

**Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions**

Journal:	<i>Hydrological Processes</i>
Manuscript ID:	HYP-09-0277.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	
Complete List of Authors:	McMillan, Hilary; National Institute of Water and Atmospheric Research Ltd, - Freer, Jim; University of Bristol, School of Geographical Sciences Pappenberger, Florian; European Centre For Medium Range Weather Forecasts (ECMWF), - Krueger, Tobias; University of East Anglia, School of Environmental Sciences Clark, Martyn; National Institute of Water and Atmospheric Research Ltd, -
Keywords:	River Flow , Model Calibration, Rating Curve, Uncertainty, Stage-Discharge, Gravel-Bed River



view

# Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions

Hilary McMillan<sup>\*1</sup>, Jim Freer<sup>2</sup>, Florian Pappenberger<sup>3</sup>, Tobias Krueger<sup>4</sup> and Martyn Clark<sup>1</sup>

<sup>1</sup> National Institute of Water and Atmospheric Research Ltd, Christchurch, New Zealand

<sup>2</sup> School of Geographical Sciences, University of Bristol, UK

<sup>3</sup> European Centre for Medium-Range Weather Forecasts, Reading, UK

<sup>4</sup> University of East Anglia, UK

\* h.mcmillan@niwa.co.nz

## Abstract

In order to quantify total error affecting hydrological models and predictions, we must explicitly recognise errors in input data, model structure, model parameters and validation data. This paper tackles the last of these: errors in discharge measurements used to calibrate a rainfall-runoff model, caused by stage-discharge rating curve uncertainty. This uncertainty may be due to several combined sources, including errors in stage and velocity measurements during individual gaugings, assumptions regarding a particular form of stage-discharge relationship, extrapolation of the stage-discharge relationship beyond the maximum gauging, and cross-section change due to vegetation growth and/or bed movement.

A methodology is presented to systematically assess and quantify the uncertainty in discharge measurements due to all of these sources. For a given stage measurement, a complete PDF of true discharge is estimated. Consequently new model calibration techniques can be introduced to explicitly account for the discharge error distribution. The method is demonstrated for a gravel-bed river in New Zealand, where all the above uncertainty sources can be identified, including significant uncertainty in cross-section form due to scour and re-deposition of sediment. Results show that rigorous consideration of uncertainty in flow data results in significant improvement of the model's ability to predict the observed flow.

## 1. Introduction

Conceptual hydrological models are important tools for understanding and predicting catchment responses to measured or modelled climate and land-use scenarios. However, the necessary gross simplifications which occur when translating a complex perceptual model of catchment behaviour into a conceptual model lead to recognised model structural omissions and model parameters which cannot be directly related to measured physical properties (Beven, 2006). Calibration methods must therefore be used to identify model parameters, based on measured data. The most commonly used calibration methods, based on minimisation of squared errors, make the implicit assumption that the only source of error is a Gaussian 'measurement error'. In truth, there are many different sources of error including uncertainties in input data (e.g. precipitation, temperature), calibration/validation data (e.g. streamflow), model structure and parameters. Where the incidence and distribution of each of these error sources is not explicitly recognised (a difficult task in very many cases; Beven *et al*, 2008), the calibration process may yield biased parameter estimates (e.g. Kavetski *et*

1  
2  
3  
4 48 *al.*, 2002a,b; Thyer *et al.*, 2009; Vrugt *et al.*, 2008). This in turn leads to biased model  
5 49 predictions, and a loss of the potential opportunity to learn more about model error  
6 50 sources and methods to mitigate these.

7 51 Our aim in designing model calibration techniques must therefore be to properly  
8 52 account for each uncertainty source and appropriately quantify or parameterise the  
9 53 resulting error distribution, which may be non-stationary in time (e.g. see Freer *et al.*,  
10 54 2004). This paper takes one step towards that goal, by presenting a methodology to  
11 55 explicitly quantify one of these uncertainty sources: namely errors in the computed  
12 56 discharge series used to calibrate a rainfall-runoff model, caused by uncertainty in the  
13 57 rating curve used to transform continuously measured stage data into discharge. This  
14 58 uncertainty in turn derives from a combination of sources, including errors in stage  
15 59 and velocity measurements during gaugings, assumption of a particular form of stage-  
16 60 discharge relationship, extrapolation of the stage-discharge relationship beyond the  
17 61 maximum gauging, and cross-section change due to vegetation growth or bed  
18 62 movement. This paper demonstrates how, using knowledge of each of these factors, a  
19 63 complete PDF of true discharge may be estimated for a given measured stage value.

20  
21  
22  
23 64 Several previous studies have investigated methods of including uncertainties in the  
24 65 stage-discharge relationship, from the first to suggest a statistical framework for those  
25 66 uncertainties (Venetis, 1970) to many modern studies (Di Baldassarre and Montanari,  
26 67 2009; Krueger *et al.*, 2009; Liu *et al.*, 2009; Moyeed and Clark, 2005; Pappenberger *et al.*,  
27 68 2006; Petersen-Øverleir, 2004; Reitan and Petersen-Øverleir 2006; 2009). These  
28 69 all rely on fitting a single set of gaugings (i.e. measured stage/discharge points) using  
29 70 a single rating curve of specified form, and investigate the uncertainty in the  
30 71 parameters of that rating curve. For example, Pappenberger *et al.* (2006) use eight  
31 72 data points to fit a power-law curve (Manning equation formulation), and hence  
32 73 determine an ‘envelope curve’: upper and lower acceptable limits on discharge  
33 74 prediction. Krueger (2009) fits stage-discharge relationships to two weirs at  
34 75 experimental field sites, with the power-law form and bed level defined by the  
35 76 appropriate weir equation, and again models are scored as having ‘perfect fit’ within  
36 77 the resulting envelope curve, with linear decline in performance measure outside of  
37 78 this.

38  
39  
40  
41 79 Other studies have considered the possible effects of uncertainty in the stage-  
42 80 discharge relationship on calibration of, and predictions from, rainfall-runoff models.  
43 81 Aronica *et al.* (2006) calibrated a conceptual linear-nonlinear rainfall-runoff model  
44 82 using upper and lower bounds for multipliers of a rating curve, and demonstrated the  
45 83 resulting change in prediction limits. Montanari (2004) simulated uncertainty in the  
46 84 measured discharge by adding Gaussian errors (bounds calculated by consideration of  
47 85 uncertainty sources). Optimised parameter sets using different error realisations were  
48 86 then compared to show induced parameter uncertainty. However, these two studies  
49 87 are both restricted to sequential consideration of alternative rating curves, as opposed  
50 88 to admission of uncertainty during model calibration.

51  
52  
53  
54 89 This paper sets out to build on these previous methods in three ways. Firstly to extend  
55 90 the ‘envelope curve’ method suitability to rivers where there is significant uncertainty  
56 91 in cross-section form due to scour and re-deposition of sediment, and hence sequential  
57 92 gauging measurements may not all belong to a single rating curve. Secondly to  
58 93 produce an explicit PDF of discharge for any given stage, as opposed to upper and  
59 94 lower limits on acceptable discharge. Lastly, to demonstrate how this empirical  
60 95 discharge PDF can be used to form a likelihood function, and used within a Markov

96 Chain Monte Carlo method for parameter calibration with full consideration of  
97 uncertainty in the stage-discharge relationship.

## 98 2. Catchment

99 The method is demonstrated for a gravel-bed river in New Zealand, the Wairau River  
100 in the northern South Island, New Zealand (Figure 1). The Wairau drains an area of  
101 3,825 km<sup>2</sup>, and elevations in the catchment range from sea level to 2,309 m.  
102 Vegetation in the Wairau includes pasture throughout the southern hills, native ever-  
103 green beech forest in the mountains to the west and southwest, a mix of native beech  
104 forest and exotic pine forest on the northern ranges, and vineyards on the Wairau  
105 plains (Figure 1). The Wairau River is a braided gravel-bed river that is  
106 approximately 100 m wide in the lower reaches. Rainfall in the Wairau is lowest over  
107 the Wairau plains and southern hills (600 mm / year) and highest over the western  
108 ranges (5,000 mm / year). There is a small hydropower scheme in the middle reaches  
109 of the Wairau and some irrigation on the Wairau plains, but these have only minor  
110 effects on catchment streamflow.

---

112 **Figure 1.** The Wairau River basin, showing (left) location; (middle) elevation, digital  
113 river network, location of discharge gauging sites (circles) and rainfall stations  
114 (triangles); and (right) land cover. For TopNet simulations the Wairau basin is  
115 disaggregated into 380 sub-catchments, linked with the digital river network (middle).  
116 Figure reproduced from Clark et al. (2008).

---

117  
118 The Wairau is managed for water allocation and flood mitigation purposes; these  
119 applications require long and short term estimates of discharge data statistics  
120 (Williman, 1995; Rae and Wadsworth, 1990; Rae, 1987). In addition to the input and  
121 model structural uncertainty which are usually assumed to dominate more stable river  
122 systems (e.g. Kavetski *et al.*, 2002a,b), scour and re-deposition of the bed gravels (and  
123 additionally anthropogenic gravel extraction) are known to introduce additional errors  
124 into models of the system. Current rainfall-runoff models in use in the catchment use  
125 a deterministic rating curve established from gauging data and adjusted over time to  
126 include new data points and discard older points which are no longer deemed  
127 representative (Ibbitt and Wild, 2005). This is problematic in the case of flood-stage  
128 gaugings which are rarely collected. The process relies on expert judgement to  
129 determine the frequency and extent to which the curve should be updated (Whalley *et*  
130 *al.*, 2001), and implicitly on an assessment of the balance between gauging errors and  
131 rating curve change. The Wairau therefore presents a good example of a catchment  
132 where the assumption of zero uncertainty in the rating curve is unjustified and a  
133 rainfall-runoff model calibration technique that is able to account for rating curve  
134 estimation errors would be a valuable tool.

## 135 3. Data and model

### 136 Flow Data

137 Flow gauging has been undertaken at various locations on the Wairau river since  
138 1937, using stage recorders backed up by gaugings to determine the rating curve (Rae,  
139 1987). The catchment outlet site at Barnett's Bank is used in this study, and represents  
140 the longest and most reliable record for the Wairau. Despite this, there is considerable  
141 scatter in the stage-discharge relationship: refer to Figure 2 for a photograph of the

1  
2  
3 142 gauged cross-section and Figure 4 which shows individual gaugings. This scatter in  
4 143 part represents the difficulties associated with flow gauging in braided, gravel-bed  
5 144 rivers. For the majority of its length the Wairau has a mobile gravel bed where  
6 145 frequent movement of gravel changes the cross-section of the river and determines the  
7 146 relative flow in each of the river braids. This includes the Barnett's Bank gauging site  
8 147 where records show that between 2005 and 2008 the river thalweg switched from the  
9 148 braid nearest the true left bank where the stage recorder is located, to the true right  
10 149 bank, and back again. The impact of river bed movement at this location outweighs  
11 150 any hysteresis effects which are minor due to the relatively steep gradient of the  
12 151 Wairau. The site is also used for gravel extraction which alters the channel cross-  
13 152 section.

---

17 153 Figure 2: Photograph of the gauged cross-section at Barnett's Bank

---

18  
19 154 Frequent gaugings go some way to identifying such changes in flow regime and hence  
20 155 in the required rating curve, however undertaking a new set of gaugings at a full range  
21 156 of river flows is an extended process which may not keep pace with river bed  
22 157 movement. Gaugings are taken by wading at low flows (stage height less than 3m on  
23 158 the gauge) and the exact cross-section used varies depending on braid locations to  
24 159 ensure the safety of the field team. At high flows gaugings are taken from the road  
25 160 bridge crossing the Wairau close to the gauge, and hence may record cross-section  
26 161 changes due to scour around the bridge piers. For flood flows (stage heights over 5m),  
27 162 ADCP gauging from a jet-boat is the preferred method, although gaugings from the  
28 163 bridge are still used in some cases. It is particularly hard to identify ratings for high  
29 164 flow events where scour and fill is continuously occurring during the event (the  
30 165 effects of this can be seen as 'sawtooth' patterns in stage recordings relating to waves  
31 166 of gravel passing the recorder, and are also recorded as a non-zero bed velocity during  
32 167 ADCP (Acoustic Doppler Current Profiler) gaugings; not shown). Clearly multiple  
33 168 gaugings would be required to fully characterise the uncertainty at flood flows;  
34 169 however practical constraints mean that a limited number of such gaugings can be  
35 170 collected.

## 39 171 40 172 **Model**

41  
42  
43 173 The distributed rainfall-runoff model TopNet was used in this study to provide flow  
44 174 predictions in the Wairau. TopNet was developed by combining TOPMODEL (Beven  
45 175 and Kirkby, 1979; Beven et al., 1995), which is most suited to small watersheds, with  
46 176 a kinematic wave channel routing algorithm (Goring, 1994) so as to have a modeling  
47 177 system that can be applied over large watersheds, using smaller sub-basins as model  
48 178 elements (Ibbitt and Woods, 2002; Bandaragoda et al., 2004; Clark et al., 2008).  
49 179 TopNet uses TopModel concepts for the representation of sub-surface storage  
50 180 controlling the dynamics of the saturated contributing area and baseflow recession,  
51 181 with additional components for evapotranspiration, interception (based on the work of  
52 182 Ibbitt, 1971), infiltration (using a Green-Ampt mechanism; Mein and Larsen, 1973)  
53 183 and soil zone. Kinematic wave routing moves the sub-basin inputs through the stream  
54 184 channel network. Complete model equations are provided by Clark et al. (2008).

55  
56  
57 185 The model uses input precipitation and climate data from Tait et al. (2006) who  
58 186 interpolated data from over 500 climate stations in New Zealand across a regular  
59 187 0.05° latitude-longitude grid (approximately 5 km \* 5 km), including data from 12  
60 188 climate stations within the Wairau catchment. These data are provided at daily time

189 steps, and are disaggregated to hourly data before use in the model, based on an  
 190 interpolation of the sub-daily distribution at the climate stations. In this study we use  
 191 data from the winter months of 2004 and 2006, in both cases including flood peaks  
 192 where discharge exceeded the mean annual flood.

193 To apply TopNet in the Wairau, TopNet requires information on catchment  
 194 topography, physical and hydrological properties. This information is available from a  
 195 variety of sources. The New Zealand River Environment Classification (REC; Snelder  
 196 and Biggs, 2002) includes a digital network of approximately 600,000 river reaches  
 197 and related sub-basins for New Zealand. A 30 m Digital Elevation Model (DEM)  
 198 provides topographic properties. Land cover and soil data is available from the New  
 199 Zealand Land Cover Database (LCDB) and the New Zealand Land Resource  
 200 Inventory (LRI; Newsome et al., 2000). The river basin was first disaggregated into  
 201 individual sub-catchments, each one of which becomes a model element. We use the  
 202 Strahler 3 sub-catchments from the REC, which have a typical size of 10 km<sup>2</sup>, and  
 203 split the Wairau Basin into 380 elements. The REC also provides the geometrical  
 204 parameters of the river network. Frequency distributions of the topographic wetness  
 205 index and distance to streams are calculated from the DEM. The wetness index is  
 206 formulated as  $\ln\left(\frac{a}{\tan \beta}\right)$  where  $a$  is the contributing upstream area and  $\beta$  is surface  
 207 slope (Beven and Kirkby, 1979). Average soil and landcover parameters are derived  
 208 from the LRI and LCDB respectively. In total, 12 parameters are required for each  
 209 sub-catchment, of which 6 may be specified using the information described above;  
 210 the remaining 6 must be calibrated (Refer to Table 1 for descriptions of all the  
 211 parameters). In addition, the Manning's  $n$  value for the sub-catchment channel section  
 212 must also be calibrated.

213

#### 214 4. Uncertainty Quantification

215 As previously described, discharge is derived at Barnett's Bank using a rating curve to  
 216 transform stage measurements into discharge estimates (Ibbitt and Wild, 2005). In  
 217 order to quantify discharge uncertainty, we use the concept of an 'uncertain rating  
 218 curve' which decomposes into a PDF of discharge for any given stage measurement.  
 219 To create the uncertain rating curve, we account for the three components of  
 220 uncertainty that were considered most important at this site:

- 221 1. Lack of knowledge as to the current cross-section state due chiefly to bed  
 222 movement, but also possibly affected by seasonal growth of vegetation.
- 223 2. Uncertainty in individual gauging measurements, via inaccuracies of stage and  
 224 velocity measurement, and interpolation between point velocity  
 225 measurements.
- 226 3. Uncertainty as to the correct form of the rating curve, leading to its  
 227 approximation by a functional type, e.g. power law.

228 We now explore each of these components in more detail to define our methodology:

#### 229 Cross-Section State

230 The first component relates to the scatter in the set of {stage, discharge} data points.  
 231 These data represent snapshots of river state during the continuous process of bed  
 232 movement and channel cross-section change, and hence cannot be lumped into a

233 single rating curve. Instead we assume that the most significant changes in bed form  
 234 occur during flood events, and hence divide the complete gauging series into coherent  
 235 sets between major events, each of which represents a more stable phase in the bed  
 236 evolution. We used a 0.5-year return period as the threshold to define a ‘major event’;  
 237 however this measure is subjective, and should be set with knowledge of the  
 238 individual gauging site. These gauging sets are each assumed to represent a possible  
 239 state of the current cross-section, and hence are each used to construct possible rating  
 240 curves. Low flows may be additionally affected by sedimentation between floods but  
 241 this would be captured by a spread of gaugings and hence higher uncertainty within  
 242 the gauging set. The number of individual points in each gauging set varies from 4 to  
 243 12, depending on the length of the stable phase and the frequency of gauging during  
 244 that time (refer to Figure 4 which differentiates the gauging sets). Phases where no  
 245 high stage measurement is made contribute to greater uncertainty at high flows; this is  
 246 in contrast to the previous deterministic rating curves used at Barnett’s Bank which  
 247 were all forced through the highest recorded gauging. These greater uncertainties are  
 248 retained when the uncertain rating curve is constructed, accounting for the component  
 249 of uncertainty relating to rating curve extrapolation. Similarly, the uncertainty  
 250 component relating to shifts in hydraulic geometry is captured through the multiple  
 251 low-flow gauging sets. When making future predictions, no distinction is made as to  
 252 which gauging set is most representative, as it is recognized that rapid changes in bed  
 253 form due to gravel transportation during a flood event may significantly alter the cross  
 254 section in a short period of time. In rivers where bed form changes only over long  
 255 timescales, it might be more appropriate to weight recent gaugings more highly than  
 256 past gaugings.

### 257 **Uncertainty in gauging measurement and rating curve form**

258 These two points are considered together. To account for the uncertainty in gauging  
 259 measurements and hence rating curve shape, our method builds on the idea of a fuzzy  
 260 rating curve developed by Pappenberger *et al.* (2006) and Krueger *et al.* (2009). We  
 261 accept the common assumption that the gauging discharge measurements are  
 262 corrupted by an error of size proportional to the discharge magnitude, here  
 263 approximated as a truncated Gaussian distribution centred on the true discharge  
 264 (Equation 1)

$$\begin{aligned}
 Q_{\text{Measured}} &\propto N(Q_{\text{True}}, \sigma^2) \quad \text{where } |Q_{\text{Measured}} - Q_{\text{True}}| < 3\sigma \\
 &= 0 \quad \text{where } |Q_{\text{Measured}} - Q_{\text{True}}| \geq 3\sigma \quad (\text{Eq 1})
 \end{aligned}$$

266 Where  $Q_{\text{Measured}}$  is the measured discharge,  $Q_{\text{True}}$  is the true discharge. The variance of  
 267 the distribution is chosen so as to give a 95% confidence interval at 8% of the true  
 268 discharge, a typical value for discharge uncertainty which is individually calculated  
 269 for each Barnett’s Bank gauging by the hydrometrists taking into account equipment  
 270 and method accuracy. Hence we set the standard deviation  $\sigma = 0.04 * Q_{\text{True}}$ . It is  
 271 preferable to set the variance according to site-based knowledge, as here, however  
 272 alternatively standard values could be used such as those provided by Pelletier (1988)  
 273 or Whalley (2001) which are comparable with the value used here.

274 The distribution is truncated at  $3\sigma$  (12%) error, which captures > 99% of the  
 275 distribution while avoiding very large error values which are not considered  
 276 reasonable (the 12% bound only represents possible error for a single gauging and  
 277 does not include error due to rating curve interpolation/extrapolation or cross section  
 278 change). Given this error form, the probability distribution for  $Q_{\text{True}}$  can then be

279 calculated numerically for a given gauging measurement of  $Q_{\text{Measured}}$ ; this results in a  
 280 skewed distribution due to the assumption that error magnitude increases with  
 281 discharge. Accepting the common assumption that stage error is invariable with stage  
 282 value, and again using typical uncertainty values recorded for Barnett's Bank  
 283 gaugings by the hydrometrists on-site, true stage is modelled using a Gaussian error  
 284 centred on the measured stage and with standard deviation of 0.02m. As with  
 285 discharge uncertainty, this value is comparable or conservative with respect to  
 286 previous studies (Van der Made, 1982; Petersen-Øverleir and Reitan, 2005).

287 Given error PDFs for discharge and stage, random samples may be drawn from these  
 288 distributions to give many possible pairs of 'true discharge' and 'true stage' for each  
 289 gauging point in the rating set. Using a Monte Carlo approach, multiple sample sets  
 290 are taken to approximate the true joint distribution of the gauging points. Each sample  
 291 set now becomes the basis for fitting of a rating curve.

292 To fit the rating curve, a variation on the method proposed by Krueger *et al.* (2009) is  
 293 used. The method is illustrated in Figure 3 and relies on using each combination of  
 294 three sample points in the set in turn (allowing exact fitting via a three-parameter  
 295 power law) to produce multiple possible rating curves:

- 296 1. Loop through all combinations of three sample points.
- 297 2. Fit the power law equation  $Q = a(h + b)^c$  exactly to these three points (this is  
 298 solved numerically).
- 299 3. Retain fitted rating if the curve intersects the error PDFs for all remaining  
 300 sample points in the gauging set other than the three chosen in step 1.

---

301 Figure 3: Illustration of the proposed Monte Carlo sampling method used to fit  
 302 possible rating curves. (1) Select three points from the gauging set (2) Take a random  
 303 sample of true stage/discharge (3) Fit the power law rating curve and check  
 304 consistency with remaining points. (4) Repeat for multiple samples.

---

305 Through this approach, we limit the individual rating curves to a single-segment,  
 306 power law form. This low-complexity approach is commensurate with the often  
 307 limited number of sample points in a single sample set. Were a denser set of gauging  
 308 measurements available, formulations such as alternative models or multi-segment  
 309 curves could be considered. If prior information were available as to the expected  
 310 rating curve parameter values, it could also be included here via a Bayesian approach.  
 311 Although the individual curve form is thus restricted, the final uncertain rating curve  
 312 may take a free and un-parameterised shape as it combines thousands of individual  
 313 curves (refer to following section and Figure 4 for an example), accounting for  
 314 uncertainty in rating curve form.

### 315 **Constructing the uncertain rating curve**

316 The fitted power law curves for each sample set and for each gauging set are  
 317 combined to produce the uncertain rating curve, as follows. First the rating curves are  
 318 weighted such that each gauging set has equal total weight (as previously stated, no  
 319 distinction is made as to which gauging set is most representative) and all rating  
 320 curves within the same gauging set have equal weight. Then, for each value of stage  
 321 for which discharge estimation is required, the discharge values given by all the rating  
 322 curves are ordered. The constraint  $b > -h$  is enforced for the rating curve parameters,  
 323 i.e. gaugings that were taken at a higher bed level where the gauged height  $h$  is lower  
 324 than the current bed are ignored. There are approximately  $10^5$  individual rating

1  
2  
3 325 curves. Finally, the weighted discharge values combine to provide a CDF for  
4 326 discharge. The form of the CDF hence represents the likelihood of each discharge  
5 327 value based on the distribution of the Monte Carlo samples.

## 8 328 **5. Results of Rating Curve Estimation**

9 329 The method described was used to estimate the form of the uncertain rating curve for  
10 330 the Wairau catchment outlet stage recorder at Barnett's Bank. The results are  
11 331 summarised in Figure 4, which shows quantiles of the estimated true discharge,  
12 332 plotted against the recorded stage.

---

15 333 Figure 4: Quantiles of estimated discharge at Barnett's Bank gauging site at the  
16 334 catchment outlet of the Wairau River, shown in linear (A) and log (B) space. Markers  
17 335 show gauging points used, the symbols identifying discrete gauging sets.

---

19 336 The results show that the combination of gauging sets gives a unique uncertain rating  
20 337 curve with a form tailored to the Barnetts Bank site. For example the log space plot  
21 338 (B) shows two preferred states at low flows, probably corresponding to different river  
22 339 braids carrying the flow thalweg. At flood flows uncertainty is high due to the limited  
23 340 number of gaugings, up to  $\pm 23\%$  of the median discharge.

## 27 341 **6. Model Calibration**

### 28 342 **6.1 Evaluation Measure**

30 343 In order to assess model flow predictions against uncertain validation data, a  
31 344 performance measure must be chosen which reflects the discharge uncertainty in  
32 345 additional to parametric and structural uncertainty. Previous studies have used a  
33 346 variety of methods to do this. Pappenberger and Beven (2004) use a 'multicomponent  
34 347 mapping' technique where an expected observation error structure is used to define  
35 348 membership values according to the distance between observed and modelled  
36 349 hydrographs. Krueger *et al.* (2009) define a timestep-based performance measure  
37 350 which scores any model prediction within the discharge envelope curve (min/max  
38 351 limits of the uncertainty estimation) as an exact match, and otherwise calculates the  
39 352 ratio of the distance between the prediction and the envelope curve to the width of the  
40 353 envelope curve; contrastingly Liu *et al.* (2009) use a triangular performance measure  
41 354 defined via a 'limits of acceptability' approach.

44 355 The result of the rating curve estimation for Barnett's Bank gauging station produced  
45 356 relatively wide uncertainty bounds due to the mobile nature of the river cross-section  
46 357 at the gauging site (Figure 4). However, to avoid overstating the uncertainty in the  
47 358 observed discharge we require a performance evaluation method which retains  
48 359 maximal information content from the gauging data. The method should therefore  
49 360 discriminate between values within the envelope of possible true discharge, using the  
50 361 CDF produced by the rating curve estimation procedure. Hence we use a timestep-  
51 362 based method, which stores the conditional probability of the modeled flow given the  
52 363 observed flow for each timestep: this is an empirical function based on the gauging  
53 364 data and does not have an analytical form. The CDF for estimated discharge at  
54 365 Barnett's Bank is shown in Figure 5 for various measured stage values, and Figure 6  
55 366 shows how this information translates into discharge bound quantiles for an example  
56 367 section of flow record. The conditional probabilities are unique to the gauging record  
57 368 at the Barnett's Bank site: note, for example, that the lower density of gauging points  
58 369 in the lower quantiles of the uncertain discharge curve (Figure 4) lead to skewed

370 CDFs (Figure 5) which in turn lead to lower uncertainty limits that are wider than the  
371 upper limits in Figure 6.

372

---

373 Figure 5: Empirical CDF for discharge at Barnett's Bank: examples at six stage  
374 values.

---

375

---

376 Figure 6: Illustration of discharge median and confidence bounds at Barnett's Bank,  
377 compared with discharge calculated using deterministic rating curve

---

378

379 A variety of methods could be used to aggregate the timestep-based likelihoods over  
380 the modeled time period: in classical Bayesian inference the product of the individual  
381 probabilities ( $p$ ) would be used:

$$382 \quad p(\theta|\mathbf{y}) = \prod_1^N p(\theta|y) \quad (\text{Eq 2})$$

383 Where  $\mathbf{y}$  is a vector of observed data of length  $N$ , and  $\theta$  are the model parameters. We  
384 assume here a uniform bounded prior. However, this method assumes independence  
385 of observed data between timesteps which is unlikely for hydrological time series.  
386 Instead, we use here a modification of this product of probabilities which accounts for  
387 the reduction of information content of the data due to such autocorrelation:

$$388 \quad \hat{p}(\theta|\mathbf{y}) = \left[ \prod_1^N p(\theta|y) \right]^{\frac{ESS}{N}} \quad (\text{Eq 3})$$

389 Where ESS is the 'Effective Sample Size': a measure of the information content of  
390 the data series. Several explanatory notes are required:

391 1. To illustrate the coherence of this form of the conditional probability distribution,  
392 consider the classical assumption of independent Gaussian residuals, combined with a  
393 Jeffrey's prior on  $\sigma^2$ . Box and Tiao (1973) derive the likelihood function:

$$394 \quad p(\theta|\mathbf{y}) \propto M(\theta)^{-N/2} \quad (\text{Eq 4})$$

395 where  $M(\theta)$  is the sum of squared errors and  $N$  is the number of data points.  
396 Following their derivation, but substituting the revised definition  $\hat{p}(\theta|\mathbf{y})$  as above (Eq  
397 3), it can be shown that the likelihood function takes the form:

$$398 \quad \hat{p}(\theta|\mathbf{y}) \propto M(\theta)^{-ESS/2} \quad (\text{Eq 5})$$

399 hence showing that this revision of the product of probabilities gives the expected  
400 likelihood function when used with standard assumptions.

401 2. We use Effective Sample Size (Thiebaut and Zwiers, 1995; Wilks, 1997) as a  
402 measure of information content. This measure is designed to represent the equivalent  
403 number of independent data points and uses autocovariance to quantify the degree of  
404 time coherence in the data series. The ESS is calculated from the true sample size  $N$   
405 as follows:

$$ESS = \frac{N}{\sum_{\tau=-(N-1)}^{\tau=+(N-1)} \left(1 - \frac{|\tau|}{N}\right) \cdot \rho(\tau)} \quad (\text{Eq 6})$$

Where  $\tau$  is time lag and  $\rho(\tau)$  is the corresponding autocorrelation function.

3. While the methods in this study account for rating curve error, we recognize that additional uncertainty components (input uncertainty, model structural uncertainty), not explicitly characterized here, also affect model response. This has important implications for the multiplicative form of the likelihood function, as prediction for a single time-step outside the empirical discharge envelope (which is likely due to these additional error components) would give a total probability of zero. Therefore we use the simplest method possible to incorporate those effects: a uniform (small) error constant  $\varepsilon$  is added to the response surface before the multiplicative step.

$$p(\theta|y) = \max(p(\theta|y), \varepsilon) \quad (\text{Eq 7})$$

This addition has the result that simulations which lie outside the flow uncertainty bounds in some timesteps are disadvantaged but not rejected completely. This step is also important to improve the convergence speed of the MCMC algorithm (Section 6.2) by allowing simulations to be properly ranked. In the extreme case where no predictions lie within the uncertainty bounds, the simulation would be ranked lower than any other realization, and quickly rejected by the MCMC algorithm.

## 6.2 MCMC Parameter Search Method

As has been extensively discussed by Beven (1993; 2005; Beven and Binley, 1992) and others (Wagener and Gupta, 2005), the many sources of uncertainty in a hydrological model application, including but not limited to the measurement uncertainty discussed in this paper, lead to equifinality of parameter sets in providing acceptable model performance. Performance is judged with reference to the observed data, here using the evaluation measure described in Section 6.1. The aim of our calibration technique is to enable an efficient search of the parameter space, identifying those regions where model performance is considered satisfactory in the light of observation error on the discharge. The task is made more difficult by the typically complex nature of the model response surface (Duan *et al.*, 1992; Sorooshian *et al.*, 1993) which may be exacerbated by artefacts of model timestep and solution techniques (Kavetski *et al.*, 2006a,b).

In response to these difficulties, Markov Chain Monte Carlo (MCMC) methods have gained increasing popularity, providing targeted sampling of the parameter space and hence considerable efficiency savings over uniform random sampling (Blasone *et al.*, 2008). These methods enable simulation of complex multivariate distributions by casting them as the invariant distribution of a Markov Chain. We use here a popular version of the original Metropolis-Hastings MCMC algorithm: the adaptive SCEM-UA algorithm (Vrugt *et al.*, 2003) which combines the Metropolis-Hastings sampler with the SCE-UA optimisation method (Duan *et al.*, 1992), using information exchange between multiple sampler chains to improve convergence rates.

MCMC algorithms have traditionally been used to sample posterior distributions derived from classical statistical likelihood functions (e.g. Thiemann *et al.*, 2001; Vrugt *et al.*, 2003). However, the careful use of ‘informal’ likelihood (performance)

measures chosen using modeler judgement can improve the ability of the algorithm to fully explore the response surface (McMillan and Clark, 2009). Here the evaluation measure described in Section 6.1 best reflects our knowledge of the information and uncertainty contained within the observed flow data, and hence is used to describe the response surface.

The method used within the MCMC algorithm to adjust the model parameters is via parameter multipliers. In this approach, the default TopNet model parameters (which vary spatially within the river basin) are adjusted uniformly throughout the river basin using a spatially constant set of parameter multipliers. In this method all sub-catchments receive the same multiplier, i.e. we assume that the spatial distribution of default TopNet parameters is suitable. While this approach represents a simplification, it is a valuable tool to reduce the dimensionality of the parameter estimation problem using prior knowledge of the spatial variation in catchment characteristics. The resulting estimation problem uses seven parameter multipliers, and accordingly the MCMC algorithm is run using seven parallel chains. A burn-in period of 2000 iterations is followed by a parameter estimation period of 1000 iterations. Calculation of the Gelman-Rubin convergence statistic during the burn-in period was used to confirm that the Markov Chain had converged to the stationary distribution representing the model posterior distribution.

## 7. Flow Modelling Results

We now demonstrate how inclusion of explicit discharge uncertainty information can offer additional insights into model calibration, by applying the method described above to calibration of the TopNet model in the Wairau catchment. The calibration is done in two parts. First TopNet is calibrated using flow data produced from the deterministic rating curve recommended for the Wairau to serve as a benchmark with which to compare the new method. Secondly, the calibration is repeated using the uncertain rating curve previously derived. Refer to Figure 6 for an example comparison of the resulting flow data in the two cases.

### Deterministic Rating Curve

Model calibration was carried out for a 6-month period in winter 2004 (1 April 2004 – 1 October 2004). The MCMC algorithm was run using the likelihood function derived in Section 6.2 under the classical assumption of Gaussian error:  $\hat{p}(\theta|\mathbf{y}) \propto M(\theta)^{-ESS/2}$  (i.e. Eq 5) where  $M(\theta)$  is the sum of squared errors and ESS is the effective sample size.

The resulting uncertainty bounds on the flow hindcast are shown in Figure 7A; note that despite the use of the effective sample size measure which reduces the peakedness of the objective function, the uncertainty bounds have a very narrow range and are barely visible except during times of high flood.

### Uncertain Rating Curve

The model calibration was repeated with the same MCMC algorithm, but using the performance measure described in Section 6.1 which incorporates rating curve uncertainty. The resulting flow hindcast with 90% confidence intervals is shown in Figure 7B.

### Comparison of Results

1  
2  
3 494 Both parameter estimation techniques are shown to underestimate the total uncertainty  
4 495 during some periods, demonstrated for example during the recession of the flood peak  
5 496 in late June: for both techniques the estimated flow bounds lie completely outside the  
6 497 estimation flow uncertainty (Figure 7). However, when the uncertain rating curve is  
7 498 used, the flow uncertainty bounds are wider and the model is more successful in  
8 499 predicting the flood peaks, especially during the wetting-up period of early winter. It  
9 500 is especially noticeable that during the highest flood peak, the calibration using the  
10 501 uncertain rating curve includes the median modelled flow within the uncertainty  
11 502 bounds, whereas the calibration using the deterministic rating curve underestimates  
12 503 the flood peak by almost 50%, even at the 90% confidence level.

13 504 It is also desirable to provide a more objective measure of the ability of the two  
14 505 models to span the observed discharge data. For example, consider the percentage of  
15 506 time that the median gauged discharge lies within the bounds of the modelled  
16 507 discharge CDF. During the calibration period, this figure is 68% for the model using  
17 508 the deterministic rating curve vs. 86% for the model using the uncertain rating curve,  
18 509 suggesting an improvement in performance in the latter case. However this measure  
19 510 could be criticised as it favours models with overly wide uncertainty bounds. To  
20 511 overcome this, we consider a generalisation of the rank histogram, which measures  
21 512 how well the spread of a model forecasts represents the true variability of the  
22 513 observations. A rank histogram, usually based on deterministic observation data, is  
23 514 derived by tallying, for each timestep, the quantile at which the observed data lies  
24 515 within the model forecast. A perfect result gives a flat histogram. A 'u' shape  
25 516 histogram indicates an underdispersive model, with many observations lying outside  
26 517 the extremes of the model prediction; conversely a dome shape indicates that the  
27 518 model spread is too large. We extend this to the case of uncertain observed data, by  
28 519 tallying the model quantile at which each observed data quantile lies, for each  
29 520 observed data quantile and for each forecast timestep.

30 521 Such generalised rank histograms are shown in Figure 8, for both deterministic and  
31 522 uncertain rating curves, for the winter 2004 calibration period. The rank histograms  
32 523 show additionally the proportion of the values in the lower and upper quantiles where  
33 524 the observed data lie outside the model bounds. It is clear that both models are  
34 525 underdispersive, i.e. the uncertainty bounds are not wide enough to capture the errors  
35 526 between modelled and measured discharge data. However, the underdispersion is less  
36 527 severe in the case of the uncertain rating curve. This is consistent with the fact that our  
37 528 method has taken into account one source of uncertainty in the modelling procedure,  
38 529 i.e. rating curve uncertainty; but there are still many uncertainty sources not  
39 530 considered which contribute to the underdispersion.

40 531

---

41 532 Figure 7: Parameter estimation incorporating rating curve uncertainty: 90%  
42 533 confidence interval for streamflow at Barnett's Bank during example section of model  
43 534 calibration period. Comparison of results using deterministic rating curve (A) vs.  
44 535 rating curve including uncertainty (B).

---

45 536

---

46 537 Figure 8: Generalised Rank Histogram showing spread of model predictions  
47 538 compared with variability of uncertain discharge data, during calibration period.  
48 539 Results shown for both deterministic and uncertain rating curve cases.

---

49 540

### 541 **Effects on Behavioural Model Parameter Sets**

542 In order to compare the effects of the deterministic vs. uncertain rating curve on  
 543 model behaviour, we examine the differences in the distribution of model parameters  
 544 between the two methods. Figure 9 presents histograms showing the marginal  
 545 posterior probability density function for each of the TopNet model parameters, for  
 546 both the deterministic and the uncertain rating curve calibration run. We observe that,  
 547 in general, parameter distributions are less constrained when the uncertain rating  
 548 curve is used. This result reflects the wider range of model behaviour considered  
 549 behavioural when the errors on the discharge measurements are not artificially  
 550 constrained. The result suggests that the identifiability of parameters such as the  
 551 Topmodel f parameter may be a consequence of the artificially peaked response  
 552 surface due to the use of a deterministic rating curve.

553

---

554 Figure 9: Parameter estimation: comparison of results using deterministic rating curve  
 555 vs. rating curve including uncertainty. Plots show marginal posterior probability  
 556 density functions for each of the TopNet model parameters.

---

557

### 558 **Validation**

559 The model calibration process was tested by running the model with deterministic and  
 560 uncertain rating curve calibrations for an independent validation time period. Again a  
 561 6-month winter period was used (1 April 2006 – 1 October 2006) where significant  
 562 flood events occurred in the Wairau and were recorded at Barnett's Bank gauge. The  
 563 model predictions for these flood events are shown in Figure 10, with the model  
 564 uncertainty bounds compared as before with the 5%, median and 95% quantiles of the  
 565 measured flow data series. The results show that when using the deterministic rating  
 566 curve calibration, the uncertainty in the model predictions is severely underestimated  
 567 in the validation phase, especially during flood events (Figure 10A). When using the  
 568 uncertain rating curve calibration, there is a significant improvement in the model's  
 569 ability to bracket the observed flow (Figure 10B). This improvement is at the cost of  
 570 increased uncertainty in the predictions, however the results suggest that the wider  
 571 uncertainty bounds are warranted due to discrepancies between modelled and  
 572 measured flow in the deterministic rating curve case. While increased forecast  
 573 uncertainty may be unwelcome for decision makers, previous research has  
 574 demonstrated that model predictions including significant uncertainty can be  
 575 successfully accommodated within a flood forecasting framework, using techniques  
 576 such as probabilistic assessment of threshold exceedance (de Roo *et al.*, 2003;  
 577 Pappenberger *et al.*, 2008).

578

---

579 Figure 10: Model validation results: 90% confidence interval for streamflow at  
 580 Barnett's Bank during model validation period. Comparison of results from model  
 581 calibrated using a deterministic rating curve (A) vs. rating curve including uncertainty  
 582 (B).

---

583

---

584 Figure 11: Generalised Rank Histogram showing spread of model predictions  
 585 compared with variability of uncertain discharge data, during calibration period.  
 586 Results shown for both deterministic and uncertain rating curve cases.

---

1  
2  
3 587  
4

5 588 As before, a rank histogram is calculated to show how well the model spread captures  
6 589 the variability in the observed data (Figure 11). Similar to the calibration period, the  
7 590 underdispersion of the model forecast is less severe when the uncertain rating curve is  
8 591 used. In part this measure demonstrates conditional bias in the model, which tends to  
9 592 underpredict during flood peaks and overpredict during recession periods; with the  
10 593 bias being more severe in the case of the deterministic rating curve. Overpredictions  
11 594 also arise from errors in timing of model predictions for the second flood peak in the  
12 595 validation phase, which may signal the influence of other unaccounted-for uncertainty  
13 596 sources.

16  
17 597 **8. Discussion**

18 598 This paper demonstrates the improvements in model performance, and particularly in  
19 599 uncertainty estimation, that can be gained by explicit recognition of uncertainty in the  
20 600 stage-discharge relationship embodied in the rating curve. Application of the model to  
21 601 a ‘validation’ time period showed that ignoring rating curve uncertainty could lead to  
22 602 significant underestimation of the uncertainty associated with the model flow  
23 603 predictions, particularly during flood events. While the improvement is particularly  
24 604 pronounced in mobile-bed rivers, such as the Wairau River considered here, all rivers  
25 605 gauged using a stage-discharge relationship are subject to rating curve uncertainty.  
26 606 Perhaps the most important advance demonstrated by our method was the ability to  
27 607 produce an explicit PDF of discharge as opposed to upper and lower limits on  
28 608 acceptable discharge. This allowed the discharge PDF to be used to form a likelihood  
29 609 function which could be used within a conventional uncertainty estimation method.  
30 610 However subjective choices were not completely removed from the method (for  
31 611 example, choice of return period for temporal segregate of the time series; choice of  
32 612 error distribution for individual measurements (Gaussian used); weighting of gauging  
33 613 sets by time), and these choices might be investigated further in future applications of  
34 614 the method.

35  
36  
37  
38  
39 615 It must be noted that this paper takes only one step towards the goal of total error  
40 616 quantification in hydrological modelling. To achieve that aim, the type of analysis  
41 617 suggested here must be combined with methods to quantify uncertainty due to input  
42 618 (precipitation) error, initial and boundary condition error, structural error, and others.  
43 619 Until then, unaccounted-for uncertainties are implicitly mapped onto parameter  
44 620 uncertainty, which can lead to bias, under- or over-estimation of uncertainty in model  
45 621 predictions. Recognition and evaluation of rating curve uncertainty magnitude may  
46 622 also help to define those situations where the uncertainty could be reduced, for  
47 623 example by increasing the number of verticals in manual discharge measurements.  
48 624 However this study has also demonstrated that the contribution of such “measurement  
49 625 uncertainties” is often small when compared to “natural uncertainties” such as shifts  
50 626 in hydraulic geometry occurring during flood events.

51  
52  
53  
54 627 Note also that the Wairau, despite its challenges, is regularly gauged; in contrast there  
55 628 are many parts of the world where large river systems are remote and difficult to  
56 629 access or monitor. In these types of environments, gaugings will be infrequent or non-  
57 630 existent, and discharge estimation may necessarily be undertaken using remote-  
58 631 sensing methods (Bjerklie *et al.*, 2003; 2005). Such methods introduce new sources of  
59 632 discharge measurement error; and additional difficulties, such as ice cover at the  
60 633 gauging site, may occur in some regions (Shiklomanov *et al.*, 2006). In such areas,

1  
2  
3 634 integrated measures such as mean annual discharge may be required rather than  
4 635 continuous discharge measurement; however these are also strongly affected by rating  
5 636 curve error (Clarke, 1999; Clarke *et al.*, 2000). Many large rivers of the world have  
6 637 complex and unstable morphology affected by multiple channel-changing  
7 638 mechanisms including floods, landslides and changes in sediment supply (e.g. Sarma,  
8 639 2005; Goswamu *et al.*, 1999; Ashworth *et al.*, 2000). Globally, significant discharge  
9 640 uncertainty may be the norm rather than the exception.

10  
11  
12 641 More general recognition of uncertainty in measured flow data will have implications  
13 642 for other hydrological modelling techniques which rely on flow data as input. For  
14 643 example, when data assimilation is used to update model states based on observed  
15 644 flow data (as it has been in the Wairau: Clark *et al.*, 2008), errors in the flow data  
16 645 must be explicitly specified. Good performance of the data assimilation method relies  
17 646 on accurate error estimates; hence it is essential to take in account the multiple  
18 647 sources of rating curve uncertainty such that the error estimates are valid even during  
19 648 flood events. An analysis such as that suggested here would allow those errors to be  
20 649 confidently specified.

21  
22  
23 650 Another example of the implications of uncertainty in flow data is in the recent  
24 651 suggestion that integrated performance measures, which evaluate a range of aspects of  
25 652 model behaviour via a single number, should be replaced by more meaningful  
26 653 'diagnostic signatures' (Gupta *et al.*, 2008). These signatures would use a specific  
27 654 interpretation of model output to focus the evaluation on a particular component of  
28 655 model structure or parameterisation, and to identify deficiencies and suggest  
29 656 improvements to the conceptual model structure. As an example, an analysis of  $dQ/dt$   
30 657 vs.  $Q$  could be used to study the form of the catchment storage-discharge relationship.  
31 658 Where uncertainty in the stage-discharge relationship is recognised, it follows that the  
32 659 'true catchment behaviour' used to define the diagnostic signature ( $Q$  or  $dQ/dt$  in this  
33 660 case) is not known exactly. Hence the true 'diagnostic signatures' will become a fuzzy  
34 661 quantity, with consequences for the methods used to compare them with model  
35 662 output. In whatever form that observed data is used for model calibration, whether via  
36 663 diagnostic signatures, 'soft data' or expert knowledge (Seibert and McDonnell, 2002),  
37 664 manual or automatic calibration (Boyle *et al.*, 2000); it is essential that the  
38 665 information, uncertainty and error within that data is evaluated, so that models are not  
39 666 incorrectly forced to fit uncertain data treated as though it were deterministic.

40  
41  
42  
43  
44 667

## 45 46 668 **9. Conclusions**

47  
48 669 This paper presents a method to quantify uncertainty in river discharge measurements  
49 670 caused by uncertainties in the rating curve used to transform stage values into  
50 671 discharge values. The method was designed to assess the combined uncertainty  
51 672 caused by errors in stage and velocity measurements, rating curve interpolation or  
52 673 extrapolation and cross-section change due to vegetation growth of bed movement.

53  
54 674 We demonstrate how the method can be used to provide a complete PDF (and hence  
55 675 also confidence bounds) for measured discharge, and how this PDF can be used to  
56 676 form a likelihood function to enable model calibration allowing for rating curve  
57 677 uncertainty. The method is tested on the Wairau River in New Zealand, and results for  
58 678 calibration and validation periods using both deterministic and uncertain rating curves  
59 679 are compared. We show that explicit consideration of the uncertainty in flow  
60 680 measurements leads to a flatter response surface with higher parameter uncertainty

1  
2  
3 681 and hence wider uncertainty bounds for flow predictions. Use of the uncertain rating  
4 682 curve therefore provides model predictions with confidence bounds which are more  
5 683 successful at enclosing the measured flow during model validation, and hence we  
6 684 suggest provide a more realistic estimate of model uncertainty.

7  
8 685

## 9 686 **Acknowledgements**

10 687 This work was supported by Foundation for Research, Science and Technology  
11 688 (FRST) (grant C01X0812) and UK Natural Environment Research Council (NERC),  
12 689 Flood Risk from Extreme Events (FREE) programme (grant number NE/E002242/1).  
13 690 We thank the NIWA field team for collecting the Wairau gauging data.

14  
15  
16 691

17 692

## 18 693 **References**

19 694 Aronica GT, Candela A, Viola F, Cannarozzo M, 2006. Influence of rating curve  
20 695 uncertainty on daily rainfall-runoff model predictions, in *Predictions in Ungauged*  
21 696 *Basins: Promise and Progress, IAHS Publ. 303*. Eds. Sivapalan M, Wagner T,  
22 697 Uhlenbrook S, Liang X, Lakshmi V, Kumar P, Zehe E, Tachikawa Y. pp 116 – 124.

23 698 Ashworth PJ, Best JL, Roden JE, Bristow CS, Klaassen GJ, 2000. Morphological  
24 699 evolution and dynamics of a large, sand braid-bar, Jamuna River, Bangladesh.  
25 700 *Sedimentology*, 47(3) 533-555.

26 701 Bandaragoda C, Tarboton DG, Woods R, 2004. Application of TOPNET in the  
27 702 distributed model intercomparison project. *Journal of Hydrology*, 298 (1-4) pp 178-  
28 703 201.

29 704 Beven KJ (1993) Prophecy, reality and uncertainty in distributed hydrologic  
30 705 modelling. *Advances in Water Resources*. 16:41–51

31 706 Beven KJ (2006) A manifesto for the equifinality thesis. *Journal of Hydrology*. 320  
32 707 (1-2) 18-36

33 708 Beven KJ, Binley AM (1992) The future of distributed models: model calibration and  
34 709 uncertainty in prediction. *Hydrological Processes*. 6:279–298

35 710 Beven KJ and Kirkby MJ, 1979. A physically based variable contributing area model  
36 711 of basin hydrology. *Hydrological Sciences Bulletin* 24 1, pp. 43–69.

37 712 Beven KJ, Lamb R, Quinn P, Romanowicz R and Freer J, 1995. Topmodel. In: Singh,  
38 713 V.P., Editor, 1995. *Computer Models of Watershed Hydrology* 18, Water Resources  
39 714 Publications, Highlands Ranch, CO (Chapter 18); pp. 627–668.

40 715 Beven KJ, Smith PJ, Freer JE. 2008. So just why would a modeller choose to be  
41 716 incoherent? *Journal of Hydrology* 354: 15-32

42 717 Blasone R-S, Vrugt JA, Madsen H, Rosbjerg D, Robinson MA, and Zyvoloski GA,  
43 718 2008. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov  
44 719 Chain Monte Carlo sampling. *Advances in Water Resources*, 31, 630-648.

45 720 Bjerklie DM, Dingman SL, Vorosmarty CJ, Bolster CH, Congalton RG, 2003.  
46 721 Evaluating the potential for measuring river discharge from space. *Journal of*  
47 722 *Hydrology*, 278(1-4): 17-38.

- 1  
2  
3 723 Bjerklie DM, Moller D, Smith LC, Dingman SL, 2005. Estimating discharge in rivers  
4 724 using remotely sensed hydraulic information. *Journal of Hydrology*, 309(1-4): 191-  
5 725 209.
- 7 726 Box GEP and Tiao GC, 1973. *Bayesian Inference in Statistical Analysis*. Addison-  
8 727 Wesley.
- 10 728 Boyle DP, Gupta HV, Sorooshian S. 2000. Toward improved calibration of  
11 729 hydrologic models: Combining the strengths of manual and automatic methods. *Water*  
12 730 *Resources Research* 36: 3663-3674
- 14 731 Clark MP, Rupp DE, Woods RA, Zheng X, Ibbitt RP, Slater AG, Schmidt J and  
15 732 Uddstrom MJ, 2008. Hydrological data assimilation with the ensemble Kalman filter:  
16 733 Use of streamflow observations to update states in a distributed hydrological model.  
17 734 *Advances in Water Research*, 31(10): 1309-1324.
- 19 735 Clarke RT, 1999. Uncertainty in the estimation of mean annual flood due to rating-  
20 736 curve indefiniton. *Journal of Hydrology*, 222(1-4): 185-190.
- 22 737 Clarke RT, Mendiondo EM, Brusa LC, 2000. Uncertainties in mean discharges from  
23 738 two large South American rivers due to rating curve variability. *Hydrological*  
24 739 *Sciences Journal-Journal Des Sciences Hydrologiques*, 45(2): 221-236.
- 26 740 De Roo, APJ and 20 others, 2003, Development of a European flood forecasting  
27 741 system, *Int. J. River Basin Management*, 1(1), 49-59
- 29 742 Di Baldassarre G, Montanari A (2009) Uncertainty in river discharge observations: a  
30 743 quantitative analysis. *Hydrol. Earth Syst. Sci. Discuss.*, 6: 39-61.
- 32 744 Duan, Q., S. Sorooshian, and V. Gupta, 1992. Effective and Efficient Global  
33 745 Optimization for Conceptual Rainfall-Runoff Models, *Water Resources Research*.  
34 746 28(4), 1015–1031.
- 36 747 Freer JE, McMillan H, McDonnell JJ, Beven KJ. 2004. Constraining dynamic  
37 748 TOPMODEL responses for imprecise water table information using fuzzy rule based  
38 749 performance measures. *Journal of Hydrology* 291: 254-277
- 40 750 Goring DG, 1994. Kinematic shocks and monoclinal waves in the Waimakariri, a  
41 751 steep, braided, gravel-bed river. *Proceedings of the International Symposium on*  
42 752 *Waves: Physical and Numerical Modelling*, University of British Columbia,  
43 753 Vancouver, Canada, 21–24 August, 1994 pp. 336–345
- 45 754 Goswami U, Sarma JN, Patgiri AD, 1999. River channel changes of the Subansiri in  
46 755 Assam, India. *Geomorphology*, 30(3) 227-244.
- 48 756 Gupta HV, Wagener T, Liu Y, 2008. Reconciling theory with observations : elements  
49 757 of a diagnostic approach to model evaluation. *Hydrological Processes* 22(18): 3802-  
50 758 3813.
- 52 759 Ibbitt RP, 1971: Development of a conceptual model of. interception. Hydrological  
53 760 research progress report. No. 5. Wellington, Ministry of Works.
- 55 761 Ibbitt R and Wild M, 2005. *Flood Forecasting for the Wairau River using a TopNet*  
56 762 *Model*. NIWA client report CHC2005-020 prepared for Marlborough District Council.
- 58 763 Ibbitt RP and Woods R, 2002. Towards rainfall-runoff models that do not need  
59 764 calibration to flow data. In *Friend 2002 – Regional Hydrology: Bridging the Gap*  
60 765 *between Research and Practice*. IAHS Publication no. 274. Eds van Lanen, HAJ and  
766 Demuth, S. pp 189 -196.

- 1  
2  
3 767 Kavetski D, Kuczera G, Franks SW, 2006a. Calibration of conceptual hydrological  
4 768 models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology*. 320 (1-2)  
5 769 173-186.
- 7 770 Kavetski D, Kuczera G, Franks SW, 2006b. Calibration of conceptual hydrological  
8 771 models revisited: 2. Improving optimisation and analysis. *Journal of Hydrology*. 320  
9 772 (1-2) 187-201.
- 11 773 Krueger T, 2009. Uncertainties in Modelling Agricultural Phosphorus Transfers. PhD  
12 774 thesis, Lancaster University
- 14 775 Krueger T, Quinton JN, Freer J, Macleod CJA, Bilotta GS, Brazier RE, Butler P,  
15 776 Haygarth PM, 2009. Uncertainties in data and models to describe event dynamics of  
16 777 agricultural sediment and phosphorus transfer. *Journal of Environmental Quality*  
17 778 38(3): 1137-1148. 10.2134/jeq2008.0179.
- 19 779 Liu YL, Freer J, Beven K, Matgen P. 2009. Towards a limits of acceptability  
20 780 approach to the calibration of hydrological models: Extending observation error.  
21 781 *Journal of Hydrology* 367: 93-103 10.1016/j.jhydrol.2009.01.016
- 23 782 Liu YL, and Gupta HV, 2007: Uncertainty in hydrologic modeling: Toward an  
24 783 integrated data assimilation framework. *Water Resources Research*. 43, W07401,  
25 784 doi:10.1029/2006WR005756.
- 27 785 Mein RG and Larson CL, 1973. Modeling infiltration during steady rain. *Water*  
28 786 *Resources Research* 9, pp. 384–394
- 30 787 Montanari A, 2004. An attempt to quantify uncertainty in observed river flows: effect  
31 788 on parameterisation and performance evaluation of rainfall-runoff models. *AGU Fall*  
32 789 *Meeting Abstracts, 2004*.  
33 790 [http://adsabs.harvard.edu/cgi-bin/nph-bib\\_query?bibcode=2004AGUFM.H11G..03M](http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2004AGUFM.H11G..03M)
- 35 791 Moyeed RA, Clark RT, 2005. The use of Bayesian methods for fitting rating curves,  
36 792 with case studies. *Advance in Water Resources*, 28: 807:818.
- 38 793 Newsome PFJ, Wilde RH, Willoughby EJ, 2000. *Land Resource Information System*  
39 794 *Spatial Data Layers*. Technical Report. Landcare Research NZ Ltd, Palmerston  
40 795 North, NZ.
- 42 796 Pappenberger F, Matgen P, Beven KJ, Henry JB, Pfister L, Fraipont P, 2006.  
43 797 Influence of uncertain boundary conditions and model structure on flood inundation  
44 798 predictions. *Advances in Water Resources* 29: 1430-1449.
- 46 799 Pappenberger F and Beven K, 2004. Functional classification and evaluation of  
47 800 hydrographs based on Multicomponent Mapping. *International Journal of River*  
48 801 *Basin Management* 2(2).
- 50 802 Pappenberger F, Bartholmes J, Thielen J, Cloke HL, Buizza R, de Roo A, 2008. New  
51 803 dimensions in early flood warning across the globe using grand-ensemble weather  
52 804 predictions. *Geophysical Research Letters* 35(10) : L10404
- 54 805 Pelletier, P.M. 1988. Uncertainties in the single determination of river discharge: a  
55 806 literature review. *Canadian Journal of Civil Engineering*. 13, 834 – 850.
- 57 807 Petersen-Øverleir A, 2004. Accounting for heteroscedasticity in rating curve  
58 808 estimates. *Journal of Hydrology*. 292(1-4): 173-181.
- 59  
60

- 1  
2  
3 809 Petersen-Øverleir, A., and Reitan, T. 2005. Uncertainty in flood discharges from  
4 810 urban and small rural catchments due to inaccurate head measurement, *Nordic*  
5 811 *Hydrology*, 36, 245-257.
- 7 812 Rae SN and Wadsworth V, 1990. Wairau river catchment flood forecasting. *NZ*  
8 813 *Journal of Hydology* 29(1): 1:17.
- 10 814 Rae SN, 1987. *Water and soil resources of the Wairau*. Marlborough catchment and  
11 815 regional water board, Blenheim. pp 301.
- 13 816 Reitan T and Petersen-Øverleir A, 2006. Existence of the frequentistic estimate for  
14 817 power-law regression with a location parameter, with applications for making  
15 818 discharge rating curves. *Stochastic Environmental Research and Risk Assessment*,  
16 819 20(6), 445–453.
- 18 820 Reitan T and Peterson-Øverleir A, 2009. Bayesian methods for estimating multi-  
19 821 segment discharge rating curves. *Stochastic Environmental Research and Risk*  
20 822 *Assessment* DOI 10.1007/s00477-008-0248-0.
- 22 823 Sarma JN, 2005. Fluvial process and morphology of the Brahmaputra River in Assam,  
23 824 India. *Geomorphology* 70 (3-4): 226-256.
- 25 825 Schmidt AR, 2002. *Analysis of Stage-Discharge Relations for Open-Channel Flow*  
26 826 *and their Associated Uncertainties*. Unpublished Thesis, University of Illinois.
- 28 827 Seibert J, McDonnell JJ. 2002. On the dialog between experimentalist and modeler in  
29 828 catchment hydrology: Use of soft data for multicriteria model calibration. *Water*  
30 829 *Resources Research* 38: art. no.-1241
- 32 830 Shiklomanov AI, Yakovleva TI, Lammers RB, Karasev IP, Vorosmarty CJ, Linder E,  
33 831 2006. Cold region river discharge uncertainty - estimates from large Russian rivers.  
34 832 *Journal of Hydrology*, 326(1-4): 231-256.
- 36 833 Smith P, Beven KJ, Tawn JA, 2008. Informal likelihood measures in model  
37 834 assessment: Theoretic development and investigation. *Advances in Water Resources*  
38 835 31(8): 1087:1100
- 40 836 Snelder TH; Biggs BJB, 2002. Multi-scale river environment classification for water  
41 837 resources management. *Journal of the American Water Resources Association* 38(5):  
42 838 1225–1240.
- 44 839 Sorooshian, S., Q. Duan, and V. K. Gupta (1993), Calibration of Rainfall-Runoff  
45 840 Models: Application of Global Optimization to the Sacramento Soil Moisture  
46 841 Accounting Model, *Water Resources Research*, 29(4), 1185–1194.
- 48 842 Tait AB, Henderson RD, Turner RW, Zheng X, 2006. Thin plate smoothing spline  
49 843 interpolation of daily rainfall for New Zealand using a climatological rainfall surface.  
50 844 *International Journal of Climatology*. 2097-2115.
- 52 845 Thiebaut HJ and Zwiers FW, 1995. The interpretation and estimation of effective  
53 846 sample size. *Journal of Climate and Applied Meteorology* 23: 800-811.
- 55 847 Thiemann M, Trosset M, Gupta H, Sorooshian S, 2001. Bayesian recursive parameter  
56 848 estimation for hydrologic models. *Water Resources Research*. 37 (10) 2521-2535.
- 58 849 Thyer M, Renard B, Kavetski D, Kuczera G, Franks SW, Srikanthan S, 2009. Critical  
59 850 evaluation of parameter consistency and predictive uncertainty in hydrological  
60 851 modeling: A case study using Bayesian total error analysis, *Water Resources*  
852 *Research*, 45, W00B14, doi:10.1029/2008WR006825.

- 1  
2  
3 853 Van der Made, J.E. 1982. Determination of the accuracy of water-level observations.  
4 854 Advances in Hydrometry (Proceedings of the Exeter Symposium, July 1982). IAHS  
5 855 Publ. no. 134, 172-184.
- 7 856 Venetis C, 1970. A note on the estimation of the parameters in logarithmic stage-  
8 857 discharge relationships with estimation of their error. *Hydrological Sciences Bulletin*,  
9 858 IASH. 15, 105-111.
- 11 859 Vrugt JA, Gupta HV, Bouten W, Sorooshian S, 2003. A Shuffled Complex Evolution  
12 860 Metropolis algorithm for optimisation and uncertainty assessment of hydrologic  
13 861 model parameters. *Water Resources Research*, 39 (8) 1201.
- 15 862 Whalley N, Iredale RS, Clare AF, 2001. Reliability and uncertainty in flow  
16 863 measurement techniques - Some current thinking. *Physics and Chemistry of the Earth*  
17 864 Part C-Solar-Terrestrial and Planetary Science, 26(10-12): 743-749
- 19 865 Williman EB, 1995. Flood frequency analysis for the Wairau river, Marlborough. *NZ*  
20 866 *Journal of Hydrology* 33(2): 71:93.
- 22 867 Wilks DS, 1997. Resampling Hypothesis Tests for Autocorrelated Fields. *Journal of*  
23 868 *Climate*, 10(1): 65-82.
- 25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

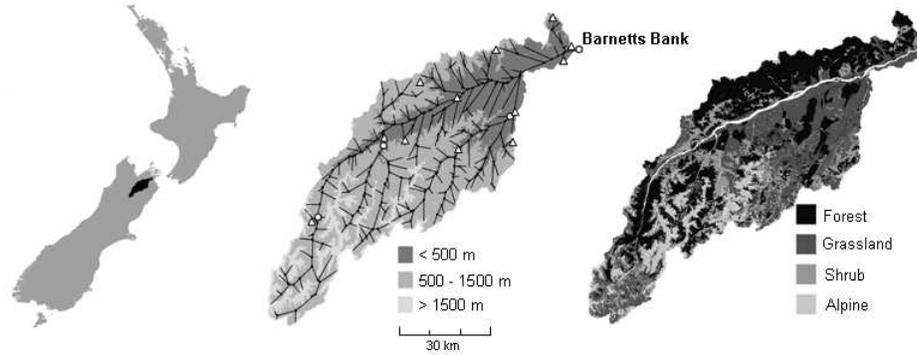
869  
870  
871

Table 1

TOPNET model parameters

	Name	Estimation
<b>Sub-basin Parameters</b>		
$f$ ( $m^{-1}$ )	Saturated store sensitivity	Constant = 12.4 <b>(multiplier calibrated)</b>
$K_0$ (m/h)	Surface saturated hydraulic conductivity	Constant = 0.01 <b>(multiplier calibrated)</b>
$\Delta\theta_1$	Drainable porosity	From soils <b>(multiplier calibrated)</b>
$\Delta\theta_2$	Plant available porosity	From soils <b>(multiplier calibrated)</b>
D (m)	Depth of soil zone	Depth $\frac{1}{4} l=f$ from soils <b>(multiplier calibrated)</b>
C	Soil zone drainage sensitivity	1
$\phi$ (m)	Wetting front suction	From soils
V (m/s)	Overland flow velocity	Constant = 0.1 <b>(multiplier calibrated)</b>
CC (m)	Canopy capacity	From vegetation
Cr	Intercepted evaporation enhancement	From vegetation
A	Albedo	From vegetation
Lapse ( $^{\circ}C/m$ )	Lapse rate	0.0065
<b>Channel parameters</b>		
N	Mannings n	Constant = 0.024 <b>(multiplier calibrated)</b>
A	Hydraulic geometry constant	0.00011
B	Hydraulic geometry exponent	0.518
<b>State variables</b>		
$z'$ (m)	Average depth to water table	Saturated zone drainage matches initial observed flow
SR (m)	Soil zone storage	0.02
CV (m)	Canopy storage	0.0005

872  
873



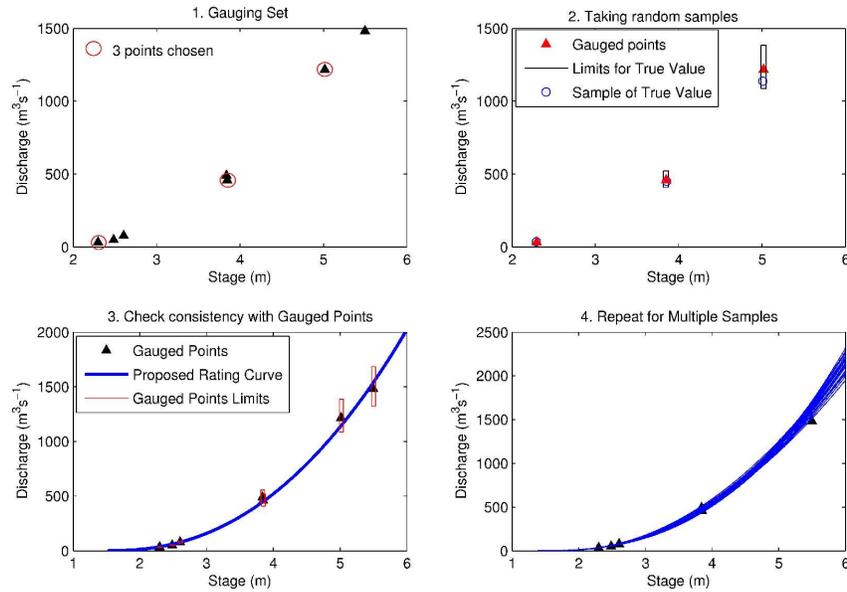
The Wairau River basin, showing (left) location; (middle) elevation, digital river network, location of discharge gauging sites (circles) and rainfall stations (triangles); and (right) land cover. For TopNet simulations the Wairau basin is disaggregated into 380 sub-catchments, linked with the digital river network (middle). Figure reproduced from Clark et al. (2008).  
219x100mm (96 x 96 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



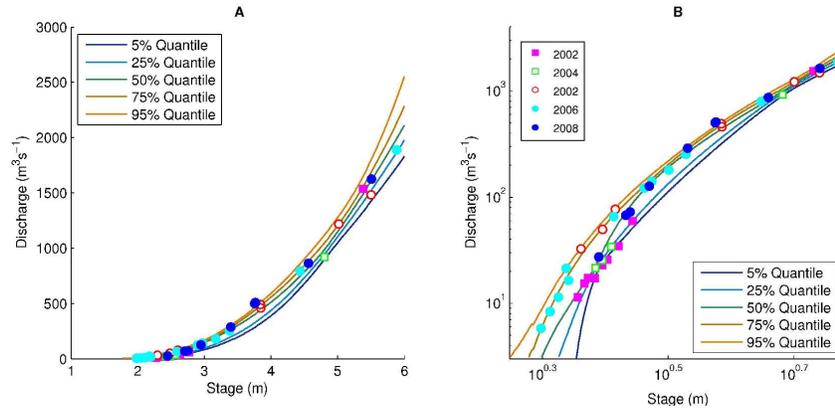
Photograph of the gauged cross-section at Barnett's Bank  
651x504mm (72 x 72 DPI)

Review



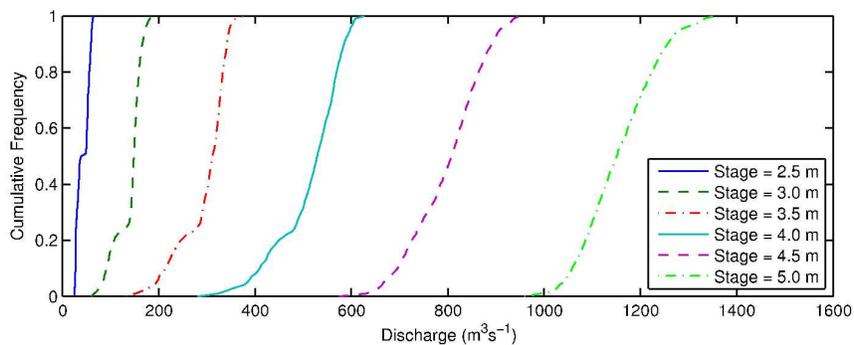
: Illustration of the proposed Monte Carlo sampling method used to fit possible rating curves. (1) Select three points from the gauging set (2) Take a random sample of true stage/discharge (3) Fit the power law rating curve and check consistency with remaining points. (4) Repeat for multiple samples.

243x156mm (600 x 600 DPI)



Quantiles of estimated discharge at Barnett's Bank gauging site at the catchment outlet of the Wairau River, shown in linear (A) and log (B) space. Markers show gauging points used, the symbols identifying discrete gauging sets.

250x103mm (600 x 600 DPI)

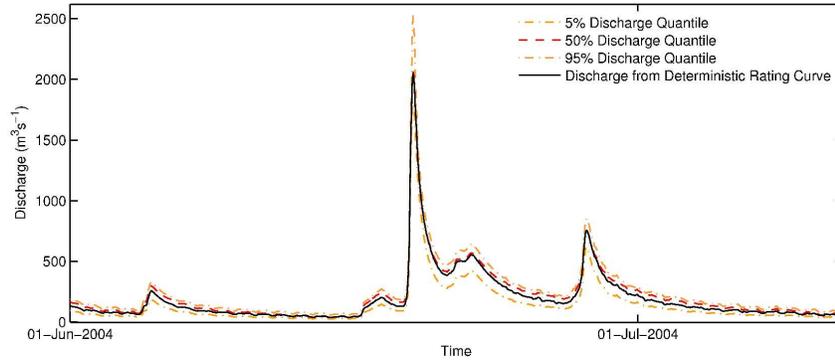


Empirical CDF for discharge at Barnett's Bank: examples at six stage values.  
208x74mm (600 x 600 DPI)

Peer Review

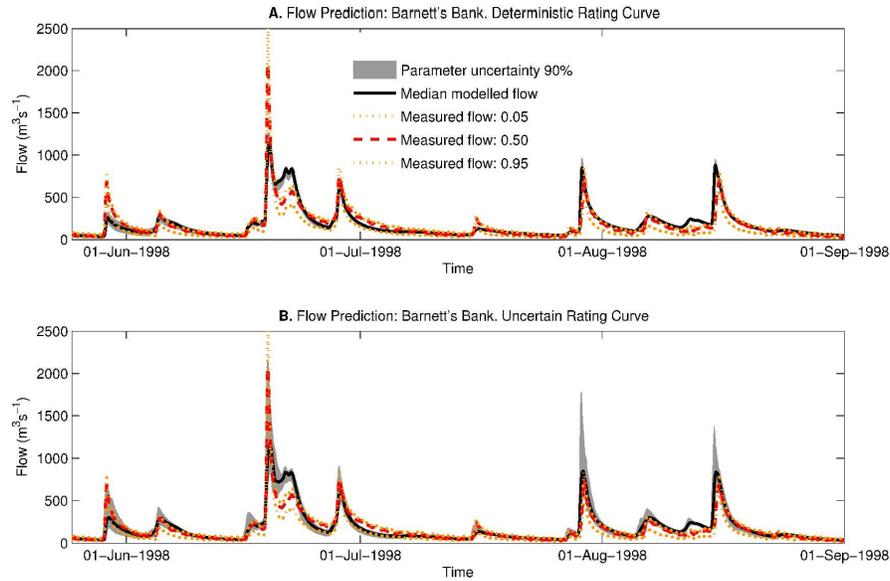
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



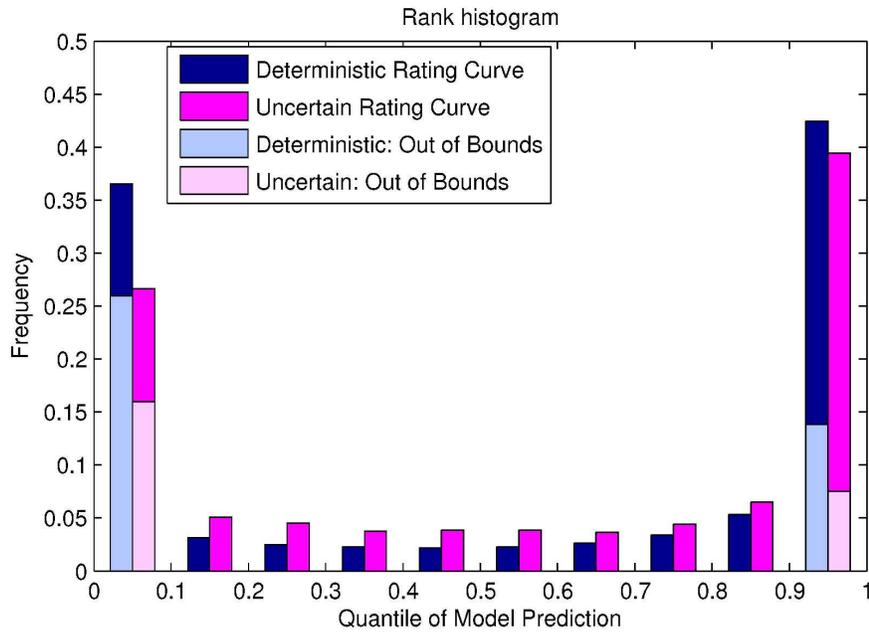
: Illustration of discharge median and confidence bounds at Barnett's Bank, compared with discharge calculated using deterministic rating curve  
250x101mm (600 x 600 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Parameter estimation incorporating rating curve uncertainty: 90% confidence interval for streamflow at Barnett's Bank during example section of model calibration period. Comparison of results using deterministic rating curve (A) vs. rating curve including uncertainty (B).  
247x158mm (600 x 600 DPI)

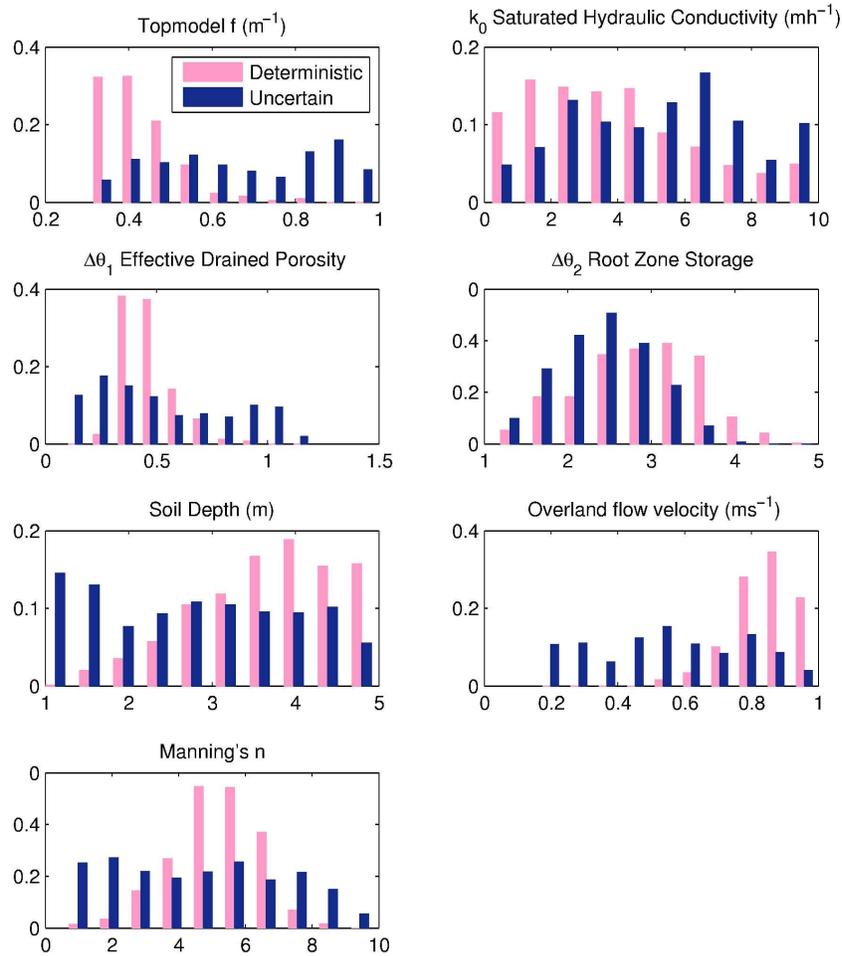
Review



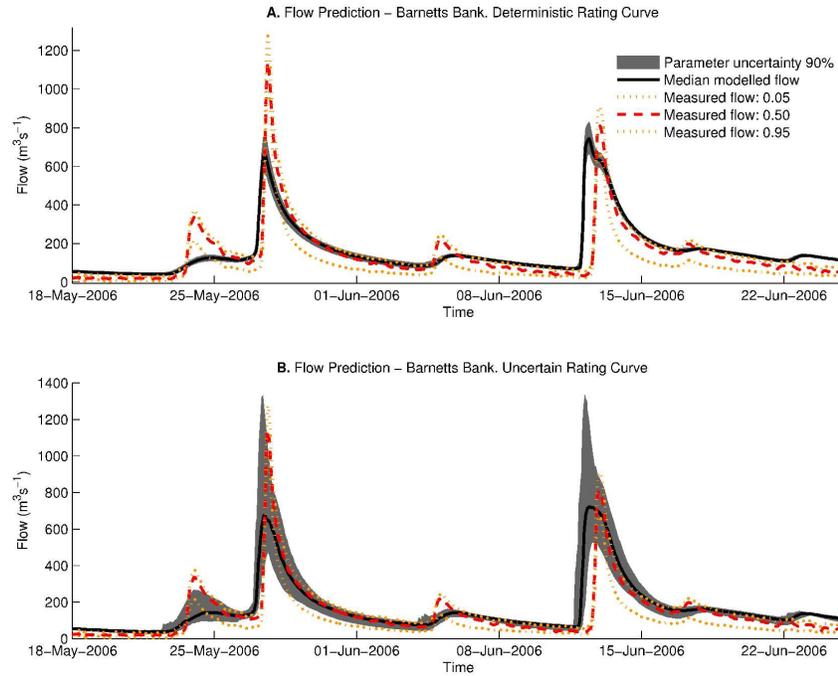
Generalised Rank Histogram showing spread of model predictions compared with variability of uncertain discharge data, during calibration period. Results shown for both deterministic and uncertain rating curve cases.  
155x101mm (600 x 600 DPI)

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

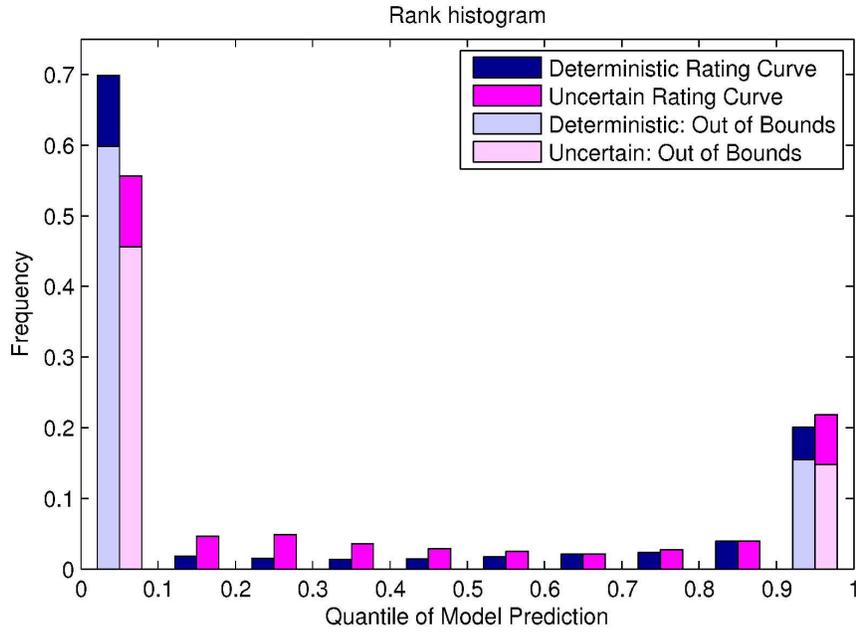


Parameter estimation: comparison of results using deterministic rating curve vs. rating curve including uncertainty. Plots show marginal posterior probability density functions for each of the TopNet model parameters.  
178x198mm (600 x 600 DPI)



33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Model validation results: 90% confidence interval for streamflow at Barnett's Bank during model validation period. Comparison of results from model calibrated using a deterministic rating curve (A) vs. rating curve including uncertainty (B).  
250x189mm (600 x 600 DPI)



Generalised Rank Histogram showing spread of model predictions compared with variability of uncertain discharge data, during calibration period. Results shown for both deterministic and uncertain rating curve cases.  
155x101mm (600 x 600 DPI)

Review