Hydrol. Earth Syst. Sci. Discuss., 7, 2747–2780, 2010
www.hydrol-earth-syst-sci-discuss.net/7/2747/2010/
doi:10.5194/hessd-7-2747-2010

Hydrology and
Earth System
Sciences
Discussions

Bayesian discharge rating curves based on B-spline smoothing functions

K. M. Ingimarsson[1,4], B. Hrafnkelsson[2], S. M. Gardarsson[3], and A. Snorrason[4]

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

Received: 19 February 2010 – Accepted: 9 March 2010 – Published: 5 May 2010

Correspondence to: B. Hrafnkelsson (birgirhr@hi.is)

Published by Copernicus Publications on behalf of the European Geosciences Union.

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 extrapolating large discharge values.

1 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 measurements 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 is true of many types of infrastructures, such as transportation structures, roads and bridges, or flooding of houses in urban areas due to overtopping of levees. Accurate prediction of large discharge, where in general the least data is available as it is hard to obtained reliable data during extreme events, usually involves an extrapolation of the rating curve beyond largest measured data points. 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

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

(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 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 Eq. (1) 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 Sect. 6). 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, 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 over Lohani et al. (2006) approach is that it has a structure that allows for prediction of discharge above the largest observed water level.

In Sect. 2, a description of the 61 discharge and water level data sets is given. Section 3 gives a brief overview of the quantities listed in the section's title, in Sect. 4, the two statistical models for discharge and water level measurements are introduced. In Sect. 5, a description of the prior distributions and posterior distribution is given. The two models are applied to these data sets in Sect. 6 and a comparison between the models is made. Finally, in the last Sect. 7 are drawn.

2 Data

The data which are analyzed in this paper were collected by the IMO water level measuring system and are from 61 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 Skjálfandafljöt by Aldeyjarfoss. The rivers were chosen such that Model 1 will fit reasonably well in one case (Jökulsá á Fjöllum), in two cases Model 1 is insufficient (Norðurá and Skjálfandafljót) and in one case where Model 1 is obviously

performing poorly (Jökulsá á Dal). The data sets contain pairs of discharge measurements (q), in m^3/s, and water level measurements (w) in m.

3 Deviance information criterion and Bayes factor

To evaluate quantitatively the quality of a fit of a model to a data set, a criterion called the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) may be 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 following 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 Sect. 6).

Bayes factors can be used to calculate the posterior probability for each of two or more proposed models conditioned on the data. The following notation is used. In case of two models for the data $\boldsymbol{y}$, the i-th model is denoted by M_i, $p_i(\boldsymbol{y}|\boldsymbol{\theta}_i)$ is the data model, $\boldsymbol{\theta}_i$ are the model parameters, $p_i(\boldsymbol{\theta}_i)$ is the prior for $\boldsymbol{\theta}_i$, $\boldsymbol{\Theta}_i$ 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

$$P(M_1|\boldsymbol{y}) = \frac{P(M_1)\int_{\boldsymbol{\Theta}_1} p_1(\boldsymbol{y}|\boldsymbol{\theta}_1)p_1(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\sum_{j=1,2} P(M_j)\int_{\boldsymbol{\Theta}_j} p_j(\boldsymbol{y}|\boldsymbol{\theta}_j)p_j(\boldsymbol{\theta}_j)d\boldsymbol{\theta}_j} = \left(1 + \frac{P(M_2)}{P(M_1)} \times \frac{1}{B_{12}}\right)^{-1}$$

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

$$B_{12} = \frac{\int_{\boldsymbol{\Theta}_1} p_1(\boldsymbol{y}|\boldsymbol{\theta}_1)p_1(\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1}{\int_{\boldsymbol{\Theta}_2} p_2(\boldsymbol{y}|\boldsymbol{\theta}_2)p_2(\boldsymbol{\theta}_2)d\boldsymbol{\theta}_2}.$$

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, are transformed to $P(M_2|\boldsymbol{y})$ (rounding the numbers from Kass and Raftery (1995) slightly) then the categories presented in Table 1 arise.

Here the evidence against Model 1 is preferred to be strong or decisive ($P(M_2|\boldsymbol{y}) > 0.90$) 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 $\int_{\Theta_i} p_i(\boldsymbol{y}|\boldsymbol{\theta}_i)p_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$, $i = 1,2$, with the formula

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

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

4 Models

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

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

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

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

where a, b and c are as in Eq. (1), 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,

$$a = \exp(\alpha_0 + \alpha_1 b + \varphi) \tag{2}$$

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 (Arnason, 2005).

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

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

where ϵ_i is an error term such that

$$\epsilon_i \sim \mathrm{N}\left(0, \eta^2(w_i - c_2)^{2b_2}\right), \quad i = 1,\ldots,n, \tag{3}$$

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 $q_i > 0$. The variance of Model 2 was chosen to be essentially the same as the variance in Model 1. Note that b_2 and c_2 play a similar role as ψb and c in the variance of 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

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

where $\lambda_L = 0$, and the parameter space of a, b, c and λ is such that $\mathrm{E}(q(w)) \ge 0$. 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

$$G_l(w) = B_l\left(\frac{w - w_{\mathrm{low}}}{w_{\mathrm{upp}} - w_{\mathrm{low}}}\right), \quad w_{\mathrm{low}} \le w \le w_{\mathrm{upp}},$$

for $l = 1,\ldots,L$. 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 a particular 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 Eq. (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 choice 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 ($w_{(n-1)}$), the third largest ($w_{(n-2)}$) and the fourth largest ($w_{(n-3)}$) 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. Table 2 shows the percentage 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 differences 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 $w_{\mathrm{low}} = w_0$ 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:

$$w_0 = w_{\mathrm{low}} \leq w_{\mathrm{min}} < w_{\mathrm{upp}} < w_{\mathrm{max}}.$$

The B-spline parameters in $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_L)$ are unknown (with the constraint that $\boldsymbol{\lambda}_L = 0$) 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 3 it was found that choosing L equal to nine captures the potential improvements gained by Model 2 compared to Model 1. Table 3 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.

5 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 Eq. (2). 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 $\tau^2\mathbf{D}(\mathbf{I} - \phi\mathbf{C})^{-1}\mathbf{MD}$ is assumed for the B-spline coefficients, $\boldsymbol{\lambda}$, (see also in Appendix). This prior works as a penalty for $\boldsymbol{\lambda}$. 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 $\boldsymbol{\lambda}$. 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

$$\mathbf{M}_{ll} = 0.5, \quad ,l = 2,\ldots,\mathbf{L}-1, \quad \mathbf{M}_{11} = \mathbf{M}_{LL} = 1.$$

The neighborhood matrix **C** is such that

$$\mathbf{C}_{l,l-1} = \mathbf{C}_{l,l+1} = 0.5, \quad ,l = 2,\ldots,L-1,$$

and $\mathbf{C}_{12} = \mathbf{C}_{L,L-1} = 1$. The posterior distribution of $\boldsymbol{\theta} = (\varphi, b, c, \eta^2, b_2, c_2, \boldsymbol{\lambda}, \tau^2, \phi)$ given the data $\boldsymbol{q} = (\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n)$, $\boldsymbol{w} = (\boldsymbol{w}_1,\ldots,\boldsymbol{w}_n)$, is given by

$$p(\boldsymbol{\theta} \| \boldsymbol{q}, \boldsymbol{w}) \propto \prod_{i=1}^{n} p(q_i|\boldsymbol{\theta}, w_i) \times p(\varphi)p(b)p(c)p(\eta^2) \times p(b_2)p(c_2)p(\boldsymbol{\lambda}|\tau^2,\phi)p(\tau^2)p(\phi) \tag{5}$$

where $p(q_i|\boldsymbol{\theta}, w_i)$ is a normal density with mean as in Eq. (4) and variance as in Eq. (3). The part $\prod_{i=1}^{n} p(q_i|\boldsymbol{\theta}, w_i)$ 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 $\boldsymbol{\lambda}$ is a multivariate normal distribution where $\boldsymbol{\lambda}$ 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 equal to constant values, namely, their posterior medians, which were found by using 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.

6 Results

In this section the two models introduced in Sect. 4 are applied to the 61 data sets from IMO database for comparison between the two models. As mentioned in Sect. 2 four of the data sets are analyzed in detail. These four data sets are from Norðurá, Jökulsá á Fjöllum, Jökulsá á Dal and Skjálfandafljót. Figure A 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.

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 2 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 four 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 3 shows the roles that the standard power-law part and the B-spline part play in Model 2 for the four rivers. 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 Fig. 3. 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 4, a comparison between the two models is made through DIC and Bayes factor (see Sect. 3). Table 4 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 $\boldsymbol{\lambda}$.

Table 4 shows that in all cases except Jökulsá á Fjöllum, Model 2 has considerably lower DIC than Model 1 indicating better fit for Model 2. 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 Fig. 3. The results in Table 4 and Figs. 2, 3 and 4 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 4. 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. The posterior probability of Model 2 is close to 0.5 for Jökulsá á Fjöllum implying that Model 1 and Model 2 give similar results.

Figure 4 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 Sect. 3). 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. It is however noted that Model 2 never yields worse results than Model 1. Model 2 only fails to improve Model 1 in some cases, so Model 2 could be used in all cases.

Table 5 shows estimates of the parameters a, b and 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 5.

In Table 6, a posterior interval is given for rest of the parameters in Model 1 and in Model 2 except for $\boldsymbol{\lambda}$. 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.

As discussed in Sect. 1, 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 5 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. From these four cases it can be concluded that Model 2 performs considerably better or equally good in predicting discharge for extrapolated water level values greater than w_{max}.

Model 1 and Model 2 (with w_{upp} equal to the third largest water level observation) are compared in terms of prediction. For that comparison only 48 data sets could be used since, as described above, the three highest water level were excluded in estimation of parameters. Model 2 performed better than Model 1 for 29 data sets out of 48 or in case of 60% of data sets. However, in terms of fitting the data, 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.

7 Conclusions

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 61 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 16 of 61 cases (DIC difference greater than ten and posterior probability of Model 2 greater than 0.9). In 36 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 gives an adequate fit, as in the case of Jökulsá á Fjöllum, 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 exception is when the data sets contain few data pairs so there may not be enough information to estimate the B-spline part successfully. Based on the experience gained in the present analysis 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 visually an 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 A

Prior distributions

The following prior distributions are proposed for the unknown parameters.

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

$$p(b) \propto \mathrm{N}(b|\mu_b = 2.15, \sigma_b^2 = 0.4^2)\mathbf{I}(0.5 < b < 5)$$

$$p(c) \propto \mathrm{N}(c|\mu_c = 75, \sigma_c^2 = 50^2)\mathbf{I}(c < w_0)$$

$$p(\psi) \propto \mathrm{N}(\psi|\mu_\psi = 0.8, \sigma_\psi^2 = 0.25^2)\mathbf{I}(0 < \psi < 1.2)$$

$$p(b_2) \propto \mathrm{N}(b_2|\mu_{b2} = 2.15, \sigma_{b2}^2 = 0.4^2)\mathbf{I}(1 < b < 6)$$

$$p(c_2) \propto \mathrm{N}(c_2|\mu_{c2} = 75, \sigma_{c2}^2 = 50^2)\mathbf{I}(c_2 < w_0)$$

$$p(\eta^2) \propto \mathrm{Inv} - \chi^2(\eta^2|\nu_\eta = 10^{-12}, S_\eta^2 = 1)$$

$$p(\phi) = \mathrm{Beta}(\phi|\alpha_\phi = 20, \beta_\phi = 0.5)$$

$$p(\tau^2) \propto \mathbf{I}nv - \chi^2(\tau^2|\nu_\tau = 10^{-12}, S_\tau^2 = 1)$$

$$p(\boldsymbol{\lambda}|\tau^2, \phi) \propto \mathrm{N}(\boldsymbol{\lambda}|0, \tau^2\mathbf{D}(\mathbf{I} - \phi\mathbf{C})^{-1}\mathbf{MD})$$

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

References

Arnason, S.: Estimating nonlinear hydrological rating curves and discharge using the Bayesian approach., Masters thesis, Faculty of Engineering, University of Iceland, 2005. 2749, 2750, 2754, 2758

Clarke, R.: Uncertainty in the estimation of mean annual flood due to rating-curve indefinition, J. Hydrol., 222, 185–190, 1999. 2749

Di Baldassarre, G. and Montanari, A.: Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13, 913–921, 2009, http://www.hydrol-earth-syst-sci.net/13/913/2009/. 2749

Gelman, A., Carlin, J., Stern, H., and Rubin, D.: Bayesian Data Analysis, Chapman and Hall/CRC, 2nd edn., Boca Raton, Florida, USA, 2004. 2752, 2760

ISO: Liquid flow measurements in open channels, Handbook 16, International Standards Organization, 1983. 2748

Jeffreys, H.: Theory of probability, University Press, Oxford, UK, 3rd edn., 1961. 2753

Kass, R. and Raftery, A.: Bayes factors and model uncertainty, J. Am. Stat. Assoc., 90, 773–795, 1995. 2753

Lambie, J.: Measurement of flow-velocity-area methods, in: Hydrometry: Principles and Practices, edited by: Herschy, R. W., John Wiley & Sons Ltd, Chichester, UK, 1–52, 1978. 2749

Lohani, A., Goel, N., and Bhatia, K.: Takagi–Sugeno fuzzy inference system for modeling stage–discharge relationship, J. Hydrol., 331, 146–160, 2006. 2750, 2751

Marx, B. and Eilers, P.: Multidimensional penalized signal regression, Technometrics, 47, 13–22, 2005. 2759

Mosley, M. and McKerchar, A.: Streamflow, in: Handbook of Hydrology, edited by: Maidment, D. R., McGraw Hill, New York, USA, 8.1–8.39, 1993. 2748, 2749

Moyeed, R. and Clarke, R.: The use of Bayesian methods for fitting rating curves, with case studies, Adv. Water Resour., 28, 807–818, 2005. 2749

Pelletier, P.: Uncertainties in the single determination of river discharge: a literature review, Can. J. Civil Eng., 15, 834–850, 1988. 2749

Petersen-Øverleir, A.: Accounting for heteroscedasticity in rating curve estimates, J. Hydrol., 292, 173–181, 2004. 2750, 2754

Petersen-Øverleir, A. and Reitan, T.: Objective segmentation in compound rating curves, J. Hydrol., 311, 188–201, 2005. 2750

Reitan, T. and Petersen-Øverleir, A.: Bayesian methods for estimating multi-segment discharge rating curves, Stoch. Env. Res. Risk A., 23, 1–16, 2008a. 2750

Reitan, T. and Petersen-Øverleir, A.: Bayesian power-law regression with a location parameter, with applications for construction of discharge rating curves, Stoch. Env. Res. Risk A., 22, 351–365, 2008b. 2749

Robert, C.: The Bayesian choice: from decision-theoretic foundations to computational implementation, Springer Verlag, New York, USA, 2007. 2754

Rue, H. and Held, L.: Gaussian Markov Random Fields: Theory and Applications, Chapman & Hall/CRC, Boca Raton, Florida, USA, 2005. 2758

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A.: Bayesian measures of model complexity and fit, J. Roy. Stat. Soc. B, 64, 583–639, 2002. 2752

Wasserman, L.: All of Nonparametric Statistics, Springer Verlag, New York, USA, 2006. 2755

Table 1. Categories for evidence against Model 1.

$P(M_2 \mid y)$	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

Table 2. The prediction performance of Model 2 when w_{upp} was set equal to the second largest ($w_{(n-1)}$), the third largest ($w_{(n-2)}$) and the fourth largest ($w_{(n-3)}$) water level measurement. The table shows the percentage of times selected value of w_{upp} yields the best prediction, the second best prediction and the third best prediction. The total number of datasets used was 48.

w_{upp}	$w_{(n-1)}$	$w_{(n-2)}$	$w_{(n-3)}$
Best prediction	39.6%	29.2%	31.2%
2nd best prediction	22.9%	47.9%	29.2%
3rd best prediction	37.5%	22.9%	39.6%

Table 3. The values of p_d, DIC, for different number of L in Model 2 for the four rivers.

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

Table 4. Model number ($i = 1,2$), values of D_{avg}, $D_{\hat{\theta}}$, p_D, DIC, DIC difference ($\Delta\text{DIC} = \text{DIC}_1 - \text{DIC}_2$) and posterior probability ($p_i = P(M_i|\boldsymbol{q},\boldsymbol{w})$) for Model 1 and Model 2 for the four rivers.

	i	D_{avg}	$D_{\hat{\theta}}$	p_D	DIC	ΔDIC	p_i
Norðurá	1	136.1	131.5	4.56	140.7		0.25
	2	117.6	113.3	4.33	122.0	18.7	0.75
Jökulsá	1	592.1	587.6	4.46	596.6		0.44
á Fjöllum	2	590.8	587.7	3.07	593.8	2.8	0.56
Jökulsá	1	768.9	764.0	4.88	773.8		0.00
á Dal	2	666.8	658.2	8.68	675.5	98.3	1.00
Skjálf-	1	271.1	266.6	4.48	274.6		0.01
andafljót	2	244.2	236.5	7.75	252.0	22.6	0.99

Table 5. Posterior estimates of a, b and c in Model 1 and Model 2.

	Mod. 1			Mod. 2		
	a	b	c	a	b	c
N.á						
Post. med.	15.7	2.16	0.88	9.6	2.45	0.69*
2.5 perc.	11.7	2.06	0.77	6.5	1.94	
97.5 perc.	17.5	2.36	0.93	21.1	2.75	
J. á Fj.						
Post. med.	69.9	2.13	0.29	66.0	2.16	0.25*
2.5 perc.	49.4	1.94	0.11	59.4	2.03	
97.5 perc.	92.2	2.34	0.43	75.4	2.27	
J. á Dal						
Post. med.	112.7	1.68	0.74	107.9	1.48	0.44*
2.5 perc.	92.6	1.52	0.64	73.6	1.22	
97.5 perc.	135.0	1.87	0.82	151.1	1.76	
Skj.fl.						
Post. med.	7.6	3.01	0.06	24.4	2.39	0.58*
2.5 perc.	4.1	2.85	−0.18	20.7	2.23	
97.5 perc.	10.1	3.36	0.18	29.0	2.54	

* c is pre-estimated and therefore a constant, as explained in Sect. 5.

Table 6. Posterior estimates of $\psi \times b$ and η^2 in Model 1 and b_2, c_2, η^2, τ^2 and ϕ in Model 2.

	Mod. 1 $\psi \times b$	 η^2	Mod. 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

* c_2 is pre-estimated and therefore a constant as explained in Sect. 5.

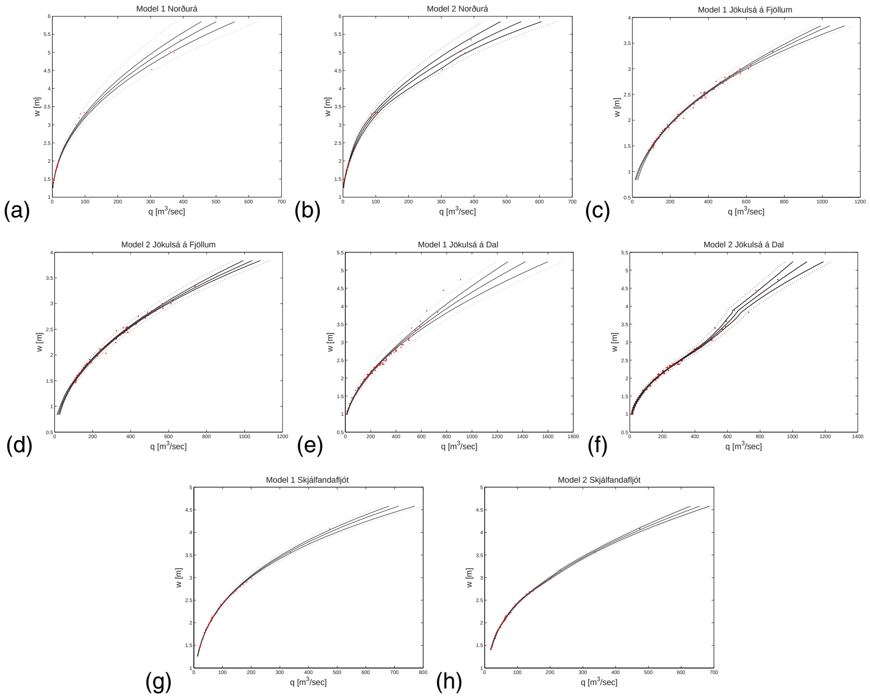

Fig. 1. The fit of Model 1 (**(a)**, **(c)**, **(e)**, **(g)**) and of Model 2 (**(b)**, **(d)**, **(f)**, **(h)**) to the four selected data sets. The vertical axes shows water level (w) in meters while the horizontal axes shows the discharge (q), in m^3/s. The black solid curves show the posterior median of E(q) and the 95% posterior interval of E(q). The dotted curves show prediction intervals.

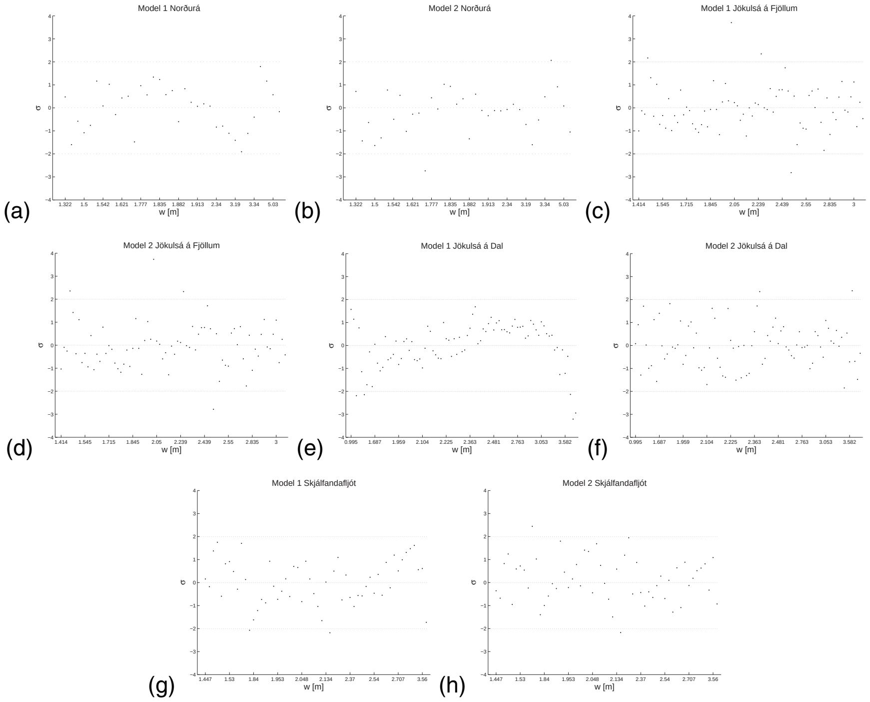

Fig. 2. Standardized residuals for the four selected data sets (vertical axes). Water level is on the horizontal axes (cm) but the scale is nonlinear. Standardized residuals for Model 1 (**(a)**, **(c)**, **(e)**, **(g)**) and Model 2 (**(b)**, **(d)**, **(f)**, **(h)**).

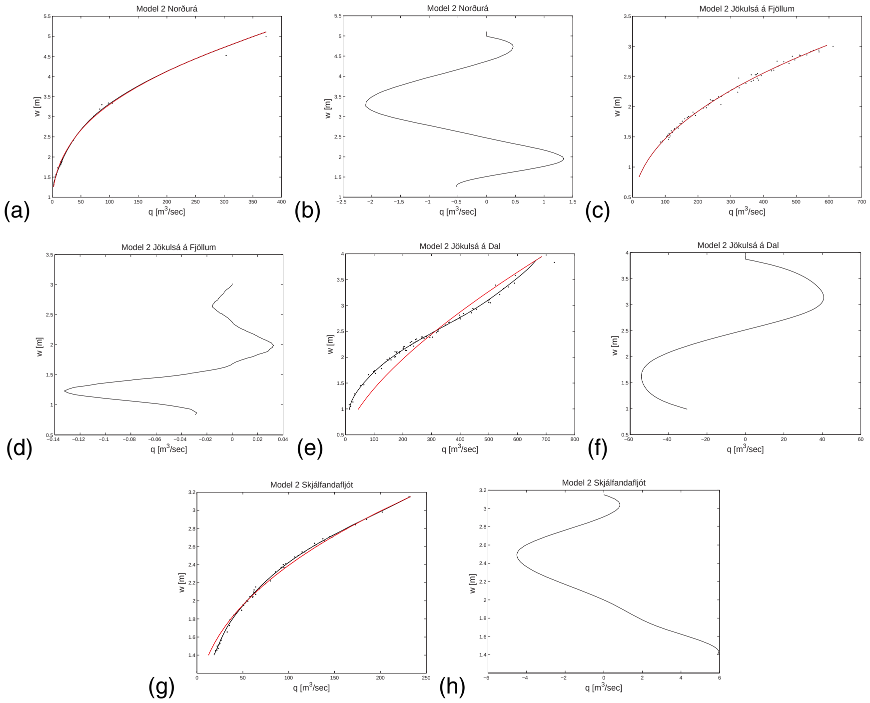

Fig. 3. Figures **(a)**, **(c)**, **(e)** and **(g)** show shows the standard power-law part (solid red curves) of Model 2 and the sum of standard power-law part and the B-spline part of Model 2 (solid black curves) for the four selected data sets. Figures **(b)**, **(d)**, **(f)** and **(h)** show the 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/s).

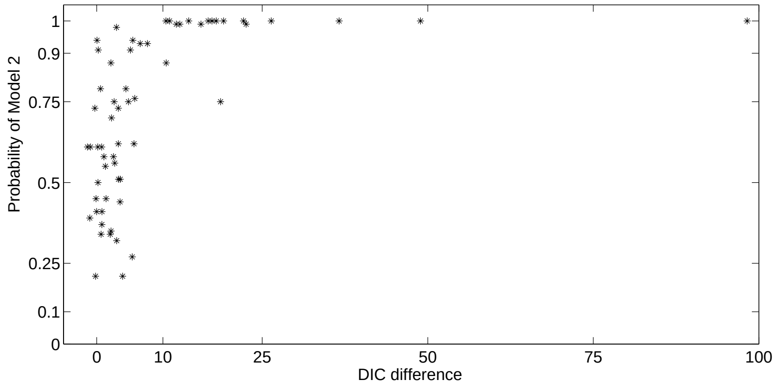

Fig. 4. The difference in DIC between the two models is on the horizontal axis and the posterior probability of Model 2 (based on Bayes factor) is on the vertical axes.

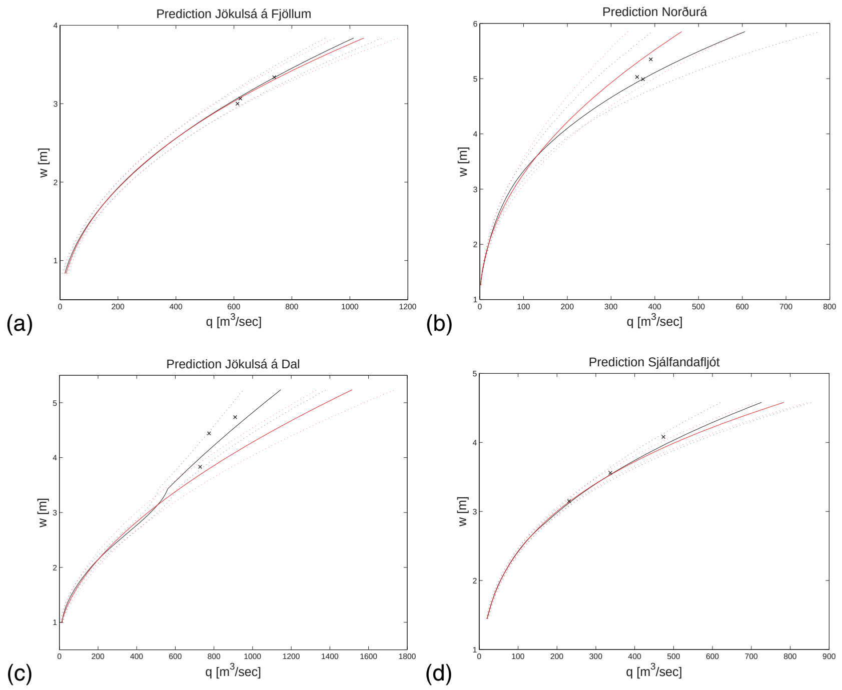

Fig. 5. The solid curves show the posterior median of E(q), red for Model 1 and black for Model 2 for the four selected data sets. The dotted curves show prediction intervals, red for Model 1 and black for Model 2. Water level is on the vertical axes (m) while discharge is on the horizontal axes (m^3/s).