No-U-Turn sampler (NUTS) (Hoffman & Gelman, 2014).

Table 6. Overview of prior choice

Parameter	Level	Distribution	Notation
σ_m	Fixed	Fixed	x
σ_r	Non-informative	Half-Normal	$ N(\mu, \sigma) $

Table 7. Prior distributions.

Parameter	BR	WL	PK
σ_m	3%, 5%	3%, 5%	3%, 5%
σ_r	$ N(0, 2) $	$ N(0, 2) $	$ N(0, 2) $

2.5. RQ3: Water balance in rating curves

For this research question a new method is proposed for establishing rating curves based on a closed water balance. Next, it is shown how this method influences the uncertainty bands of the rating curves, as compared to the rating curves of research question 1. In scientific literature, no method can be found that considers the water balance in the establishment of rating curves.

In Section 1.2 it was shown that the original rating curves give a systematic error in water balance. However, there is no way to determine which of the three rating curves is most reliable and which of the three contribute(s) most to the water balance error. Therefore, all three locations are treated as equally reliable and a new method is proposed in which for each location data from the other two locations is used to adjust for water balance-offset in the rating curves. Novel in this method is that discharge measurements from other locations are included. In current practice, similar to the rating curves of research question 1, rating curves are constructed by only using locally measured stage and discharge as input data.

To illustrate this method, Lobith is picked as example location. First, the data previously used for RQ1 (Section 2.3) is filtered on the prerequisite that a water level and discharge measurement is available in each of the three branches within the same day. This allows a same-day comparison of the water balance of the discharge measurements for each data point. The resulting filtered dataset contains 292 same-day discharge and water level measurements for each of the three locations. Second, for the upstream location of Lobith, the same-day discharge measurements of the two downstream bifurcating branches are summed and coupled with the same-day water levels at Lobith. The resulting 292 calculated data points are added to the 292 locally measured same-day stage-discharge data at Lobith. Per water level point, there are now have two same-day discharges values, one local and one non-local, yielding a total of 584 data points. This same approach was used to construct 584 local and calculated data points for the two downstream locations. See Table 8 for an overview of how for each station additional data points were calculated based on a water balance. Finally, similar to the method used in Section 2.3.3, Bayesian rating curves were constructed for all three locations.

Table 8. Overview of how for each station additional non-local same-day data points were calculated based on a water balance

Considered location	Discharge	Water level
BR	$Q_{WL} + Q_{PK}$	H_{Lobith}
WL	$Q_{BR} - Q_{PK}$	$H_{Pann. Kop}$
PK	$Q_{BR} - Q_{WL}$	$H_{Pann. Kop}$

By using same-day measurements, variation in discharge and water level in time and space is neglected. It is assumed that the temporal variation is small, because the river Rhine is a delta river. It is estimated that the maximum rate of change of discharge is roughly $20 \text{ m}^3\text{s}^{-1}$ per hour. Also, the same-day measurements are taken in a small time window, namely between 08:00 and 17:00. Assuming that the measurements are distributed evenly throughout the day and given the rate of change of $20 \text{ m}^3\text{s}^{-1}$ per hour and the time window of 9 hours, there would be an average deviation of $9 / 2 * 20 = 90 \text{ m}^3\text{s}^{-1}$. This corresponds with roughly 5 cm water level deviation at Lobith. Next, it is assumed that the variation in space is small, since the distance between Lobith and Pannerdense Kop is only 5 km, see Figure 3. Furthermore, for the water level, there exists a strong linear correlation, namely $r^2=0.9989$, between the same-day water level measurements at Lobith and Pannerdense Kop, which proves the little variation in discharge and water level in time and space.

Finally, another method is explored for establishing rating curves based on a closing water balance, in which an attempt was made to impose the water balance as a boundary condition in the rating curve function. This attempted method did not work. Since three rectangular subsections are used to schematize a branch, each rating curve contains two kinks. To impose the water balance between the upstream branch and the two downstream branches, the rating curves of the two downstream branches are summed and thereby creates a rating curve with 4 kinks. However, it is not possible to compare this rating curve with 4 kinks with the rating curve of the upstream branch, which contains 2 kinks.

3. Results

3.1. Homogenized data

A single rating curve implies a steady open channel flow and a permanent control (the reference hydraulic situation) (Le Coz, 2012). However, the measured stage-discharge data contains temporal variation and weir effects. Therefore, the data is homogenized, by:

- selecting data without weir effects (Section 3.1.1); and
- detrending water levels for river bed subsidence (Section 3.1.2).

3.1.1. Selecting data without weir effects

In the entire research period (1988-2018), the weir program of Driel has been as follows: as soon as the water level at Lobith dropped below 10.00 m +NAP the weir was put into operation (HKV, 2016; Reeze et al., 2017). Therefore, a water level of 10.00 m +NAP at Lobith is therefore the threshold for selecting data without weir effects, which corresponds with an upstream discharge of roughly $2500 \text{ m}^3\text{s}^{-1}$. By selecting data without weir effects, the length of the measurement data has been reduced from 1303 to 568, 1202 to 464 and 1520 to 533, for BR, WL and PK, respectively. Figure 9 shows the data points with and without weir effects. The data points with weir effects were excluded from the dataset.

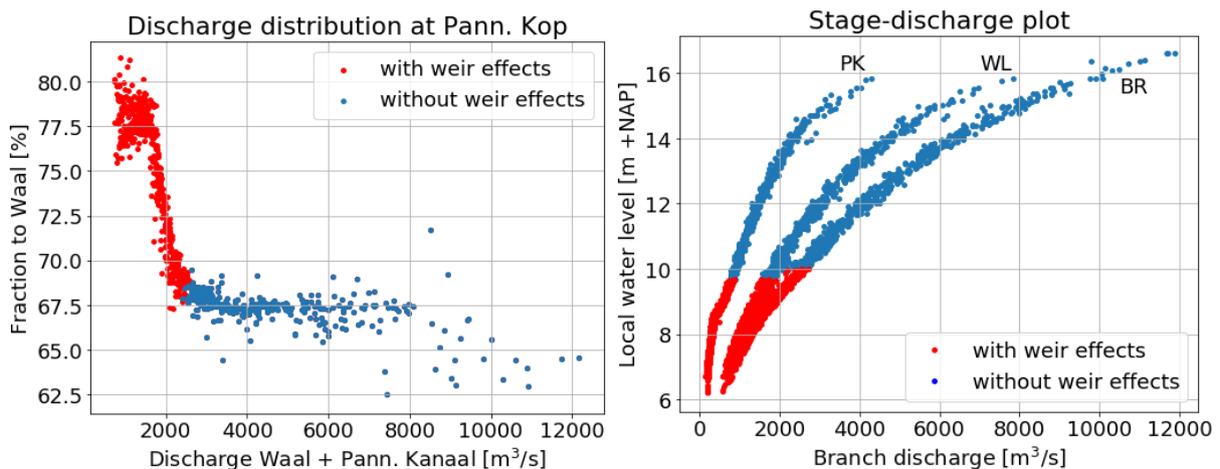


Figure 9. Discharge distribution at Pannerdense Kop (left) and stage discharge plot (right) with indication of measurements that are not influenced by weirs.

3.1.2. Detrending water levels for river bed subsidence

For all three branches, water levels have lowered at equal discharges due to river bed subsidence (up to 0.6 m within the entire study period). To homogenize the stage-discharge data for river bed subsidence, measured water levels are detrended. The measured discharges remain unchanged.

The first step in detrending the measured water levels, was to divide the dataset without weir effects into subperiods. The dataset was divided into 4 adjacent subperiods: 1988-1995; 1996-2003; 2004-2011; 2012-2018. The second step was to construct a rating curve through the subsets, see Figure 10. The third step was to determine the linear trends over time of the water levels at 50 discharge values. For proper visualization of the trend lines, only 4 of the 50 trend lines are presented in Figure 11. To determine how well the linear models fit the data points, an R-squared value is determined per discharge, see Figure 12. R-squared gives the percentage of the variable variation

that is explained by a linear model: 'R-squared = Explained variation / Total variation'. The R-squared scores are good (>0.8) for the Waal and the Pannerdensch Kanaal, but less good for the Bovenrijn for high discharges. However, it is assumed that this is not a problem, because for high discharges the correction of the data points is relatively smaller.

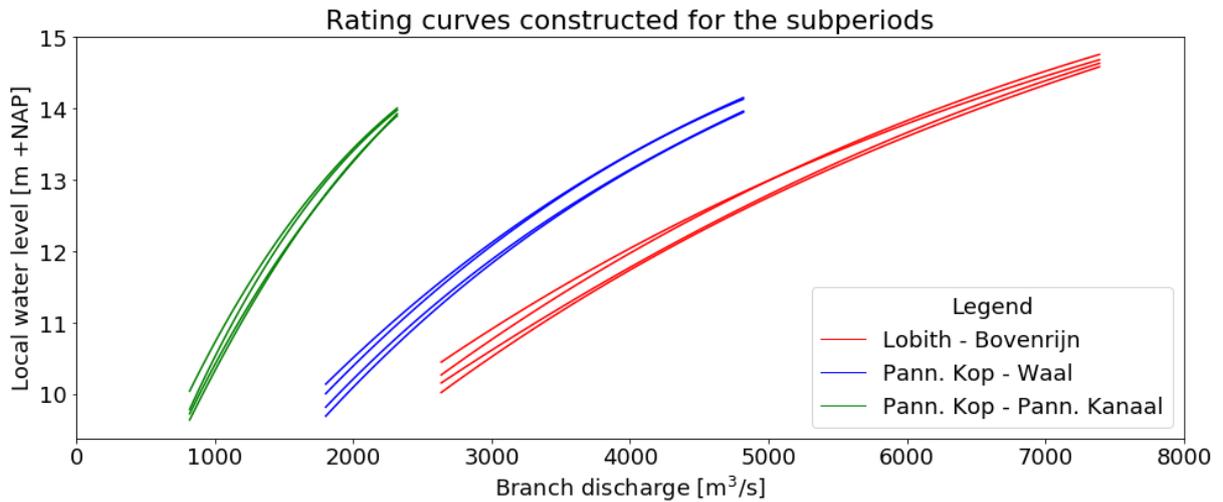


Figure 10. Rating curves constructed for the subperiods. Only the domain for which all subperiods hold data points are used in the rating construction.

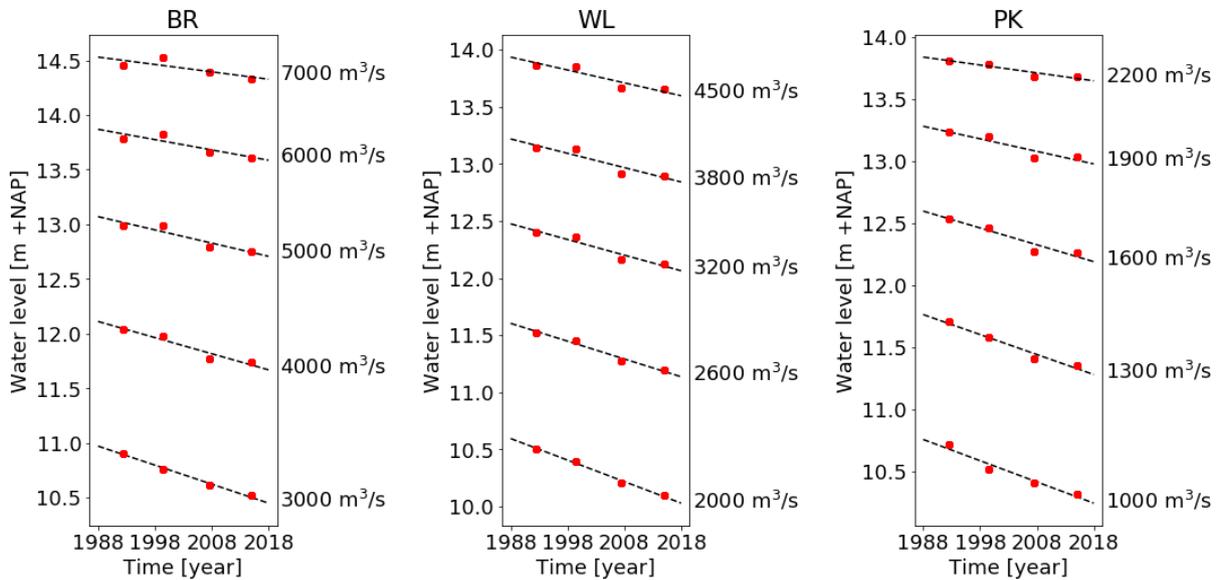


Figure 11. Water level trends different discharge values. The red dots indicate where the discharge value corresponds with the water level of Figure 10. For higher discharges the water level trends become less steep. Note: a total of 50 discharges have been used in the detrending process.

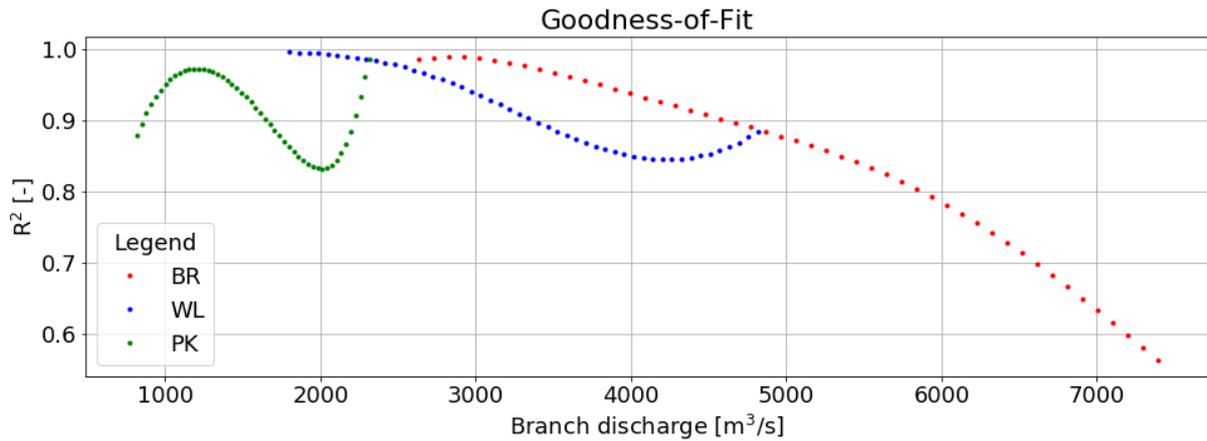


Figure 12. Goodness of fit of linear trends in water levels of Figure 11.

The final step was to determine the relationship between the water level trend and the branch discharge. In Figure 13, it can be seen that the shape of the relationships of the data points are different per branch. It is important to know that all data points in Figure 13 fall within the discharge domain of main channel flow, for all three branches. It is assumed that the difference in shape of the relationship shown by the data points, results from the different cross-sectional main channel geometries of the three branches, see Figure 14. Constant river bed subsidence rates are assumed, based on Berends (2019). In theory, when the cross-sectional geometry of the main channel is rectangular, a constant river bed subsidence rate results in a constant water level decrease rate, which results in a linear shape of the relationship between the water level trend and branch discharge. Since the cross-sectional main channel geometry of the Bovenrijn is nearly rectangular at Lobith, the shape of the relationships of the data points is linear, see Figure 13. Next, the non-linearity of the shape of relationships of the WL and PK could be explained by the nearly triangular cross-sectional geometries of the main channels of the Waal and Pannerdensch Kanaal at Pannerdense Kop, see Figure 14. Also, due to the triangular cross-sections, river bed subsidence is more dominant towards the deeper part of the cross-sections, which could also contribute to non-linearity of the shape of the relationships. For WL at a discharge of 2000 m³s⁻¹, Berends (2019) obtained a trend of 1.7 cm/year. This compares well with the result of 1.8 cm/year, see Figure 13.

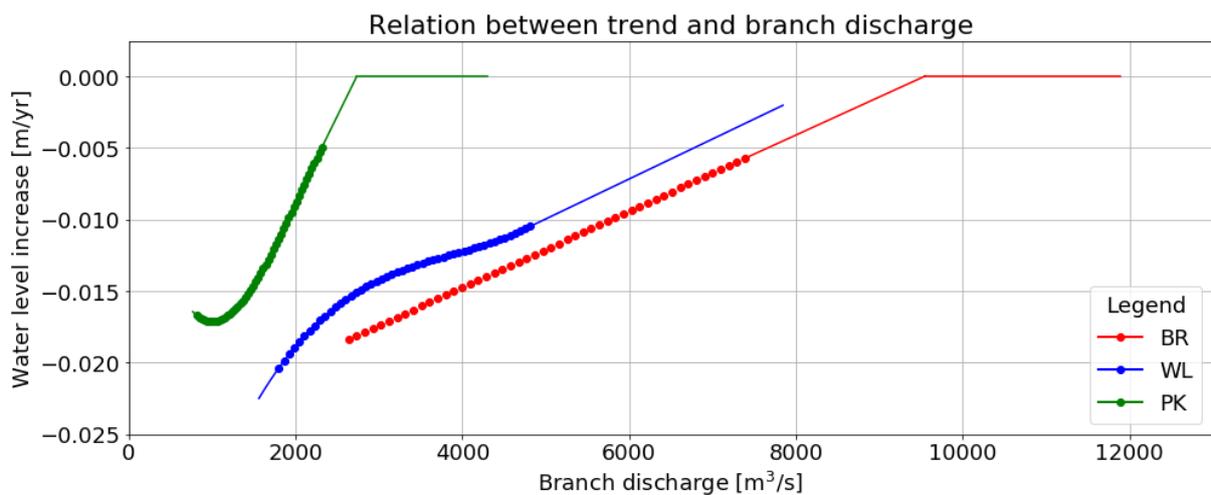


Figure 13. Relationship between water level trend and branch discharge resulting from Figure 11.

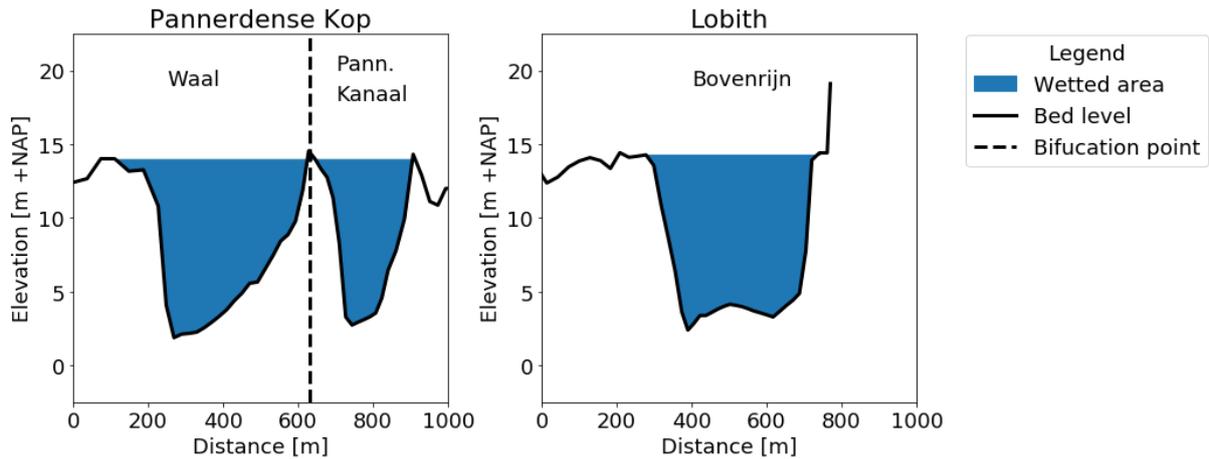


Figure 14. Main channel cross-sectional geometry of the three locations. The wetted area is representative for the maximum water level in the domain for which all subperiods hold data points, see Figure 10. The chainage of the cross-section at Lobith is 862.2 km and the chainage of the cross-section at Pannerdense Kop is 868.3km (source: Baseline 5.3.3., 2018)

The data points in Figure 13 are interpolated using a linear function for BR and a third order polynomial function for WL and PK. Furthermore, extrapolations were needed since there are data points that need to be detrended and contain discharge values outside the discharge domain, for which all subperiods hold values. The data points in Figure 13 are linearly extrapolated until intersection with the x-axis, where the water level increase remains constant. Linear extrapolation is used, since there is no information to choose for a different extrapolation approach. Also, since the correction of the data points at high discharges is relatively small, it is assumed that this is not a problem. Given Figure 13 and the time difference with the reference date '2019-01-01 00:00:00', the water level corrections are obtained, see Figure 15. Detrending the stage-discharge dataset without weir effects by applying the corrections of Figure 15 gives the detrended stage-discharge dataset of Figure 16. In Figure 16 it can be seen that the spread has decreased, especially for lower water levels due to the relatively larger correction, and that the position of data points have become independent of time.

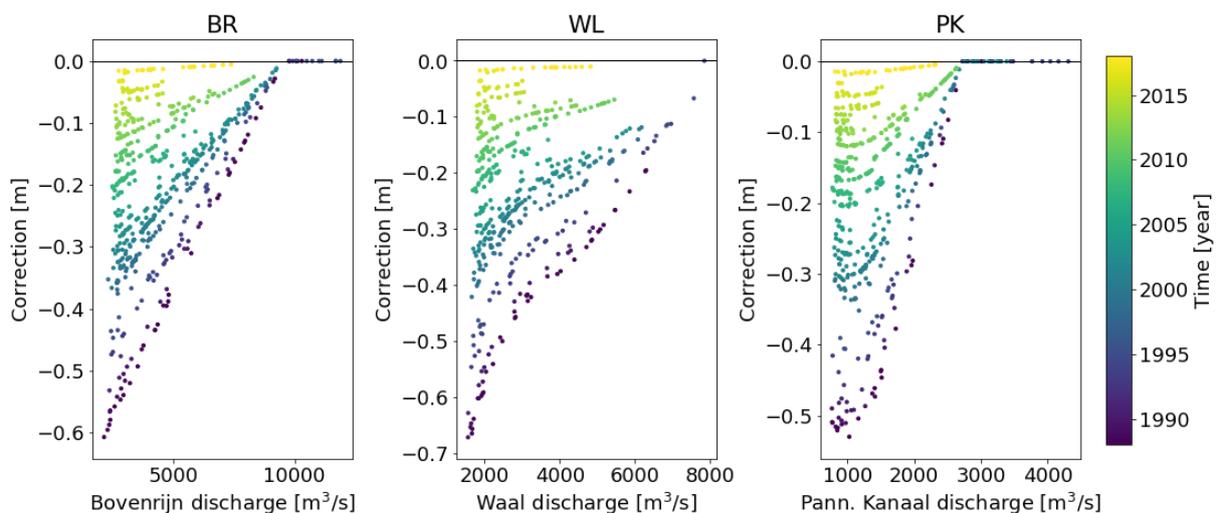


Figure 15. Water level corrections obtained from Figure 13 and the time difference with the reference date '2019-01-01 00:00:00'.

Detrending long term temporal change in stage-discharge relationship

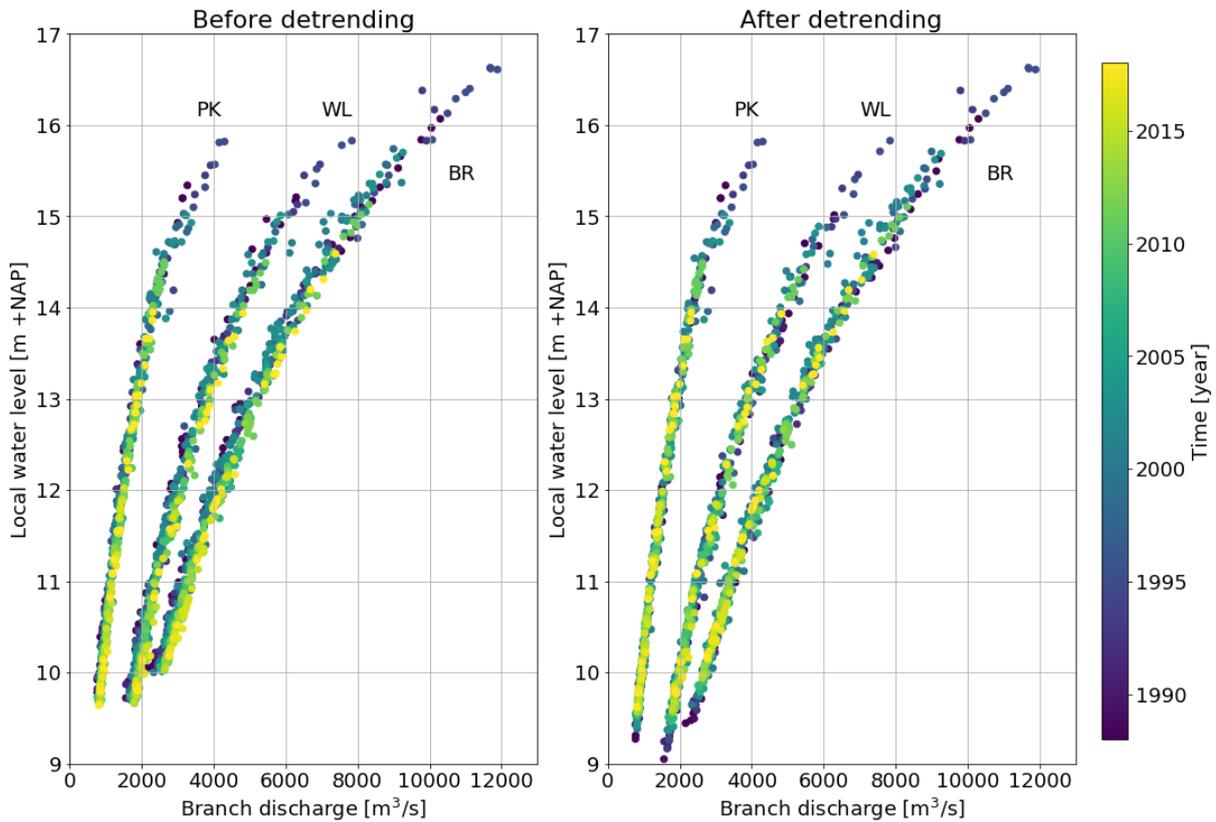


Figure 16. Result of detrending the long-term temporal change in stage discharge relationship. The left plot shows the stage-discharge data before detrending and the right plot shows the stage-discharge data after detrending. This removed the time-dependency and reduced the spread.

3.2. RQ1: Rating curve uncertainty

3.2.1. Deterministic optimized rating curves

Per branch, a deterministic optimized rating curve is needed to derive the prior distributions of a_i for the Bayesian rating curves (Section 2.3.3). Furthermore, deterministic optimized rating curves are compared with the probabilistic Bayesian rating curves. The result of deterministic optimization of the rating curve parameters is presented in Table 9. Filling in these parameters into the general rating curve form (Equation 1) gives the rating curves for the branches, see Figure 17. To check whether optimizing the values of b_1 and b_2 resulted realistic values, a comparison can be made with the values of b_1 and b_2 of HKV (2009), see Table 9. HKV (2009) estimated values of b_1 and b_2 based on the cross-sectional geometries of the locations for 2008-2009 and in this thesis the values of b_1 and b_2 are optimized for 2018. Assuming that values of b_1 and b_2 do not change much over a period of 10 year, our values of b_1 and b_2 compare well with HKV (2009). Furthermore, next to obtaining realistic values for b_1 and b_2 , realistic values for p_i were maintained by fixing these parameters on $5/3$ (≈ 1.667). HKV (2009) chose to optimize p_i , which in their case resulted in unrealistic hydraulic exponents larger than 2, see Table 9. Therefore, it is a better optimization strategy to optimize b_1 and b_2 , and fix values for p_i on 1.667.

Table 9. Deterministic optimized rating curve parameters.

Parameter	BR		WL		PK	
	Result RQ3	HKV (2009)	Result RQ3	HKV (2009)	Result RQ3	HKV (2009)
a_0	121.082	102.094957	86.924	97.432094	42.420	38.940582
a_1	78.004	87.006333	126.286	135.916967	31.650	9.405317
a_2	264.031	246.106698	279.510	231.1769	409.778	352.874043
b_0	3.57	3.58	3.44	3.51	3.58	3.86
b_1	11.093	11	11.680	11	10.5145	10.5
b_2	13.965	14	14.195	13.5	14.121	14.2
p_0	1.667	1.721789	1.667	1.577044	1.667	1.723856
p_1	1.667	1.720418	1.667	1.481197	1.667	2.416239
p_2	1.667	2.00681	1.667	1.929289	1.667	1.741329

3.2.2. Bayesian rating curves

Following other authors (Reitan and Petersen-Øverleir, 2011; Le Coz et al., 2014; Mansanarez et al., 2019; Berends, 2019) and given the advantages of the Bayesian approach, Bayesian inference and Markov Chain Monte-Carlo (MCMC) were used to construct stochastic rating curves. The result of Bayesian inference and Markov chain Monte Carlo simulations, as based on a homogenized measurement dataset, is presented in Figure 17. The posterior distributions of the underlying parameters (parameter uncertainty) of the rating curves are shown in Appendix B. The total uncertainty bandwidths of the rating curves increase for higher water levels. Indications of the total rating curve uncertainty bandwidths for local water levels of 12 and 16 m +NAP are shown in Table 10. The total uncertainty consists out of model uncertainty and predictive uncertainty. The model uncertainty depends on the number of measurements and increases for domains with less measurements. The predictive uncertainty depends on the spread of the measurements and increases when the spread in the measurements increases. In Figure 17 it can be seen that the model uncertainty is dominant for domains with little measurements, which in this case is for high discharges. The obtained rating curves contain standard deviations of predictive uncertainty of 2.54%, 2.90%, 3.31% for the BR, WL and PK, respectively.

Furthermore, the maximum a posteriori (MAP) rating curves have been plotted in Figure 17. Using the MAP values is a standard choice in Bayesian statistics (Mansanarez, 2016). The MAP rating curve is in fact the modus of the total uncertainty of the rating curve and is not obtained by taking the modi of the individual parameters (θ, σ). For all branches, Figure 17 shows that the MAP rating curves are equal to the deterministic rating curves for the measured discharge domain. Also, for all branches, Figure 17 shows that the MAP rating curves lie above the deterministic rating curves for higher discharges outside the measured discharge domain, especially for river Waal. The difference between the MAP- and the deterministic rating curves could be explained by the fact that the deterministic approach gives more weight to higher data points. This results in lower rating curves as compared to the Bayesian rating curves, which allows for a full uncertainty propagation leading to higher rating curves.

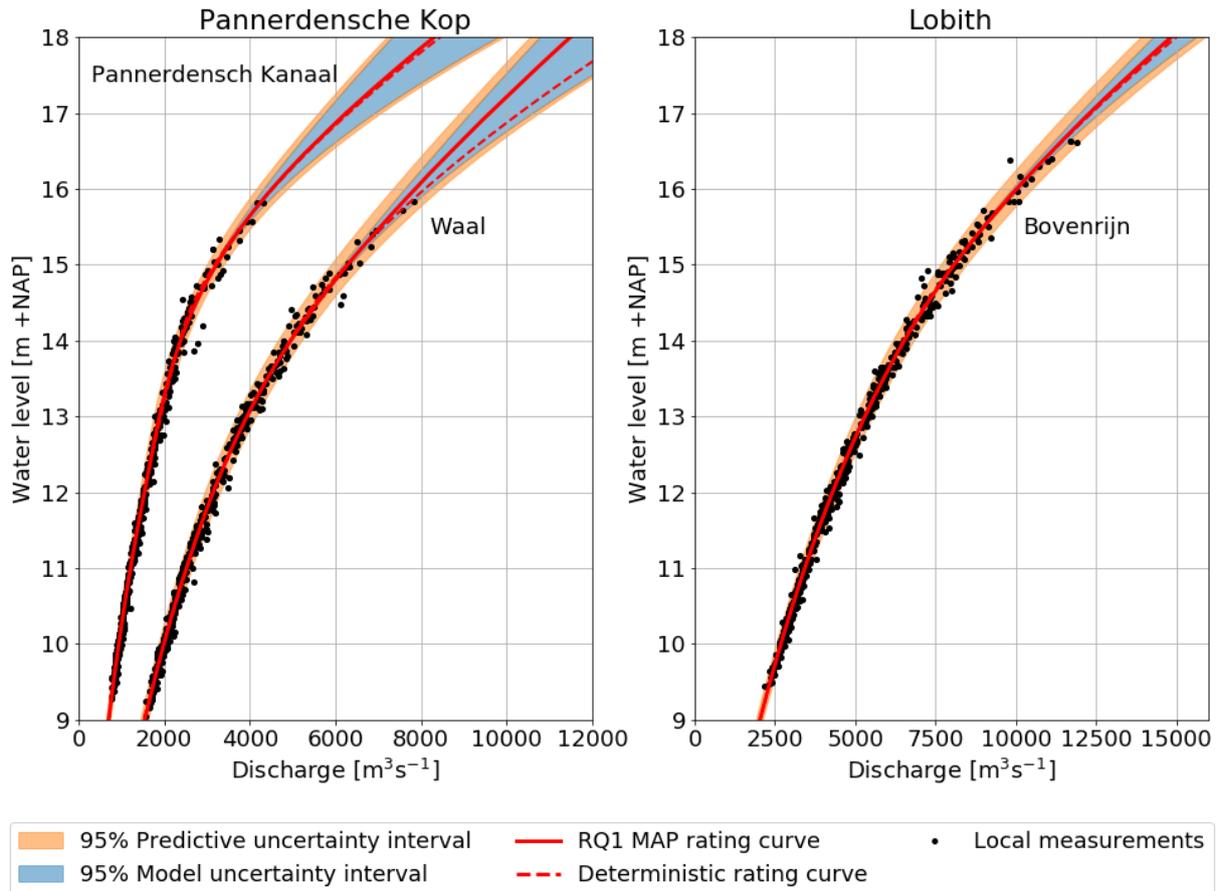


Figure 17. Bayesian rating curve uncertainty. The total uncertainty consists out of model uncertainty (depends on the number of measurements) and predictive uncertainty (depends on the spread of the measurements). The MAP rating curves indicate the modus curve of the total uncertainty.

Table 10. Total uncertainty bandwidths at local water levels of 12 and 16 m +NAP.

Location	Water level = 12 m +NAP	Water level = 16 m +NAP
	Total uncertainty bandwidth [m ³ s ⁻¹]	Total uncertainty bandwidth [m ³ s ⁻¹]
BR	430.1	1018.6
WL	364.5	1113.7
PK	199.8	727.9

3.3. RQ2: Influence of measurement errors

For the stage-discharge data two different instruments (Ott and ADCP) have been used by Rijkswaterstaat. To check whether there is a difference in spread of the discharge measurements between the instruments, the Ott-mill measurements are split from the ADCP measurements. Separate runs of Bayesian inference with MCMC simulations for separate Ott- and ADCP measurement input resulted in the posterior distributions of the predictive uncertainty, see Figure 18. The predictive uncertainty represents the measure of the spread in the data points. For all branches the Ott measurements have a higher predictive uncertainty than the ADCP measurements,

especially at PK. Note that the predictive uncertainty includes variation from measurement errors, hysteresis effects and seasonal variation. Assuming that seasonal variation and hysteresis effects are equal for both Ott- and ADCP measurement data, the measurement errors are likely higher for helical Ott mills as compared to ADCP instruments. This is in agreement with the common consideration that ADCP measurements are more accurate than Ott measurements (e.g., Le Coz et al., 2014; Mansanarez, 2016). Figure 18 shows that this difference is statistically significant for PK, since the 95% confidence intervals do not intersect. However, this is not the case for all locations and therefore no distinction is made between the two instruments. Also, in order to distinguish between the different measurements, a separate analysis would have to be done for Ott-mill measurements and ADCP measurements. Since measurement data is limited, all measurement data were needed for rating curve construction. This will be discussed in the discussion (Chapter 4).

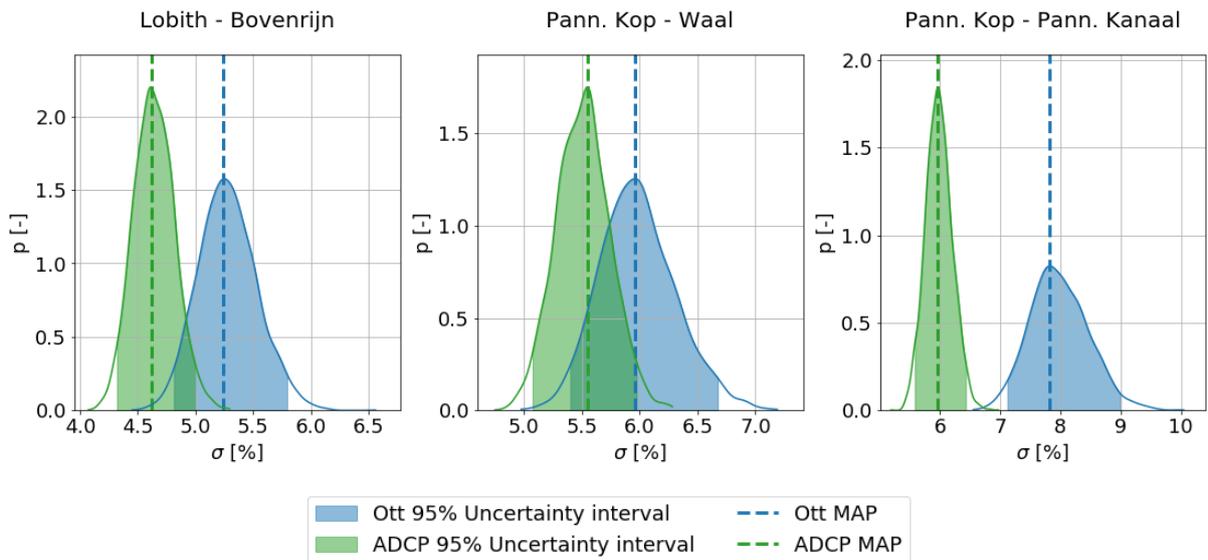


Figure 18. Predictive uncertainty of Ott- and ADCP measurements expressed in relative standard deviation.

The aim of this research question is to determine the extent into which the variation present in the measurement data can be explained by measurement errors. The variation (spread) of the measurement data is reflected by the predictive uncertainty, which includes variation of measurement errors, hysteresis effects and seasonal changes. Similar to Le Coz et al. (2014), ISO reports and scientific literature were used to determine the measurement errors (3 and 5%). These measurement errors were imposed in Bayesian inference and MCMC simulations, resulting in posterior distributions of the remaining predictive uncertainty (hysteresis effects and seasonal changes), see Appendix C. The MAP values of the posterior distributions of the remaining predictive uncertainty were used to analyze to what extent the measurement errors define the predictive uncertainty. Table 11 shows the percentages of variation of the predictive uncertainty, explained by measurement uncertainty σ_m and the remaining uncertainty σ_r . In Table 11 it can be seen that from all three branches, for equal measurement errors, the measurement uncertainty is most dominant in BR. There is a large difference in variation explained by measurement errors when assuming discharge measurement errors of 3% versus 5%, especially for BR. However, it remains unknown what the exact measurement error is per location. The measurement uncertainty is likely highest for WL, since more measurement uncertainty is introduced due to assumed floodplain flow in the relatively large and complex cross-section of the floodplain of WL. Next, even though variation due to hysteresis is still present in the data, Table 11 shows little remaining uncertainty for $\sigma_m=5\%$ (especially for BR), which indicates that variation due to hysteresis effects is small compared to variation of measurement errors. A possible explanation for this might be the validation by data owner Rijkswaterstaat, since outlier removal reduces the variation in the data.

Table 11. Variation of predictive uncertainty explained by measurement uncertainty σ_m vs. remaining uncertainty σ_r . MAP values have been used for the remaining uncertainty σ_r (Appendix C).

Location	$\sigma_m = 3\%$		$\sigma_m = 5\%$	
	σ_m	σ_r	σ_m	σ_r
BR	35.1%	64.9%	97.4%	2.6%
WL	26.6%	73.4%	74.9%	25.1%
PK	20.5%	79.5%	57.2%	42.8%

3.4. RQ3: Water balance in rating curves

For this research question a new method is proposed for establishing rating curves based on a closed water balance. Novel in this method is that discharge measurements from other locations are included. Per water level point, there are now two same-day discharges values, one local and one (calculated) non-local, yielding a total of 584 data points. In current practice, similar to the rating curves of research question 1, rating curves are constructed by only using locally measured stage and discharge as input data. Bayesian rating curves were constructed for all three locations, using the same method as RQ1 (Section 2.3.3). The results are shown in Figure 19, where among the data points the non-local calculated data points are plotted in grey. In Figure 19 it is visible that the rating curves of RQ3 have now shifted as compared to the rating curves of RQ1. The effect of the shifting rating curves on the water balance error between the three locations is shown in Figure 20, where a comparison is given between three different methods of rating curve construction. The different methods are the deterministic optimization method, the Bayesian method of RQ1 and the Bayesian method of RQ3. For Bovenrijn discharges below roughly $10,000 \text{ m}^3\text{s}^{-1}$, the rating curves of RQ3 give slightly lower discharges at equal water level for the upstream location of Lobith and slightly higher discharge values, for equal water levels, at the two downstream locations. For Bovenrijn discharges above roughly $10,000 \text{ m}^3\text{s}^{-1}$, the rating curves of RQ3 give higher discharges at equal water level for the upstream location of Lobith and lower discharge values for equal water levels at the two downstream locations. In Figure 20 it can be seen that the Bayesian method of RQ3 gives a smaller water balance error than the Bayesian method of RQ1. So, adding the calculated non-local data points to the analysis clearly reduced the systematic error in water balance.

Next, it is shown how the proposed method influences the uncertainty bands of the rating curves, as compared to the rating curves of research question 1. Figure 21 gives a comparison of the 95% uncertainty bandwidths of the RQ1- and RQ3 rating curves for all three locations. In Figure 21 it can be seen that the uncertainty bandwidths of the rating curves of RQ1 and RQ3 are nearly equal for BR and for water levels below 16 m +NAP for WL, which means that it was possible to reduce the water balance error without changing the rating curve uncertainty bandwidth. Next, for water levels above 16 m +NAP at WL, the uncertainty bandwidth RQ1 is much larger than the uncertainty bandwidth of RQ3, which means that it was possible to reduce the water balance error and the rating curve uncertainty bandwidth both. Finally, the uncertainty bandwidth RQ1 is slightly smaller than the uncertainty bandwidth of RQ3 for PK, which means that reducing the water balance error led to a slight increase in the rating curve uncertainty bandwidth.

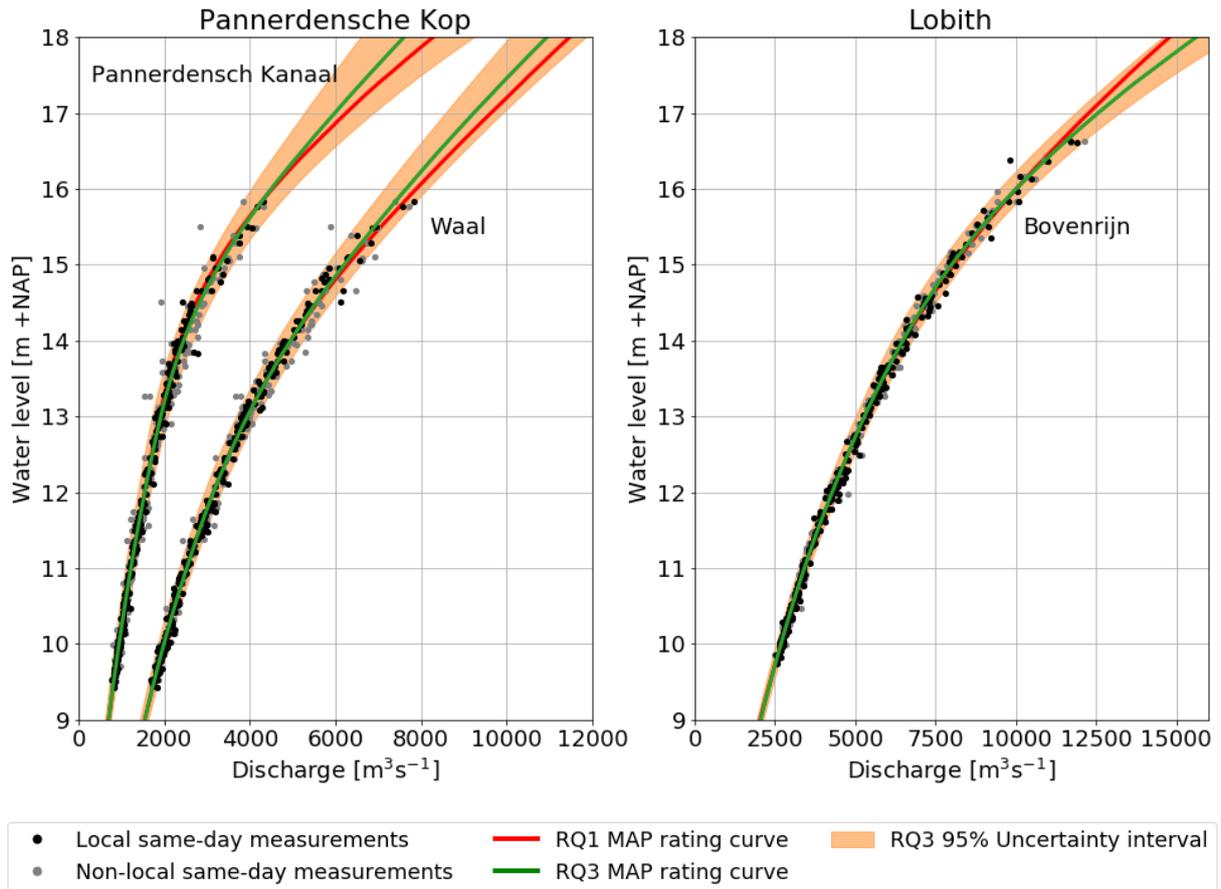


Figure 19: Bayesian rating curve uncertainty based on water balance consideration. The RQ1 MAP rating curve is plotted for comparison.

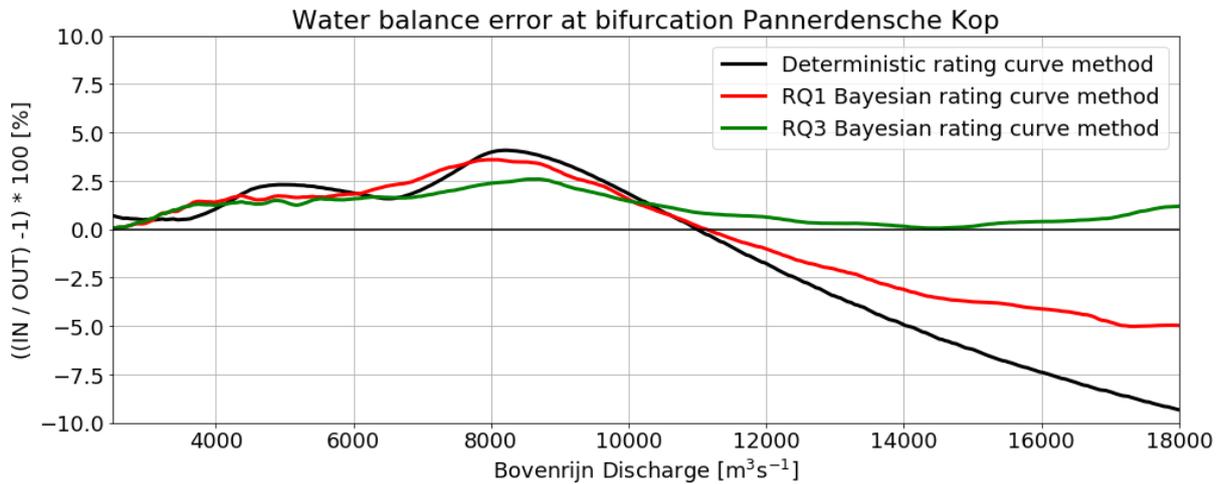


Figure 20. Comparing the water balance error at the Pannerdense Kop of the deterministic rating curves, RQ1 Bayesian rating curves and RQ3 Bayesian rating curves.

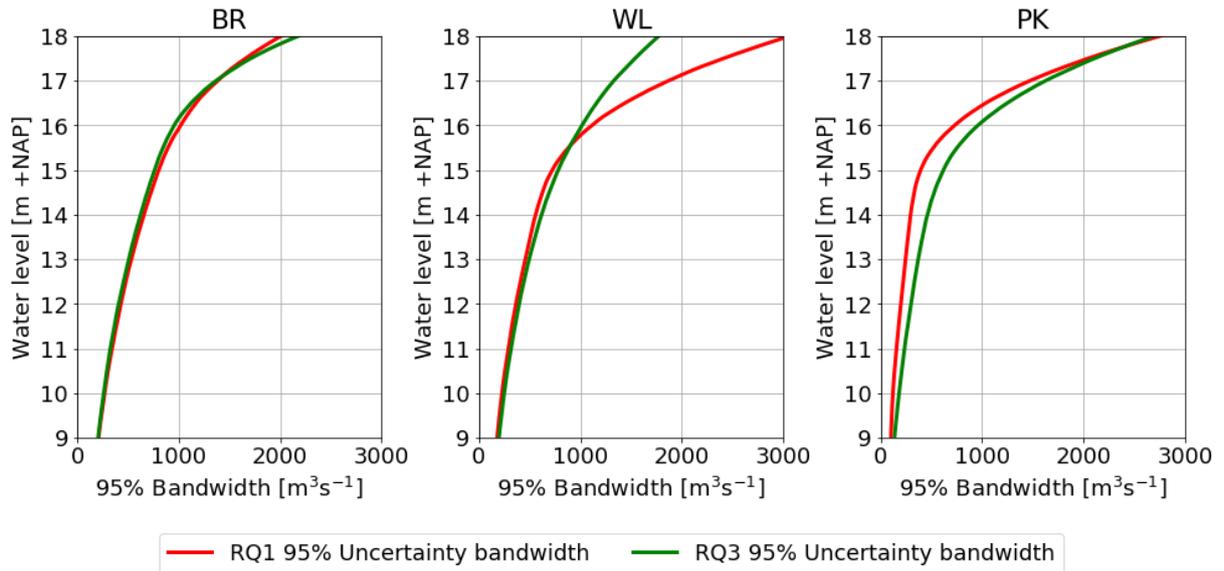


Figure 21. Comparing the 95% uncertainty bandwidths of the RQ1- and RQ3 rating curves for all three locations.

Finally, rating curve uncertainty bandwidths can be influenced by a change in the number of data points and a change in the spread of the data points. For RQ1 and RQ3, roughly the same number of data points were used. For RQ1 (Section 2.3), a number of 568, 464 and 533 data points were used for BR, WL and PK, respectively. For RQ3, per water level point, two same-day discharges values (one local and one calculated non-local) were used, yielding a total of 584 data points. Since for RQ1 and RQ3 roughly the same number of data points were used, it is assumed that this did not influence the rating curve uncertainty. Next, adding the calculated non-local data to the analysis resulted in a different spread of the data points, resulting in a slight reduction of the spread for BR, a slight increase for WL, and a large increase for PK, see Table 12. An increase in spread is introduced by a relatively larger spread of calculated non-local data. The reason why the increase in spread is especially large for PK, is because PK is a relatively small branch on which the uncertainties of the larger branches have a larger effect. The opposite holds for BR, which is the largest branch, on which the uncertainties of the two smaller branches have almost no effect.

Table 12. The influence water balance consideration on the spread (predictive uncertainty) of the data points.

	RQ1	RQ3
Location	Relative standard deviation [%]	Relative standard deviation [%]
BR	2.54%	2.51%
WL	2.90%	3.12%
PK	3.31%	4.35%

4. Discussion

4.1. Bayesian approach

The best method for assessing rating curve uncertainty is an important and open scientific issue (Le Coz et al., 2014). Following other authors (Reitan and Petersen-Øverleir, 2011; Le Coz et al., 2014; Mansanarez et al., 2019; Berends, 2019) Bayesian inference was used to construct stochastic rating curves. Unlike deterministic optimization, where some parameters must have preset fixed values due to too many unknown rating curve parameters, all rating curve parameters are described by probability distributions. This means that an uncertainty analysis can simply be done by using percentiles from the posterior distribution. Furthermore, the main advantage of Bayesian inference is that 'hydraulic knowledge of the stage-discharge relation can be explicitly translated into prior distributions of the parameters of the assumed rating curve equation' (Le Coz et al., 2014), which prevents physically unrealistic outcomes (Mansanarez, 2016). However, this main advantage is also this is also a disadvantage, since hydraulic knowledge of the stage-discharge relation is required and since it is difficult to translate this knowledge into prior distributions of the rating curve parameters. Specifying the priors too informative can lead to poor results. Therefore, it is important carefully consider the degree of how informative the priors are specified.

4.2. Data homogenization

Effects of hysteresis and river interventions were not considered in the homogenization process. Hysteresis is not considered due to time limitation. River interventions are not considered based on the work of Berends (2019). Using the same data set, Berends (2019) found that there is no visible effect of river interventions on water levels and argued that the effect of historical river interventions may be hidden in rating curve uncertainty. By not considering effects of hysteresis and river interventions in the homogenization process, their uncertainty is implicitly included in the rating curve uncertainty. If the effects of hysteresis and river interventions were considered in the homogenization process, the spread of the data points would have been smaller, resulting in less rating curve uncertainty.

4.3. Measurement instrument

For all branches the Ott measurements have a higher predictive uncertainty (representing the spread in the data points) than the ADCP measurements, especially at PK. Note that the predictive uncertainty includes variation from measurement errors, hysteresis effects and seasonal variation. Assuming that seasonal variation and hysteresis effects are equal for both Ott- and ADCP measurement data, the measurement errors are likely higher for helical Ott mills as compared to ADCP instruments. This is in agreement with the common consideration that ADCP measurements are more accurate than Ott measurements (e.g., Le Coz et al., 2014; Mansanarez, 2016). In order to distinguish between the different measurements, a separate analysis has to be done for Ott-mill measurements and ADCP measurements. However, since measurement data is limited, all measurement data were needed for rating curve construction. Therefore, no distinction is made between the different measurement instruments. The consequence of this is that the variation present in the discharge data explained by measurement errors are underestimated for Ott measurement and overestimated for ADCP measurements.

4.4. Assumed floodplain flow

For all three branches, at high discharges the floodplains contain flow. The floodplains are not accessible by measuring boats. Therefore, assumptions are made for floodplain flow which give more uncertainty to the discharge measurements. This can lead to a water balance error of the rating curves. The floodplain of the Waal is large compared to the Bovenrijn and the Pannerdensch Kanaal.

Making an assumption for floodplain flow is therefore most difficult in the Waal. So, the Waal could be (most) responsible for the water balance error of the original rating curves of 2018. However, all three locations were treated as equally reliable in the new method for establishing rating curves based on a closed water balance. The new method clearly reduces the systematic error in water balance. Further improvements in the reduction of the water balance error could be made by making corrections in the discharge measurements, by considering the uncertainty of the assumed floodplain flow. A large floodplain with a complex geometry results in more uncertainty of the assumed floodplain flow, for which a larger correction could be made in the discharge measurements of that particular branch.

4.5. Discharge distribution

Figure 22 shows the total rating curve uncertainty of RQ1 and RQ3 translated into the discharge distribution uncertainty at Pannerdense Kop bifurcation. The hydrodynamic model results of Deltares (2015), who used the 'Splittingspunten model' in WAQUA, have also been plotted in Figure 22 and compare well with the discharge distribution uncertainty resulting from the rating curves of RQ1 and RQ3. Next, Gensen et al. (2018a) concluded that "the variations in discharge distributions in m^3s^{-1} increase for an increasing upstream discharge", which can be seen in Figure 22. Furthermore, based on expert opinions, Steenblik et al. (2020) quantified the uncertainty in the discharge distribution at the Pannerdense Kop. Steenblik et al. (2020) found that for an upstream discharge at Lobith of $16,000 \text{ m}^3\text{s}^{-1}$ (old norm for flood protection in the Netherlands before the new norm was adopted in 2017), the 90% uncertainty bandwidth is $571 \text{ m}^3\text{s}^{-1}$ towards the Waal. Figure 22 shows uncertainty bandwidths at 95% confidence of $908 \text{ m}^3\text{s}^{-1}$ and $992 \text{ m}^3\text{s}^{-1}$ towards the Waal, for respectively RQ1 and RQ3 for an upstream discharge at Lobith of $16,000 \text{ m}^3\text{s}^{-1}$. Note that Steenblik et al. (2020) obtained a bandwidth at a 90% confidence level and in this thesis, bandwidths are obtained at a 95% confidence level, which does not allow a direct comparison. However, the results of Steenblik et al. (2020) still show smaller uncertainty bandwidths. Next, Gensen et al. (2018b) studied the effect of main channel roughness uncertainty the discharge distribution uncertainty of the Pannerdense Kop for upstream discharge at Lobith of $16,000 \text{ m}^3\text{s}^{-1}$. Gensen et al. (2018b) used deterministic scenarios, and for their extreme scenarios obtained a bandwidth of roughly $1000 \text{ m}^3\text{s}^{-1}$, which is similar the bandwidth of Figure 22.

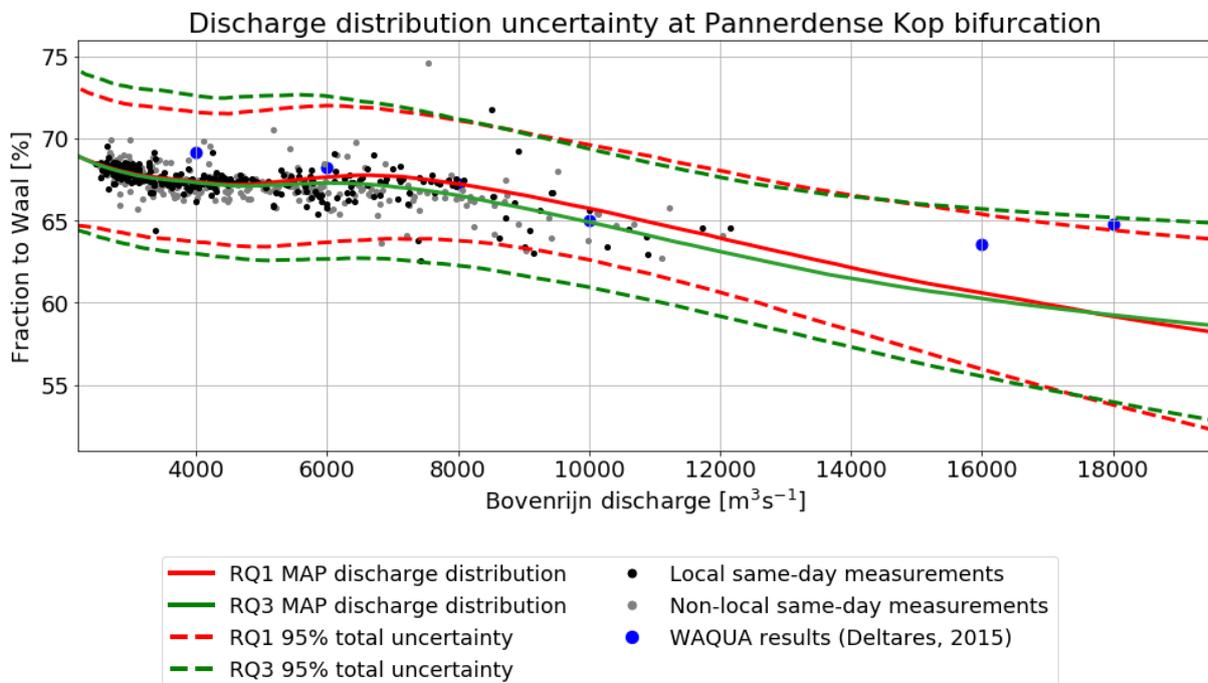


Figure 22. Discharge distribution uncertainty at Pannerdense Kop bifurcation as based on total rating curve uncertainty results from RQ1 and RQ3.

Next, rating curve uncertainty translates into discharge distribution uncertainty, which in turn translates to uncertainty of downstream water level predictions. “The effect of varying discharge distributions decreases the range of water levels along the Waal, while it increases the range for the Nederrijn and IJssel” (Gensen et al., 2018b). In the Dutch flood risk framework, the probabilities of water levels need to be calculated. This discharge distribution uncertainty presented in Figure 22 could be used in a model study, to determine the effect on the water levels along the Rhine branches for a number of discharges.

4.6. Flood management norms

Since rating curves are essential in the construction of discharge time-series from water levels and in the calibration of river models, it is important that the systematic error in rating curves are removed as much as possible. Especially if these discharge time-series and calibrated models are used to define and hydraulically model design flood events. Hydrodynamic models treat the river system as a closed water balance. To prevent systematic under- or overestimation of discharges or water levels on entire river branches, there must be no systematic error in the water balance and water levels of rating curves. In the Netherlands, the design discharge for the Rhine river network is far beyond any event that has ever been observed. Therefore, it is of utmost importance that models used for development of flood management norms and regulations do not contain systematic effects that distort realistic system behavior.

5. Conclusion and recommendations

5.1. Conclusion

5.1.1. RQ1: Rating curve uncertainty

The aim of this research question is to determine how large the uncertainty bands of the single rating curves for the three branches are. Therefore, Bayesian inference and Markov Chain Monte-Carlo (MCMC) were used to construct the total uncertainty intervals of the rating curves for the three branches, as based on homogenized measurement data. The total uncertainty bandwidths increase for higher water levels. An indication of the total rating curve uncertainty bandwidths is as follows: for local water levels of 12 m +NAP the total rating curve uncertainty bandwidths are 430.1 m³s⁻¹, 364.5 m³s⁻¹ and 199.8 m³s⁻¹, for BR, WL and PK, respectively, and for local water levels of 16 m +NAP the total rating curve uncertainty bandwidths are 1018.6 m³s⁻¹, 1113.7 m³s⁻¹ and 727.9 m³s⁻¹, for BR, WL and PK, respectively.

5.1.2. RQ2: Influence of measurement errors

The aim of this research question is to determine the extend into which the variation present in the measurement data can be explained by measurement errors. Similar to Le Coz et al. (2014), ISO reports and scientific literature are used to determine the measurement errors (3 and 5%). These measurement errors were imposed in Bayesian inference and MCMC simulations, resulting in posterior distributions of the remaining predictive uncertainty (hysteresis effects and seasonal changes). From all three branches, for equal measurement errors, the measurement uncertainty is most dominant in BR. There is a large difference in variation explained by measurement errors when assuming discharge measurement errors of 3% versus 5%, especially for BR. However, it remains unknown what the exact measurement error is per location. The measurement uncertainty is likely highest for WL, since more measurement uncertainty is introduced due to assumed floodplain flow in the relatively large and complex cross-section of the floodplain of WL. Next, even though variation due to hysteresis is still present in the data, there is little remaining uncertainty for $\sigma_m=5\%$ (especially for BR), which indicates that variation due to hysteresis effects is small compared to variation of measurement errors. A possible explanation for this might be the validation by data owner Rijkswaterstaat, since outlier removal reduces the variation in the data.

5.1.3. RQ3: Water balance in rating curves

The aim of this research question is to propose a method to establish rating curves based on water balance closure and to see how this influences the uncertainty bands of the single rating curves for the three branches. Novel in the proposed method is that discharge measurements from other locations are included. Per water level point, there are two same-day discharges values, one local and one (calculated) non-local, yielding a total of 584 data points. In current practice and in RQ1, rating curves are constructed by only using locally measured stage and discharge as input data. Similar to RQ1, Bayesian rating curves were constructed for all three locations. The rating curves of the proposed method have shifted as compared to the rating curves of RQ1, which clearly reduced the water balance error between the three locations. So, adding the calculated non-local data points to the analysis clearly reduces the systematic error in water balance and thereby it provides more consistent rating curves for the river network of the Dutch Rhine.

Next, the proposed method influenced the uncertainty bands of the rating curves, as compared to the rating curves of research question 1. The uncertainty bandwidths of the RQ1- and RQ3 rating curves are nearly equal for BR and for water levels below 16 m +NAP for WL, which means that it was possible to reduce the water balance error without changing the rating curve uncertainty bandwidth. Next, for water levels above 16 m +NAP at WL, the uncertainty bandwidth RQ1 is much larger than the uncertainty bandwidth of RQ3, which means that it was possible to reduce the water balance error and the rating curve uncertainty bandwidth both. Finally, the uncertainty bandwidth RQ1 is slightly smaller than the uncertainty bandwidth of RQ3 for PK, which means that reducing the water balance error led to a slight increase in the rating curve uncertainty bandwidth.

Finally, the rating curve uncertainty bandwidths can be influenced by a change in the number of data points and a change in the spread of the data points. For RQ1 and RQ3, roughly the same number of data points were used. Therefore, it is assumed that this did not influence the rating curve uncertainty. Next, adding the calculated non-local data to the analysis resulted in a different spread of the data points, resulting in a slight reduction of the spread for BR, a slight increase for WL, and a large increase for PK. An increase in spread is introduced by a relatively larger spread of calculated non-local data. The reason why the increase in spread is especially large for PK, is because PK is a relatively small branch on which the uncertainties of the larger branches have a larger effect. The opposite holds for BR, which is the largest branch, on which the uncertainties of the two smaller branches have almost no effect.

5.2. Recommendations

Based on this thesis on the water balance in the Dutch river Rhine and uncertainty of rating curves, recommendations can be made regarding hysteresis effects, assumed floodplain flow and measurement strategy.

For future research it is recommended to extend this work by correcting for hysteresis effects in the homogenization of the dataset. This will reduce the spread of the data points, resulting in less rating curve uncertainty and improving the quality of the results.

Next, for Rijkswaterstaat it is recommended to consider the uncertainty of the assumed floodplain flow of the different locations. This could be done by making corrections in the discharge measurements of the different locations. A large floodplain with a complex geometry results in more uncertainty of the assumed floodplain flow, for which larger corrections could be made in the discharge measurements of that particular location. This will improve the accuracy and consistency of the rating curves.

Finally, for Rijkswaterstaat it is recommended to take more same-day measurements for future measurement campaigns. In addition, it is recommended to minimize the timespan in the which the same-day measurements are taken to reduce uncertainty caused by delay. The same-day discharge measurements of the three locations can be compared to check whether they are consistent from the perspective of a closing water balance. Also, same-day measurement can be used in combination with the proposed method, to impose water balance closure in the establishment of rating curves, which will improve the accuracy and consistency of the rating curves for the river network of the Dutch Rhine.

6. Acknowledgements

The presented thesis was carried out at HKV IJN in water. The field data was provided by Susanne Quartel from the Dutch Ministry of Infrastructure and Water Management (Rijkswaterstaat). This thesis is part of the Perspectief research programme All-Risk with project number P15-21, which is (partly) financed by NWO Domain Applied and Engineering Sciences, in collaboration with the following private and public partners: Rijkswaterstaat, Deltares, STOWA, HKV consultants, Natuurmonumenten and the regional water authorities Noorderzijlvest, Vechtstromen, Fryskje Gea, HHNK.

References

- Berends, K. (2019). *Human intervention in rivers: quantifying the uncertainty of hydraulic model predictions*. Enschede: University of Twente. doi:10.3390/1.9789036548823
- Berends, K., Gensen, M., Warmink, J., & Hulscher, S. (2020). Thirty-year analysis of a changing river system using Bayesian rating curves. *River Flow 2020*.
- Buschman, F., Blom, A., Dijk, T. v., Kleinhans, M., & Mark, R. v. (2017). *Informatiebehoefte en aanbevelingen voor monitoring in de Bovendelta van de Rijn*. Deltares.
- Deltares. (2015). *5th Generation WAQUA Subdomain Models of Rijntakken*. Delft.
- DHV, & DelftHydraulics. (1999). *Hydrology Project Training Module: How to establish stage discharge rating curve, SWDP-29*. New Delhi: DHV Consultants BV & Delft Hydraulics.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian analysis 1* (3), 515-533.
- Gensen, M. R., Warmink, J. J., & Hulscher, S. J. (2018a). *Water level uncertainties due to uncertain bedform dynamics in the Dutch Rhine system*.
- Gensen, M. R., Warmink, J. J., & Hulscher, S. J. (2018b). *Propagating main channel roughness uncertainty in the bifurcating Dutch Rhine system*. Enschede: University of Twente.
- Herschly, R. (2009). *Streamflow measurement*. London: Routledge.
- HKV. (2009). *Beheer en onderhoud afvoerreksen Rijntakken, deelonderzoek A: Operationaliseren Qf-relatie Bovenrijn, Waal, Pannerdensch Kanaal, Nederrijn en IJssel, rapport PR1444.20*.
- HKV. (2016). *Vernieuwen Qf-relatie Rijntakken*.
- Hoffman, M., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 1593-1623.
- International Standards Organization (ISO). (2007). *Hydrometry – Velocity-area methods using current-meters – Collection and processing of data for determination of uncertainties in flow measurement, ISO 1088:2007*. Geneva, Switzerland.
- International Standards Organization (ISO). (2010). *Hydrometry – Measurement of liquid flow in open channels – Part 2: Determination of the stage-discharge relationship, ISO 1100-2:2010*. Geneva, Switzerland.
- International Standards Organization (ISO). (2012). *Hydrometry - Acoustic Doppler profiler - Method and application for measurement of flow in open channels, ISO/TR 24578:2012*. Geneva, Switzerland.
- Kean, J. W., & Smith, J. D. (2010). Calculation of stage-discharge relations for gravel bedded channels. *Journal of Geophysical Research*, F03020.
- Le Coz, J. (2012). *A literature review of methods for estimating the uncertainty associated with stage-discharge relations*. Lyon, France: Cemagref, Hydrology-Hydraulics.

- Le Coz, J., Renard, B., Bonnifait, L., Branger, F., & Boursicaud, R. L. (2014). Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: a Bayesian approach. 573-587. doi:10.1016/j.jhydrol.2013.11.016
- Mansanarez, V. (2016). *Non-unique stage-discharge relations : Bayesian analysis of complex rating curves and their uncertainties*. Université Grenoble Alpes: Hydrology.
- Mansanarez, V., Renard, B., Coz, J., Lang, M., & Darienzo, M. (2019). Shift happens! Adjusting Stage-Discharge Rating Curves to Morphological Changes at Known Times. *Water Resources Research*, 2876-2899. doi:10.1029/2018wr023389
- Pappenberger, F., Matgen, P., Beven, J., Henry, B., Pfister, L., & Fraipont, P. d. (2006). Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Advanced Water Resources*, 1430–1449.
- Quartel, S., Veen, R. v., Wijbenga, A., Berben, F., Kok, F., & Heinen, P. (2011). Dynamic stage-discharge relations of the Dutch Rhine branches. In *NCR days 2011: controlling the Dutch rivers* (pp. 35-36). Delft: Delft University of Technology.
- Rantz, S. (1982). *Measurement and computation of streamflow, Volume 2, Computation of discharge*. Washington: Water-Supply Paper 2175. U. S. Geological Survey.
- Reeze, B., Winden, A. v., Postma, J., Pot, R., Hop, J., & Liefveld, W. (2017). *Watersysteemrapportage Rijntakken 1990-2015*. Harderwijk: Ontwikkelingen waterkwaliteit en ecologie. Bart Reeze Water & Ecologie.
- Reitan, T. & Ø. (2011). Dynamic rating curve assessment in unstable rivers using Ornstein-Uhlenbeck processes. *Water Resources Research*. doi:10.1029/2010WR009504
- Rijkswaterstaat. (2015). *Nieuw stuurprogramma voor de Neder-Rijn/Lek*. Arnhem: Rijkswaterstaat Oost-Nederland.
- Schmidt, A. (2002). *Analysis of stage-discharge relations for open-channel flows and their associated uncertainties*. Illinois: PhD thesis, University of Illinois, Urbana-Champaign.
- Steenblik, S., Gensen, M., Warmink, J., & Hulscher, S. (2020). Quantifying uncertainties in the discharge distribution over the river Rhine branches using expert elicitation. *River Flow 2020*.
- STOWA. (2009). *Handboek debietmeten in open waterlopen*. Kruyt Grafisch Adviesbureau.
- Vervoorn, H. (1998). *QH-analyses Rijntakken: Vervolgonderzoek: Eindverslag*. Nijmegen: Royal Haskoning DHV.
- World Meteorological Organization (WMO). (2010). *Manual on Stream Gauging. Volume I – Fieldwork, No. 1044*. Geneva, Switzerland.

Appendix A: Physical background rating curve model

In this section, the rating curve model that is used in this thesis and the underlying physical basis is shown. Neglecting the physical background will lead to poor rating curve quality (Le Coz, 2012) and rating curve construction should not simply be considered a mathematical fitting operation without any physical basis (DHV & DelftHydraulics, 1999).

For a channel with a rectangular cross-section, the total discharge is given as:

$$Q = \langle u \rangle (h - b)W \quad (\text{A-1})$$

with total discharge Q [m^3s^{-1}], cross-sectionally averaged flow velocity $\langle u \rangle$ [$m s^{-1}$], water level h [$m + NAP$], bed level b [$m + NAP$] and channel width W [m]. Using the Chézy equation to describe steady uniform flow in a wide channel, the equation becomes:

$$Q = WCS_b^{1/2}(h - b)^{3/2} \quad (\text{A-2})$$

with Chézy coefficient C [$m^{1/2}s^{-1}$] and slope S_b [-]. Using the often-used empirical Manning-Strickler equation $C = (h - b)^{1/6}n^{-1}$, the equation becomes:

$$Q = Wn^{-1}S_b^{1/2}(h - b)^{5/3} \quad (\text{A-3})$$

with Manning coefficient n [$s m^{-1/3}$]. Usually, this equation forms the physical basis for empirical rating curve models. In practice this equation is simplified, which is acceptable for its hydrometric purposes, to the following commonly used one-dimensional power function [adapted from WMO, 2010; ISO, 2010]:

$$Q = a(h - b)^p, \quad h > b \quad (\text{A-4})$$

where parameter $a = Wn^{-1}S_b^{1/2}$ [$m^{4/3}s^{-1}$] is related to the channel characteristics, parameter b [$m + NAP$] is the cease-to-flow reference level (sometimes written h_0) and parameter p [-] is the hydraulic exponent related to the hydraulic control. The parameters are determined based on available stage discharge data. The power function is derived from the Manning-Strickler equation and assumes steady uniform flow in a wide rectangular channel. Therefore, the hydraulic exponent p can deviate roughly ± 0.1 from the typical value of $5/3$ (≈ 1.667) (Le Coz et al., 2014).

The cross-sectional geometry of the Rhine branches is too complex to be modelled as a single rectangular cross-section. Therefore, the divided channel approach is applied to approximate the complex geometries by multiple rectangular cross-sections. The general form is given as a summation of the power function:

$$Q = \sum_{i=0}^N a_i (h - b_i)^{+p_i}, \quad h > b_i \quad (\text{A-5})$$

with number of subsections N [-] and subsections i [-]. Following HKV (2009), three subsections are assumed, representing the equivalent of the main channel, groin fields and flood plains. These subsections represent different channel characteristics and hydraulic controls, which can be expressed in the parameters of power function of Equation A-5. Whether a subsection is active (containing flow), depends on the river stage h with respect to the cease-to-flow levels b_i of the subsection. $N = 3$ is used, where the cease-to-flow levels b_0, b_1, b_2 are related to the position of the main channel bed level, the height of the groin crests and the flood plain level respectively. The cease-to-flow levels are schematizations. Using $N < 3$ will result in high hydraulic exponents p_i , giving a weak physical basis for the rating curve model. Therefore, $N = 3$ is required.

Appendix B: Rating curve parameter uncertainty

In this section the parameter uncertainties of RQ1 are presented for the three locations, to give an indication of the sensitivity of the parameters. Figure B-1 presents the parameter uncertainty of Lobith – Bovenrijn, Figure B-2 presents the parameter uncertainty of Pannerdense Kop – Waal, and Figure B-3 presents the parameter uncertainty of Pannerdense Kop – Pannerdensch Kanaal.

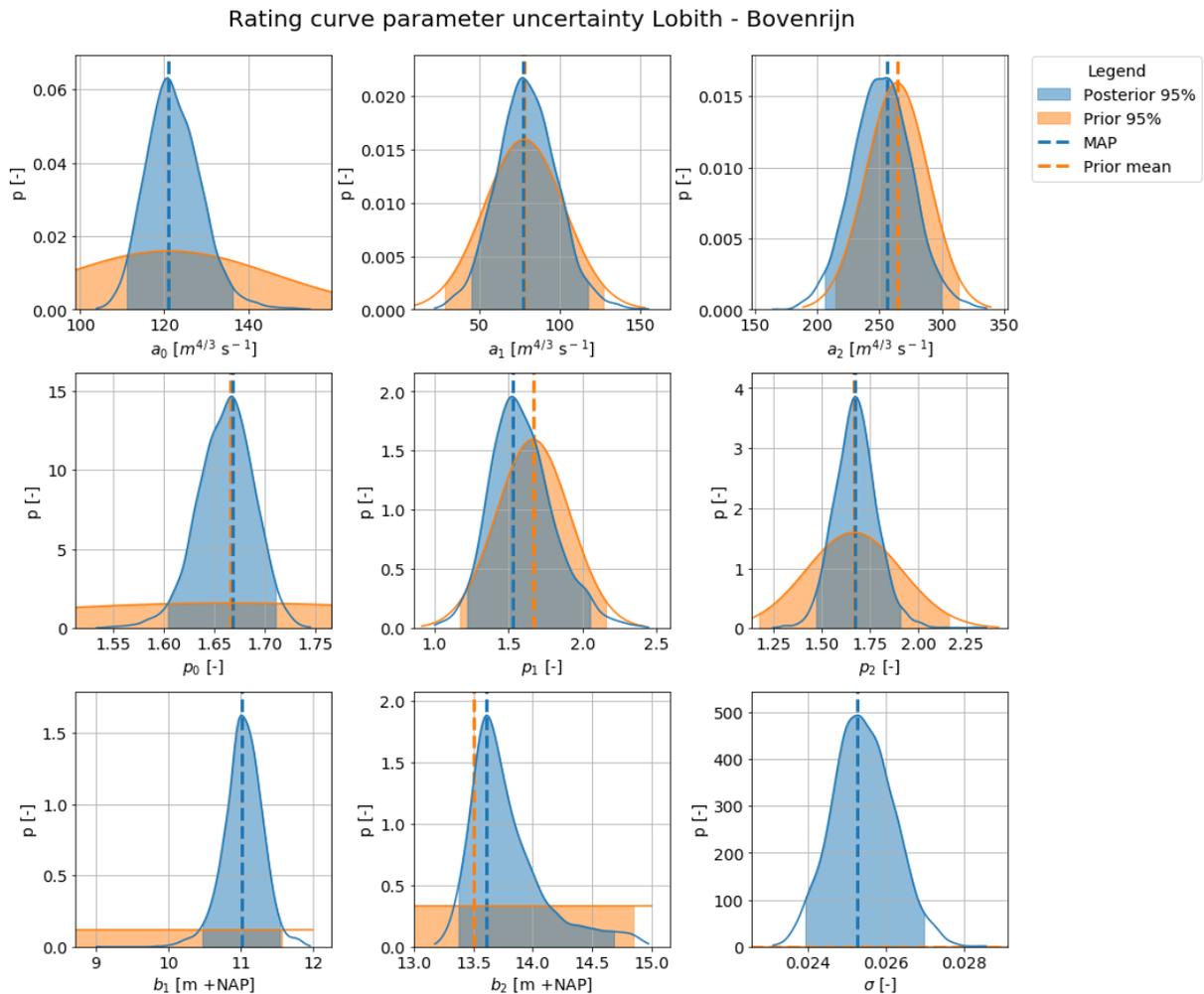


Figure B-1. Prior- and posterior distributions of the rating curve parameter uncertainty for Lobith – Bovenrijn.

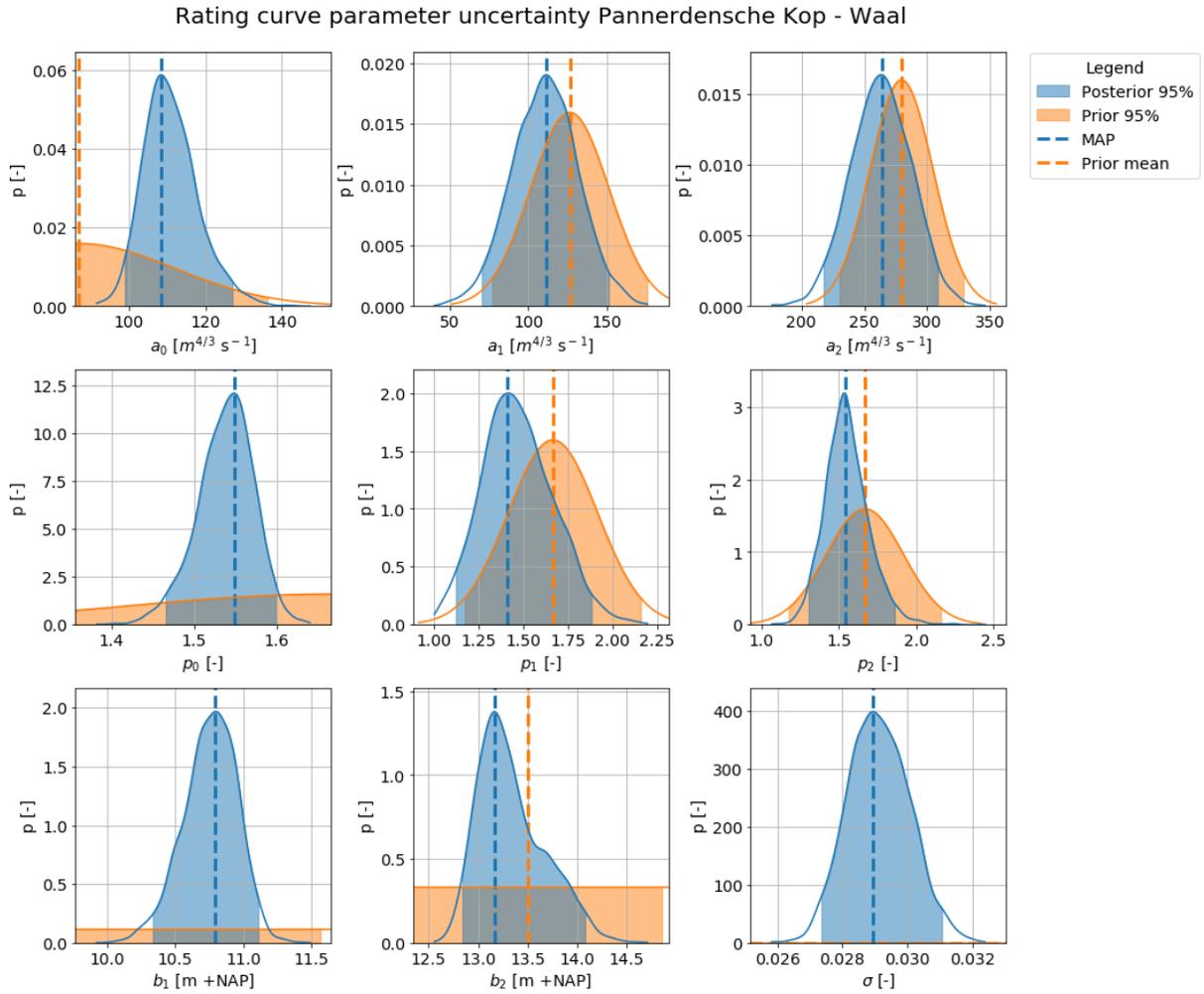


Figure B-2. Prior- and posterior distributions of the rating curve parameter uncertainty for Pannerdense Kop – Waal.

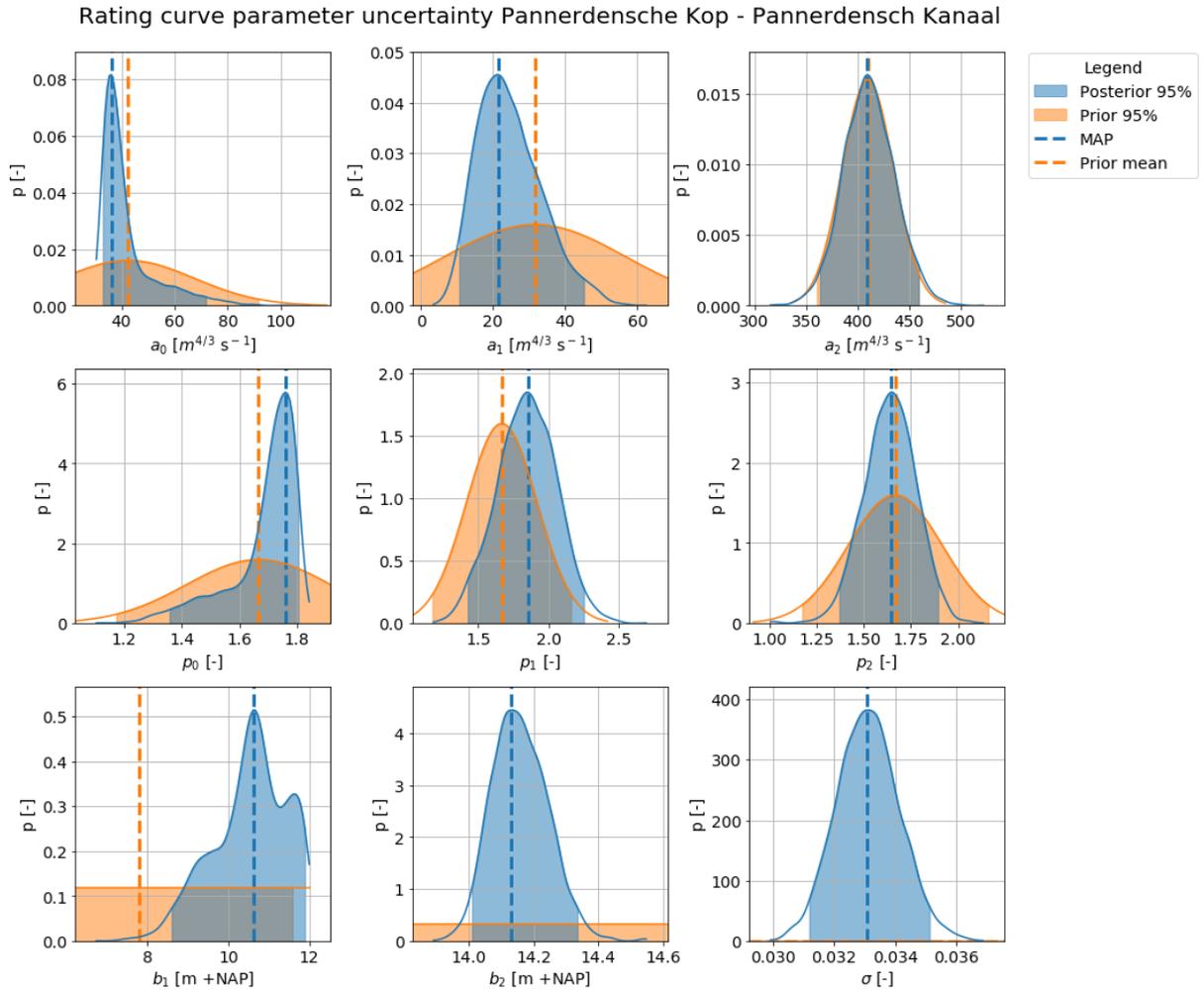


Figure B-3. Prior- and posterior distributions of the rating curve parameter uncertainty for Pannerdensch Kop – Pannerdensch Kanaal.

Appendix C: Remaining predictive uncertainty

In Figure C-1 the posterior distributions of the remaining uncertainty of RQ2 are presented for the three locations for the outer values (3 and 5%) of measurements errors, to give an indication of the sensitivity.

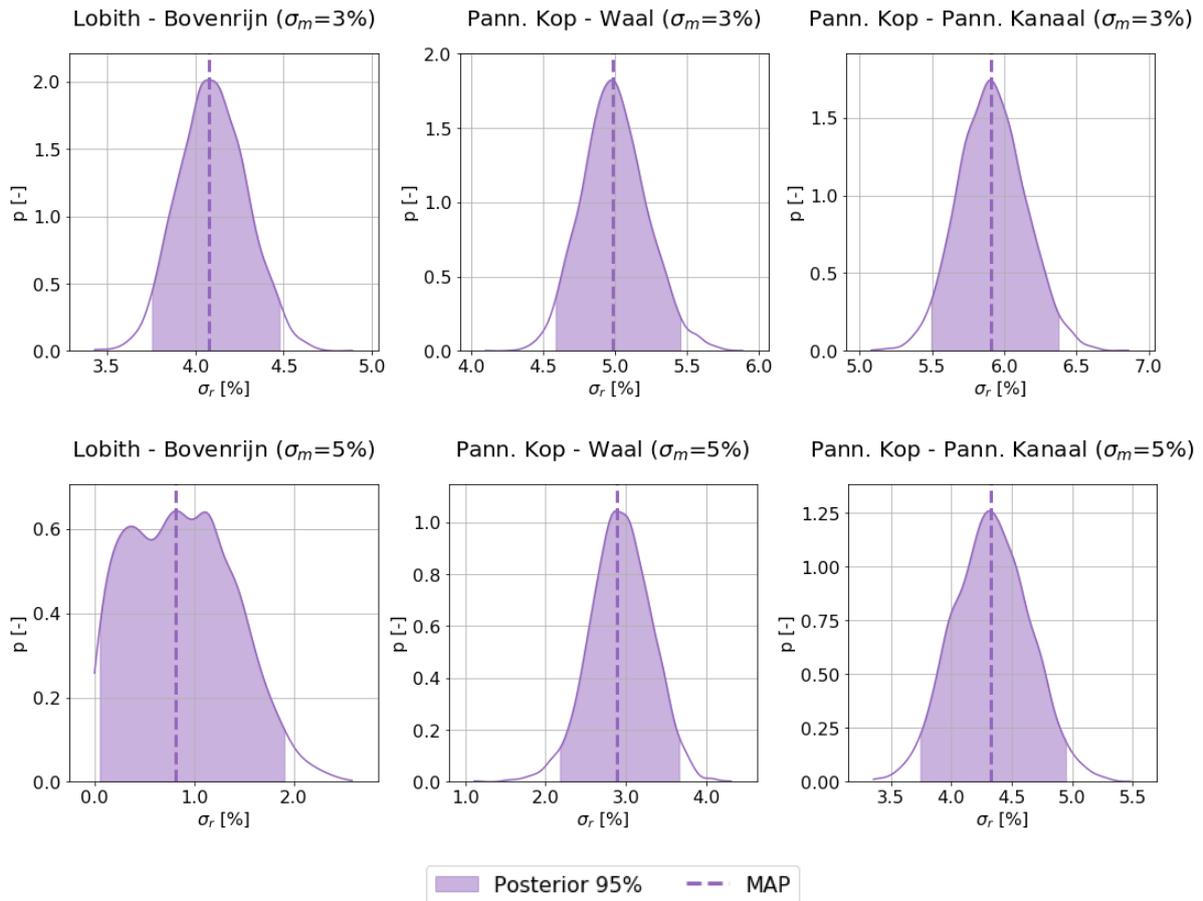


Figure C-1. Posterior distributions of the remaining uncertainty.