Discharge Rating Curve Using Bayesian Statistics

Kristinn Mar Ingimarsson

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
University of Iceland
2010

DISCHARGE RATING CURVE USING BAYESIAN STATISTICS

Kristinn Mar Ingimarsson

60 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Industrial Engineering

Advisor
Birgir Hrafnkelsson
Sigurdur M. Gardarsson
Olafur P. Palsson

Faculty Representative
Daniel F. Gudbjartsson

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, January 2010

Discharge Rating Curve Using Bayesian Statistics
60 ECTS thesis submitted in partial fulfillment of a M.Sc. degree in Industrial Engineering

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
VRII, Hjardarhagi 2-6
107, Reykjavik, Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Kristinn Mar Ingimarsson, 2010, Discharge Rating Curve Using Bayesian Statistics, M.Sc. thesis, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland.

ISBN XX

Printing: Háskólaprent, Fálkagata 2, 107 Reykjavík
Reykjavik, Iceland, January 2010

ABSTRACT

Discharge in rivers is commonly estimated by the use of rating curve constructed from pairs of water level and discharge measurements. Water level measurements are collected continuously from each river while pairs of discharge and water level are only collected couple times a year due to high cost. The need for accurate estimation of discharge is important for constructions as bridges, hydroelectric power plants as well as for hydrological models. The methodology currently used by the Icelandic Meteorological Office is based on the standard power-law. The power-law is derived from a theoretical basis and serves as an appropriate model in most cases. However, in some natural settings deviations from this form arise. The new methodologies presented in this thesis account for the deviations from the standard power-law by extending it with a smooth B-spline function or by assuming two of the are a function of water level and modeled them with B-splines. These methodologies have shown to perform equally well or better than the current methodology used by the Icelandic Meteorological Office.

ÚTDRÁTTUR

Rennsli í ám er oft metið útfrá rennslislyklum sem eru smíðaður útfrá pari af vatnshæðar og rennslismælingum. Vatnshæðarmælingum er safnað samfellt frá hverri á en par af rennslis- og vatnshæðarmælingum eru einungis safnað nokkrum sinnum á ári vegna mikils kostnaðar. Milkil þörf er á nákvæmu mati á rennsli við hönnun á byggingum á borð við brýr, vatnsaflsvirkjanir og einnig við gerð vatnalíkana. Aðferðafræðin sem er notuð á Veðurstofu Íslands er byggð á standard power-law jöfnunni. Standard power-law er dregið úr fræðilegum grunni og er viðeigandi líkan í flestum tilvikum, en í náttúrulegum aðstæðum geta frávik frá þessari jöfnu komið upp. Hinar nýju aðferðir sem kynntar eru í þessari ritgerð gera ráð fyrir fráviki frá standard power-law jöfnunni með því að útvíkka hana með B-spline föllum. En það er gert með því að leggja B-spline föllin við standard power-law jöfnuna eða með því að taka B-spline föllin inn í parametrana. Þessar aðferðir hafa sýnt að þær eru jafn góðar eða betri en núverandi aðferðir sem eru notaðar á Veðurstofu Íslands.

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

This M.Sc. project was carried out at the Faculty of Industrial Engineering, Mechanical Engineering and Computer Science in the School of Engineering and Natural Sciences, University of Iceland.

First of all I would like to thank my supervisors, associate research professor Birgir Hrafnkelsson at the Science Institute, professor Sigurdur M. Gardarsson at Faculty of Civil and Environmental Engineering and professor Ólafur P. Pálsson at Faculty of Industrial, all at the University of Iceland, for their time, co-operation and support throughout this work. I would specially like to thank Birgir for great times at the conferences in Australia and all the extra hours. Furthermore would I like to thank the staff at The Icelandic Meteorological Office for all the help and the deputy for School of Engineering and Natural Sciences: Daníel F. Guðbjartsson at deCODE genetics. In addition I would like to thank my family, friends and my partner, Hrafnhildur Ólafsdóttir, for their support and encouragement throughout my studies. The study was sponsored by Rannís - Research fund and during my time of study I was employed by The Icelandic Meteorological Office.

Reykjavik, August 2009.
Kristinn Már Ingimarsson.

1 INTRODUCTION

1.1 Goals of the Project

The main goal of the project described in this thesis is to create an objective methodology for establishing hydrological rating curves based on the Bayesian approach which can be applied to data from all rivers in Icelandic Meteorological Office (IMO) data base. A secondary goal is to make the estimation of discharge rating curves accessible through an user-friendly computer program.

1.2 Rating Curve

Discharge in rivers is commonly calculated by mapping water surface elevations, measured at a specific location in the river, to discharge by means of a rating curve. The rating curve is usually an equation that describes a curve that is fitted through data points of measured water surface elevation against measured discharge at a location where downstream hydraulic control assures a stable, sensitive and monotonic relationship between water surface elevation and discharge (Mosley and McKerchar, 1993; ISO, 1983). This methodology is applied as direct measurements of discharge are expensive compared to measurement of water surface elevation that are relatively straightforward and inexpensive undertaking and often well suited for automation. The sources of uncertainty in the discharge obtained by a rating curve methodology are several; both due to uncertainty in river discharge measurements and uncertainty in the rating curve (Pelletier, 1988; Clarke, 1999; Moyeed and Clarke, 2005; Di Baldassarre and Montanari, 2009).

In this thesis, a methodology for improved fit to the data points and improved extrapolation of the rating curve for large discharges is proposed, based on the Bayesian approach and B-spline functions. Based on hydraulic principles, the relationship between discharge and water level is given by the standard power-law

$$q = a(w-c)^b \tag{1}$$

(Lambie, 1978; Mosley and McKerchar, 1993) where q is discharge, w is water level, a is a positive scaling parameter, b is a positive shape parameter and c is the water level when the discharge is zero. These parameters are usually estimated from paired measurements of water level and discharge. The Bayesian approach has been successfully applied to the estimate of discharge rating curve (Moyeed and Clarke (2005), Arnason (2005) and Reitan and Petersen-Øverleir (2008b)). However, an application of this method has shown that it can not handle all data sets in the IMO database. That is due to the fact that in natural setting the shape of the riverbed can change with rising water level. To demonstrate this behavior a riverbed is plotted in Figure 1. The parameter b in equation (1) represent the shape of the riverbed, for example it takes the values 1.5 and 2.5 for rectangular and v-shaped sections, respectively.

In Figure 1 it can be seen how the shape of the riverbed can be different from a rectangular of v-shaped section for example. A common practice has been to use multi-segment discharge rating curves where the shape parameter is different between segments which has been modeled by (Petersen-Øverleir and Reitan, 2005; Reitan and Petersen-Øverleir, 2008a). Here different approaches are introduced which assume that changes in the rating curve occur gradually.

Figure 1: *An example of a riverbed.*

All methods that are presented here use the Bayesian approach. In the Bayesian approach all unknown parameters are treated as random variables. Prior information about unknown parameters based on previously collected data and/or scientific knowledge can be combined with new data for parametric inference. The advantage is that all uncertainty can be taken into account which allows for an accurate inference about the unknown parameters. Prediction intervals for discharge can be evaluated accurately and thus it is possible to have a criteria that shows if the new measurements are in line with the behavior of the river or if the riverbed has changed.

1.3 Computer Program

The calculations are conducted with the software Matlab. The Matlab programs can be time consuming to read through and work with, especially if the staff has not worked with Matlab before. In addition people working with the rating curves should only be allowed to change few parameters in the program and should not be allowed to change the algorithms. Therefore a user-interface was created that connects the user to the Matlab programs. Therefore a user-interface was created that connects the user to the Matlab codes. Matlab then runs the calculations and stores the results in a well defined folder. The interface is shown in Figure 2. The user-interface allows the user to plot up the data set and see how well the starting values for the parameters in the standard power-law fit the data set before running the calculations. The user can choose the starting values by filling in for a, b, w_0 or by filling only in for one or two of them and let the program find the optimal solution for the rest of the parameters, that is if the optimization gives an optimal values. The user can choose from three different types of Bayesian models which are the models presented in this thesis. It can be necessary to change the prior distributions however the user is only allowed to change the priors for the parameters b and c which is marked as w_0 in the interface. The user can also decide on the length of the run, the number of B-spline kernels and at what water level the B-spline affects the rating curve. Another important property is that the user can skip newest pairs to see if they fit within the prediction interval of the data excluding them, than a rating curve is plotted and the excluded measurements are plotted as x-is to separate them from the others. The output of the program is a report with all information desired by the IMO. All the changes can be made from the user-interface shown in Figure 2 and user never has to code in Matlab.

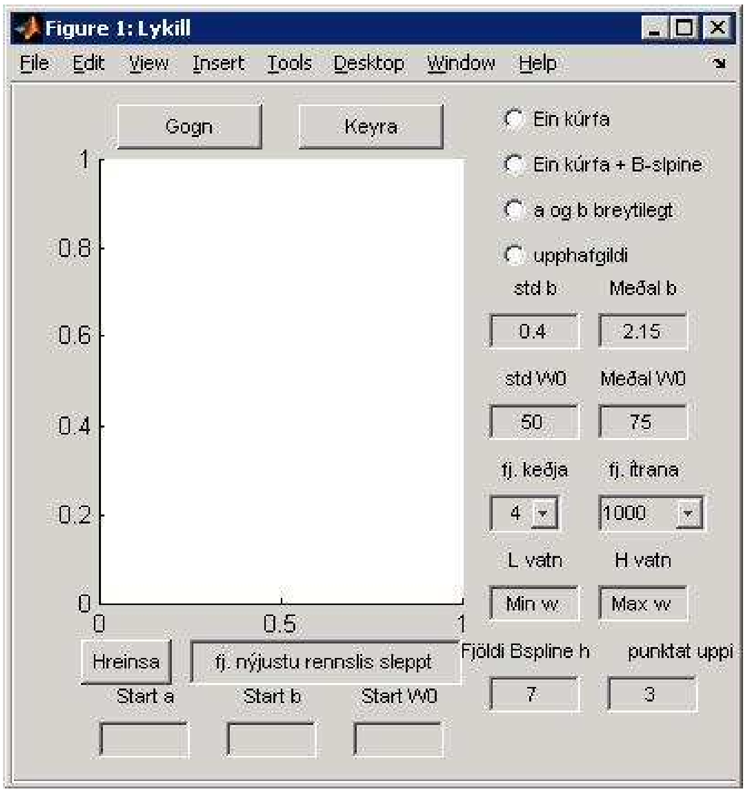

Figure 2: *The window of the Matlab user-interface.*

1.4 Scope of the Work

Three papers are presented in this thesis. The paper by Ingimarsson et al. (2008) is a conference paper and was published in XXV Nordic Hydrological Conference, Nordic Association for Hydrology, 2008, Volume 1, 308-317. In Ingimarsson et al. (2008) the standard power-law is extended by a smooth B-spline function and compared to the model presented in Moyeed and Clarke (2005) and Arnason (2005). The paper Ingimarsson et al. (2010a) was submitted to Hydrology and Earth Systems Sciences. In this paper the same models as in Ingimarsson et al. (2008) are compared, however the B-spline model has been modified from Ingimarsson et al. (2008). The extended model provides substantially better fit than the standard power-law model for about 30% of the data sets and when the standard power-law appears to give an adequate fit, the extended model imitates the standard power-law model. The extended model also performs better for 60% of the rivers when predicting large discharge values. The models and the results in Ingimarsson et al. (2008) and Ingimarsson et al. (2010a) have been presented in the following conferences and workshops. They were presented in a poster at Bayesian Environmetrics Workshop, Brisbane, Australia, at the 9th International Society for Bayesian Analysis (ISBA) conference 2008, Hamilton Island, Australia and at the XXV Nordic Hydrological Conference 2008 Reykjavik, Iceland. They were presented as a talk in Statistics colloquium at the University of Iceland, April 2008, at the Norwegian Computing Center at the Statistics for Innovation, Oslo, Norway, 24 September, 2008 and Nordic-Baltic Biometric Conference, 10-12 June 2009, Tartu, Estonia. Due to the fact that Ingimarsson et al. (2008) is an earlier version of Ingimarsson et al. (2010a), the paper Ingimarsson et al. (2008) is given in appendix. In the paper Ingimarsson et al.

(2010b) the B-spline model in Ingimarsson et al. (2010a) is compared to three models in which B-splines are used to estimate the parameters of *a* and *b* as a function of water level. These three models differ only in the variance function. The variance function of only one of the new models has the flexibility to give fit that is similar to that of the B-spline model presented in Ingimarsson et al. (2010a) for all data sets. There are however cases were the models in Ingimarsson et al. (2010b) clearly outperforms the B-spline model in Ingimarsson et al. (2010a) and vice versa.

2 BAYESIAN DISCHARGE RATING CURVES BASED ON B-SPLINE SMOOTHING FUNCTIONS

Kristinn Mar Ingimarsson[1,4], Birgir Hrafnkelsson[2],
Sigurdur M. Gardarsson[3] and Arni Snorrason[4]

[1]Faculty of Industrial Engineering, Mechanical Engineering and Computer Sciences,
University of Iceland,
[2]Science Institute, University of Iceland,
[3]Faculty of Civil Engineering and Environmental Engineering, University of Iceland,
[4]Icelandic Meteorological Office

Corresponding author: Birgir Hrafnkelsson, Science Institute, University of Iceland, Dunhagi 5, 107 Reykjavik, Iceland, Tel: +354-5254669, Fax:+354-5254632, email: birgirhr@hi.is

ABSTRACT

Discharge in rivers is commonly estimated by the use of a rating curve constructed from pairs of measured water elevations and discharges at a specific location. The Bayesian approach has been successfully applied to estimate discharge rating curves that are based on the standard power-law. In this paper the standard power-law model is extended by adding a B-spline function. The extended model is compared to the standard power-law model by applying the models to discharge data sets from sixty one different rivers. In addition four rivers are analyzed in detail to demonstrate the benefit of the extended model. The models are compared using two measures, the Deviance Information Criterion (DIC) and Bayes factor. The former provides robust comparison of fit adjusting for the different complexity of the models and the latter measures the evidence of one model against the other. The extended model captures deviations in the data from the standard power-law but reduces to the standard power-law when that model is adequate. The extended model provides substantially better fit than the standard power-law model for about 30% of the rivers and performs better for 60% of the rivers when predicting large discharge values.

INTRODUCTION

Discharge in rivers is commonly calculated by mapping water surface elevations, measured at a specific location in the river, to discharge by means of a rating curve. The rating curve is usually an equation that describes a curve that is fitted through data points of measured water surface elevation against measured discharge at a location where downstream hydraulic control assures a stable, sensitive and monotonic relationship between water surface elevation and discharge (Mosley and McKerchar, 1993; ISO, 1983). This methodology is applied as direct measurements of discharge are expensive compared to measurement of water surface elevation that are relatively straightforward and inexpensive undertaking and often well suited for automation. The sources of uncertainty in the discharge obtained by a rating curve methodology are several; both due to uncertainty in river discharge measurements and uncertainty in the rating curve (Pelletier, 1988; Clarke, 1999; Moyeed and Clarke, 2005; Di Baldassarre and Montanari, 2009).

In many instances, such as in engineering design, there is a great interest in an accurate estimate of large discharges as in many cases property and even human life can depend on obtaining reliable estimate of extreme discharges. This relates, e.g., to transportation structures such as roads and bridges or flooding of houses in urban areas due to over topping of levees. Accurate prediction of large discharge, in which case the least data is available in general as it is hard to obtained reliable data during extreme events, usually involves an extrapolation of the rating curve beyond largest measured data points. An accurate estimation of the discharge below the largest measured data point is also important to get an accurate estimation of the annual discharge. Which can be used in engineering design for hydrological power plants. In this paper, a methodology for improved extrapolation of the rating curve for large discharges is proposed, based on the Bayesian approach and B-spline functions.

Based on hydraulic principles, the relationship between discharge and water level is given by the standard power-law

$$q = a(w - c)^b \tag{2}$$

(Lambie, 1978; Mosley and McKerchar, 1993) where q is discharge, w is water level, a is a positive scaling parameter, b is a positive shape parameter and c is the water level when the discharge is zero. These parameters are usually estimated from paired measurements of water level and discharge.

The Bayesian approach has been successfully applied to discharge rating curves (Moyeed and Clarke, 2005; Reitan and Petersen-Øverleir, 2008b; Arnason, 2005). In the Bayesian approach all unknown parameters are treated as random variables. Prior information about unknown parameters based on previously collected data and/or scientific knowledge can be combined with new data for parametric inference. For example, the fact that the parameter b in equation (2) takes the values 1.5 and 2.5 for rectangular and v-shaped sections, respectively, is an example of prior knowledge that can be used to form the prior distribution for one of the unknown parameters. Combination of the prior distributions and the model for the data results in the posterior distribution which can be used to obtain point estimates and interval estimates of the parameters. Icelandic Meteorological Office (IMO) runs a water level measuring system which collects water level data continuously from rivers in Iceland, while the discharge is only measured a few times a year due to high cost. IMO has applied the Bayesian approach successfully to data on discharge and water level for discharge rating curve estimation as presented in Arnason (2005), which is based on the model introduced in Petersen-Øverleir (2004). This model will be referred to as Model 1. Model 1 is not sufficient for about 30% of the data sets at IMO which calls for modifications (see the section Results). The common practice would be to use multi-segment discharge rating curves (Petersen-Øverleir and Reitan, 2005; Reitan and Petersen-Øverleir, 2008a). Reitan and Petersen-Øverleir (2008a) present a Bayesian approach to multi-segment discharge rating curves which results in stable estimation while non-Bayesian methods can have problems with stability (Petersen-Øverleir and Reitan, 2005). Other methods like Takagi–Sugeno fuzzy inference system which is a nonparametric estimation method, have been applied to discharge rating curves (Lohani et al., 2006).

The power-law is derived from a theoretical basis and serves as an appropriate model in most cases. However, in some natural settings deviations from this form arise. For example, the river bed can change from a v-shape to a rectangular shape as the water level increases. Changes of this type are likely to occur gradually as opposed to occurring at a single point with a sharp change or a jump around the breaking points (Petersen-Øverleir and Reitan, 2005). This motivates the use of a smooth function to describe deviation from the power-law instead of using one or more segmentations. A new model, that is an extension of Model 1 is proposed. This model, referred to as Model 2, captures the main trend in discharge as a function of water level through the power-law part, $a(w-c)^b$. To model the remaining variability in the mean response a B-splines function is added which allows for more flexibility than in Model 1. The B-spline part is set equal to zero above a specified water level so the fitted curve is only based on the power-law above this value and the power-law alone is used to extrapolate discharge for large water level. The power-law part of Model 2 plays a similar role as the curve in the segment for the highest water level values in a segmented rating curve model.

The proposed method is similar to Lohani et al. (2006) since both methods rely on the nonparametric approach to estimation. However, it has a few advantages over Lohani et al. (2006) approach. It gives measures of uncertainty in parameters and fit. The complexity and the fit of the model can be evaluated and compared with another Bayesian model with a model criterion. This model criteria penalizes for the number of effective parameters which is a measure of model complexity in the Bayesian setting against the fit of the model. An important advantage of the model introduced here is that it has a structure that allows for prediction of discharge above the largest observed water level.

In the section Data, a description of the sixty one discharge and water level data sets is given. In the following section, Deviance information criteria and Bayes factor, a brief overview of the quantities listed in the section's title is given. In the section Models the two statistical models for discharge and water level measurements is introduced. In the section Bayesian inference a description of the prior distributions and posterior distribution is given. The two models are applied to these data sets in the section Results and a comparison between the models is made. Finally, in the last section, conclusions are drawn.

DATA

The data which are analyzed in this paper were collected by the IMO water level measuring system and are from sixty one different rivers in Iceland. For each river, time series of water level measurement are available. The time series give information about the range of the water level for each river. Detailed analyses are performed for four rivers. They are Norðurá in Borgarfjörður by Stekk, Jökulsá á Fjöllum by Grímsstaðir, Jökulsá á Dal by Brú and Skjálfandafljót by Aldeyjarfoss. The rivers were chosen such that Model 1 will fit reasonably well in one case (Jökulsá á Fjöllum) and two cases where Model 1 is insufficient (Norðurá and Skjálfandafljót) and one case where Model 1 is obviously performing poorly (Jökulsá á Dal). The data sets contain pairs of discharge measurements (q), in $\mathrm{m}^3/\mathrm{sec}$, and water level measurements (w) in m.

DEVIANCE INFORMATION CRITERION AND BAYES FACTOR

To evaluate quantitatively the quality of a fit of the model to a data set a criterion called the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is employed. The deviance information criterion is defined as

$$\mathrm{DIC} = D_{\mathrm{avg}} + p_D,$$

where $p_D = D_{\mathrm{avg}} - D_{\hat{\theta}}$. The quantity p_D is the effective number of parameters and measures the complexity of the model. The quantities D_{avg} and $D_{\hat{\theta}}$ are based on the likelihood function which arises from the proposed probability model of the data. Both D_{avg} and $D_{\hat{\theta}}$ measure the fit of the model to the data. As the complexity of the model (p_D) increases the fit of the model as measured by D_{avg} becomes smaller. Hence, DIC weights the fit of the model against the complexity of the model. It is also noted that the prior distributions restrict the unknown parameters with the effect that the effective number of parameters becomes less than the actual number of parameters. The actual numbers of parameters in Model 1 and Model 2 are five and $L+8$, respectively, where L is the number of B-spline kernels as is discussed in the next section. DIC is used to compare two or more models which are applied to the same data in terms of their fit. In such a comparison the model with the lowest DIC is considered as the first candidate out of the evaluated models. The candidate model needs to be evaluated further in terms of goodness of fit. For details on DIC, D_{avg}, $D_{\hat{\theta}}$ and p_D, see Spiegelhalter et al. (2002) and Gelman et al. (2004). In this paper, if DIC of Model 2 is smaller than DIC of Model 1 by ten or more, then Model 2 is deemed as significantly better than Model 1. The decision of selecting ten as a cut-off value is supported by calculations of Bayes factor (see section Results).

Bayes factors can be used to calculated the posterior probability for each of two or more proposed models conditioned on the data. In case of two models for the data y the following notation is used. The i-th model is denoted by M_i, $p_i(y|\theta_i)$ is the data model, θ_i are the model parameters, $p_i(\theta_i)$ is the prior for θ_i, Θ_i is the parameter space and $P(M_i)$ is the prior probability of model i, $i = 1,2$. The posterior probability of Model 1 is given by

$$P(M_1|y) = \frac{P(M_1)\int_{\Theta_1} p_1(y|\theta_1)p_1(\theta_1)d\theta_1}{\sum_{j=1,2} P(M_j)\int_{\Theta_j} p_j(y|\theta_j)p_j(\theta_j)d\theta_j} = \left(1 + \frac{P(M_2)}{P(M_1)} \times \frac{1}{B_{12}}\right)^{-1}$$

where B_{12} is Bayes factor for the comparison of models M_1 and M_2 (Kass and Raftery, 1995), given by

$$B_{12} = \frac{\int_{\Theta_1} p_1(y|\theta_1)p_1(\theta_1)d\theta_1}{\int_{\Theta_2} p_2(y|\theta_2)p_2(\theta_2)d\theta_2}.$$

Kass and Raftery (1995) presented a table to categorize the evidence against a null model (based on a table from Jeffreys (1961)). Here the null model and the alternative model would be Model 1 and Model 2, respectively. If the Bayes factor values which mark the categories in this table are transformed to $P(M_2|y)$ (rounding the numbers slightly) then the categories presented in Table 1 arise.

Table 1: *Categories for evidence against Model 1.*

$P(M_2\|y)$	Evidence against Model 1
0.50 to 0.75	Barely worth mentioning
0.75 to 0.90	Substantial
0.90 to 0.99	Strong
0.99 to 1.00	Decisive

Here the evidence against Model 1 is preferred to be strong or decisive ($P(M_2|y) > 0.90$) along with a DIC difference of ten or more, favoring Model 2, for the selection of Model 2 over Model 1. The prior probabilities are selected as $P(M_i) = 0.5$, $i = 1,2$.

One way to compute B_{12} is by evaluating the integrals $\int_{\Theta_i} p_i(y|\theta_i)p_(\theta_i)d\theta_i$, $i = 1,2$, with the formula

$$\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{p_i\left(y|\theta_i^{(t)}\right)}\right\}^{-1}$$

where $\theta_i^{(t)}$ is the t-th posterior sample of θ_i. See Robert (2007) for details.

MODELS

A Bayesian model for discharge rating curves based on the standard power-law is given by

$$q_i = a(w_i - c)^b + \varepsilon_i, \quad i = 1,...,n,$$

where n is the number of observations for a given site, (w_i, q_i) denotes the i-th pair of observations and ε_i is a mean zero measurement error such that

$$\varepsilon_i \sim \mathrm{N}(0, \eta^2(w_i - c)^{2b\psi}),$$

where a, b and c are as in (2), the parameter ψ controls how the error variance behaves as a function of the expected value of q, and η^2 is a scaling parameter for the variance. In essence this is the same model as the one presented by Petersen-Øverleir (2004) and it is currently used at IMO. The parameter a is a function of φ and b, that is,

$$a = \exp(\alpha_0 + \alpha_1 b + \varphi) \tag{3}$$

where $\alpha_0 = 4.9468$ and $\alpha_1 = -0.7674$. This reparametrization is motivated by correlation between estimates of $\ln(a)$ and b, denoted by $\ln(\hat{a})$ and $\hat{b}$, which are based on data from IMO, and the values for α_0 and α_1 are selected such that the correlation between $\ln(\hat{a})$ and $\ln(\hat{a}) - \alpha_0 - \alpha_1\hat{b}$ is zero, see Arnason (2005).

A new model referred to as Model 2 is proposed. The form of this model is given by

$$q_i = \mathrm{E}(q(w_i)) + \varepsilon_i, \quad i = 1,...,n,$$

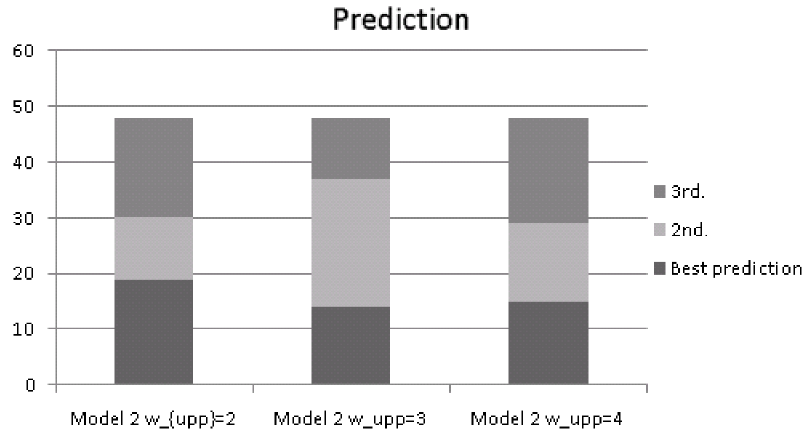

Figure 3: *The prediction performance of Model 2 when w_{upp} was set equal to the second largest, third largest and fourth largest water level measurement. The bars show the frequency of best prediction, second best prediction and third best prediction.*

where ε_i is an error term such that

$$\varepsilon_i \sim \mathrm{N}\left(0, \eta^2 (w_i - c_2)^{2b_2}\right), \quad i = 1, \ldots, n,$$

and η^2, b_2 and c_2 are unknown parameters. The observed discharge is always positive so q_i is normally distributed under the constraint $q_i > 0$. Note that b_2 and c_2 play a similar role as ψb and c in the variance of Model 1, so the variance of Model 2 is essentially the same as the variance in Model 1. However, the variance of Model 2 does not include parameters of the mean function. This is done to simplify the conditional distributions of the Gibbs sampler and obtain more stable simulation from the posterior distribution. The expected value of $q(w)$ is given by

$$\mathrm{E}(q(w)) = \begin{cases} a(w-c)^b, & w > w_{\mathrm{upp}} \\ a(w-c)^b + \sum_{l=1}^{L} \lambda_l G_l(w), & w_{\mathrm{low}} < w \leq w_{\mathrm{upp}},\ \lambda_L = 0 \\ a(w-c)^b + \lambda_1, & w_0 \leq w \leq w_{\mathrm{low}} \end{cases} \tag{4}$$

where the parameter space of a, b, c and λ is such that $\mathrm{E}(q(w)) \geq 0$. Note that $\mathrm{E}(q(w))$ is not defined for $w < w_0$. The coefficient λ_L is set equal to zero to ensure continuity at w_{upp}. The terms $G_l(w)$ are such that

$$G_l(w) = B_l\left(\frac{w - w_{\mathrm{low}}}{w_{\mathrm{upp}} - w_{\mathrm{low}}}\right), \quad l = 1, \ldots, L, \quad w_{\mathrm{low}} \leq w \leq w_{\mathrm{upp}}.$$

The terms $B_l(z)$, $l = 1, \ldots, L$, are cubic B-splines (Wasserman, 2006) which have support on the interval $z \in [0,1]$, w_{low} and w_{upp} are the lower and upper points, respectively, of the interval influenced by the B-splines. For a given river the quantities w_{min} and w_{max} are the smallest and the largest observed water level, respectively, within the pairs (w_i, q_i), $i = 1, \ldots, n$. Based on time series for that river, the smallest water level ever observed is found and is denoted by w_0.

The quantity w_{upp} should be selected close to w_{max} as the data points above w_{upp} have little influence on the B-spline part but mainly influence the power-law curve and thus strengthen the estimation of the parameters of the power-law. For values above w_{upp} the fitted curve is only based on the power-law curve as equation (4) indicates. However, w_{upp} should be smaller than $w_{\max}$ as leaving no data points above w_{upp} will take away information from the parameters of the power-law curve, in particular if the amplitude of the B-spline part is large. This would also result in less accurate prediction of discharge above w_{upp}. However, there is always some information on the power-law parameters in the data points below w_{upp}, especially in the data points that are close to w_{upp}. This is partly due to the fact that λ_L is set equal to zero. Selecting w_{upp} much smaller than $w_{\max}$ results in less flexibility of the model since the B-spline part is then effective over a smaller range of water level values. If that is done the power-law alone is used to fit over a larger range of water level values which may result in a biased fit if there are substantial deviations from a single power-law curve above the selected w_{upp}. Hence, when selecting w_{upp}, there is a trade-off between a good fit below $w_{\max}$ and certainty in prediction intervals for water level above $w_{\max}$. Here, a good fit is preferred at the cost of certainty in prediction. However, w_{upp} is not set equal to $w_{\max}$ but a few data points are left to direct the power-law curve for values above w_{upp}. In order to evaluate the appropriate choose of w_{upp} the ability of the model to predict discharge above $w_{\max}$ was evaluated for three choices of w_{upp}. The quantity w_{upp} was set equal to the second largest, the third largest and the fourth largest water level measurement but these three choices of w_{upp} where deemed to be the ones leading to good prediction properties and good fit. To evaluate these three choices of w_{upp} all data sets with fourteen or more pairs of observations were analyzed. In each case, the three observations with the largest observed water level were omitted in estimation of the rating curve and predicted with the fitted rating curve. The sum of squared residuals was used to compare the three choices of w_{upp} in Model 2. Figure 3 shows the number of times the three models give the best prediction, the second best prediction and the third best prediction. The choice with w_{upp} equal to the third largest water level measurement gave predictions that were the best and the second best in most cases. Since the difference between the best and the second best prediction were usually small, w_{upp} is set equal to the third largest water level observation.

The lower end of the effective range of the B-spline, w_{low}, is set equal to w_0 to ensure that the fitted curve is influenced by the B-spline for all water level values below w_{upp} and down to the smallest water level for which discharge is predicted. If w_{low} would be set equal to a value greater than $w_{\min}$ the same power-law curve alone would apply to both large and small water level values and restrict the flexibility of the model. The choice $w_{\text{low}} = w_0$ will minimize the effect of the data points with the smallest water level observations on the parameters of the power law. The coefficient corresponding to the first B-spline kernel, λ_1, is allowed to be non-zero to introduce more flexibility to the model. Hence at w_{low} the fitted curve deviates by amount equal to λ_1 from the power-law. The above selection of w_{low} and w_{upp} leads then to the following ordering: $w_0 = w_{\text{low}} \leq w_{\min} < w_{\text{upp}} < w_{\max}$.

The B-spline parameters in $\lambda = (\lambda_1, ..., \lambda_L)$ are unknown (with the constraint that $\lambda_L = 0$) where L is the number of B-spline kernels. For simplicity reasons the number of B-splines kernels is fixed (the value of L) and the spacing between the interior knots is also fixed. Equally spaced B-splines are used to obtain consistent smoothness over the entire B-spline interval as well as to reduce computational complexity. It is not optimal to have fixed number of B-spline kernels but a reasonable number can be deduced by using DIC as a measure. Based on evaluation of the four discharge data sets shown in Table 2 it was found that choosing L equal to nine captures the potential improvements gained by Model 2 compared to Model 1. Table 2 shows that there is a small difference in the DIC for values of

L between seven and fifteen in favor of adding kernels. In the case of Norðurá with L equal to five the model needs extra kernels to be able to fit the data accurately and it needs more than seven kernels to become stable. However, it is of course possible to select a number different from nine for individual data sets by optimizing DIC or applying some other criteria.

Table 2: *The values of p_d, DIC, for different number of L in Model 2 for the four rivers.*

		J. Fjöllum		Norðurá		J. Dal		Skjálf.
L	p_d	*DIC*	p_d	*DIC*	p_d	*DIC*	p_d	*DIC*
5	3.27	594.35	3.00	134.30	6.32	676.96	5.63	255.04
7	3.12	594.23	0.95	118.12	7.80	673.95	6.95	253.79
9	3.07	593.84	4.33	121.95	8.68	675.50	7.75	251.98
11	3.08	593.65	4.99	123.19	10.00	675.92	8.29	252.98
13	3.14	593.27	5.94	124.65	10.91	675.48	8.95	251.67
15	3.07	592.97	5.95	125.17	11.53	674.57	9.50	251.57

BAYESIAN INFERENCE

The Bayesian approach requires specification of prior distributions for each of the unknown parameters. The normal prior distributions selected for φ, b, c and ψ in Model 1 are the same (with one exception) as those in Arnason (2005) where point estimates of a, b and c calculated from several data set at IMO were used to construct a prior for these parameters. The exception is the standard deviation in the normal density for b. Arnason (2005) used $\sigma_b = 0.75$ but in this paper $\sigma_b = 0.4$ is used. It is considered safe to decrease the value of σ_b since the previous value was based on point estimates which included sampling error. This prior is reasonable in terms of sensible values of b. The prior of a was then transformed to the prior of φ according to equation (3). The prior distributions for φ, b and c are specified in Appendix. Note that the prior density for b, denoted by $p(b)$, is a truncated normal density between 0.5 and 5 so values below 0.5 and above 5 are assumed invalid. The posterior density of c will be influenced by its prior density which is denoted by $p(c)$ and also by w_0. Since c is the water level at which discharge is zero, values of c above w_0 are invalid. A vague but a proper prior is chosen for η^2 since the mean function for q is fairly well determined by the priors for the parameters in the mean function and the deviation of the data from the mean curve is allowed to form the posterior distribution. An inverse–χ^2 prior distribution for η^2 results in an inverse–χ^2 conditional posterior distribution which is convenient when using the Gibbs sampler. The hyperparameters in the prior distribution of η^2 are chosen to have a minimal effect on the posterior distribution. The prior for η^2 could be improved by collecting point estimates of η^2 based on past data sets. This improvement is left for future research.

Some of the prior distributions for the parameters in Model 2 are the same as the prior distributions of corresponding parameters in Model 1. First, b, c, φ and η^2 in Model 2 have the same prior distributions as b, c, φ and η^2 in Model 1. The parameter c_2 in Model 2 has the same prior distributions as c in Model 1. The prior distribution of b_2 is constructed such that it has a distribution that is similar to that of b times ψ in Model 1.

A normal Markov random field prior (Rue and Held, 2005) with mean zero and covariance matrix $\tau^2 D(I-\phi C)^{-1}MD$ is assumed for the B-spline coefficients, λ, (see also in Appendix). This prior works as a penalty for λ. The parameters τ^2 and ϕ are unknown. In Marx and Eilers (2005) methods for multidimensional splines using classical statistics are discussed. The authors introduce penalty terms in their objective function for the estimation of the spline coefficients. The prior distribution proposed here for the B-spline coefficients gives term in the logged posterior distribution which has a form very similar to the one dimensional penalty term in the objective function in Marx and Eilers (2005). The parameter τ^2 plays the same role as one over the smoothing parameter in Marx and Eilers (2005). The parameter ϕ needs to be one to obtain the same matrices as in Marx and Eilers (2005). But for the prior on the B-spline coefficients to be proper ϕ needs to be less than one, in fact $\phi \in [0,1)$. In order to have the prior working similarly to the penalty in Marx and Eilers (2005), the prior for ϕ is selected such that it favors values very close to one. To accomplish this a beta prior distribution with $\alpha = 20$ and $\beta = 0.5$ is selected for ϕ. This distribution has 90% of its mass between 0.93 and 1. With these prior distributions for ϕ and λ rapid changes in consecutive λ are avoided, the uncertainty in the λs is reduced and the B-spline function is smoother than if ϕ was equal to zero. It was also found that if $\phi = 0$ then the Bayesian computation becomes unstable and the λs do not converge to an optimal value.

The parameter τ^2 controls the size of the elements of λ. A vague inverse–χ^2 prior is chosen for τ^2 due to the lack of knowledge about sensible values for this parameter. This prior allows the posterior distribution to put a lot of mass close to zero which is a desirable property since in many cases τ^2 is in fact equal to zero (the B-spline part is zero). The prior for τ^2 also puts a lot of mass on larger values of τ^2. The variability in the data is bounded which in turn bounds the variability in the posterior distribution of τ^2.

The matrices D and M are diagonal with known constants on their diagonals and C is a constant first order neighborhood matrix. The role of D is to let the prior variance of the λ's decrease as the index goes from 1 to L which forces the B-spline part to become smaller as w approaches w_{upp} therefore it could be used to further force the model to be smooth at the w_{upp}. However, in this paper D is set equal to the identity matrix. The role of the matrix M is to adjust for the end points. M is such that

$$M_{ll} = 0.5, \quad ,l = 2,\dots,L-1, \quad M_{11} = 1, \quad M_{LL} = 1.$$

The neighborhood matrix C is such that

$$C_{l,l-1} = C_{l,l+1} = 0.5, \quad ,l = 2,\dots,L-1, \quad C_{12} = 1, \quad C_{L,L-1} = 1.$$

The posterior distribution of $\theta = (\varphi, b, c, \eta^2, b_2, c_2, \lambda, \tau^2, \phi)$ given the data $q = (q_1,\dots,q_n)$, $w = (w_1,\dots,w_n)$, is given by

$$\begin{aligned} p(\theta|q,w) &\propto \textstyle\prod_{i=1}^n p(q_i|\theta,w_i) \times p(\varphi)p(b)p(c)p(\eta^2)p(b_2)p(c_2) \\ &\times \; p(\lambda|\tau^2,\phi)p(\tau^2)p(\phi) \end{aligned}$$

where $p(q_i|\theta,w_i)$ is a normal density such that

$$p(q_i|\theta,w_i) = \mathrm{N}\left(q_i \middle| a(w_i-c)^b + \sum_{l=1}^{L} \lambda_l G_{li}, \eta^2 (w_i - c_2)^{2b_2} \right),$$

where $G_{li} = G_l(w_i)$. The part $\prod_{i=1}^n p(q_i|\theta,w_i)$ is the likelihood function which is used for the computation of DIC and Bayes factor.

The inference about the unknown parameters is based on samples from the posterior distribution which are generated by a Markov chain Monte Carlo (MCMC) simulation. A Gibbs sampler with Metropolis-Hastings steps is used for the MCMC simulation which consists of the conditional distributions of the unknown parameters (see Gelman et al. (2004) for further details on MCMC and the Gibbs sampler). The conditional distributions of η^2 and τ^2 are scaled inverse chi-square distributions. The conditional distributions of λ is a multivariate normal distribution where λ is first generated without any constraints then the constraint $\lambda_L = 0$ is taken into account. To generate from the conditional distributions of φ, b, c, b_2, c_2 and ϕ, a Metropolis–Hastings steps is needed in each case. However in Model 2 the values for the parameters c and c_2 are set as constants after they have been estimated in the Gibbs sampler. Other parameters in the model are estimated again with c and c_2 fixed, resulting in more reliable estimates.

For the unknown parameters of Model 1 and Model 2 four separate chains of iterations are used. Each chain takes a number of iterations to converge. Those iterations are thrown away and referred to as burn-in period. The decision on the length of the burn-in period is based on the data set that took the longest time to converge. Both models rely the same total number of iterations or 450 thousand. Model 1 than has a burn-in period of 390 thousand iterations. Every fourth value of each chain was stored after the burn-in period to reduce correlation between iterations, yielding four chains of length 15 thousand for posterior inference. For Model 2 the first burn-in period covers the first quarter of each chain. The parameters c and c_2 are estimated from the iterations in the second quarter of each chain. A second burn-in period starts after the first half of each chain. Out of the 60 thousand remaining iterations every fourth value of each chain is stored as in Model 1. Posterior simulations for both Model 1 and Model 2 were stable and the simulated chains converged in all cases. However, it is worth mentioning that in many cases both models converge when the total number of iterations is 160 thousand.

RESULTS

In this section the two models introduced in the section Models are applied to the sixty one data sets from IMO database for comparison between the two models. Analysis of four of the data sets is shown here in details. As mentioned in the Data section these four data sets come from Norðurá, Jökulsá á Fjöllum, Jökulsá á Dal and Skjálfandafljót. Figure 4 shows the fitted discharge rating curves of the two models for these four data sets, along with prediction intervals and posterior intervals for the discharge rating curves.

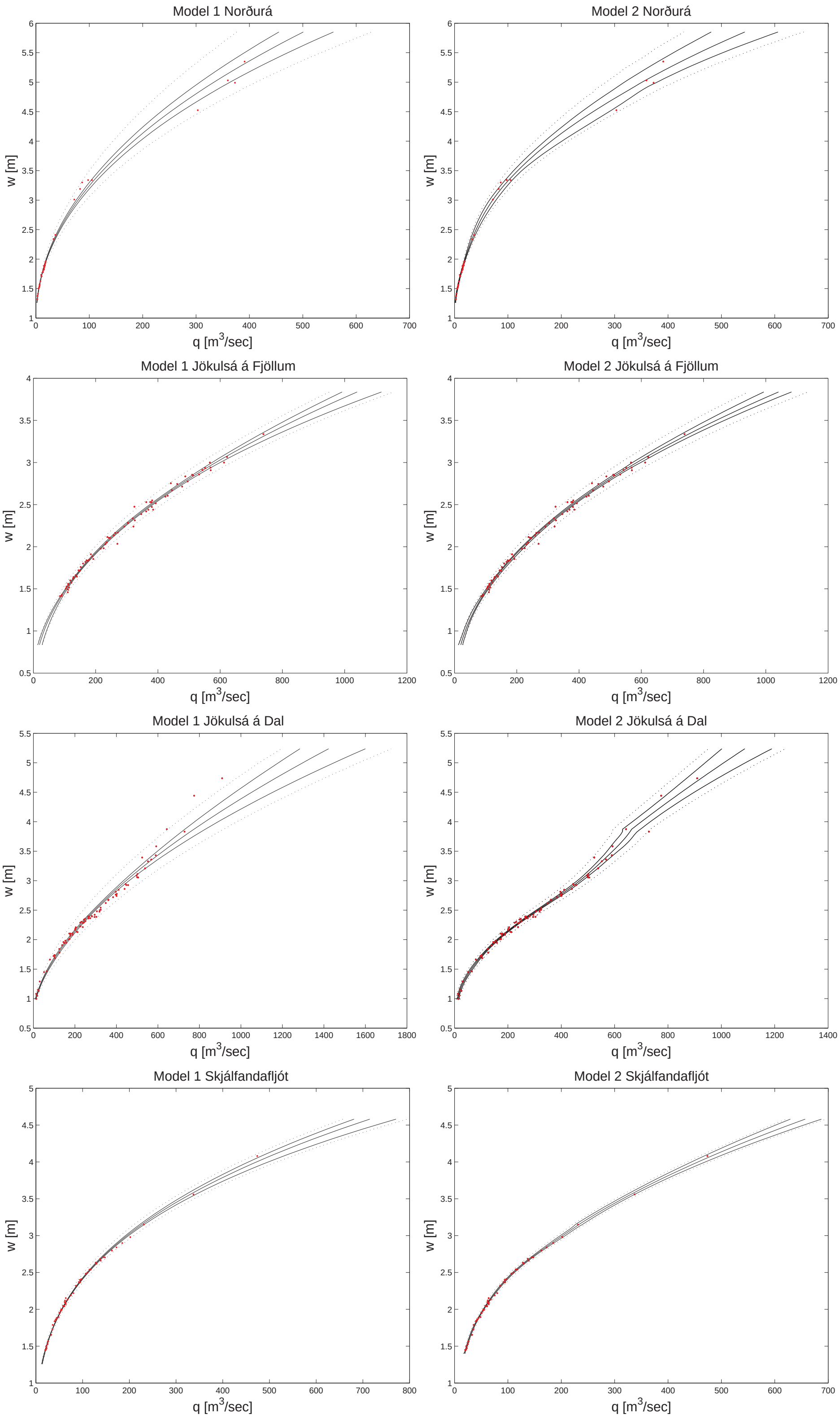

Figure 4: *The fit of Model 1 (left panel) and of Model 2 (right panel) to the four selected data sets. The vertical axes shows water level (w) in m while the horizontal axes shows the discharge (q), in* m^3/sec*. The black solid curves show the posterior median of* $\mathrm{E}(q)$ *and the 95% posterior interval of* $\mathrm{E}(q)$*. The dotted curves show prediction intervals.*

In all cases, except for Jökulsá á Fjöllum, the 95% prediction intervals are wider for larger values of water level in Model 1 than in Model 2. This is mainly due to the fact that if the fit through the observations is adequate then the variability around the fitted curve is smaller when compared to the variability around a poorer fit, this in turn results in narrower prediction intervals.

Figure 5 shows the standardized residuals of the two models versus water level. In general, when an adequate model is used then the standardized residuals should not show any trend and appear to have the same variance for all values of the water level. In the case of Norðurá, Model 2 yields more convincing standardized residuals than Model 1, which shows a trend in the standardized residuals while that is not the case for Model 2. In the case of Jökulsá á Fjöllum there is no visible difference in the standardized residuals which indicates that Model 2 imitates Model 1 when Model 2 does not provide significant improvement over Model 1. For Jökulsá á Dal the trend in the standardized residuals of Model 1 is obvious, while the standardized residuals of Model 2 show no trend. In the case of Skjálfandafljót, there appears to be a trend in the standardized residuals of Model 1 for water level values lower than 1.84 m and greater than 2.37 m while the standardized residuals of Model 2 show no trend. These examples demonstrate that Model 2 can provide better results than Model 1 and when Model 1 appears to be adequate, Model 2 performs as well as Model 1.

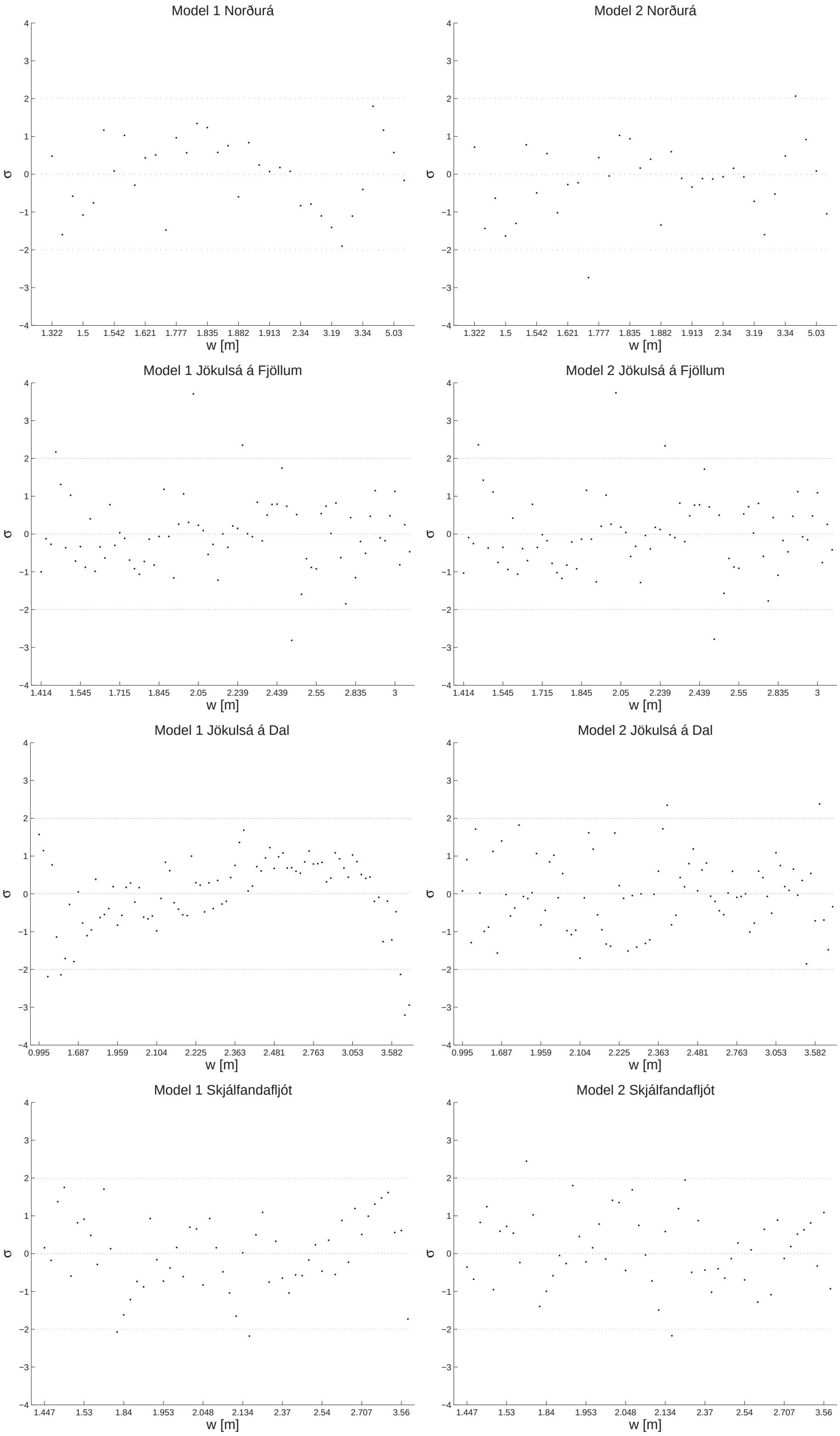

Figure 5: *Standardized residuals for the four selected data sets (vertical axes). Water level is on the horizontal axes (cm) but the scale is nonlinear. Standardized residuals for Model 1 (left panel) and Model 2 (right panel).*

Figure 6 shows the roles that the standard power-law part and the B-spline part play in Model 2. The B-spline part models the variation in the data for the values of the water level below w_{upp} that the standard power-law part can not adjust for on its own. The B-spline part is zero at and above w_{upp} and it smoothly approaches zero as w approaches w_{upp} from below. In the case of Norðurá as well as Skjálfandafljót the B-spline part allows Model 2 to give a visibly better fit. The standard power-law model (Model 1) is adequate in the case of Jökulsá á Fjöllum as is seen in the left panel of Figure 6. The right panel shows clearly the ability of the B-spline part of Model 2 to reduce to almost zero, thus, the B-spline addition has insignificant effect on the discharge rating curve for such case. In case of Jökulsá á Dal it can be seen that the B-spline part can take as large values as needed when the standard power-law part is inadequate for the data set. In Table 3, a comparison between the two models is made through DIC and Bayes factor (see the section Deviance information criterion and Bayes factor). Table 3 shows that p_D is less than the actual number of unknown parameters in Model 1 and Model 2 which are 5 and 15 respectively. This is expected due to the fact that the prior distributions constrain the unknown parameters. It seems that the more the B-spline part is contributing, the larger the number of effective parameters. This shows the adaptive nature of the Markov random field prior for λ.

Table 3 shows that in all cases except Jökusá á Fjöllum, Model 2 has considerably lower DIC than Model 1. The difference in DIC between Model 1 and Model 2 is about 19 and 23 for Norðurá and Skjálfandafljót, respectively, and about 98 for Jökulsá á Dal. In the case of DIC these are all relatively large differences. In the case of Jökulsá á Fjöllum the difference in DIC is less than 3 which is viewed as a small difference. This is reflected in the fitted discharge rating curves of Model 1 and Model 2 which show no visible differences for Jökulsá á Fjöllum in Figure 6. The results in Table 3 and Figures 4, 5 and 6 show that the B-spline part of Model 2 either improves the fit compared to Model 1 or gives a fit equally good as that of Model 1 when Model 1 is adequate. The posterior probability of Model 2 (based on Bayes factor) is also computed for the four selected data sets in Table 3. The computed probability values confirm that the DIC differences for Norðurá, Jökulsá á Dal and Skjálfandafljót are relatively large and support selecting Model 2 over Model 1 while the DIC difference for Jökulsá á Fjöllum is small and supports selecting Model 1 over Model 2.

Table 3: *The values of D_{avg}, $D_{\hat{\theta}}$, p_D and DIC for Model 1 and Model 2 for the four rivers.*

	Model 1				Model 2			
	D_{avg}	$D_{\hat{\theta}}$	p_d	*DIC*	D_{avg}	$D_{\hat{\theta}}$	p_d	*DIC*
Norðurá	136.10	131.54	4.56	140.66	117.62	113.3	4.33	121.95
Jökulsá á Fjöllum	592.11	587.64	4.46	596.57	590.76	587.69	3.07	593.84
Jökulsá á Dal	768.87	764.00	4.88	773.75	666.83	658.15	8.68	675.50
Skjálfanda-fljót	271.07	266.57	4.48	274.55	244.23	236.47	7.75	251.98

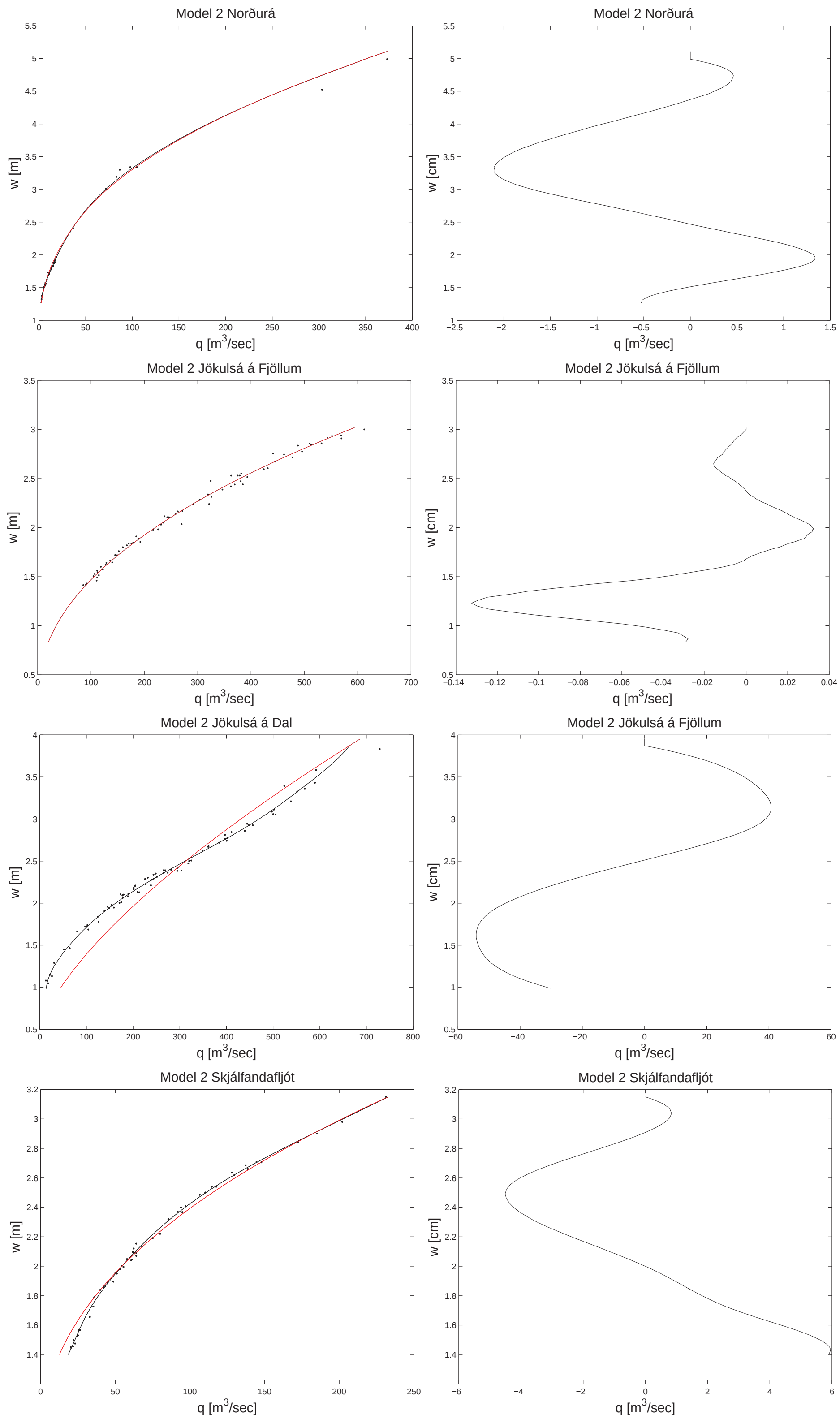

Figure 6: *The left panel shows the standard power-law part (solid red curves) of Model 2 and the sum of standard power-law part and the B-spline part of Model 2 (solid black curves) for the four selected data sets. The right panel shows the B-spline part of Model 2 for each data set. Water level is on the vertical axes (m) while discharge is on the horizontal axes* (m^3/sec).

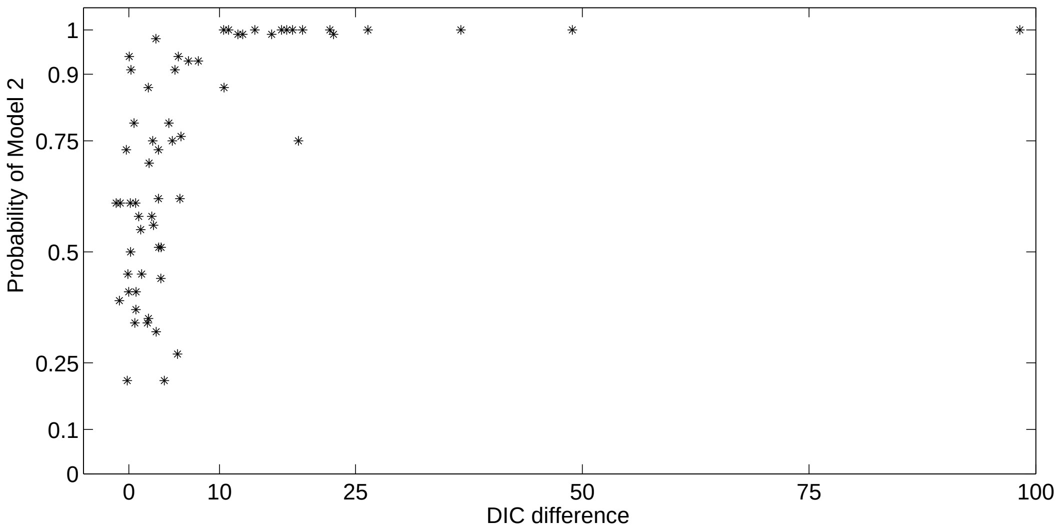

Figure 7: *The difference in DIC between the two models is on the horizontal axis and the posterior probability of Model 2 (based on Bayes factor) is on the vertical axes.*

Figure 7 shows comparison between Model 1 and Model 2 for 61 stations analyzed from the IMO database by plotting the difference in DIC between the Model 2 and Model 1 on the horizontal axis (positive if Model 2 gives a better fit) and the posterior probability of Model 2 on the vertical axis. When the DIC difference is greater than ten and the posterior probability of Model 2 is greater than 0.9, then Model 2 significantly improves the fit of Model 1 (see the section Deviance information criterion and Bayes factor). This is the case for 16 rivers which is about 26% of the data sets. When the probability of Model 2 is between 0.0 and 0.90 and the DIC diffence is less than 10 then Model 2 is not outperforming Model 1 and that Model 1 is adequate. This is the case for 36 rivers out of 61, or 59%. In case when the DIC difference is less than 10 and the posterior probability of Model 2 is greater than 0.9 (7 of 61), and in the case when the DIC difference is greater than 10 and the posterior probability of Model 2 is less than 0.9 (2 of 61), a close look at the descriptive plots and statistics is needed to determine whether Model 1 is adequate or not. This is true in general, that is, a detailed analyzes of each data set is needed before a final decision about Model 1 or Model 2 is made. The DIC difference and the posterior probability of Model 2 are important measures to support that decision.

Table 4 shows estimates of the parameters a, b, c which are sufficient to construct discharge rating curves based on standard power-law. These parameters are presented for both Model 1 and Model 2. There is a substantial difference in these parameters between Model 1 and Model 2 which is due to the extra flexibility of Model 2. The B-spline part in Model 2 has the ability to utilize information from lower values of water level in the data and therefore the standard power-law parameters can be estimated with a more focus on the higher water level when needed. This can lead to a different posterior density for a, b and c in the two models as seen in Table 4.

Table 4: *Parameter estimates of a, b and c in Model 1 and Model 2. * c is pre-estimated and therefore a constant.*

	Model 1			Model 2		
	a	b	c	a	b	c
Norðurá						
Post. median	15.65	2.16	0.88	9.64	2.45	0.69^*
2.5 percentile	11.69	2.06	0.77	6.49	1.94	
97.5 percentile	17.54	2.36	0.93	21.10	2.75	
Jökulsá á Fj.						
Post. median	69.89	2.13	0.29	65.96	2.16	0.25^*
2.5 percentile	49.42	1.94	0.11	59.44	2.03	
97.5 percentile	92.16	2.34	0.43	75.36	2.27	
Jökulsá á Dal						
Post. median	112.71	1.68	0.74	107.87	1.48	0.44^*
2.5 percentile	92.57	1.52	0.64	73.64	1.22	
97.5 percentile	134.97	1.87	0.82	151.10	1.76	
Skjálfandaflj.						
Post. median	7.61	3.01	0.06	24.37	2.39	0.58^*
2.5 percentile	4.05	2.85	−0.18	20.67	2.23	
97.5 percentile	10.10	3.36	0.18	29.04	2.54	

In Table 5, a posterior interval is given for rest of the parameters in Model 1 and in Model 2 except for λ. For Model 1 the parameter ψ is multiplied by b so it can be compared to the parameter b_2 in Model 2. The posterior median of τ^2 varies from 2.99 in Jökulsá á Fjöllum to 1180.6 in Jökulsá á Dal which shows the difference in the amplitude of the B-spline part for these data sets. The parameter ϕ is forced to be close to one through its prior distribution to ensure strong positive correlation between the elements of λ. The effect of the prior is clear in the posterior estimates of ϕ.

Table 5: *Parameter estimates of ψ and η^2 in Model 1 and b_2, c_2, η^2, τ^2 and ϕ in Model 2.* * *c_2 is pre-estimated and therefore a constant.*

	Model 1		Model 2				
	$\psi \times b$	η^2	b_2	c_2	η^2	τ^2	ϕ
Norðurá							
Post. median	2.42	0.05	2.71	0.62^*	0.22	19.80	0.95
2.5 percentile	2.12	0.02	2.30		0.12	0.32	0.80
97.5 percentile	2.63	0.15	3.12		0.47	597.33	0.99
Jökulsá á Fj.							
Post. median	1.77	1.04	2.02	0.15^*	8.87	2.99	0.95
2.5 percentile	1.17	0.06	1.48		4.14	0.0004	0.81
97.5 percentile	2.57	21.64	2.55		20.32	136.74	0.99
Jökulsá á Dal							
Post. median	1.44	4.74	1.96	-0.19^*	3.77	1180.6	0.96
2.5 percentile	1.03	0.84	1.52		1.67	437.8	0.83
97.5 percentile	1.84	45.17	2.40		9.09	4274.6	0.99
Skjálfandaflj.							
Post. median	2.75	0.024	2.03	0.11^*	0.26	17.11	0.95
2.5 percentile	1.98	0.002	1.43		0.11	3.88	0.82
97.5 percentile	3.72	0.231	2.63		0.64	79.43	0.99

As discussed in the Introduction section discharge rating curves are frequently used in extrapolation of discharge. As a demonstration, the three highest water level observations, along with corresponding discharges observations, were excluded from the data sets for the four rivers previously analysed. Then both models were used to extrapolate over the range of the three excluded water level values. Figure 8 shows the results. In all cases the three excluded discharge values are within the 95% prediction interval for Model 2 but only in two cases for Model 1, namely, Jökulsá á Fjöllum and Norðurá. For these two cases the models are similar for Jökulsá á Fjöllum but Model 2 looks better for Norðurá. For the other two cases Model 1 is considerably of the mark. Hence, it can be concluded that for these four cases Model 2 performs considerably better in predicting discharge for extrapolated water level values greater than w_{max}.

The data analysis conducted to select w_{upp} for Model 2 in the section Models was also performed for Model 1. When Model 1 and Model 2 (with w_{upp} equal to the third largest water level observation) are compared in terms of prediction then Model 2 performs better than Model 1 for 60% of the data sets. This is based on 48 data sets so 29 data sets give better results under Model 2 in terms of prediction. However, 16 data sets out of 61 are such that Model 2 is judged to give a better fit than Model 1. So, in some cases even if the fit for Model 1 is better than or equally good as that of Model 2, then Model 2 appears to perform better when predicting discharge for water level greater than w_{max}. However, in few cases Model 1 performs better when predicting discharge for water level greater than w_{max} even though Model 2 gives a better fit.

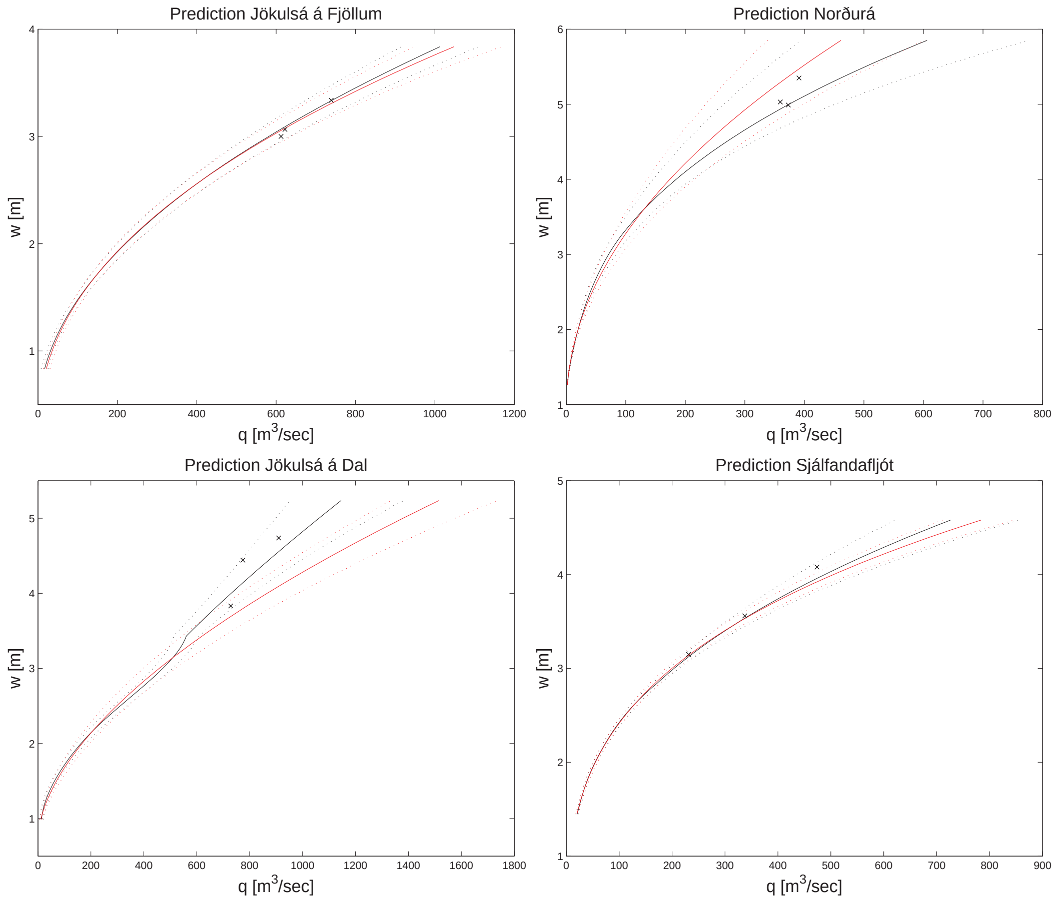

Figure 8: *The solid curves show the posterior median of* $\mathrm{E}(q)$*, red for Model 1 and black for Model 2 for the four selected data sets. The dotted curves show prediction intervals, red for Model 1 and black for Model 2. Water level is on the vertical axes (m) while discharge is on the horizontal axes (*m^3/sec*).*

CONCLUTION

A Bayesian model for discharge rating curves, labeled Model 2, was developed by extending the standard power-law model, labeled Model 1, by adding a B-spline function. Comparison of these two models based on analysis of sixty one data sets from IMO shows that Model 2 outperforms or performs as well as Model 1. One of the most important properties of Model 2 is the capacity of the B-spline part to catch deviation in the data from the standard power-law model when that model is inadequate. In these cases, Model 2 achives a more convincing fit to the data than Model 1. This is confirmed with calculations of DIC and Bayes factor where Model 2 yields a substantially lower DIC values and higher posterior probabilities than Model 1 in sixteen of sixty one cases (DIC difference greater than ten and posterior probability of Model 2 greater than 0.9). In thirty six cases the DIC difference is less than ten and the posterior probability of Model 2 less than 0.9 and it is debatable whether the added complexity of Model 2 leads to an improvement. Another important property of Model 2 is that when Model 1 appears to give an adequate fit as in the case of Jökulsá á Fjöllum then Model 2 imitates Model 1 by reducing the amplitude of the B-spline almost down to zero. Model 2 performs better than Model 1 when it comes to prediction of discharge for water level above w_{max} as it gives better results for 60% of the analyzed data sets, which supports the use of Model 2.

It is concluded that Model 2 can be used to fit discharge rating curves regardless of whether the standard power-law model is adequate or not. The exceptional cases are when the data sets contain a few data pairs and there may not be enough information to estimate the B-spline part successfully. Based on the experience gained here at least ten data pairs are needed.

Finally, it is noted that segmentation has been commonly used in estimating discharge rating curves and it could be argued that maybe it is more appropriate than Model 2 for data sets where there is a visually apparent shift. A direct comparison between segmentation models and Model 2 is needed to compare their performance. A joint use of multi-segment discharge rating curves and B-splines could potentially be beneficial for such cases.

APPENDIX

The following prior distributions are proposed for the unknown parameters.

$$p(\varphi) = \mathrm{N}(\varphi|\mu_\varphi = 0, \sigma_\varphi^2 = 0.82^2)$$

$$p(b) \propto \mathrm{N}(b|\mu_b = 2.15, \sigma_b^2 = 0.4^2)I(0.5 < b < 5)$$
$$p(c) \propto \mathrm{N}(c|\mu_c = 75, \sigma_c^2 = 50^2)I(c < w_0)$$
$$p(\psi) \propto \mathrm{N}(\psi|\mu_\psi = 0.8, \sigma_\psi^2 = 0.25^2)I(0 < \psi < 1.2)$$
$$p(b_2) \propto \mathrm{N}(b_2|\mu_{b2} = 2.15, \sigma_{b2}^2 = 0.4^2)I(1 < b < 6)$$
$$p(c_2) \propto \mathrm{N}(c_2|\mu_{c2} = 75, \sigma_{c2}^2 = 50^2)I(c_2 < w_0)$$
$$p(\eta^2) \propto \text{Inv-}\chi^2(\eta^2|\nu_\eta = 10^{-12}, S_\eta^2 = 1)$$
$$p(\phi) = \mathrm{Beta}(\phi|\alpha_\phi = 20, \beta_\phi = 0.5)$$
$$p(\tau^2) \propto \text{Inv-}\chi^2(\tau^2|\nu_\tau = 10^{-12}, S_\tau^2 = 1)$$
$$p(\lambda|\tau^2, \phi) \propto \mathrm{N}(\lambda|0, \tau^2 D(I - \phi C)^{-1} MD)$$

where $I(A)$ is such that $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. In the prior distribution for λ, I is an identity matrix, D and M are diagonal matrices and C is a neighborhood matrix with constants on the first off-diagonals, other elements are equal to zero.

3 ESTIMATION OF DISCHARGE RATING CURVES USING THE STANDARD POWER-LAW WITH VARYING COEFFICIENTS

Kristinn Mar Ingimarsson[1,4], Birgir Hrafnkelsson[2],
Sigurdur M. Gardarsson[3] and Arni Snorrason[4]

[1]Faculty of Industrial Engineering, Mechanical Engineering and Computer Sciences, University of Iceland,
[2]Science Institute, University of Iceland,
[3]Faculty of Civil Engineering and Environmental Engineering, University of Iceland,
[4]Icelandic Meteorological Office

Corresponding author: Birgir Hrafnkelsson, Science Institute, University of Iceland, Dunhagi 5, 107 Reykjavik, Iceland, Tel: +354-5254669, Fax:+354-5254632, email: birgirhr@hi.is

ABSTRACT

Bayesian methodology for estimating discharge rating curves that is based on the standard power-law where the parameters are not constant with water level. The model uses B-splines to estimate the parameters in the power-law equation. This model is compared to the standard power-law model with a B-spline smoothing function added to it. These models are compared by using paired discharge and water level measurements data sets from forty nine rivers.

INTRODUCTION

Hydrological rating curves give discharge as a function of water level. Based on hydraulic principles, the relationship between discharge and water level is given by the standard power-law

$$q = a(w-c)^b$$

(Lambie, 1978; Mosley and McKerchar, 1993) where q is discharge, w is water level, a is a positive scaling parameter, b is a positive shape parameter and c is the water level when the discharge is zero. These parameters are usually estimated from paired measurements of water level and discharge.

The power-law is derived from a theoretical basis and serves as an appropriate model in most cases. The Bayesian approach has been successfully applied to discharge rating

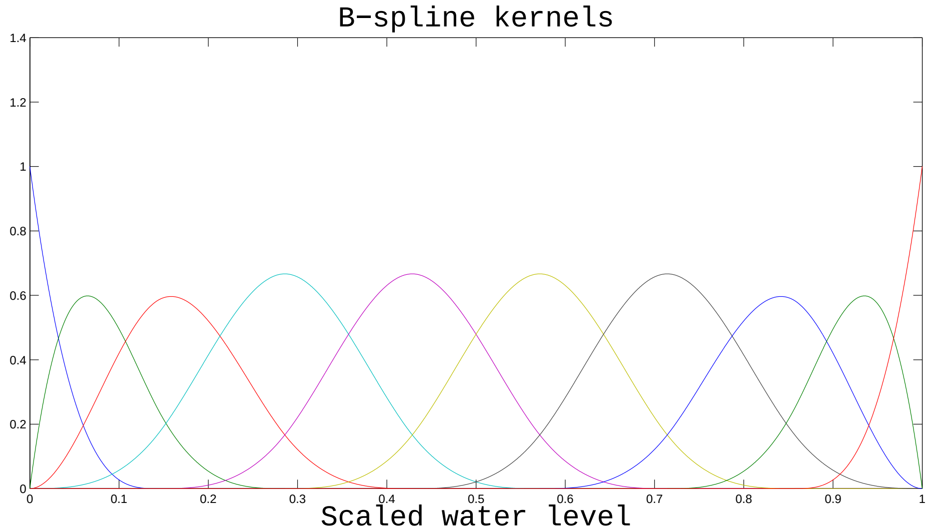

Figure 9: *Ten b-spline kernels.*

curves(Moyeed and Clarke (2005), Reitan and Petersen-Øverleir (2008b) and Arnason (2005)). However, in some natural settings the stream bed changes with water level and therefore when using the power-law equation it becomes essential to allow the parameters to change with water level according to the change in the stream bed. This can be executed in various ways the classical way is to use the power-law with constant parameters for a given water level interval and then change the parameters at the next interval if needed, this is called segmentation as in (Petersen-Øverleir and Reitan, 2005; Reitan and Petersen-Øverleir, 2008a). Another approach is to use smoothing function on top of the power-law as in Ingimarsson et al. (2010a) where a B-spline smoothing function is used to describe deviation from the power-law. Figure 9 demonstrate how the B-spline kernels work. Here a new method is proposed that estimates the parameters in the power-law continuously as the water level raises. This approach allows for easier interpretation of the discharge rating curve than for the model introduced in Ingimarsson et al. (2010a).

In the section Data, a description of the forty nine discharge and water level data sets is given. In the section DIC an brief overview of the model criteria DIC is given. In the section Models the two statistical models for discharge and water level measurements is introduced. In the section Bayesian Inference a description of the prior and posterior distributions is given. The four models are applied to these data sets in the section Results and a comparison between the models is made. Finally, conclusions are drawn in the last section Conclusions.

DATA

Estimation of discharge rating curves requires a data set with water level and corresponding discharge measurement is needed. The IMO has collected this type of data form rivers around Iceland. Forty nine data sets from equally many rivers are analyzed in this paper. For each of these rivers, time series of water level measurement are available. These time series give valuable information about the range of the water level for each river. For each river the smallest observed water level within the time series is useful for the estimation of the discharge rating curve. The role of this value will be introduced in the chapter Models. The data sets contain pairs of discharge measurements (q), in $\mathrm{m}^3/\mathrm{sec}$, and water level measurements (w) in m. One river is analyzed in more detailed due to apparent deviation from the standard power-law with fixed parameters.

DEVIANCE INFORMATION CRITERION

To evaluate quantitatively the quality of a fit of the model to a data set a criterion called the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is employed. The deviance information criterion is defined as DIC = D_{avg} + p_D, where $p_D = D_{\mathrm{avg}} - D_{\hat{\theta}}$. The quantity p_D is called the effective number of parameters and D_{avg} and $D_{\hat{\theta}}$ are based on the likelihood function. Hence, the measure DIC penalizes for effective number of parameters. Also noted that the prior distributions restrict the unknown parameters with the effect that the effective number of parameters becomes less than the actual number of parameters but the actual numbers of parameters in Model 1 and Model 2 are five and L + 6, respectively, where L is the number of B-spline kernels. DIC is used to compare two or more models which are applied to the same data. In such a comparison the model with the lowest DIC is considered as the first candidate out of the evaluated models. The candidate model needs to be evaluated further in terms of goodness of fit. For details on DIC, D_{avg}, $D_{\hat{\theta}}$ and p_D, see Spiegelhalter et al. (2002) and Gelman et al. (2004).

MODELS

Here new models for discharge rating curves is introduced. Inference for this model is based on the Bayesian approach. The Bayesian model for discharge rating curves that is used as a comparison to the new model. This model is introduced in Ingimarsson et al. (2010a) and the form of this model is given by

$$q_i = \mathrm{E}(q(w_i)) + \varepsilon_i, \quad i = 1, ..., n.$$

where ε_i in an error terms such that

$$\varepsilon_i \sim \mathrm{N}\left(0, \eta^2 (w_i - c_2)^{2b_2}\right), \quad i = 1, ..., n,$$

where η^2, b_2 and c_2 are unknown parameters. The observed discharge is always positive so q_i is normally distributed under the constraint $q_i > 0$. The expected value of $q(w)$ is given by

$$\mathrm{E}(q(w)) = \begin{cases} a(w-c)^b, \quad w > w_{\mathrm{upp}} \\ a(w-c)^b + \sum_{l=1}^{L} \lambda_l G_l(w), \quad w_{\mathrm{low}} < w \leq w_{\mathrm{upp}}, \lambda_L = 0 \\ a(w-c)^b + \lambda_1, \quad w_0 \leq w \leq w_{\mathrm{low}} \end{cases} \tag{5}$$

where the parameter space of a, b, c and λ is such that $\mathrm{E}(q(w)) \geq 0$. Note that $\mathrm{E}(q(w))$ is not defined for $w_0 < w$. The coefficient λ_L is set equal to zero to ensure continuity at w_{upp}. The terms $G_l(w)$ are such that

$$G_l(w) = B_l\left(\frac{(w - w_{\mathrm{low}})}{(w_{\mathrm{upp}} - w_{\mathrm{low}})}\right), \quad l = 1, ..., L, \quad w_{\mathrm{low}} \leq w \leq w_{\mathrm{upp}}. \tag{6}$$

The terms $B_l(z)$ are cubic B-splines (Wasserman, 2006) which have support on the interval $z \in [0,1]$, w_{low} and w_{upp} are the lower and upper points of the interval influenced by the B-splines, respectively. For a given river the quantity $w_{\min}$ and $w_{\max}$ are the smallest and the largest observed water level, respectively, within the pairs (w_i, q_i), $i = 1,...,n$. Based on time series for that river, the smallest water level of the time series for that river is denoted by w_0. Here w_{low} is set equal to w_0. The quantity w_{upp} is set equal to the third largest water level observation with a corresponding discharge measurement. Therefore, the order of these quantities here is such that $w_0 = w_{\text{low}} \leq w_{\min} < w_{\text{upp}} < w_{\max}$. The B-spline parameters in $\lambda = (\lambda_1,...,\lambda_L)$ are unknown (with the constraint $\lambda_L = 0$) and L is the number of B-spline kernels. For simplicity it is decided to use fixed number of B-splines kernels (value of L) and to use fixed spacing between the interior knots. Equally spaced B-splines are used to obtain consistent smoothness over the entire interval B-spline interval as well as to reduce computational complexity. This Model will be referred to as Model 1.

Model 1 has the ability to pick up deviations from the power-law. This model is compared to the standard power-law equation, this has been tested in the (ingimarsson.et.al) where the parameters a and b were constants and then model lacked the flexibility that is needed in many data sets. Here a model based on the standard power-law are proposed where the parameters a and b are not constant with water level. This model has mean function is given by

$$\log(q_i) = \mathrm{E}(\log(q(w_i))) + \varepsilon_i, \quad i = 1,...,n. \tag{7}$$

$$\mathrm{E}(\log(q(w))) = a(w) + b(w) * \log(w - c) \tag{8}$$

where c is a unknown parameter, a is given by

$$a(w) = \begin{cases} a_1 + \omega_L & w > w_{\text{upp}} \\ a_1 + \sum_{l=1}^{L} \omega_l X_l(w) & w_{\min} < w \leq w_{\max} \\ a_1 + \omega_1 & w > w_{\min} \end{cases} \tag{9}$$

and where b is given by

$$b(w) = \begin{cases} b_1 + \xi_K & w > w_{\text{upp}} \\ b_1 + \sum_{k=1}^{K} \xi_k U_k(w) & w_{\min} < w \leq w_{\max} \\ b_1 + \xi_1 & w > w_{\min} \end{cases} \tag{10}$$

The functions X and U are the same as G in Equation 6 with the exception that $w_{\max}$ is replaced with w_{upp}.

This model requires variance function that is different from that in Model 1 and therefore three different structures are tested for the error term ε_i in (7). These models will be referred to as Models 2-4 and will all have the same expected value shown in Equation (8). The error term ε_i in Equation (7) for Model 2 is given by

$$\varepsilon_i \sim \mathrm{N}\left(0, \sigma_2^2\{1 + r_1 \exp[(w - w_{\text{fix}})/r_2]\right), \quad i = 1,...,n,$$

where r_1, r_2 and σ_0^2 are unknown constants and w_{fix} is a constant set equal to $w_{\min}$. The error term ε_i in Equation (7) for Model 3 is given by

$$\varepsilon_i \sim \mathrm{N}\left(0, \sigma_3^2\{1 + \exp[d(w)]\right), \quad i = 1,...,n,$$

where σ_0^2 is a unknown constant and $d(w)$is given by

$$d(w) = \begin{cases} \kappa_K & w > w_{\text{upp}} \\ \sum_{o=1}^{K} \kappa_o U_o(w) & w_{\min} < w \leq w_{\max} \\ \kappa_1 & w > w_{\min} \end{cases} \tag{11}$$

where κ is a unknown vector.

The error term ε_i in (7) for Model 4 is given by

$$\varepsilon_i \sim \mathrm{N}\left(0, \sigma_4^2\{1 + s_1(w - w_{\text{fix}})^{s_2}\right), \quad i = 1, \ldots, n,$$

Where σ_4^2, s_1 and s_2 are a unknown constants.

BAYESIAN INFERENCE

The Bayesian approach requires specification of prior distributions for each of the unknown parameters. The prior distributions for Model 1 are the same as in Ingimarsson et al. (2010a). For Models 2, 3 and 4 the priors for the parameters in the expected value are the same. For a_1 and b_1 a normal prior distributions are selected as in Ingimarsson et al. (2010a) however due to the re-parametrization of a in that paper the prior distributions for a_1 becomes normal with mean as -6.60 and variance as 4.10. The prior distributions for ω and ξ are a Gaussian Markov random field distributions (Rue and Held, 2005) with mean zero and covariance matrix $\tau_j^2 D(I - \phi_j C)^{-1} MD$ where j is either 1 or 2 depending on whether it is in the prior distribution for ω or ξ. This prior works as a penalty for ω and ξ. The parameters τ_j^2 and ϕ_j are unknown. The parameter τ_j^2 controls the size of the elements of ω and ξ and acts similar to a smoothing parameter in a non-Bayesian approach. A vague inverse–χ^2 prior is chosen for τ^2 due to the lack of knowledge about sensible values for this parameter. This prior allows the posterior distribution to put a lot of mass close to zero which is a desirable property since in many cases τ^2 is in fact equal to zero (the B-spline part is zero). The prior for τ_j^2 also puts a lot of mass on positive non-zero values of τ^2. The variability in the data is bounded which in turn bounds the variability in the posterior distribution of τ^2. The parameter ϕ is in the interval $[0, 1)$. To obtain a strong positive correlation between the λ coefficients the value of ϕ needs to be close to 1, but that is preferred here to avoid rapid changes in the λ's which can lead to lack of smoothness in the B-spline part and thus in the rating curve. To accomplish this a beta prior distribution with $\alpha = 20$ and $\beta = 0.5$ is selected for ϕ. This distribution has 90% of its mass between 0.93 and 1. The prior distributions for the σ^2 parameters in Models 2, 3 and 4 are inverse–χ^2 distributions. Other parameters in Model 2 have normal distributed priors. A prior for r_1 has the form $\mathrm{N}(r_1|0, 0.5)$ and for r_2 $\mathrm{N}(r_1|0, 0.5)$. In Model 3 the prior for a_3 has the form $\mathrm{N}(a_{3i}|0, 2)$ where $\quad i = 1, \ldots, L$. In Model 4 the prior distributions for s_1 ands_2 have the following form $\mathrm{N}(s_1|0, 0.5)$ and $\mathrm{N}(s_2|2, 1)$.

The posterior distribution of $\theta = (a_1, b_1, c, \omega, \xi, \tau_1^2, \tau_2^2, \phi_1, \phi_2, \sigma_2^2, r_1, r_2)$ given the data $q = (q_1, \ldots, q_n)$, $w = (w_1, \ldots, w_n)$, for Model 2 is given by

$$\begin{aligned} p(\theta|q, w) \quad &\propto \quad \prod_{i=1}^{n} p(q_i|\theta, w_i) \times p(a_1)p(b_1)p(c)p(\sigma_2^2)p(r_1)p(r_2) \\ &\times \quad p(\omega|\tau_1^2, \phi_1)p(\tau_1^2)p(\phi_1) \\ &\times \quad p(\xi|\tau_2^2, \phi_2)p(\tau_2^2)p(\phi_2) \end{aligned}$$

where $p(q_i|\theta, w_i)$ is a normal density such that

$$p(q_i|\theta, w_i) = \mathrm{N}\left(q_i \middle| (a'(w)) + b(w)\log(w_i - c), \sigma_2^2\{1 + r_1 \exp[(w_i - w_{\text{fix}})/r_2]\right),$$

and the part $\prod_{i=1}^{n} p(q_i|\theta, w_i)$ is the likelihood function. The posterior distribution is derived similarly for Model 3 and 4.

A Gibbs sampler with Metropolis–Hastings steps is used to generate samples from the posterior distribution. The conditional distributions of η^2 and τ^2 are scaled inverse–χ^2 distributions. The conditional distributions of ω and ξ are a multivariate normal distribution where ω and ξ are first generated without any constraints then the constraint of the first and last values in the vector of ω and ξ are taken into account. To generate from the conditional distributions of a_1, b_1, c, r_1, r_2, a_3, s_1, s_2, and ϕ_j where $j = 1, 2$, a Metropolis–Hastings steps is needed in each case. However in all models the value for the parameters c is set as constant after it has been estimated with the Gibbs sampler. The parameters in the model are estimated again with c fixed. This is to strengthen the estimate of other parameters.

RESULTS

In this section a comparison between the models introduced in the section Models is made. For this comparison forty nine data sets from the data base of IMO are used. To compare how well the models are performing for different types of data sets DIC values were calculated for all forty nine data sets and the difference between Model 1 and the other models is computed, see Figure 10.

As can be seen from Figure 10, Model 3 clearly outperforms Model 2 and 4 in six cases. However, for the rest of the DIC differences there is little difference between these models. Model 2 and Model 4 give very similar results with one exception. Due to the fact that Models 2, 3 and 4 all have the same mean function and DIC calculations are in favor of Model 3 a focus will be on Model 3. It is apparent that Model 1 is performing better in more cases than Model 3 however in 32 cases the DIC difference is less than 5 which means that there is a small difference in how the models are performing. In fifteen cases Model 3 is performing with a smaller DIC values than Model 1 where five of these cases have DIC difference grater than 5.

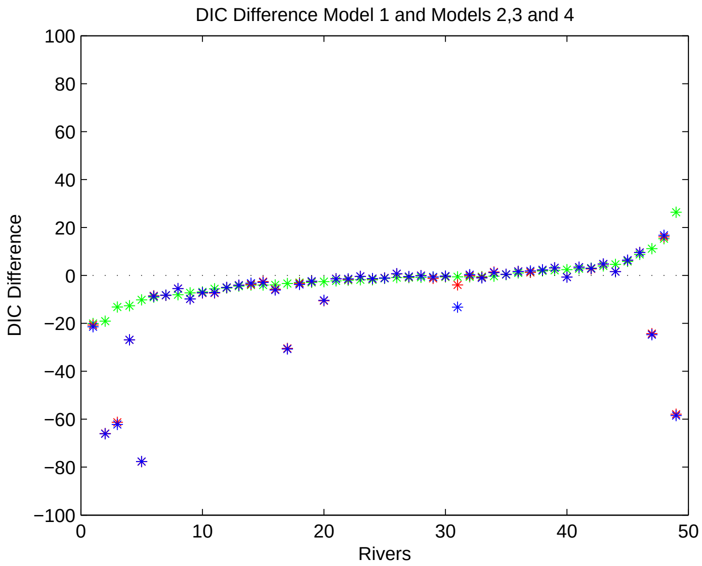

Figure 10: *On the vertical axis the difference in DIC between the Model 1 and Models 2, 3 and are denoted by stars and number of rivers is on the horizontal axes. The difference between Model 1 and Model 2 is denoted by red stars, green stars show the difference between Model 1 and Model 3 and finally blue stars show the difference in DIC between Model 1 and Model 4.*

The river Jökulsá á Dal is used to show how the models proposed in this paper perform. In Figure 11 the behavior of the parameters a and b are shown for Model 3. By allowing the parameters a and b to be a function of water level Model 2, 3 and 4 can give a better fit to the data better than if the parameters are fixed. It is of interest to see how little both parameters change with water level since this is the data set that gives the worst fit in the hole IMO data base if a standard power-law model is used with fixed parameters and no segmentation.

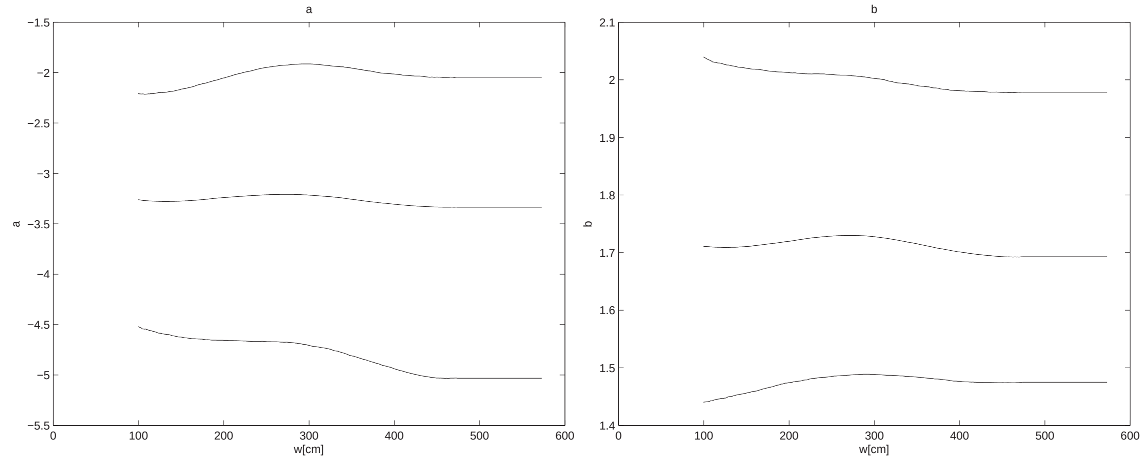

Figure 11: *The 95% posterior interval and the posterior median for the parameters a (left panel) and b (right panel) in Model 3 for Jökulsá á Dal are presented. The vertical axes shows water level (w) in cm while the horizontal axes shows how the parameters in the standard power-law vary with water level.*

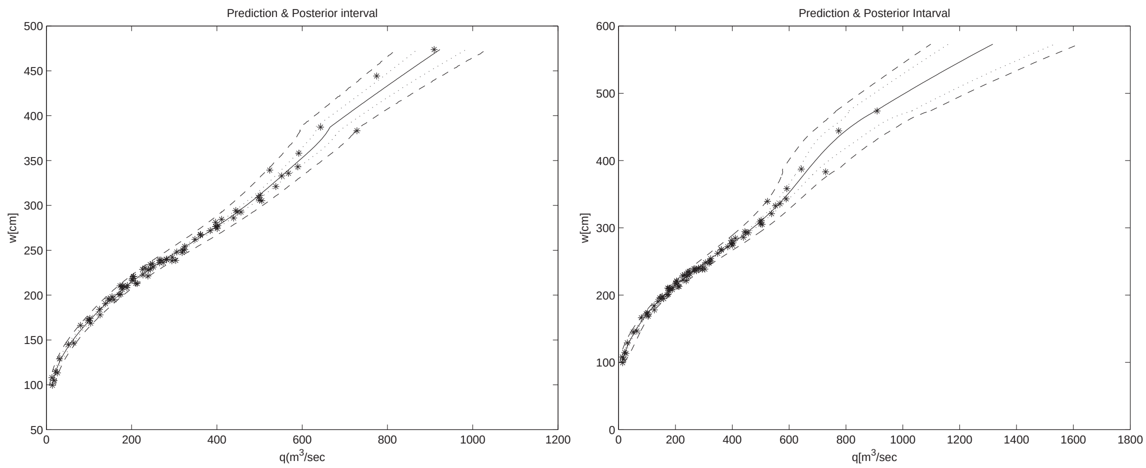

Figure 12: *The fit of Model 1 to data (left panel), and that of Model 3 (right panel). The vertical axes shows water level (w) in cm while the horizontal axes shows the discharge (q), in m^3/sec. The black solid curves show the posterior median of $\mathrm{E}(q)$ and the 95% posterior interval of $\mathrm{E}(q)$ is displayed by the dotted curves. The broken lines show prediction intervals.*

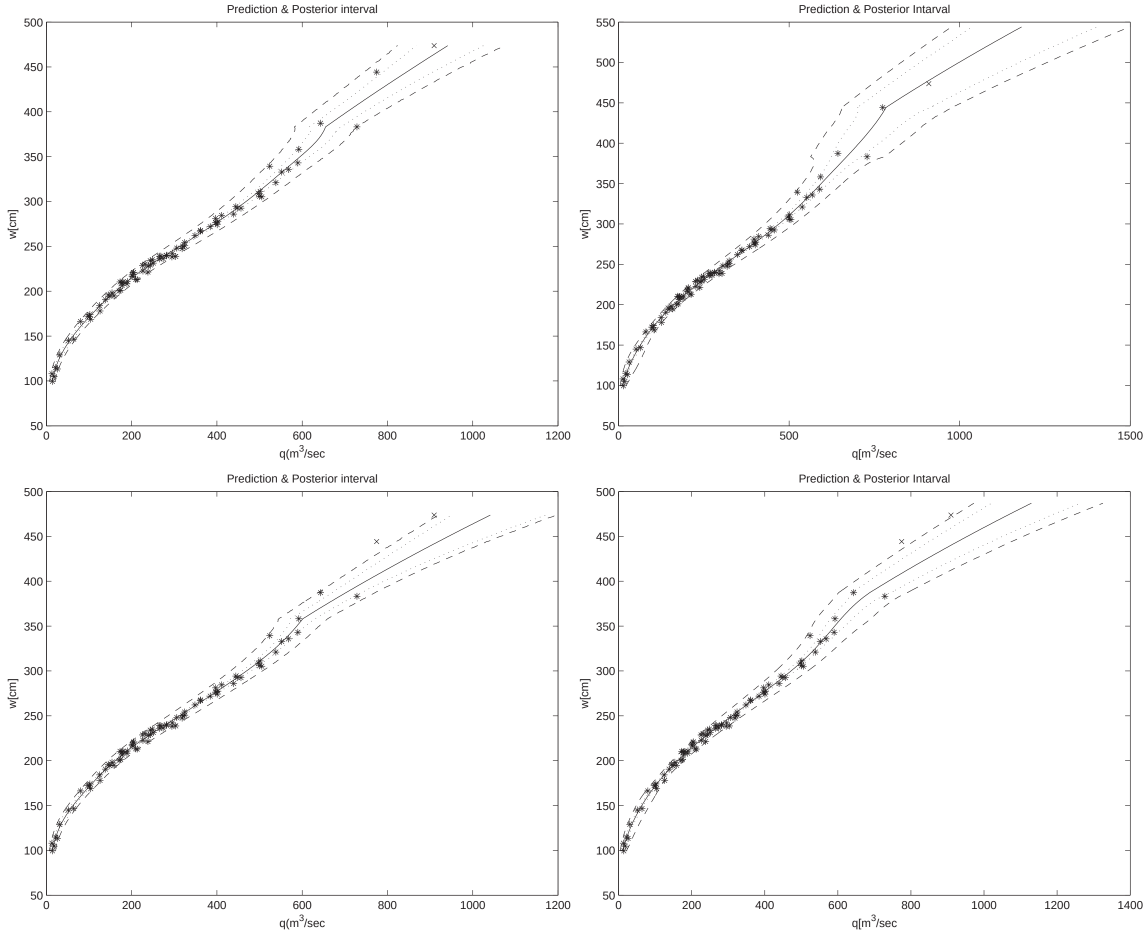

Figure 13: *The fit of Model 1 to data (left panel), and that of Model 3 (right panel). The vertical axes shows water level (w) in cm while the horizontal axes shows the discharge (q), in* m^3/sec. *The black solid curves show the posterior median of* $\mathrm{E}(q)$ *and the 95% posterior interval of* $\mathrm{E}(q)$ *is displayed by the dotted curves. The broken lines show prediction intervals and the x show the excluded data points.*

Figure 12 shows how well Model 1 and Model 3 fit the data set from Jökulsá á Dal. Both models fit the data well. However, the posterior interval is wider for Model 3 and DIC calculation which has difference of the magnitude of 6.67 are in the favor of Model 1 which suggests the use of Model 1 gives a better rating curve.

The fit alone may not be the only thing of interest when deciding which rating curve to use in most cases the extrapolation is as important. Discharge rating curves are frequently used in extrapolation of discharge. To test how well the two models extrapolate then a few of the highest discharge observation are excluded from the data sets and the models are used to predict these excluded discharge values. This is done for highest, second, fourth and fifth highest discharge observations. Figures 13 and 14 show these test results.

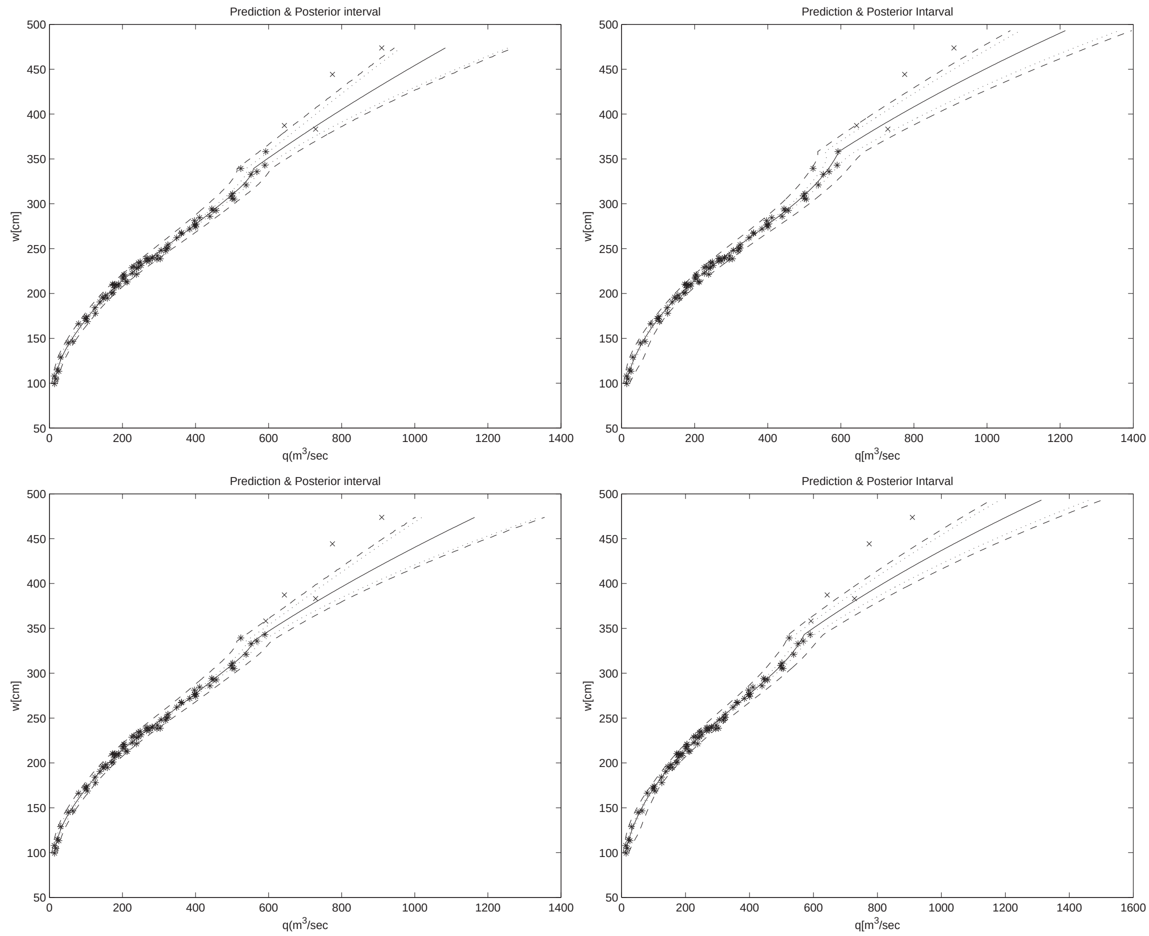

Figure 14: *The fit of Model 1 to data (left panel), and that of Model 3 (right panel). The vertical axes shows water level (w) in cm while the horizontal axes shows the discharge (q), in* m^3/sec*. The black solid curves show the posterior median of* $\mathrm{E}(q)$ *and the 95% posterior interval of* $\mathrm{E}(q)$ *is displayed by the dotted curves. The broken lines show prediction intervals and the x show the excluded data points.*

Figure 13 shows that both models can predict the data if there is one data point excluded from this data set, however, both models overestimate the discharge if two data points are excluded. In Figure 14 both models continue to overestimate the discharge when four and five data points are excluded. There is not an obvious difference from the Figures 13 and 14 in the way Model 1 and Model 3 extrapolate.

Table 6: *The values of b for Model 1 and Model 2.*

	Model 1	Model 3
parameter *b*, 1 excluded point	1.81	1.94
parameter *b*, 2 excluded points	1.60	1.89
parameter *b*, 3 excluded points	1.50	1.82
parameter *b*, 4 excluded points	1.75	1.83
parameter *b*, 5 excluded points	1.42	1.8

From Table 1 and the fact that both models overestimate the discharge when extrapolation then Model 1 more likely perform better when dealing with even greater water level than the largest data point, this is due to lower value of the parameter *b* in Model 1.

CONCLUSIONS

Even though Models 2, 3 and 4 have shown a lot of potential Model 1 performs better over all. However there are data sets which Model 1 gives the worst fit of the four models and for most data sets or 32 out of 49 the difference in DIC is not greater than 5. It would be of interest to find out the characteristics of these data sets. It also could be beneficial to test the models also by using the Bayes factor which combined with the DIC could give a more accurate estimate of the performance of the models.

APPENDIX

BAYESIAN ESTIMATION OF DISCHARGE RATING CURVES

Kristinn Mar Ingimarsson[1], Birgir Hrafnkelsson[2],
Sigurdur Magnus Gardarsson[1] and Arni Snorrason[3]

[1]Faculty of Engineering, University of Iceland, Hjardarhagi 6, 107 Reykjavik, Iceland, e-mail: kmi@hi.is
[2]Division of Applied Mathematics, Science Institute, University of Iceland, Dunhagi 5, 107 Reykjavik, Iceland, e-mail: birgirhr@hi.is
[3]Hydrological Service in Iceland, Orkugardur, Grensasvegur 9, 108 Reykjavik, Iceland, email: asn@os.is

ABSTRACT

The Bayesian approach has been successfully applied to the estimation of discharge rating curves which are based on the standard power-law. Here the standard power-law model is extended by adding a B-spline function to it. The extended model is compared to the standard power-law model through discharge data from the direct run stream Norðurá in Borgarfjörður in the Western part of Iceland. The extended model provides a substantially better fit to these data than the standard power-law model.

INTRODUCTION

Hydrological Service in Iceland (HSI) runs a water level measuring system which collects water level data continuously from rivers around the country while the discharge is only measured a few times a year due to high cost. Hydrological rating curves give discharge as a function of water level. Based on hydraulic principles, the relationship between discharge and water level is given by the standard power-law relationship

$$q = a(w-c)^b \tag{12}$$

(Lambie, 1978; Mosley and McKerchar, 1993) where q is discharge, w is water level, a is a positive scaling parameter, b is a positive shape parameter and c is the water level when the discharge is zero. These parameters are usually estimated from paired measurements of water level and discharge.

The Bayesian approach has been successfully applied to discharge rating curves, see Moyeed and Clarke (2005), Reitan and Petersen-Øverleir (2008b) and Arnason (2005). In the Bayesian approach all unknown parameters are treated as random variables. Prior information about unknown parameters based on previously collected data and/or scientific knowledge can be combined with new data for parametric inference. For example, the fact that the parameter b in equation (12) takes the values 1.5 and 2.5 for rectangular and v-shaped sections, respectively, is an example of prior knowledge that can be used to form the prior distribution for one of the unknown parameters. Combination of the prior distributions and the model for the data results in the posterior distribution which can be used to obtain point estimates and interval estimates for the parameters. HSI has applied the Bayesian approach successfully to data on discharge and water level for discharge rating curve estimation.

In the section Models two statistical models for discharge and water level measurements are introduced. In the section Data a description of discharge and water level data is given. The two models are applied to the data in the section Results and a comparison between the models is made. Finally, in the last section conclusions are drawn.

MODELS

The Bayesian model for discharge rating curves currently used at HSI is given by

$$q_i = a(w_i - c)^b + \varepsilon_i, \quad i = 1, \ldots, n$$

where n is the number of observations for a given site, (w_i, q_i) denotes the i-th pair of observations, ε_i is a mean zero measurement error such that

$$\varepsilon_i \sim \mathrm{N}(0, \eta^2 (w_i - c)^{2b\psi}).$$

In essence this is the same model as the one presented by Petersen-Øverleir (2004). The parameter a is a function of φ and b, that is,

$$a = \exp(\alpha_0 + \alpha_1 b + \varphi)$$

where $\alpha_0 = 4.9468$ and $\alpha_1 = -5.3726$. This reparametrization is motivated by correlation between values of $\ln(\hat{a})$ and $\hat{b}$ which are based on data from HSI, and the values for α_0 and α_1 are selected such that there is no correlation between $\ln(\hat{a})$ and $\ln(\hat{a}) - \alpha_0 - \alpha_1 \hat{b}$, see Arnason (2005). The parameter ψ controls how the error variance behaves as a function of the expected value of q, and η^2 is a scaling parameter for the variance. This model will be referred to as Model 1.

Model 1 is not sufficient for about 5% of the data sets at HSI which calls for modifications. Here, a model is proposed that is an extension of Model 1. It captures the main trend in discharge as a function of water level through the power-law part, $a(w-c)^b$, but a linear combination of B-splines is added, which allows for more flexibility than in Model 1. The form of this model is given by

$$q_i = a(w_i - c)^b + \sum_{l=1}^{L} \lambda_l B_{li} + \varepsilon_i$$

where

$$B_{li} = B_l((w_i - w_{\mathrm{low}})/r), \quad l = 1, \ldots, L, \quad i = 1, \ldots, n,$$

and the terms $B_l(z)$ are cubic B-splines (Wasserman, 2006) which have support on the interval $[0,1]$, $r = w_{\mathrm{upp}} - w_{\mathrm{low}}$, w_{low} and w_{upp} are the lower and upper points of the interval influenced by the B-splines, respectively. Usually $w_{\text{low}} = w_{\text{min}}$, i.e., where w_{min} is the smallest observed water level. The quantity w_{upp} is selected as the 90th percentile of the water level observations or the number such that at least three water level observations are above it. The parameters in $\lambda = (\lambda_1, ..., \lambda_L)$, are unknown and L is the number of B-spline kernels. Further, the error terms are such that

$$\varepsilon_i \sim \mathrm{N}\left(0, \eta^2 (w_i - c_2)^{2b_2}\right), \quad i = 1, ..., n,$$

where b_2 and c_2 are unknown parameters. Note that b_2 plays a similar role as ψ in Model 1. This model will be referred to as Model 2. Further, Model 2 is such that $\lambda_L = 0$ to avoid a jump at $w = w_{\text{upp}}$, and for $w < w_{\text{low}}$, $\mathrm{E}(q) = a(w-c)^b + \lambda_1$.

The Bayesian approach requires specification of prior distributions for each unknown parameter. The same normal prior distributions as used in Arnason (2005) are used here for φ, b and c, see details in Appendix. The prior distribution for b is a truncated normal distribution between 0.5 and 5. The posterior distribution of c will be influenced by its prior distribution but also by the smallest water level measurement, denoted by w_{min} since $c < w_{\text{min}}$. A normal Markov random field prior (Rue and Held, 2005) is assumed for λ, see details in Appendix.

The posterior distribution of $\theta = (\varphi, b, c, \eta^2, b_2, c_2, \lambda, \tau^2, \phi)$ given the data $q = (q_1, ..., q_n)$, $w = (w_1, ..., w_n)$ and w_{min}, is given by

$$\begin{aligned} p(\theta | q, w, w_{\text{min}}) \quad &\propto \quad \textstyle\prod_{i=1}^n p(q_i | \theta, w_i) \times p(\varphi) p(b) p(c) p(\eta^2) p(b_2) p(c_2) \\ &\times \quad p(\lambda | \tau^2, \phi) p(\tau^2) p(\phi) \end{aligned}$$

where $p(q_i | \theta, w_i)$ is a normal density such that

$$p(q_i | \theta, w_i) = \mathrm{N}\left(q_i \middle| a(w_i - c)^b + \sum_{l=1}^{L} \lambda_l B_{li}, \eta^2 (w_i - c_2)^{2b_2}\right),$$

and the part $\prod_{i=1}^n p(q_i | \theta, w_i)$ is the likelihood function.

DATA

The data which are analyzed in this paper were collected by HSI water level measuring system and are from Norðurá in Borgarfjörður by Stekk. The river is located in the Western part of Iceland. The water level of Norðurá has been measured continuously since 1965. The data contain 35 pairs of discharge measurements (q), in m^3/sec, and water level measurements (w) in cm. Norðurá is a direct run stream with 500 km^2 drainage basin above Stekk. In direct run streams the discharge dependents heavily on the season and rainfall.

RESULTS

Here the two models introduced in the section Models are applied to the data from Norðurá in Borgarfjörður for comparison between the two models. Figure 15 shows the fit of the two models to the data. There is a clear difference between the two models. Both models fit the data very well for smaller values of water level while for larger values of water level Model

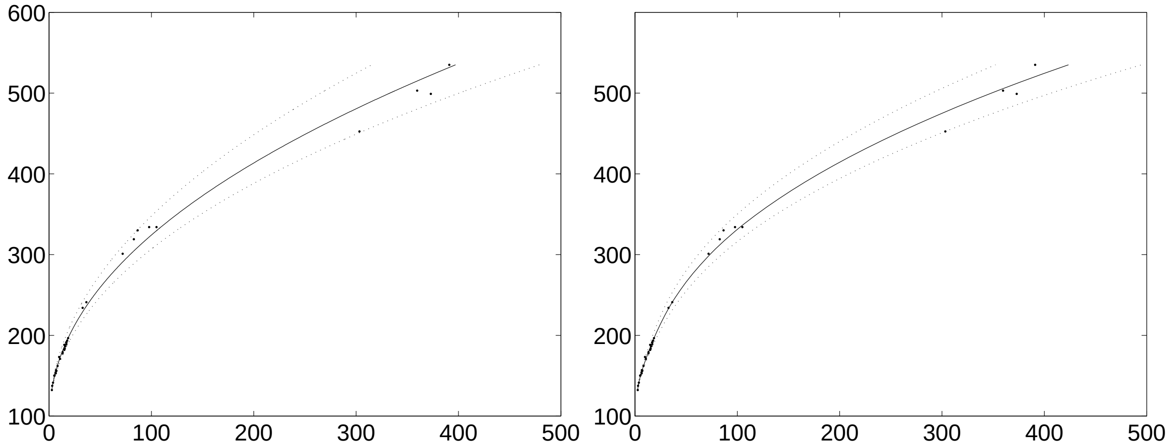

Figure 15: *Water level is on the vertical axes (cm) while discharge is on the horizontal axes (m^3/sec). The points show the observed data from Norðurá in Borgarfjörður and the fit of Model 1 to these data (left panel), and the fit of Model 2 to the same data (right panel). The solid curves show the posterior median of* $\mathrm{E}(q)$ *while the dotted curves show prediction intervals.*

2 seems to perform better. This is due to the lack of flexibility of Model 1, it is not flexible enough to give a good fit to the few observations with large values of water level. The 95% prediction interval is wider for larger values of water level in Model 1 than in Model 2. This is mainly due to the fact that η^2, the parameter controlling the variance of the errors, is smaller in Model 2 than in Model 1. Figure 16 shows the standardized residuals of the two models versus water level.

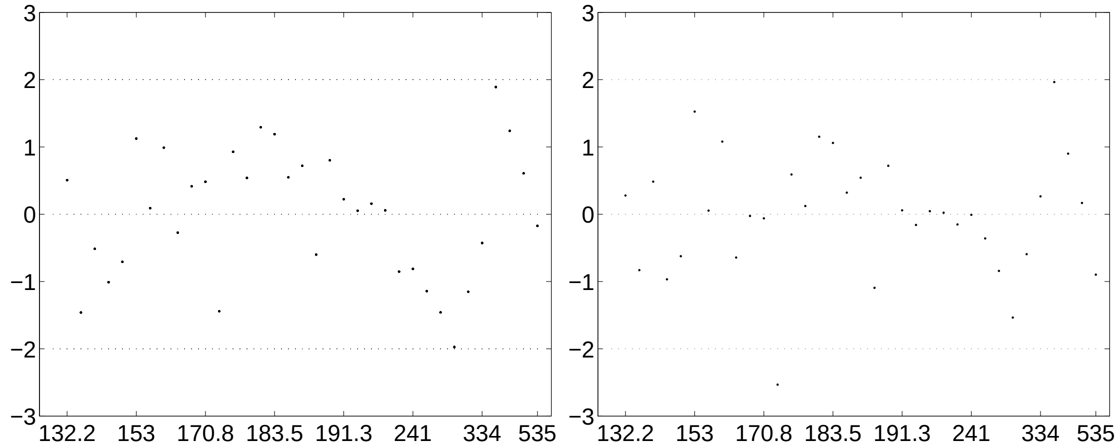

Figure 16: *Water level is on the horizontal axes (cm) but the scale is nonlinear, standardized residuals are on the vertical axes. Standardized residuals for Model 1 (left panel) and Model 2 (right panel).*

In Figure 16, it can be seen that Model 1 (left panel) is not flexible enough to handle the trend found in the standardized residuals while Model 2 yields more convincing standardized residuals. In general standardized residuals should not show any trend and appear to have the same variance for all values of the water level.

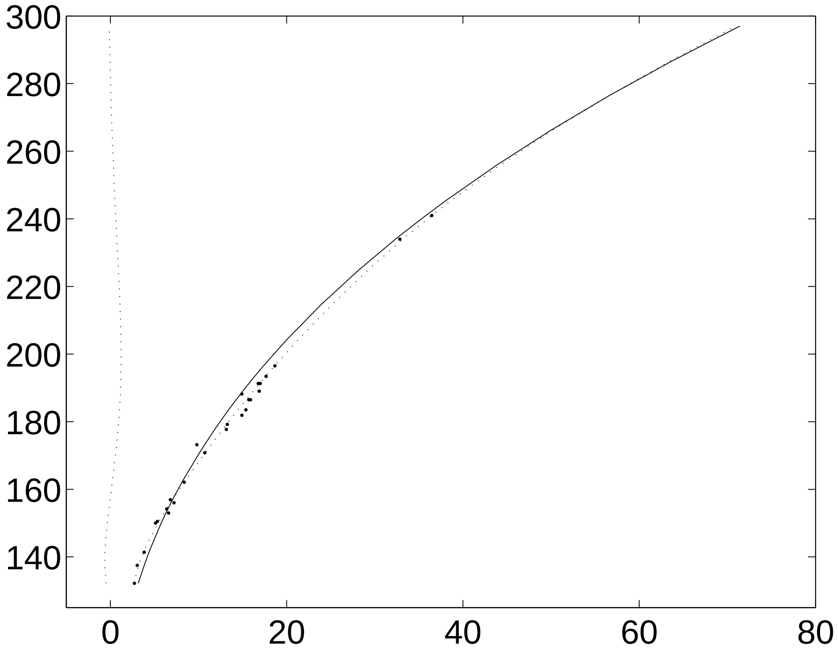

Figure 17: *Water level is on the vertical axes (cm) while discharge is on the horizontal axes (m^3/sec). The standard power-law part (solid line) and the B-spline part (dotted line) of Model 2. The dotted line next to the solid line shows the sum of the two parts. The figure shows the water level values where B-splines have an effect on the model.*

Figure 17 shows the roles that the B-spline part (dotted line) and the standard power-law part (solid line) play in Model 2. The B-spline part picks up the extra trend in the data for the values of the water level below w_{upp} that the standard power-law part can not adjust for to the same extent by itself. This, in turn, allows the standard power-law part in Model 2 to give a better fit above w_{upp}. The B-spline part slowly dies out with increased water level and is practically zero above a value much smaller than w_{upp}. This behavior of the fit for Model 2 indicates that there is no breaking point in the discharge rating curve.

A model criterion called the deviance information criterion (DIC) (Spiegelhalter et al., 2002) is used to further compare the two models. Three other quantities are computed for each of the two models, namely, D_{avg} and $D_{\hat{\theta}}$ which are based on the likelihood function, and p_D, where $p_D = D_{\text{avg}} - D_{\hat{\theta}}$. The quantity p_D is the effective number of parameters. Further, $\text{DIC} = D_{\text{avg}} + p_D$, where DIC is such that the lower it is, the better is the fit of the model to the data. For details on DIC, D_{avg}, $D_{\hat{\theta}}$ and p_D, see Spiegelhalter et al. (2002) and Gelman et al. (2004).

The values of D_{avg}, $D_{\hat{\theta}}$, p_D and DIC for Models 1 and 2 are shown in Table 7. Here Model 2 has lower DIC than Model 1, the difference is more than twelve which is a substantial difference while a difference of size four or less leads to inconclusive results. This confirms that the added complexity of Model 2 does improve the fit.

Model 1 has five effective parameters, however, in the simulation the estimate of p_D is 4.65 which is slightly different from five but this difference can be explained by the stochastic nature of the simulation. The estimated number of effective parameters in Model 2 is 7.66. The number of parameters in Model 2, if counted directly, is 23 since here $L = 15$, however, since the λ's are penalized through the prior of the λ's, the addition of c_2, λ, τ^2 and ϕ in Model 2 compared to the five parameters in Model 1 is equivalent to two or three unconstrained parameters.

Table 7: *DIC and p_D for Model 1 and Model 2 along with D_{avg} and $D_{\hat{\theta}}$.*

	D_{avg}	$D_{\hat{\theta}}$	p_D	DIC
Model 1	137.11	132.47	4.65	141.76
Model 2	120.95	113.29	7.66	128.62

Table 8 shows estimates of the parameters φ, b, c, η^2 and ψ in Model 1 while Table 9 shows estimates of the parameters φ, b, c, η^2, b_2 and c_2 in Model 2. The posterior mean of b is 2.17 for Model 1 while it is 2.51 for Model 2, so, the added flexibility of Model 2 yields a larger shape parameter. Yet, this increase in b will result in a large increase in discharge prediction for water level larger than w_{max}. The precision of these five parameters is better in Model 1 than in Model 2. For example, the 95% Bayesian confidence intervals for b and c are about three times and six times wider in Model 2 than in Model 1, see the 2.5 and 97.5 percentiles for b and c in Tables 2 and 3.

Table 8: *Parameter estimates for Model 1.*

	φ	b	c	ψ	η^2
Post. mean	-0.54	2.17	88.0	1.08	0.004
Post. median	-0.54	2.16	88.4	1.09	0.004
2.5 percentile	-0.66	2.03	78.7	0.94	0.002
25 percentile	-0.57	2.11	85.5	1.05	0.003
75 percentile	-0.51	2.21	91.0	1.13	0.005
97.5 percentile	-0.45	2.33	95.0	1.18	0.012

The smaller precision seen in Model 2 results in less precision in the estimated discharge curve, $\mathrm{E}(q)$, than in Model 1, see Figure **??**. For the larger values of water level the width of the posterior interval for $\mathrm{E}(q)$ is around 35% greater in Model 2 when compared to Model 1. However, for water level values greater than 200 cm Model 1 appears to lack the curvature that the data suggest. The smaller precision in Model 2 is due to its complexity relative to Model 1 but what is gained is a better fit to the data.

Table 9: *Parameter estimates for Model 2.*

	φ	b	c	b_2	c_2	η^2
Post. mean	-0.86	2.51	66.2	2.52	82.5	6.97e-009
Post. median	-0.88	2.53	63.0	2.55	82.4	4.01e-011
2.5 percentile	-1.38	2.05	24.6	1.96	51.1	4.42e-013
25 percentile	-1.05	2.35	50.1	2.34	71.7	5.30e-012
75 percentile	-0.66	2.68	81.5	2.71	93.5	4.58e-010
97.5 percentile	-0.30	2.94	118.0	2.94	113.0	4.45e-008

Posterior simulations for both Model 1 and Model 2 are stable and the simulated chains convergence in all in both cases. In case of Model 1 four chains of length 50 thousand are sufficient while for Model 2 four chains of length 100 thousand are needed to obtain adequate convergence.

CONCLUSIONS

Model 2 shows promising results when fitting rating curves in cases where Model 1 lacks the flexibility needed. The B-spline part of Model 2 is small relative to the standard power-law part but it catches the small deviation from the standard power-law model which results in a more convincing fit for Model 2 than Model 1. This is confirmed with DIC calculations where Model 2 yields a substantially lower value than Model 1.

Model 2 is formulated such that if Model 1 is the correct model or the adequate model then the B-spline part will be close to zero. However, further research is required to test the performance of Model 2 when Model 1 is the correct model.

APPENDIX

The following prior distributions are proposed for the unknown parameters.

$$p(\varphi) = \mathrm{N}(\varphi|\mu_\varphi = 0, \sigma_\varphi^2 = 0.82^2)$$

$$p(b) \propto \mathrm{N}(b|\mu_b = 2.15, \sigma_b^2 = 0.75^2)I(0.5 < b < 5)$$
$$p(c) \propto \mathrm{N}(c|\mu_c = 75, \sigma_c^2 = 50^2)I(c < w_{\min})$$
$$p(\psi) \propto \mathrm{N}(\psi|\mu_\psi = 0.8, \sigma_\psi^2 = 0.25^2)I(0 < \psi < 1.2)$$
$$p(b_2) \propto \mathrm{N}(b_2|\mu_{b2} = 2.15, \sigma_{b2}^2 = 0.75^2)I(1 < b < 6)$$
$$p(c_2) \propto \mathrm{N}(c_2|\mu_{c2} = 75, \sigma_{c2}^2 = 50^2)I(c_2 < w_{\min})$$
$$p(\eta^2) \propto \text{Inv-}\chi^2(\eta^2|\nu_\eta = 10^{-12}, S_\eta^2 = 1)$$
$$p(\phi) = \mathrm{Beta}(\phi|\alpha_\phi = 1, \beta_\phi = 20)$$
$$p(\tau^2) \propto \text{Inv-}\chi^2(\tau^2|\nu_\tau = 10^{-12}, S_\tau^2 = 1)$$
$$p(\lambda|\tau^2, \phi) \propto \mathrm{N}(\lambda|0, \tau^2 D(I - \phi C)^{-1} MD)$$

where $I(A)$ is such that $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. In the prior distribution for λ, I is an identity matrix, D and M are diagonal matrices and C is a neighborhood matrix with constants on the first off-diagonals, other elements are equal to zero.

REFERENCES

Arnason, S., 2005. Estimating nonlinear hydrological rating curves and discharge using the Bayesian approach. Masters thesis, Faculty of Engineering, University of Iceland.

Clarke, R., 1999. Uncertainty in the estimation of mean annual flood due to rating-curve indefinition. Journal of Hydrology 222 (1-4), 185–190.

Di Baldassarre, G., Montanari, A., 2009. Uncertainty in river discharge observations: a quantitative analysis. Hydrol. Earth Syst. Sci 13, 913–921.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2004. Bayesian Data Analysis, 2nd Edition. Chapman and Hall/CRC.

Ingimarsson, K. M., Hrafnkelsson, B., Gardarsson, S. M., Snorrason, A., 2008. Bayesian estimation of discharge rating curves. XXV Nordic Hydrological Conference, 308–317.

Ingimarsson, K. M., Hrafnkelsson, B., Gardarsson, S. M., Snorrason, A., 2010a. Bayesian discharge rating curves based on b-spline smoothing functions. Submitted for publication to Hydrology and Earth System Sciences.

Ingimarsson, K. M., Hrafnkelsson, B., Gardarsson, S. M., Snorrason, A., 2010b. Estimation of discharge rating curves using the standard power-law with varying coefficients. An unpublished technical report.

ISO, 1983. Liquid flow measurements in open channels. Handbook 16. International Standards Organization.

Jeffreys, H., 1961. Theory of probability, 3rd Edition. Oxford: Oxford University Press.

Kass, R., Raftery, A., 1995. Bayes factors and model uncertainty. Journal of the American Statistical Association 90 (430).

Lambie, J., 1978. Measurement of flow-velocity-area methods. Hydrometry: Principles and Practices.

Lohani, A., Goel, N., Bhatia, K., 2006. Takagi–Sugeno fuzzy inference system for modeling stage–discharge relationship. Journal of Hydrology 331 (1-2), 146–160.

Marx, B., Eilers, P., 2005. Multidimensional penalized signal regression. Technometrics 47 (1), 13–22.

Mosley, M., McKerchar, A., 1993. Streamflow. Handbook of Hydrology.

Moyeed, R., Clarke, R., 2005. The use of Bayesian methods for fitting rating curves, with case studies. Advances in Water Resources 28 (8), 807–818.

Pelletier, P., 1988. Uncertainties in the single determination of river discharge: a literature review. Canadian Journal of Civil Engineering 15 (5), 834–850.

Petersen-Øverleir, A., 2004. Accounting for heteroscedasticity in rating curve estimates. Journal of Hydrology 292, 173–181.

Petersen-Øverleir, A., Reitan, T., 2005. Objective segmentation in compound rating curves. Journal of Hydrology 311 (1-4), 188–201.

Reitan, T., Petersen-Øverleir, A., 2008a. Bayesian methods for estimating multi-segment discharge rating curves. Stochastic Environmental Research and Risk Assessment, 1–16.

Reitan, T., Petersen-Øverleir, A., 2008b. Bayesian power-law regression with a location parameter, with applications for construction of discharge rating curves. Stochastic Environmental Research and Risk Assessment 22 (3), 351–365.

Robert, C., 2007. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Verlag.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC.

Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal Of The Royal Statistical Society Series B 64 (4), 583–639.

Wasserman, L., 2006. All of Nonparametric Statistics. Physica-Verlag.