

1 **Rainfall-Runoff Model Calibration using Informal Likelihood Measures within a**
2 **Markov Chain Monte Carlo Sampling Scheme**

3

4 Hilary McMillan* and Martyn Clark

5

6 *h.mcmillan@niwa.co.nz

7 National Institute of Water and Atmospheric Research Ltd, 10 Kyle St, Riccarton, Christchurch, new

8 Zealand

9 **Abstract**

10 This paper considers the calibration of a distributed rainfall-runoff model in a
11 catchment where heterogeneous geology leads to a difficult and high-dimensional
12 calibration problem, where the response surface has multiple optima and strong
13 parameter interactions. These characteristics render the problem unsuitable for
14 solution by uniform Monte Carlo sampling and require a more targeted sampling
15 strategy. MCMC methods, using the SCEM-UA algorithm, are trialled using both
16 formal and informal likelihood measures. Each method is assessed in its success at
17 predicting the catchment flow response and capturing the total uncertainty associated
18 with this prediction. The comparison is made at both the catchment outlet and at
19 internal catchment locations with distinct geological characteristics. Lastly, we
20 demonstrate how information gained from the exploration of the response space, in
21 conjunction with qualitative knowledge of system behaviour, can be used to constrain
22 the Markov Chain trajectory.

23

24

25 **1 Introduction**

26 Improving availability and coverage of spatial data has driven developments in
27 distributed, process-based catchment modelling; however, despite the correspondence
28 between modeled and observed processes, it is not usually possible to determine
29 model parameters values directly from field measurements. Instead, the values
30 required are those of the ‘effective parameters’ which represent integrated behaviour
31 at the model element scale. These values must therefore be determined through a
32 calibration methodology, via some search of the parameter space. As has been
33 extensively discussed by Beven (1993; 2005; Beven and Binley, 1992) and others
34 (Wagener and Gupta, 2005), the many sources of uncertainty in a hydrological model
35 application lead to equifinality of parameter sets in providing acceptable model
36 performance with reference to some observed data. These uncertainty sources may
37 include, but are not limited to, input data uncertainty, initial condition uncertainty,
38 model structural error, observed data uncertainty (Liu and Gupta, 2007). Indeed, since
39 it is certain that our hydrological model does not fully represent the complexity of the
40 natural catchment and is therefore ‘wrong’, we must expect that any calibration
41 technique is a process of identifying some subset of model parameterisations which
42 produce reasonable approximations to some aspects of true catchment behaviour
43 under some circumstances.

44 The aim of a calibration technique should therefore be to enable an efficient search of
45 the parameter space, identifying those regions where model performance is considered
46 satisfactory. The task is made more difficult by the typically complex nature of the
47 model response surface (Duan *et al.*, 1992; Sorooshian *et al.*, 1993) which may be
48 exacerbated by artefacts of model timestep and solution techniques (Kavetski *et al.*,

49 2006a,b). Difficulties encountered may include multiple local optima in multiple
50 regions of attraction, discontinuous derivatives, parameter interaction and flat areas
51 (Duan *et al.*, 1992). The nature of these surfaces prohibits standard search
52 mechanisms such as simplex- and Newton- type schemes. Alternative methods such
53 as uniform random sampling suffer from a lack of sampling efficiency and can be
54 extremely costly in terms of model evaluations. They also typically specify the sample
55 space using minimum and maximum values for each parameter, based usually on
56 expert judgement, physical interpretation of the parameter and previous model use.
57 However with good model performance often occurring up to the boundary of the
58 sample region, this technique may unjustifiably restrict the search.

59 In recent years, Markov Chain Monte Carlo (MCMC) methods have gained increasing
60 popularity, in particular the Metropolis-Hastings (MH) Algorithm (e.g., Chib and
61 Greenberg, 1995). These methods enable simulation of complex multivariate
62 distributions by casting them as the invariant distribution of a Markov Chain. By
63 finding an appropriate transition kernel which converges to this distribution, samples
64 with the desired posterior distribution can be drawn from the Markov Chain. A
65 popular version of the MH algorithm is the adaptive SCEM-UA algorithm (Vrugt *et*
66 *al.*, 2003) which combines the MH sampler with the SCE-UA optimisation method
67 (Duan *et al.*, 1992), using information exchange between multiple sampler chains to
68 improve convergence rates.

69 All search techniques require a definition of the model response surface to be
70 searched: this is usually couched in terms of ‘probability of model correctness given
71 observed data’ and is assessed via a likelihood measure. The debate continues on the
72 relative advantages of the informal likelihood measures used in the GLUE framework
73 compared with parameter estimation via formal statistical likelihood estimation (e.g.

74 Mantovan and Todini, 2006; Beven *et al.*, 2007; Mantovan *et al.*, 2007; Thiemann *et*
75 *al.*, 2001; Beven, 2003; Gupta *et al.*, 2003; Clarke 1994). If statistical likelihood
76 theory is to be used, the error model between model predictand and observed variable
77 must be specified exactly; this may include information on heteroscedasticity and
78 autocorrelation (e.g. Sorooshian, 1981; Sorooshian and Dracup, 1980) and may rely
79 on hierarchical error models (Kuczera *et al.*, 2006). Under GLUE, the concept of a
80 true model (and error model) against which to compare observations is rejected and it
81 is accepted that many interacting sources of error, without well-defined formulations,
82 combine to give total model error. Models are instead judged against informal
83 likelihood measures, chosen by the hydrologist, which represent their expert
84 perception of model performance in prediction of observed data (Beven, 2006).

85 Although MCMC methods have traditionally used formal likelihood measures to
86 define the response surface (e.g. Arhonditsis *et al.*, 2008; Marshall *et al.*, 2004; Vrugt
87 *et al.*, 2006; Vrugt *et al.*, 2003; Thiemann *et al.*, 2001), it is also possible to use
88 informal likelihoods (e.g., Engeland and Gottschalk, 2002; Blasone *et al.*, 2008; Vrugt
89 *et al.*, 2008). When informal likelihoods are used in MCMC methods, the main
90 difference between MCMC methods and GLUE is that MCMC methods provide
91 targeted sampling of the parameter space. Blasone *et al.* (2008) compared
92 performance of the informal likelihoods in the SCEM-UA method with the traditional
93 GLUE method and demonstrated that the targeted sampling resulted in better
94 predictions of the model output (and that the uncertainty limits were less sensitive to
95 the number of retained solutions). Vrugt *et al.* (2008) compared a formal Bayesian
96 approach that attempts to explicitly quantify the individual sources of uncertainty in
97 the hydrological modelling process with the traditional GLUE method that maps all
98 sources of uncertainty onto the parameter space. They showed that while the

99 estimates of total uncertainty were similar in both methods, the GLUE method
100 produced large estimates of parameter uncertainty which can lead to erroneous
101 conclusions on the identifiability of model parameters.

102 The formal Bayesian approaches for explicitly quantifying the individual sources of
103 uncertainty suffer from two important limitations. First, as formulated by Vrugt et al.
104 (2008) and Kavetski et al. (2006a; 2006b), the formal Bayesian methods require
105 solving a high-dimensional optimization problem (i.e., separate multipliers for each
106 storm); a problem that is intractable for distributed hydrological models where it is
107 necessary to quantify uncertainty in the spatial pattern of precipitation events.
108 Second, current methods for quantifying error in model structure are poorly
109 developed—indeed, Vrugt et al. (2008) and Kavetski et al. (2006a; 2006b) essentially
110 combine error in model inputs and model structure into a single error term. Informal
111 likelihood measures therefore remain an attractive option.

112 This paper considers the calibration of a distributed rainfall-runoff model (described
113 in Section 2.2) in an interesting case study catchment, the Rangitaiki in New Zealand
114 (described in Section 2.1), where heterogeneous geology leads to a difficult and high-
115 dimensional calibration problem, where the response surface has multiple optima and
116 strong parameter interactions. These characteristics render the problem unsuitable for
117 solution by uniform Monte Carlo sampling (as per standard GLUE) and require a
118 more targeted sampling strategy. MCMC methods, using the SCEM-UA algorithm,
119 are trialled using both formal (Section 3.1) and informal (Section 3.2) likelihood
120 measures, and assessed in their success at full coverage of the response surface.

121 **2 Model and Data**

122 **2.1 Catchment**

123 The Rangitaiki River is located in the central North Island of New Zealand. It has a
124 length of 155 km and mean flow in the lower reaches of around $25 \text{ m}^3\text{s}^{-1}$. The river
125 flows along a series of fault-angle valleys which define a structural geological
126 boundary. To the west are Quaternary volcanic rocks, comprising a series of partially
127 overlapping, rhyolitic, welded ignimbrite sheets, overlain by thick tephra and pumice
128 sequences; to the east are uplifted Jurassic basement greywackes and meta-
129 greywackes (Beanland and Haines, 1998; Manville *et al.*, 2004). These two parts of
130 the catchment have strikingly different hydrological regimes: the porous tephra have
131 a characteristic high stable baseflow regime and subdued flood peaks; the steep and
132 relatively impermeable greywacke responds quickly to rainfall with a peaked runoff
133 pattern.

134 **Subcatchment Geology**

135 Each subcatchment is classified according to its substrate geology as recorded in the
136 New Zealand Land Resource Inventory (NZLRI). For the purposes of this study, a
137 simple binary division was made between impermeable (greywacke, argillite, lava)
138 and permeable (pumice, lapilli, tephra) geology. Although the two categories are
139 broadly divided East and West of the Rangitaiki river in the upper catchment, there is
140 some local variation (Figure 1).

141 **2.2 Data**

142 Gauging data for the Rangitaiki is available at Te Teko, at the entrance to the coastal
143 Rangitaiki Plains. The gauging station has a catchment area of 2890 km^2 and
144 represents the combined flow of the pumice and greywacke areas: a relatively

145 sustained baseflow is overlain by significant flood peaks. The contrasting
146 subcatchment flow regimes can be compared through the discharge records of two
147 internal gauging stations at Murupara and Galatea. Murupara is situated on the main
148 branch of the Rangitaiki, with a catchment of 1140 km² of the Kaingaroa Plateau. The
149 average annual mean flow is 21 m³s⁻¹ and the mean annual flood is 40 m³s⁻¹. Galatea
150 is sited on the Whirinaki, and drains a 509 km² area of the greywacke ranges. Here the
151 average annual mean flow is 14.5 m³s⁻¹, and the mean annual flood is 109 m³s⁻¹
152 (McKerchar and Pearson, 1989).

153 The model uses input precipitation and climate data from Tait *et al.* (2006) who
154 interpolated data from over 500 climate stations in New Zealand across a regular
155 0.05° latitude-longitude grid (approximately 5 km * 5 km). These data are provided at
156 daily time steps, and are disaggregated to hourly data before use in the model. In this
157 study we use data from the year 1998 when a large flood event occurred in the
158 Rangitaiki catchment, allowing a test of the model response over a full range of
159 discharge magnitudes.

160 To apply TopNet in the Rangitaiki, TopNet requires information on catchment
161 topography, physical and hydrological properties. This information is available from a
162 variety of sources. The New Zealand River Environment Classification (REC; Snelder
163 and Biggs, 2002) includes a digital network of approximately 600,000 river reaches
164 and related sub-basins for New Zealand. A 30 m Digital Elevation Model (DEM)
165 provided topographic properties. Land cover and soil data is available from the New
166 Zealand Land Cover Database (LCDB) and the New Zealand Land Resource
167 Inventory (LRI; Newsome *et al.*, 2000). The river basin was first disaggregated into
168 individual subcatchments, each one of which becomes a model element. We use the
169 Strahler 3 subcatchments from the REC, which have a typical size of 10 km², and split

170 the Rangitaiki Basin into 308 elements. The REC also provides the geometrical
171 parameters of the river network. Frequency distributions of the topographic wetness
172 index and distance to streams are calculated from the DEM. Average soil and
173 landcover parameters are derived from the LRI and LCDB respectively. In total, 12
174 parameters are required for each subcatchment, of which 6 may be specified using the
175 information described above; the remaining 6 must be calibrated (Refer to Table 1 for
176 descriptions of all the parameters). In addition, the Manning's n value for the
177 subcatchment channel section must also be calibrated.

178

179 **2.3 TOPNET**

180 TOPNET was developed by combining TOPMODEL (Beven et al., 1979; Beven et
181 al., 1995), which is most suited to small watersheds, with a kinematic wave channel
182 routing algorithm (Goring, 1994) so as to have a modeling system that can be applied
183 over large watersheds using smaller sub-basins within the large watershed as model
184 elements (Ibbitt and Woods, 2002; Bandaragoda *et al.*, 2004; Clark et al., 2008).
185 TOPNET uses TOPMODEL concepts for the representation of sub-surface storage
186 controlling the dynamics of the saturated contributing area and baseflow recession. To
187 form a complete model, potential evapotranspiration, interception (based on the work
188 of Ibbitt, 1971), infiltration (using a Green-Ampt mechanism; Mein and Larsen, 1973)
189 and soil zone components were added. Kinematic wave routing moves the sub-basin
190 inputs through the stream channel network. Complete model equations are provided
191 by Clark et al. (2008) and are not repeated here.

192 2.4 Calibration via Parameter Multipliers

193 In distributed rainfall-runoff models, the calibration problem is greatly complicated by
194 the large number of model parameters: multiple model parameters for each model
195 spatial element. Experience suggests that the integrated variables typically available to
196 evaluate model performance, such as streamflow series, may hold insufficient
197 information to determine all model parameter values (Beven, 2001). Various
198 approaches have been applied to ease this discrepancy. Many studies assume that
199 several parameters are spatially constant over the model domain, using a value
200 determined either by expert opinion or by directly using values measured at point
201 locations. Another popular approach is to apply a set of “parameter multipliers” to *a-*
202 *priori* model element parameter values, significantly reducing the dimensionality of
203 the calibration problem (Clark et al., 2008). However, due to the reliance on a
204 previously determined spatial distribution of model parameters, there is a danger that
205 distributed hydrological models calibrated using integrated data such as catchment
206 outlet discharge may fail to properly represent the range of hydrological behaviours.
207 Poor forecasts would then be produced at internal catchment locations (Clark et al.,
208 2008).

209 This paper presents a model calibration strategy that provides correct representation of
210 internal catchment processes. The calibration method is applied in the Rangitaiki,
211 where two sub-regions of the catchment have significantly different hydrological
212 characteristics. Our knowledge of catchment geology cannot be translated directly
213 into values for model parameters; instead we seek to use the qualitative information to
214 inform our calibration strategy.

215 Figure 1: Geology of the Rangitaiki River basin, classified according to permeability.

216 Gauging Locations are marked.

217 The method used is to classify each Strahler 3 sub-catchment as either ‘permeable’ or
218 ‘impermeable’ (refer to Section 2.1; note that in other catchments, three or more
219 qualitative categories may be appropriate). *A-priori* model parameters are specified in
220 each individual subcatchment using topography, soils and land-cover data (Table 1).
221 Two sets of parameter multipliers are then allowed, one for each category. The
222 optimisation process allows all multipliers to be calibrated simultaneously, such that
223 the optimum combination of process descriptions in the two categories is found.

224 The Rangitaiki provides an ideal test location, as the model calibration can be
225 implemented using only the outlet discharge gauged at Te Teko (Figure 1), but tested
226 for diverse internal process representation using the two gauges at Murupara (pumice
227 subcatchment) and Galatea (greywacke catchment). This internal check allows a test
228 of model conditioning and parameter identification success; an important
229 consideration due to the increased number of parameters used with this method.

230 **3 MCMC technique (Bayesian Uncertainty Framework)**

231 **3.1 Metropolis and Adaptive Metropolis Algorithms**

232 Markov Chain Monte Carlo provides a general approach to sampling from the
233 posterior distribution. Classical Markov Chain theory specifies the transition kernel
234 $P(x,A)$ which gives the probability from moving from the point x to any point in the
235 set A . A common question is then to determine whether the chain has an invariant
236 distribution π which is unchanged by applying the transition kernel. The MCMC
237 technique reverses the problem: the required posterior distribution is taken as the
238 invariant π ; instead we seek the appropriate transition kernel $P(x,A)$ such that a chain
239 using this kernel provides samples from the posterior. The Metropolis-Hastings
240 algorithm, one of the most popular MCMC methods, provides a method for finding

241 the required transition kernel. At each step of the Markov Chain, a new sample is
242 drawn from a ‘proposal distribution’ $q(x,y)$. However the chain only moves to this
243 sample point according to a ‘probability of move’ $\alpha = \pi(y)/\pi(x)$, otherwise it remains
244 at the previous sample point.

245 The choice of proposal distribution $q(x,y)$ has important consequences for the
246 algorithm behaviour. Where $q(x,y)$ is too diffuse or does not properly represent
247 interactions between parameters, α is often small and many candidate points are
248 rejected, slowing the chain evolution. Where $q(x,y)$ is too compact, the chain will
249 move inefficiently around the search space, causing particular problems with spatially
250 distal optima. The SCEM-UA algorithm (Vrugt *et al.*, 2003) seeks to avoid these
251 problems by continually updating the proposal distribution using information gained
252 about the nature of the posterior distribution. The proposal distribution becomes a
253 multivariate normal with mean and covariance structure taken as the sample mean and
254 sample covariance of different ‘complexes’ of points in the high-density region of the
255 sample space. Although it is not proven that the SCEM-UA algorithm with adaptive
256 proposal distribution provides an ergodic Markov Chain with the correct invariant
257 distribution (Haario *et al.*, 1999; 2001), experimental investigations have shown that
258 the algorithm performs well (Vrugt *et al.*, 2003).

259 **3.2 Formal Bayesian Likelihood**

260 The MCMC method is first carried out using a formal Bayesian Likelihood derivation
261 for the posterior density. Following Thiemann *et al.* (2001), Vrugt *et al.* (2003), Bates
262 and Campbell (2001), Marshall *et al.*, (2004) and others, we assume that measurement
263 errors can be transformed via a one-to-one transformation to have the exponential
264 power density $E(\sigma,\beta)$, and hence the conditional posterior density can be derived to be
265 of the form (Box and Tiao, 1973)

266
$$p(z | \theta, \sigma, \beta) = \left[\frac{\omega(\beta)}{\sigma} \right]^T \cdot \exp \left[-c(\beta) \cdot \sum_{t=1}^T \left| \frac{v(t)}{\sigma} \right|^{\frac{2}{1+\beta}} \right] \quad (\text{Equation 1})$$

267 Where

268
$$c(\beta) = \left[\frac{\Gamma[3(1+\beta)/2]}{\Gamma[(1+\beta)/2]} \right]^{1/(1+\beta)}, \omega(\beta) = \frac{\{\Gamma[3(1+\beta)/2]\}^{1/2}}{(1+\beta) \cdot \{\Gamma[(1+\beta)/2]\}^{3/2}}$$

269 β is a scale parameter, σ is the standard deviation of the measurement errors, T is the
 270 number of timesteps, and $v(t)$ are the transformed errors.

271 **3.3 Informal Likelihood Measures**

272 Secondly, the MCMC sampling is repeated using an informal likelihood measure as
 273 used under the philosophy of the GLUE system (Beven and Binley, 1992). This
 274 technique also requires the selection of a ‘behaviourability threshold’ such that when
 275 the likelihood measure falls below this value, the model is rejected. Although
 276 typically the choice of threshold has been based on the expert judgement of the
 277 modeller as to the error magnitude that is acceptable for the particular application, it
 278 may also be chosen objectively such that a set proportion of the observed values lie
 279 within the uncertainty bounds (Blasone *et al.*, 2008; Montanari, 2005).

280 **3.3.1 Nash-Sutcliffe Likelihood**

281 The Nash-Sutcliffe index of model efficiency (NSE; Equation 1) is one of the most
 282 commonly used descriptors of rainfall-runoff model performance Hall (2001).

283
$$\text{NSE} = 1 - \frac{\sigma_{\varepsilon}^2}{\sigma_o^2} \quad (\text{Equation 2})$$

284 Where σ_{ε}^2 is the error variance and σ_o^2 is the variance of the observed flow series.

285 Hence the NSE takes a value of 1 for a perfect model fit, a value of 0 for a model no
 286 better than the constant mean of the observed data. The Nash-Sutcliffe index is often
 287 used in the GLUE framework as an informal likelihood measure. In order for it to be
 288 used in SCEM-UA, it must be non-negative and monotonically increasing with
 289 improved performance. To meet the former condition, the NSE is set to zero when
 290 negative values are returned. The NSE is only used via the posterior density ratio R of
 291 two samples, which can be expressed in the following form:

$$292 \quad R = \frac{1 - \frac{\sigma_{\varepsilon 1}^2}{\sigma_o^2}}{1 - \frac{\sigma_{\varepsilon 2}^2}{\sigma_o^2}} = \frac{\sigma_o^2 - \sigma_{\varepsilon 1}^2}{\sigma_o^2 - \sigma_{\varepsilon 2}^2} = \frac{K - SSE_1}{K - SSE_2} \quad (\text{Equation 3})$$

293 Where SSE_1 and SSE_2 are the sums of squared errors for the two samples and K is a
 294 constant.

295 After initial trials of a MCMC method using this index, it was found that the chain
 296 was initially slow to migrate to high performance regions of the sample space. This
 297 was hypothesised to be due to two factors:

298 the lack of ability to order poor model fits (as the NSE was set to zero whenever σ_o^2
 299 $> \sigma_{\varepsilon}^2$) which prevented the chain from gradual movement towards high performance
 300 regions.(1) Poor representation of relative model performance, e.g. a NSE of 0.9
 301 would typically be considered a significant improvement relative to a NSE of 0.8,
 302 however in this method there would be a high probability of move from 0.9 down to
 303 0.8 as the posterior density ratio is $0.8/0.9 = 0.89$.

304 (2) Lack of ability to order poor model fits (as the NSE was set to zero whenever σ_o^2
 305 $> \sigma_{\varepsilon}^2$) which prevented the chain from gradual movement towards high performance
 306 regions.

307 In order to address this issue (1), the constant K may be adjusted to mimic the effect
308 of the behavioural threshold and alter the ratio R ; i.e. reducing K causes higher weight
309 to be placed on small improvements in NSE. To address issue (2), the exact sum of
310 squared error scores were retained such that all model fits could be correctly ordered,
311 even though this information was not used to calculate the ratio R . A combination of
312 these two measures This was found to significantly improve the Markov Chain
313 efficiency.

314 **3.3.2 Extended Nash-Sutcliffe**

315 Despite the perennial popularity of error variance measures such as the Nash-Sutcliffe
316 score, there are occasions when an approach base on the sum-of-squared-errors is
317 likely to produce counterintuitive results when assessing the fit of modelled and
318 observed hydrographs. Of particular concern is the relative importance assigned to
319 discharge magnitude errors versus timing errors. It is a common occurrence for
320 rainfall-runoff models to incorrectly predict the timing of a flood peak; however due
321 to the timestep-by-timestep comparison in an SSE analysis, timing errors can cause
322 extremely poor performance measure values (Figure 2).

323 Figure 2: A synthetic example of hydrographs in which a model with minor (2 hour)
324 timing error is graded as having poorer performance than a model with 40% discharge
325 error

326 A generalised version of the Nash-Sutcliffe likelihood is suggested in order to address
327 these concerns, by allowing discrepancies between observed and modelled data points
328 to be considered as a combination of discharge and timing errors. This is achieved by
329 using the modeller's judgment on relative importance of discharge and timing errors
330 to determine the shape of an oval search window (Figure 3). The error at each

331 timestep is defined as the minimum distance from the oval centre to the point on the
332 oval boundary which intersects the opposing discharge curve. The squared error
333 values are then summed and substituted directly into the standard Nash-Sutcliffe
334 equation. Standard NS appears as a special case within the Extended NS when timing
335 errors are considered infinitely worse than discharge errors and the search oval
336 becomes a vertical line. A procedural description of calculation of the new error
337 measure can be found in Appendix A.

338 Figure 3: Error magnitudes for the Extended Nash-Sutcliffe are found using an oval
339 search window.

340

341 **4 Results**

342 **4.1 Flow prediction**

343 **Formal Bayesian Likelihood**

344 Model calibration was carried out using data from the year 1998, using the MCMC
345 method described in Section 3.1 and a formal likelihood measure based on an
346 exponential error distribution (Section 3.2). Ten parallel Markov Chains are run for a
347 total of 5000 iterations; the first 1000 iterations are discarded as a ‘burn-in’ period for
348 the chain. Gelman-Rubin convergence statistics are calculated to check the Markov
349 Chain has converged to the stationary distribution representing the model posterior
350 distribution.

351 Figure 4: 90% Uncertainty bounds on flow at Te Teko using formal likelihood
352 measure to control MCMC search algorithm.

353 The resulting uncertainty bounds on the flow hindcast are shown in Figure 4; note that
354 the bounds are sufficiently narrow to be hardly visible as distinct from the median
355 calibrated prediction.

356

357 **Informal Likelihood**

358 The model calibration was repeated using the same Markov Chain set-up, but using in
359 turn the Nash-Sutcliffe and Extended Nash-Sutcliffe likelihood measures. The
360 resulting flow hindcasts are shown in Figures 5 and 6 respectively. It is clear that
361 using an informal likelihood measure suggests a much greater uncertainty in the flow
362 forecast, with uncertainties greatest during peak flow periods.

363 Figure 5: 90% Uncertainty bounds on flow at Te Teko using Nash-Sutcliffe informal
364 likelihood measure to control MCMC search algorithm.

365 Figure 6: 90% Uncertainty bounds on flow at Te Teko using Extended Nash-Sutcliffe
366 informal likelihood measure to control MCMC search algorithm.

367 A study of the Markov Chain behaviour can be used to provide additional information
368 about the model response surface, and the success of the MCMC algorithm in fully
369 exploring the surface (Vrugt *et al.*, 2003). Figures 7 and 8 allow a comparison of the
370 sequential values of the Topmodel f parameter when using formal vs. informal
371 likelihood measures. Figure 7 shows that in the case of the formal likelihood measure,
372 the distribution quickly collapses to a single optimum, and the remainder of the
373 parameter space is not explored. In contrast, Figure 8 shows that the informal
374 likelihood measure produces a continuing wide dispersal of behavioural parameter
375 values, and therefore a flatter response surface. Other model parameters showed
376 similar trends. It is also interesting to note in Figure 8 that there is a distinct higher-

377 density band for f in the range $[0, 0.1]$, coupled with a more disperse band in the range
378 $[0.2, 0.6]$. This suggests the possibility of a bi-modal distribution for f , with only the
379 more peaked lower optimum found by the formal likelihood measure: this issue is
380 discussed more fully in the following section.

381 Figure 7: Topmodel 'f' parameter value over successive iterations of the MCMC
382 search algorithm using formal Bayesian (exponential error model) likelihood measure.

383 Figure 8: Topmodel 'f' parameter value over successive iterations of the MCMC
384 search algorithm using informal 'Extended Nash-Sutcliffe' likelihood measure.

385

386

387 **4.2 Calibration Constraints using Qualitative Geological Information**

388 **Internal Catchment Flow Gauging**

389 By using the informal likelihood measure (Section 4.1) the Markov Chain revealed a
390 dispersed posterior response surface, with the possibility of dual optima suggested by
391 distinct bands in the parameter mixing diagrams when using the informal Extended
392 Nash-Sutcliffe likelihood measure (Figure 8). Given the division of the catchment into
393 dual 'permeable' and 'impermeable' areas, it seemed logical that these two
394 phenomena might be related. The issue was investigated further using flow data from
395 the two internal catchment gauges which had not previously been used in model
396 calibration (Figure 9).

397 Figure 9: Comparison of Internal Flow predictions at Murupara (pumice
398 subcatchment) and Galatea (greywacke subcatchment) using formal and informal
399 likelihood measures.

400 Striking differences were seen here between the formal and informal likelihood
401 results. The informal likelihood results show a very large spread in possible internal
402 flow distribution in the catchment, where the majority of the quickflow may be
403 attributed to either pumice or greywacke areas (Figures 9C/9D). In reality, the pumice
404 subcatchment provides a steady baseflow, with the greywacke catchment providing a
405 peaked response to storm events (refer to Section 2.1) – however the unconstrained
406 model calibration may assign ‘pumice’ vs. ‘greywacke’ characteristics to the sub-
407 catchments in either order. In contrast, the calibration using a formal likelihood
408 measure has collapsed to a single parameter allocation (Figures 9A/9B) which has
409 incorrectly classified the subcatchments and in effect assigned ‘greywacke-type’
410 characteristics to the pumice sub-catchment, and vice-versa.

411

412 **Constrained Calibration**

413 It is natural to ask whether the calibration procedure may be constrained such that
414 Markov Chains converge to the correct optimum such the flow characteristics are
415 correctly assigned to the two geologically distinct sub-catchments. Although in the
416 case of the Rangitaiki this could be achieved using multi-criteria calibration with
417 additional data from the internal flow gauges, here we are interested in a strategy
418 using only the catchment outlet flow gauge, such that the methodology would be
419 transferable to other catchments with a single flow gauge.

420 The constraint process aimed to subdivide the parameter space in the simplest
421 possible way into volumes representing ‘pumice’ or ‘greywacke’ behaviour. In order
422 to be considered as constraints, parameters had to satisfy the dual criteria of having a
423 physical interpretation, such that characteristics could be accurately assigned, and

424 showing good discrimination between model realisations representing the two
425 response types. The parameters that achieved this were (1) Topmodel f parameter -
426 related to depth of soil profile and aquifer response time (2) $\Delta\theta_1$ – effective drained
427 porosity (3) $\Delta\theta_2$ – root zone storage.

428 Multiplier ranges were defined for each of these based on separation of the observed
429 marginal distribution by behavioural group. This was achieved by physical
430 interpretation of the bi-modal parameter distributions, and resulting predicted flows,
431 in the unconstrained calibration (Figure 8). Previous research in New Zealand
432 demonstrates significant behavioural differences between pumice vs. non-volcanic
433 regions, with pumice regions characterized by lower flood peaks (McKerchar and
434 Pearson, 1989) and higher yields (Hutchinson, 1990). The bi-modal form is therefore
435 compatible with an expectation that parameter multipliers for Topmodel “f”, $\Delta\theta_1$ and
436 $\Delta\theta_2$ may need to be different for the two geology types to make targeted corrections to
437 the default values. The two modes of the parameter distribution are classified as
438 providing ‘Pumice-type’ and ‘Greywacke-type’ behaviour respectively. The resulting
439 marginal distributions are shown in Figure 10: the Topmodel f parameter is seen to
440 show non-intersecting ranges for the two parameters sets, the $\Delta\theta_1$ and $\Delta\theta_2$ parameters
441 show defined ranges for the ‘greywacke-type’ parameters only. Other parameters (not
442 shown) did not show good discrimination between behavioural types.

443 Figure 10: Multiplier ranges categorised by behavioural type for parameters: (a)
444 Topmodel f (b) $\Delta\theta_1$ effective drained porosity (c) $\Delta\theta_2$ root zone storage. These plots
445 were used to define constrained parameter ranges

446 The calibration was re-run using appropriate parameter ranges for each sub-catchment
447 according to its geological classification. An informal likelihood measure was used as
448 this is consistent with the analysis suggesting the presence of behavioural simulations

449 within the constrained range: the formal likelihood measure in contrast rejected at the
450 90% level all simulations within the new constraints. The Extended Nash-Sutcliffe
451 measure was used in order to allow proper consideration of both magnitude and
452 timing errors.

453 Figure 11: Internal Flow predictions at Murupara (pumice subcatchment) and Galatea
454 (greywacke subcatchment) using informal likelihood measures under a constrained
455 calibration procedure.

456 The results for flow predictions at the two internal catchment flow gauges are shown
457 in Figure 11. These results show accurate flow prediction in each subcatchment with
458 substantially reduced uncertainty compared to the unconstrained calibration. We
459 therefore conclude that imposing constraints on the 3 parameters f , $\Delta\theta_1$, $\Delta\theta_2$ was
460 sufficient to guide the MCMC algorithm to the correct optimum.

461

462 **5 Discussion and Conclusions**

463 Where a catchment has sub-regions of contrasting hydrological behaviour, such as
464 those caused by different geologies, there is a danger that distributed hydrological
465 models calibrated using integrated data such as catchment outlet discharge may fail to
466 properly represent the range of hydrological behaviours. Due to a wide range of
467 possible distributions of flow within different branches of the catchment, the response
468 surface representing the posterior distribution may have multiple optima and flat areas
469 characteristic of complex equifinal behaviour. It is therefore important to use a
470 calibration procedure which is capable of fully capturing and describing the
471 behavioural regions of the parameter space.

472 MCMC algorithms such as the Metropolis-Hastings and its variants are popular
473 choices for efficient exploration of complex response surfaces, however this paper has
474 shown that the formal likelihood measures which are typically used within such
475 algorithms may prevent the Markov Chain from fully exploring regions of the
476 parameter space which might be considered behavioural when assessed using a
477 standard performance measure such as the Nash-Sutcliffe statistic. Such formal
478 Bayesian approaches assume that the model structure is correct, and therefore do not
479 account for cases where the parameters compensate for weaknesses in model
480 structure. This may lead to cases where, although parameter uncertainty is small, the
481 optimised parameter values are in fact ‘wrong,’ as shown in Section 4.1 in the form of
482 extremely poor flow predictions at internal locations.

483 By using instead an informal likelihood measure, we attempt to capture the total
484 uncertainty in flow predictions due a range of known and unknown error sources. This
485 methodology results in a greater volume of the parameter space being sampled, thus
486 revealing more complete information about possible multiple optima or flat areas of
487 the response surface. Of course, the posterior probability distribution to be sampled
488 must reflect the hydrologist’s best understanding of the errors present in the modeling
489 process; where these can be described very exactly a formal likelihood measure would
490 be a more appropriate choice and would better represent the information on posterior
491 parameter distribution which could be derived from the observed data. Unfortunately,
492 however, it may often be the case that a formal likelihood measure which makes
493 strong assumptions about model error distribution is used under conditions of
494 incomplete information on error form.

495 Finally, this paper has shown how the additional information gained using an
496 exploration of the response surface using an informal likelihood measure can be used

497 to improve the calibration process in order to focus the Markov Chain trajectory on
498 regions of the parameter space reflecting our qualitative knowledge of system
499 behaviour. The ability to incorporate qualitative or 'soft' data into calibration
500 algorithms is very valuable but may be more effectively deployed in conjunction with
501 a description of the response surface which identifies threshold or boundaries between
502 different response types.

503

504

505 **Appendix A**

506 **Algorithm for Calculation of Extended Nash-Sutcliffe Performance Measure**

507

- 508 1. Define ε_T as the timing error (e.g. in hours) which is considered ‘equally bad’ as a
509 discharge error of 1 unit (typically $1 \text{ m}^3\text{s}^{-1}$), and τ the maximum allowable timing
510 error.

511 For each timestep (t) in turn:

- 512 2. Identify the greater of the two discharge series (observed, modelled) at time t :

513
$$Q_1(t) = \max\{Q_{obs}(t), Q_{mod}(t)\}$$

- 514 3. Create a vector of timesteps within the allowable time window:

515
$$\bar{T} = [t - \tau, \dots, t - \Delta t, t, t + \Delta t, \dots, t + \tau]$$

- 516 4. Create a vector of discharges corresponding to these time steps:

517
$$\bar{Q}_2 = \begin{cases} [Q_{obs}(t - \tau), \dots, Q_{obs}(t), \dots, Q_{obs}(t + \tau)] & \text{where } Q_{mod}(t) \geq Q_{obs}(t) \\ [Q_{mod}(t - \tau), \dots, Q_{mod}(t), \dots, Q_{mod}(t + \tau)] & \text{where } Q_{obs}(t) > Q_{mod}(t) \end{cases}$$

- 518 5. Calculate the squared error vector relating to this set of time steps:

519
$$\overline{SE} = \left(\frac{t - \bar{T}}{\varepsilon_T} \right)^2 + (Q_1(t) - \bar{Q}_2)^2$$

- 520 6. Minimise the squared error over the time window:

521
$$Squared\ Error(t) = \min\{\overline{SE}\}$$

522 Having calculated the squared error for each timestep, return to the standard Nash-

523 Sutcliffe method:

524 7. Calculate the error variance

525
$$\sigma_{\varepsilon}^2 = \frac{1}{n-1} \cdot \sum_t \text{Squared Error}(t)$$

526 8. Calculated the Extended Nash-Sutcliffe Score:

527
$$\text{Extended NSE} = 1 - \frac{\sigma_{\varepsilon}^2}{\sigma_o^2}$$

528 where σ_o^2 is the variance of the observed flow series.

529

530 Note that at each timestep the oval search window is centred on the greater of the
531 modelled and observed discharges: this avoids the situation where narrow, high
532 discharge peaks which are not predicted correctly are not accounted for in the error
533 calculation as the search window picks up low flows before or after these events. The
534 reverse situation with a sudden trough in discharge levels would be extremely unusual
535 in either a modelled or observed flow series.

536 **References**

- 537 Arhonditsis GB., Perhar G, Zhang W, Massos E, Shi M, Das A (2008), Addressing
538 equifinality and uncertainty in eutrophication models, *Water Resources Research*, 44,
539 W01420, doi:10.1029/2007WR005862.
- 540 Bandaragoda C, Tarboton DG, Woods R, 2004. Application of TOPNET in the
541 distributed model intercomparison project. *Journal of Hydrology*, 298 (1-4) pp 178-
542 201.
- 543 Bates BC and Campbell EP, 2001. A Markov Chain Monte Carlo scheme for
544 parameter estimation and inference in conceptual rainfall-runoff modelling. *Water*
545 *Resources Research*. 37(4) 937-947.
- 546 Beanland S and Haines J, 1998. The kinematics of active deformation in the North
547 Island, New Zealand, determined from geological strain rates, *New Zealand Journal*
548 *of Geology and Geophysics* 41 (1998), pp. 311–324.
- 549 Beven KJ (1993) Prophecy, reality and uncertainty in distributed hydrologic
550 modelling. *Advances in Water Resources*. 16:41–51
- 551 Beven KJ, 2001. *Rainfall-Runoff Modelling: The Primer*. Wiley: Chichester.
- 552 Beven KJ, 2003. Comment on "Bayesian recursive parameter estimation for
553 hydrologic models" by M. Thiemann, M. Trosset, H. Gupta, and S. Sorooshian. *Water*
554 *Resources Research*, 39 (5) 1116.
- 555 Beven KJ (2006) A manifesto for the equifinality thesis. *Journal of Hydrology*. 320
556 (1-2) 18-36
- 557 Beven KJ, Binley AM (1992) The future of distributed models: model calibration and
558 uncertainty in prediction. *Hydrological Processes*. 6:279–298
- 559 Beven KJ and Kirkby MJ, 1979. A physically based variable contributing area model
560 of basin hydrology. *Hydrological Sciences Bulletin* 24 1, pp. 43–69.
- 561 Beven KJ, Lamb R, Quinn P, Romanowicz R and Freer J, 1995. Topmodel. In: Singh,
562 V.P., Editor, 1995. *Computer Models of Watershed Hydrology* 18, Water Resources
563 Publications, Highlands Ranch, CO (Chapter 18); pp. 627–668.
- 564 Beven KJ, Smith P, Freer J, 2007. Comment on "Hydrological forecasting uncertainty
565 assessment: incoherence of the GLUE methodology" by Pietro Mantovan and Ezio
566 Todini. *Journal of Hydrology* 338 (3-4) 315-318
- 567 Blasone R-S, Vrugt JA, Madsen H, Rosbjerg D, Robinson MA, and Zyvoloski GA,
568 2008. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov
569 Chain Monte Carlo sampling. *Advances in Water resources*, 31, 630-648.
- 570 Box GEP and Tiao GC, 1973. *Bayesian Inference in Statistical Analysis*. Addison-
571 Wesley-Longman: Reading, Mass.
- 572 Chib S, and Greenberg E, 1995. Understanding the Metropolis-Hastings Algorithm.
573 *The American Statistician*. 49(4): 327 – 335.
- 574 Clark MP, Rupp DE, Woods RA, Zheng X, Ibbitt RP, Slater AG, Schmidt J and
575 Uddstrom MJ, 2008. Hydrological data assimilation with the ensemble Kalman filter:
576 Use of streamflow observations to update states in a distributed hydrological model.
577 *Advances in Water Research*, in press.

578 Clarke RT, 1994. *Statistical Modelling in Hydrology*. Wiley: Chichester

579 Duan Q, Sorooshian S, and Gupta HV, 1992. Effective and Efficient Global
580 Optimization for Conceptual Rainfall-Runoff Models, *Water Resources Research*.
581 28(4), 1015–1031.

582 Engeland K and Gottschalk L, 2002. Bayesian estimation of parameters in a regional
583 hydrological model. *Hydrology and Earth System Sciences*. 6 (5): 883:898.

584 Goring DG, 1994. Kinematic shocks and monoclinal waves in the Waimakariri, a
585 steep, braided, gravel-bed river. *Proceedings of the International Symposium on*
586 *Waves: Physical and Numerical Modelling*, University of British Columbia,
587 Vancouver, Canada, 21–24 August, 1994 pp. 336–345

588 Gupta HV, Thiemann M, Trosset M, Sorooshian, S, 2003. Reply to comment by K.
589 Beven and P. Young on "Bayesian recursive parameter estimation for hydrologic
590 models", *Water Resources Research*, 39 (5) 1117.

591 Haario H, Sakman E, Tammimien J, 1999. Adaptive proposal distribution for random
592 walk Metropolis algorithm. *Computational Statistics*. 14(3): 375-395.

593 Haario H, Sakman E, Tammimien J, 2001. An adaptive Metropolis algorithm.
594 *Bernouilli*. 7(2): 223-242.

595 Hall MJ, 2001. How well does your model fit the data? *Journal of Hydroinformatics* 3
596 (1) pp 49-55.

597 Hutchinson PD, 1990. Regression Estimation of Low Flow in New Zealand.
598 Publication No. 22 of the Hydrology Centre, DSIR Marine and Freshwater,
599 Christchurch, N.Z. 51p. ISSN 0112-1197.

600 Ibbitt RP, 1971: Development of a conceptual model of. interception. Hydrological
601 research progress report. No. 5. Wellington, Ministry of Works.

602 Ibbitt RP and Woods R, 2002. Towards rainfall-runoff models that do not need
603 calibration to flow data. In *Friend 2002 – Regional Hydrology: Bridging the Gap*
604 *between Research and Practice*. IAHS Publication no. 274. Eds van Lanen, HAJ and
605 Demuth, S. pp 189 -196.

606 Kavetski D, Kuczera G, Franks SW, 2006a. Calibration of conceptual hydrological
607 models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology*. 320 (1-2)
608 173-186.

609 Kavetski D, Kuczera G, Franks SW, 2006b. Calibration of conceptual hydrological
610 models revisited: 2. Improving optimisation and analysis. *Journal of Hydrology*. 320
611 (1-2) 187-201.

612 Kuczera G, Kavetski D, Franks S, Thyer, M., 2006. Towards a Bayesian total error
613 analysis of conceptual rainfall-runoff models: Characterising model error using storm-
614 dependent parameters. *Journal of Hydrology*. 331 (1-2) 161-177.

615 Liu Y, and Gupta HV, 2007: Uncertainty in hydrologic modeling: Toward an
616 integrated data assimilation framework. *Water Resources Research*. 43, W07401,
617 doi:10.1029/2006WR005756.

618 Mantovan P, Todini E, 2006. Hydrological forecasting uncertainty assessment:
619 Incoherence of the GLUE methodology. *Journal of Hydrology*. 330 (1-2): 368-381.

620 Mantovan P, Todini E, Martina MLV, 2007. Reply to comment by Keith Beven, Paul
621 Smith and Jim Freer on "Hydrological forecasting uncertainty assessment:
622 Incoherence of the GLUE methodology". *Journal of Hydrology* 338 (3-4) 319-324.

623 Manville V, Newton EH, White JDL, 2005. Fluvial responses to volcanism:
624 Resedimentation of the 1800a Taupo ignimbrite eruption in the Rangitaiki River
625 catchment, North Island, New Zealand. *Geomorphology*, 65 (1-2), pp. 49-70

626 Marshall L, Nott D, Sharma A, 2004. A comparative study of Markov chain Monte
627 Carlo methods for conceptual rainfall-runoff modelling. *Water Resources Research*.
628 40 (2): W02501.

629 McKerchar AI and Pearson CP, 1989. Flood frequency in New Zealand, *Publication*,
630 *Hydrology Section* vol. 20, Division of Water Sciences, Department of Scientific and
631 Industrial Research, Christchurch 87 pp.

632 Mein RG and Larson CL, 1973. Modeling infiltration during steady rain. *Water*
633 *Resour. Res.* 9, pp. 384–394

634 Montanari A, 2005. Large sample behaviors of the generalized likelihood uncertainty
635 estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. *Water*
636 *Resources Research*. 41(8): W08406.

637 Newsome PFJ, Wilde RH, Willoughby EJ, 2000. *Land Resource Information System*
638 *Spatial Data Layers*. Technical Report. Landcare Research NZ Ltd, Palmerston
639 North, NZ.

640 Snelder TH; Biggs BJF, 2002. Multi-scale river environment classification for water
641 resources management. *Journal of the American Water Resources Association* 38(5):
642 1225–1240.

643 Sorooshian, S and Dracup, JA, 1980. Stochastic parameter estimation procedures for
644 hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. *Water*
645 *Resources Research*. 16 (2): 430-442.

646 Sorooshian S, 1981. Parameter estimation of rainfall run-off models with
647 heteroscedastic streamflow errors – noninformative data case. *Journal of Hydrology*.
648 52 (1-2): 127-138.

649 Sorooshian S, Duan Q, and Gupta HV (1993), Calibration of Rainfall-Runoff Models:
650 Application of Global Optimization to the Sacramento Soil Moisture Accounting
651 Model, *Water Resour. Res.*, 29(4), 1185–1194.

652 Tait AB, Henderson RD, Turner RW, Zheng X, 2006. Thin plate smoothing spline
653 interpolation of daily rainfall for New Zealand using a climatological rainfall surface.
654 *International Journal of Climatology*. 2097-2115.

655 Thiemann M, Trosset M, Gupta H, Sorooshian S, 2001. Bayesian recursive parameter
656 estimation for hydrologic models. *Water Resources Research*. 37 (10) 2521-2535.

657 Vrugt JA, Gupta HV, Bouten W, Sorooshian S, 2003. A Shuffled Complex Evolution
658 Metropolis algorithm for optimisation and uncertainty assessment of hydrologic
659 model parameters. *Water Resources Research*, 39 (8) 1201.

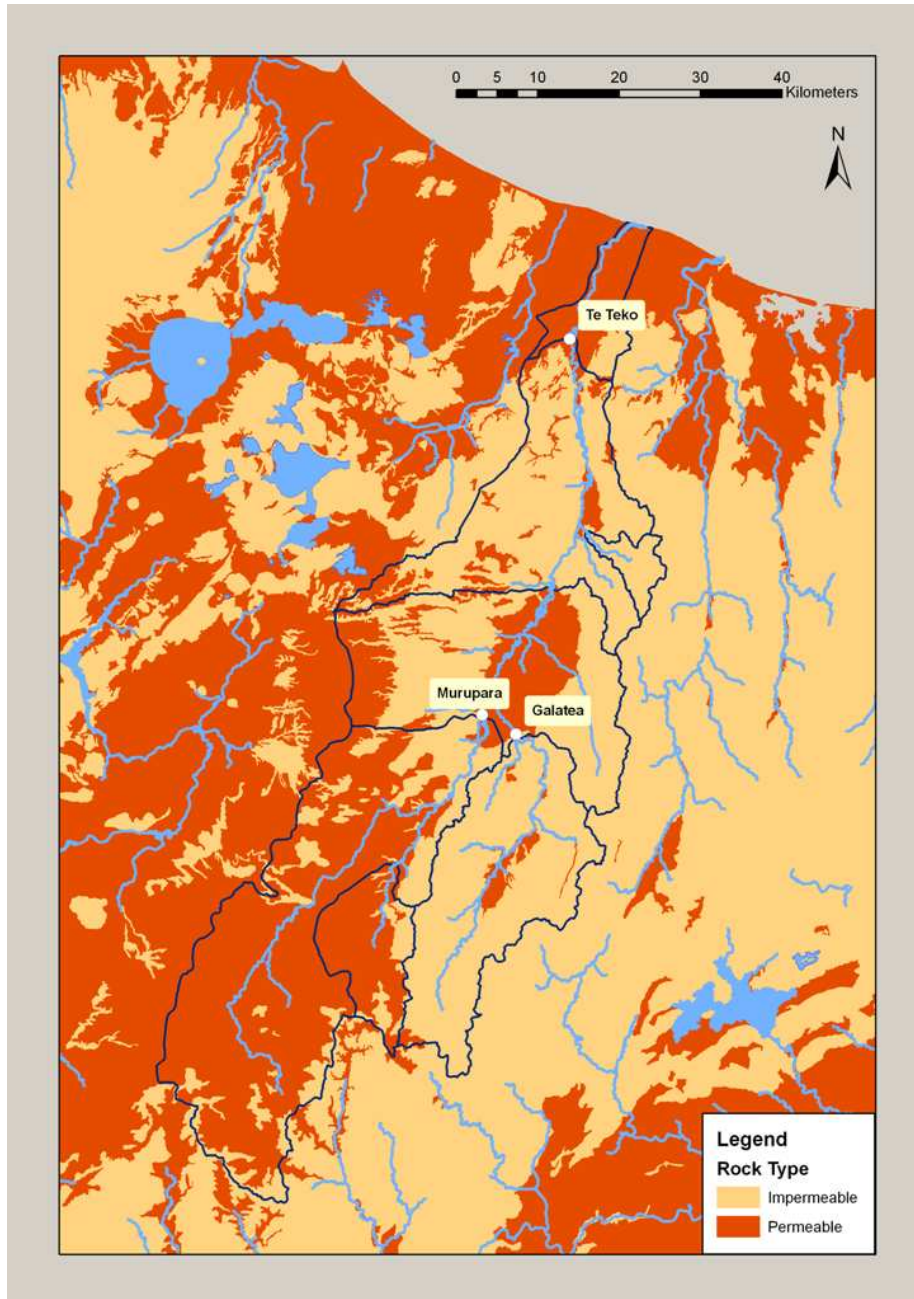
660 Vrugt JA, Gupta HV, Dekker SC, Sorooshian S, Wagener T, Bouten W, 2006.
661 Application of stochastic parameter optimization to the Sacramento Soil Moisture
662 Accounting model. *Journal of Hydrology*. 324: 288-307.

- 663 Vrugt JA, ter Braak CJF, Gupta HV, and Robinson BA, 2008: Equifinality of formal
664 (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modelling.
665 *Stochastic Environmental Research and Risk Assessment*, in review.
- 666 Wagener T, Gupta HV, 1995. Model identification for hydrological forecasting under
667 uncertainty, *Stochastic Environmental Research and Risk Assessment*, 19 (6): 378-
668 387.

669

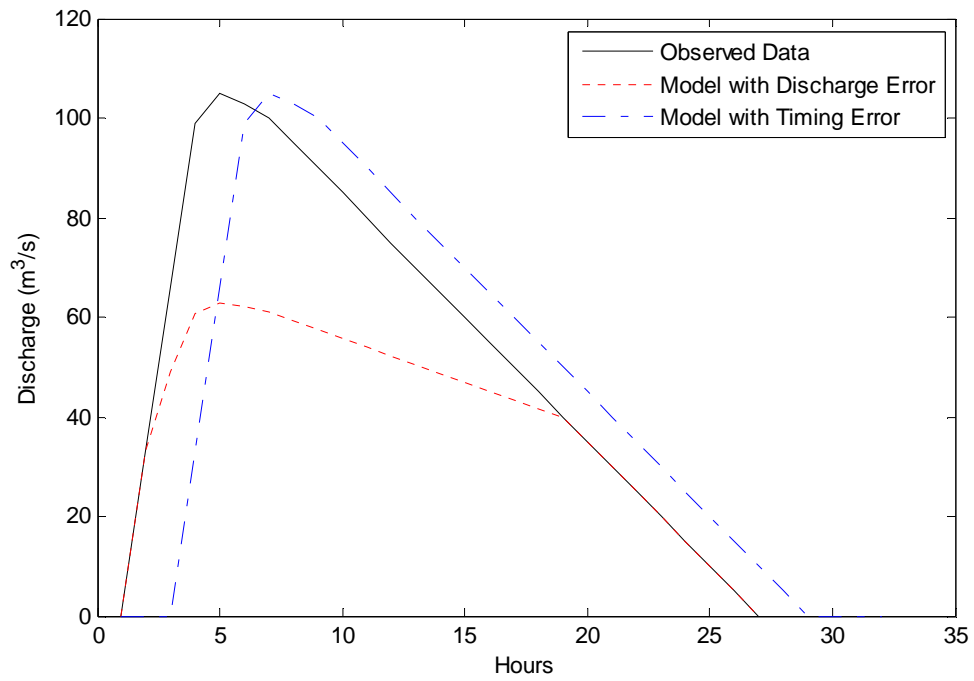
670 Figure 1: Geology of the Rangitaiki River basin, classified according to permeability.
671 Gauging Locations are marked.

672



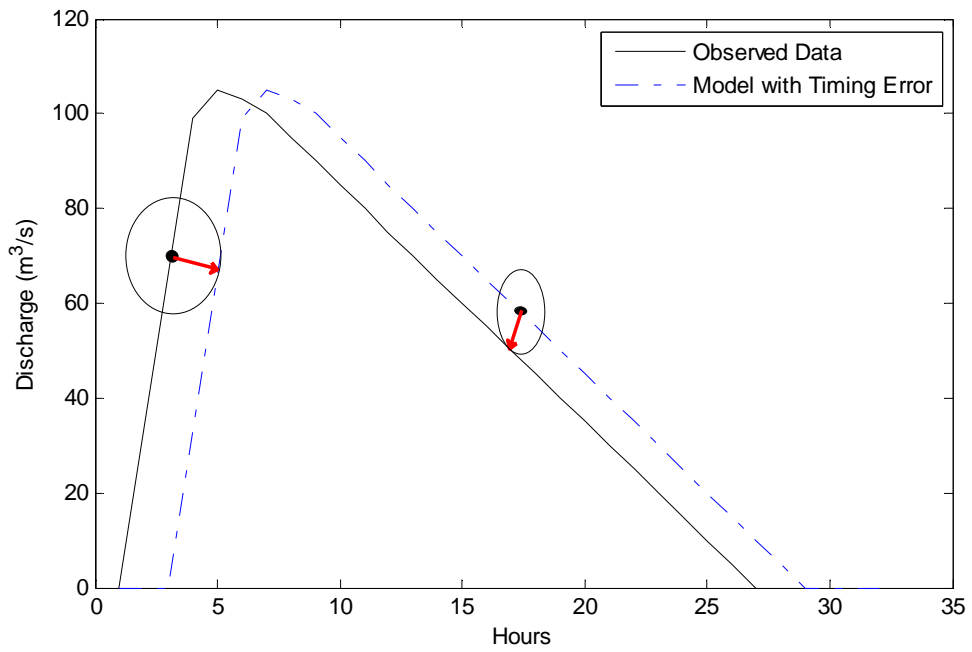
673

674 Figure 2: A synthetic example of hydrographs in which a model with minor (2 hour)
675 timing error is graded as having poorer performance than a model with 40% discharge
676 error
677



678

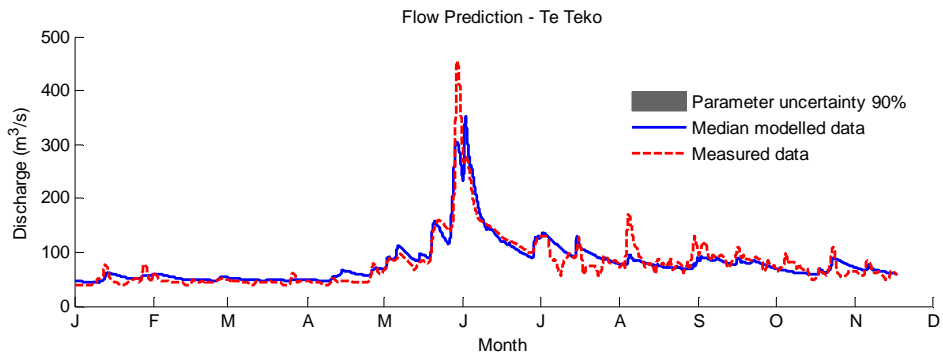
679 Figure 3: Search window to determine 'distance' between observed and predicted
680 flow values under the Extended Nash Sutcliffe likelihood measure.
681



682

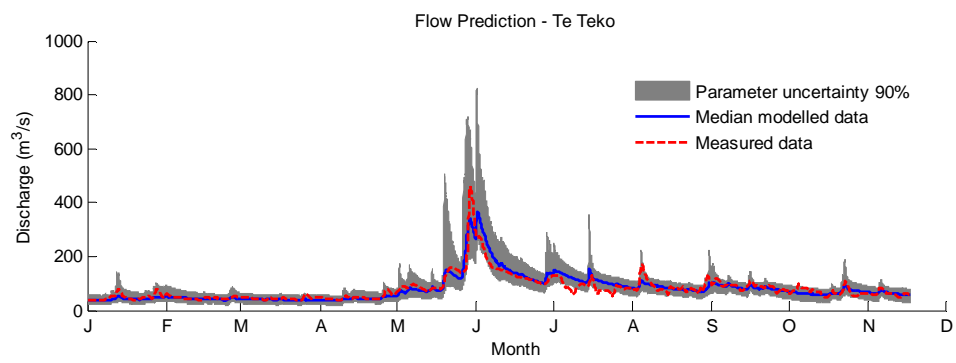
683 Figure 4: 90% Uncertainty bounds on flow at Te Teko using formal likelihood
684 measure to control MCMC search algorithm. Note that the bounds are sufficiently
685 narrow to be hardly visible as distinct from the median calibrated prediction.

686



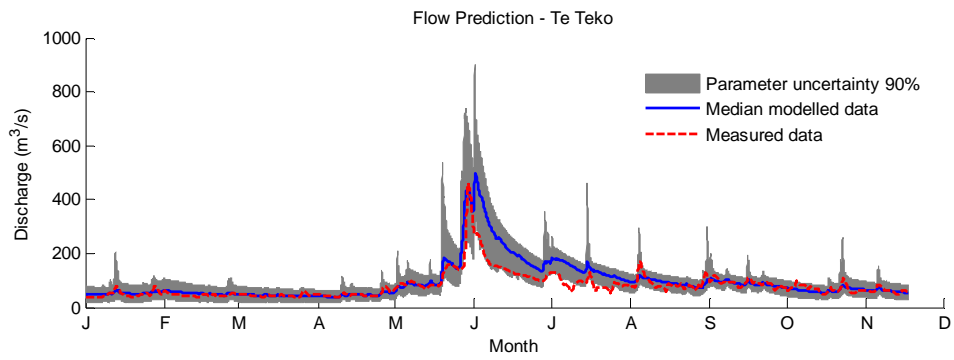
687

688 Figure 5: 90% Uncertainty bounds on flow at Te Teko using Nash-Sutcliffe informal
689 likelihood measure to control MCMC search algorithm.
690



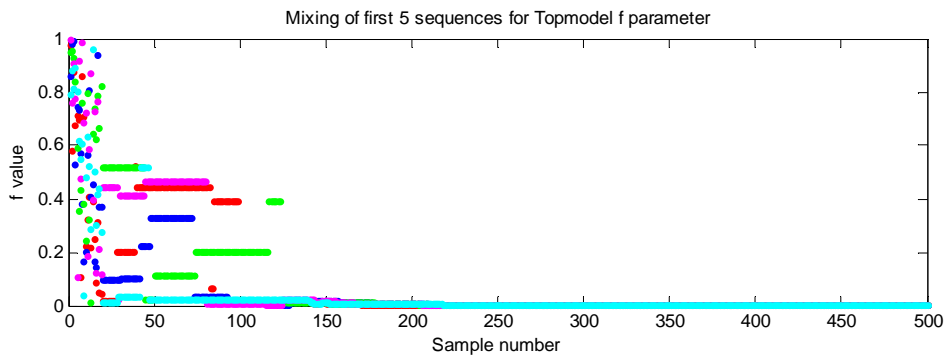
691

692 Figure 6: 90% Uncertainty bounds on flow at Te Teko using Extended Nash-Sutcliffe
693 informal likelihood measure to control MCMC search algorithm.
694



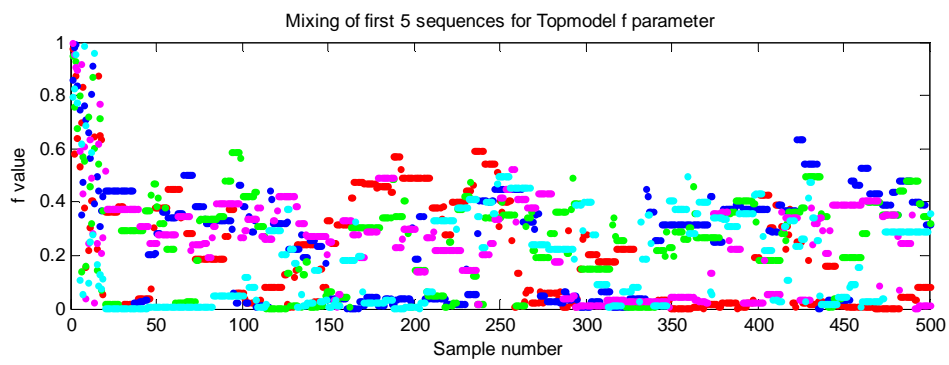
695

696 Figure 7: Topmodel 'f' parameter value over successive iterations of the MCMC
697 search algorithm using formal Bayesian (exponential error model) likelihood measure
698



699

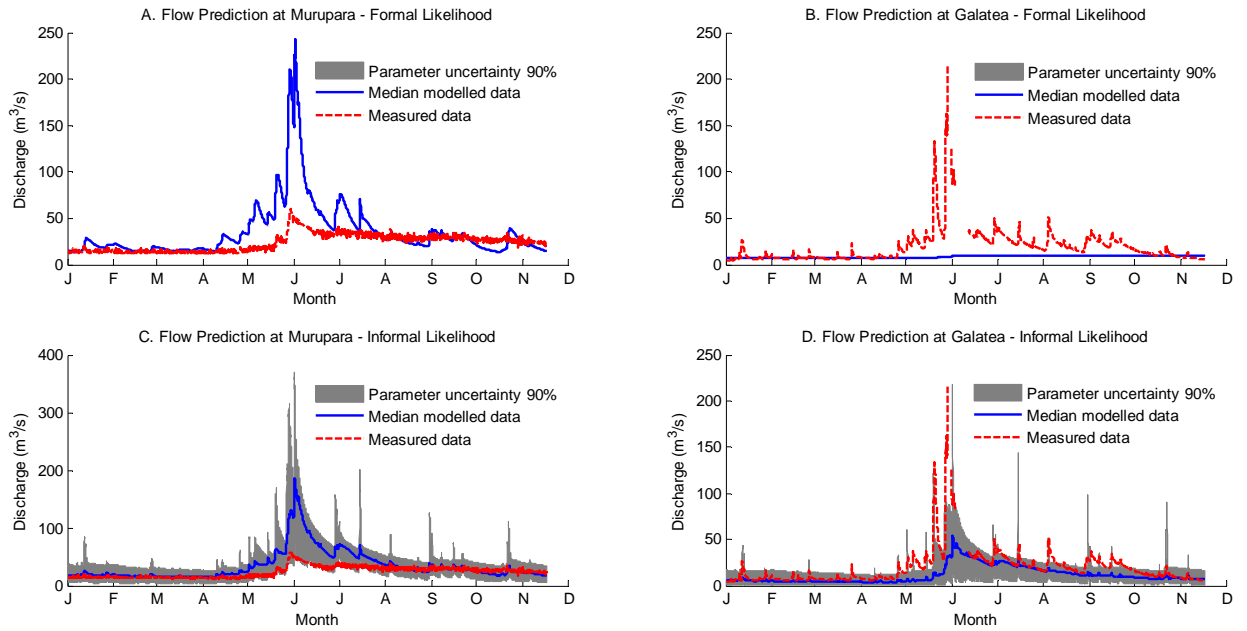
700 Figure 8: Topmodel 'f' parameter value over successive iterations of the MCMC
701 search algorithm using informal 'Extended Nash-Sutcliffe' likelihood measure
702



703

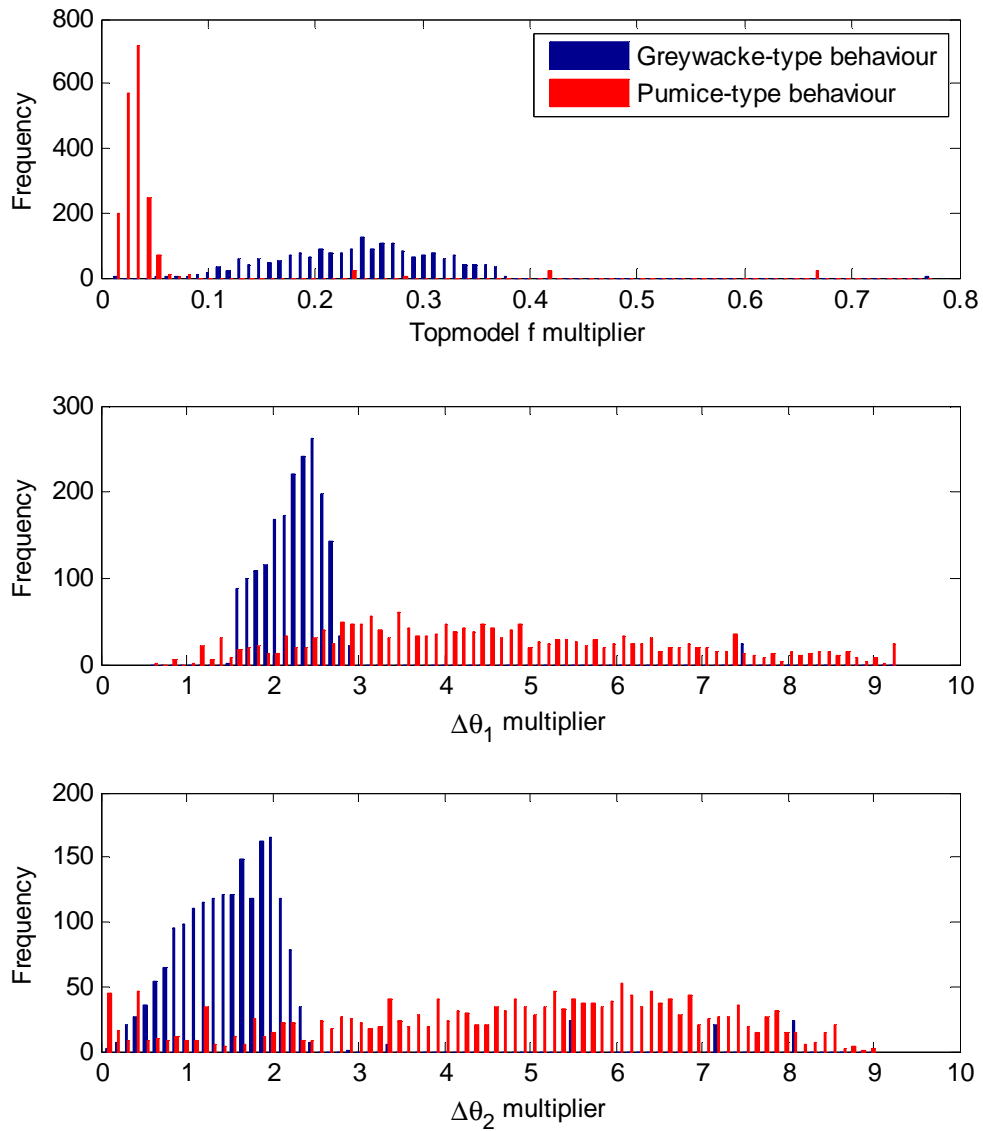
704 Figure 9: Comparison of Internal Flow predictions at Murupara (pumice
705 subcatchment) and Galatea (greywacke subcatchment) using formal and informal
706 likelihood measures

707



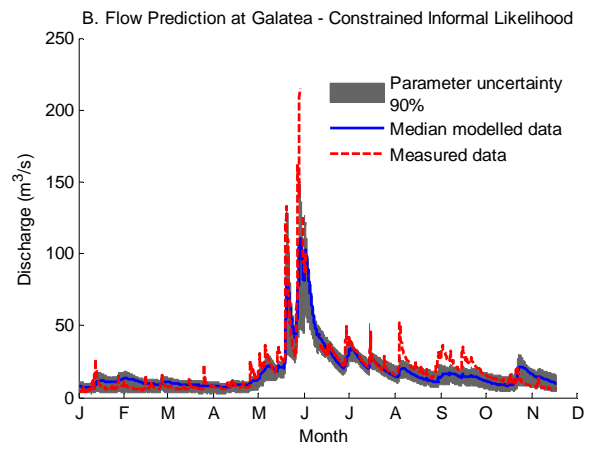
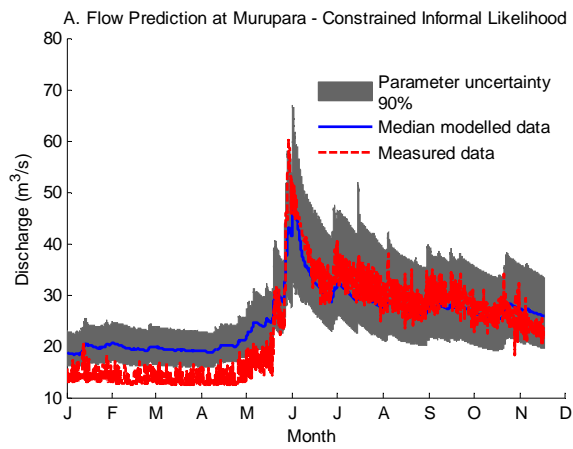
708
709

710 Figure 10: Multiplier ranges categorised by behavioural type for parameters:
 711 (a) Topmodel f (b) $\Delta\theta_1$ effective drained porosity (c) $\Delta\theta_2$ root zone storage. These
 712 plots were used to define constrained parameter ranges.
 713



714

715 Figure 11: Internal Flow predictions at Murupara (pumice subcatchment) and Galatea
716 (greywacke subcatchment) using informal likelihood measures under a constrained
717 calibration procedure
718
719



720

Table 1

TOPNET model parameters

	Name	Estimation
Sub-basin Parameters		
f (m ⁻¹)	Saturated store Sensitivity	Constant = 12.4 (multiplier calibrated)
K ₀ (m/h)	Surface saturated hydraulic conductivity	Constant = 0.01 (multiplier calibrated)
Δθ ₁	Drainable porosity	From soils (multiplier calibrated)
Δθ ₂	Plant available porosity	From soils (multiplier calibrated)
D (m)	Depth of soil zone	Depth ¼ l=f from soils (multiplier calibrated)
C	Soil zone drainage sensitivity	1
φ (m)	Wetting front suction	From soils
V (m/s)	Overland flow velocity	Constant = 0.1 (multiplier calibrated)
CC (m)	Canopy capacity	From vegetation
Cr	Intercepted evaporation enhancement	From vegetation
A	Albedo	From vegetation
Lapse (°C/m)	Lapse rate	0.0065
Channel parameters		
N	Mannings n	Constant = 0.024 (multiplier calibrated)
A	Hydraulic geometry constant	0.00011
B	Hydraulic geometry exponent	0.518
State variables		
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