AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2025WR040087

'Where the Creeks Run Dry or Ten Feet High': A Probabilistic Approach to Stage Height Forecasting in Australia

Rajitha Athukorala[1,2], Hadi Mohasel Afshar[3], Sally Cripps[3,4,5], and R. Willem Vervoort[1,2]

[1]ARC Industrial Transformation Training Centre in Data Analytics for Resources and Environments, Sydney, NSW, Australia, [2]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia, [3]Human Technology Institute, The University of Technology Sydney, Sydney, NSW, Australia, [4]School of Mathematical Sciences, The University of Technology Sydney, Sydney, NSW, Australia, [5]School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Sydney, NSW, Australia

Key Points:

- Propose a Bayesian Hierarchical Mixture of Experts (BHME) model to forecast one day ahead stage height
- The model captures sudden changes in flow regimes commonly seen in Australia

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. Athukorala,
rajitha.athukorala@sydney.edu.au

Citation:

Athukorala, R., Afshar, H. M., Cripps, S., & Vervoort, R. W. (2025). 'Where the creeks run dry or ten feet high': A probabilistic approach to stage height forecasting in Australia. *Water Resources Research*, *61*, e2025WR040087. https://doi.org/10.1029/2025WR040087

Received 28 JAN 2025
Accepted 18 AUG 2025

Abstract Streamflow prediction is a challenging but important task as it affects many processes in the environment. In Australia, the task is particularly complex because the transition from low to high flows can be rapid, and the overall variability is high. Accurate water resource management in Australia requires models that can accommodate these unique characteristics. Complicating matters further, streamflow is not directly measured but inferred from stage height measurements using rating curves. To address the challenges posed by the unique Australian hydrological features mentioned above, the present work proposes a model that: (a) directly predicts stage height rather than streamflow, (b) is sufficiently complex to learn the high variability of stage heights and (c) quantifies epistemic and aleatoric uncertainty by using a Bayesian framework. Our model is a Bayesian Hierarchical Mixture of Experts (BHME) model, with two mixture components. The two components are gamma densities in which the mean and variance are parameterized to depend on upstream stage height and rainfall from the previous day. The mixture weights are also parameterized to depend on the previous day's upstream stage height and rainfall through a logit link function. The method was tested on three gauging stations in the Orara and Richmond rivers. The BHME model provides better fit to the data and better predictive densities compared to a single component model. Notably, the BHME model excels in forecasting less frequent, high stage heights, crucial for managing high flow events. Additionally, transitions between mixture components offer insights into sudden stage height regime changes.

1. Introduction

Physics based conceptual hydrological models (hereafter referred to as hydrological models) are the most common tools available to understand and predict the behavior of hydrological systems to varying hydro-meteorological variables (Víctor et al., 2018; Wagener & Wheater, 2006). For example, they are frequently used to estimate the contribution of waterflow in unmonitored sub-catchments into a main river, for water accounting purposes. However, despite their extensive use, even state-of-the art models cannot capture many complexities in hydrological processes (Hrachowitz & Clark, 2017). This can introduce significant uncertainty in predictions, which affects estimates for water accounting. For example, in the years 2019-20, predictions from hydrological models failed to account for 28,472 mega liters of water, which is approximately 20% of the total volume of stream flow between the Mollee and Gunidgera weirs in the Namoi river, Australia (Burrell et al., 2021).

Accounting for uncertainty in both models and data as well as combining various sources of information is critical for decision makers in natural resource management–accurate predictions for natural resource ecosystems are driven by model complexity as well as data diversity, and the uncertainty associated with both needs to be considered for optimal management. The Bayesian paradigm provides a logically consistent framework for quantifying both these types of uncertainties (Cripps & Durrant-Whyte, 2023) as well as for combining prior knowledge about models and parameters, with information contained in observed data. Although the computational effort required to implement Bayesian procedures initially hindered the application of these methods (Allenby et al., 2005; Fasbender & Ouarda, 2010), the development of sampling methods (Gelfand & Adrian, 1990; Geman & Geman, 1984) and variational techniques (Jordan et al., 1999) to approximate posterior

distributions has resulted in the uptake of Bayesian methods in several areas of environmental sciences including hydrology (Green et al., 2020).

In Sections 1.1 and 1.2 we provide a more detailed review of the existing literature on deterministic and probabilistic approaches in hydrological modeling, respectively. Then in Section 1.3 we discuss the difference between modeling streamflow versus stage height.

1.1. Current Approaches to Physical Hydrological Modeling

The major limitation in current approaches, which uses a single hydrological model, is that these can fail to capture many non-linear complexities involved in hydrological processes. This in turn contributes significantly to the overall uncertainty, not only in prediction (Moges et al., 2020; Van Kempen et al., 2021), but also in inference, because models produce similar predictions with different parameter combinations (Beven, 2006; Jakeman & Hornberger, 1993; Kuczera et al., 2006).

To address the limitations of predictions from a single hydrological model, such as HBV (Bergström, 1992), GR4J (Perrin et al., 2003) and AWBM (Boughton, 2004), research has focused on combining several models (Clark et al., 2008, 2015; Fenicia et al., 2008; Mehrnegar et al., 2020; Mohammadi et al., 2021; Prieto et al., 2021, 2022), using methods like Superflexpy (Dal Molin et al., 2021). These combinations of models capitalize on the variation between model specification to capture the response of catchment runoff to rainfall. However, these approaches only provide tailored simulators for specific catchments: "one tool per problem" at the expense of generalizability and with potential overfitting on the training data (Todorović et al., 2024). This has led to the development of modeling techniques which combine both statistical and hydrological models (hereafter referred to as hybrid modeling approaches) rather than combining multiple hydrological models.

1.2. Probabilistic Approaches to Hydrological Modeling

Most hybrid modeling approaches incorporate prior hydrological knowledge through an existing hydrological model to model mean streamflow. These approaches then model the deviations from this mean via a likelihood function (Kuczera et al., 2006; D. Li et al., 2021; McInerney et al., 2017, 2018; Prieto et al., 2021; Thiemann et al., 2001; Ulzega & Albert, 2022). However, the improvement in prediction from such methods is at the expense of interpretability, as the integration of the hydrological model and statistical model can compensate for the flaws in the structure of the hydrological model. As a result, this influences the model parameter estimation process, which makes it difficult to link the inferred parameters back to the hydrological representation in the relevant model conceptualization (Evin et al., 2013). For example, the operational 7-day streamflow forecasts by the Bureau of Meteorology, Australia, uses such a hybrid model approach, (M. Li et al., 2016) combining the GR4H hydrological model with a sequence of statistical models to reduce the variance in the residuals.

Although these hybrid approaches simulate deviations from the mean stochastically, they assume that mean is known and ignore uncertainty surrounding it (Bates & Campbell, 2001; D. Li et al., 2021; Smith et al., 2015). This is not ideal when the objective is not merely to have a model which fits the data, but also to generate knowledge and insights useful for decision makers.

The alternative approach is to not use a hydrological model, but to incorporate the hydrological knowledge through statistical priors in an interpretable probabilistic framework. This is a relatively new approach in the hydrological modeling community and has the advantage of modeling both the mean and the variance of the estimated variable. Two publications provide starting points for such modeling frameworks. Ravindranath et al. (2019) proposed a Bayesian Network model to predict streamflow in the Upper Missouri river basin, utilizing the spatial dependence induced by the river network topology. The flow at the downstream station was assumed to depend on the flow at the most immediate upstream station and an additional exogenous variable which can represent the flow processes between the two gauges. Extending this method, Ossandon et al. (2021) developed a Bayesian Hierarchical Network Model (BHNM) to produce real-time forecasts of daily streamflow in a river network, for the monsoon period in India, using two main covariates: (a) the upstream streamflow and (b) the rainfall received, in the previous day.

The predictive ability of both the proposed methods relied on identifying (if available) appropriate feeder gauge/s upstream and the time taken for the upstream flow to impact downstream flow. A recent publication by Rastogi et al. (2025) extended the method proposed by Ossandon et al. (2021) to stage height predictions in the same river

network and the period. A major limitation of the method proposed by Ossandon et al. (2021) is that it is only applicable to the monsoon period, where continuous flow was guaranteed and therefore its usefulness for situations where no flow exists or where the catchment dynamics change throughout the year is questionable. Despite the mentioned shortcomings, these methods provide a valuable starting point that we rely and improve upon to formulate a data-driven Bayesian statistical model for stage height prediction in Australia.

Hierarchical Mixture of Experts (HME) models (Jacobs et al., 1991; Jordan & Jacobs, 1993) are a widely used class of models which combine the effects from multiple models in a probabilistic sense (Masoudnia & Ebrahimpour, 2014). HMEs are particularly useful when modeling data, that is generated from complex processes changing in space and time (Bertolacci et al., 2019; Holsclaw et al., 2017). Previous applications of HMEs in streamflow forecasting have relied on hydrological models with multiple parameter combinations as the experts to model the mean while using a Bayesian inference scheme to model the residuals as arising from some known density (Jeremiah et al., 2013; Marshall et al., 2007; Moges et al., 2016).

We propose a Bayesian Hierarchical Mixture of Experts (BHME) model where the mixture components are gamma densities, rather than a hydrological model, and with the mixing weights depending on the current catchment condition.

The knowledge-base here is that any given catchment will undergo changes over time due to changing and evolving hydro-climatic conditions which would then result in different responses from the catchment over time (Xiong et al., 2001). Australia's streamflow has the highest inter-annual variability globally (Chiew et al., 2000), and therefore poses challenges in streamflow prediction (Jiang et al., 2015). In this paper, we demonstrate how the proposed BHME model permits a flexible expression of the hydrological processes with fewer assumptions while providing better predictive densities and catchment specific knowledge through data, compared to its single component counterpart proposed by Ossandon et al. (2021).

1.3. Modeling a Latent Variable Versus Observed Variable

Stage height measurements are considered as proxies for streamflow discharge (Tomkins, 2014). Indeed many studies treat derived streamflow measurements as actual observed data and ignore the underlying uncertainties induced by converting stage heights to streamflow (Shao et al., 2014).

To convert stage height to streamflow, direct measurements of both quantities are taken at a gauging station to build a stage height-discharge relationship, formally known as a "rating curve." The ability of these rating curves to accurately assess streamflow discharge varies significantly across gauging stations (Westerberg et al., 2016; Westerberg & Karlsen, 2024). A study by McMahon and Peel (2019), which assessed the accuracy of ratings curves for 171 Australian Bureau of Meteorology Hydrologic reference stations, found that levels of uncertainty in the mean of the predicted streamflow ranged from +29% to −22%.

Therefore, in this paper, we model stage height (which is measured directly) instead of streamflow (which is estimated using a rating curve) since modeling stage height is particularly important in the Australian context where the stage height can change rapidly, as experienced firsthand during the recent flood events across Queensland and New South Wales in 2022, resulting in 13 deaths and 4.3 billion AUD worth of damage (Resilience, 2022).

In the next Section, we formalize our proposed BHME model. Section 3 introduces the data that we are analyzing to demonstrate the proposed method. The inference of the BHME model is explained in Section 4. Section 5 provides a comparison between the performance of a single-component model and the BHME using a simulated data set. In Section 6, we showcase the performance of the BHME on a real data set described in Section 3, followed by the discussion of results and conclusions in Sections 7 and 8, respectively.

2. Model

2.1. Notation

Throughout, we follow the common convention of denoting vectors using bold characters. The transpose of a vector, $\mathbf{v}$, is denoted by $\mathbf{v}'$. The Hadamard product (a.k.a elementwise product) of same-length vectors, $\mathbf{v}$ and $\mathbf{u}$, is represented by $\mathbf{v} \odot \mathbf{u}$.

Let $\mathbf{H} \coloneqq (\mathbf{h}_1, \mathbf{h}_2)$ be a $T \times 2$ matrix of stage height observations at 2 locations, each over T days, where $\mathbf{h}_1$ is lagged and upstream of $\mathbf{h}_2$, so that $\mathbf{h}_1 \coloneqq (h_{1,1}, \ldots, h_{T,1})'$ spans the days corresponding to $t = 1:T$, while $\mathbf{h}_2 \coloneqq (h_{2,2}, \ldots, h_{T+1,2})'$ spans the days corresponding to $t = 2:T+1$. Let $X \coloneqq (\mathbf{1}_T, \log(\mathbf{h}_1), \mathbf{r}, \log(\mathbf{h}_1) \odot \mathbf{r})$ be a $T \times 4$ design matrix, where $\mathbf{1}_T$ is a $T \times 1$ vector of ones and $\mathbf{r} \coloneqq (r_1, \ldots, r_T)'$ is a time-series of daily rainfall measured at the closest available upstream location corresponding to days $t = 1:T$. The units of stage height and rainfall values used throughout this paper are meters (m) and millimeters (mm) respectively.

2.2. Overview

We propose a Bayesian Hierarchical Mixture of Experts model for stage height, where the mixture components are gamma distributions, with parameters that depend on two covariates; the previous day's upstream stage height and the previous day's rainfall. The mixing probabilities also depend on the same two covariates, so that downstream stage height is impacted by upstream stage height and rainfall via the components in the mixture and the probabilities attached to those components.

The model proposed in this paper, extends the original model by Ossandon et al. (2021) in four ways:

1. The means and standard deviations of the component gamma distributions are not deterministic functions of upstream stage height and rainfall but rather a noisy signal of these covariates.
2. The relationship between log downstream stage height and log upstream stage height and rainfall is presumed to be linear to ensure that the positivity constraint is imposed.
3. We include an interaction term between the log upstream stage height and rainfall. This is to accommodate the fact that, for a given volume of rainfall, both the river's height and width will increase, and therefore the impact of rainfall on up/downstream stage height will decrease with up/downstream stage height.
4. The stage height is modeled as coming from a mixture of (gamma) distributions, to allow for more flexibility in estimating the underlying true distribution and to account for the potential for model misspecification, more likely in a single distribution model.

2.3. Model Conceptualization

Figure 1 outlines the three different model conceptualizations. The left panel describes a model where the downstream stage height is linearly and additively related with upstream stage height and rainfall. That is, the impact of upstream rainfall on downstream stage height does not depend on the upstream stage height. This is the assumption imposed by the model described in Ossandon et al. (2021) (The model proposed by Ossandon et al. (2021) was applied for streamflow while in our case the modeled variable is stage height). The model indicated by the middle panel describes a model where the log of the downstream stage height is linearly related to the log of the upstream stage height. In the absence of other factors this assumes a power function between upstream and downstream stage heights. In addition, an interaction term has been included. This allows the impact of upstream rainfall on downstream stage height to depend on the upstream stage height. Intuitively this makes sense, as the stage height increases so too does the width of the river. Thus rainfall will have a decreasing impact on downstream stage heights as the upstream stage height increases. However, this model assumes that the rate of the decrease in the impact of rainfall on downstream stage height is constant for all heights. The model represented by the right panel resembles the proposed BHME which accounts for varying interaction effects from upstream stage height and rainfall on downstream stage height by introducing different mixture components. Intuitively this model states that although the river widens as the log of the stage height increases, the rate at which it widens varies with height, until, in the extreme case, the banks overflow. For example, if the dependencies between the downstream stage height and the upstream stage height and rainfall are as in the model represented by the middle panel, it cannot adapt to spreading of water across floodplains in over bank flows, and thus over-predicts downstream stage heights. This corresponds to smaller increments in stage height at already high stage heights as a result of over bank flows compared to larger increments in stage height for in channel flows as a result of similar upstream stage heights and rainfall depths as shown by the dotted lines in Figure 1.

2.4. Likelihood: Bayesian Hierarchical Mixture of Experts Model (BHME)

We model the likelihood of $\mathbf{h}_2$, conditional on the lagged upstream observed stage height, $\mathbf{h}_1$, rainfall, $\mathbf{r}$, and a parameter vector $\boldsymbol{\theta}$, where $\boldsymbol{\theta} \coloneqq (\boldsymbol{\beta}_{1:2}, \boldsymbol{\phi}_{1:2}, \boldsymbol{\sigma}_{\mu_{1:2}}, \boldsymbol{\sigma}_{\eta_{1:2}}, \boldsymbol{\alpha})$, with $\boldsymbol{\beta}_{1:2} \coloneqq (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, $\boldsymbol{\phi}_{1:2} \coloneqq (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$, $\boldsymbol{\sigma}_{\mu_{1:2}} \coloneqq (\sigma_{\mu_1}, \sigma_{\mu_2})$ and $\boldsymbol{\sigma}_{\eta_{1:2}} \coloneqq (\sigma_{\eta_1}, \sigma_{\eta_2})$, as the following mixture model:

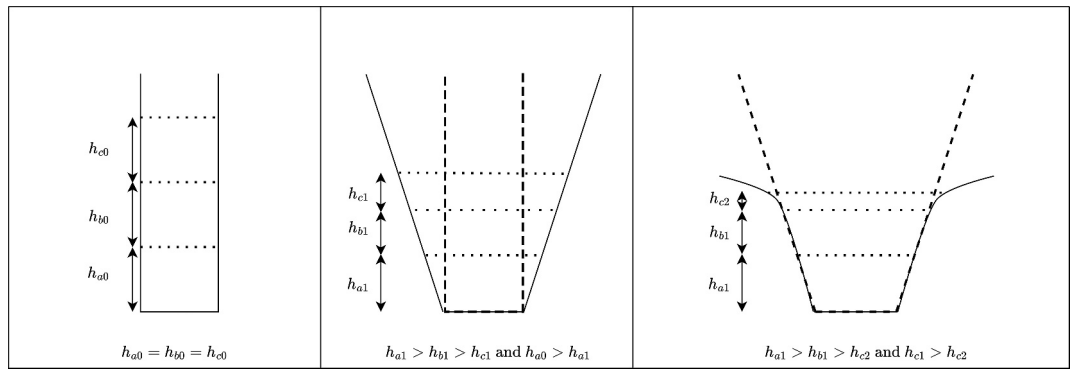

Figure 1. Different model conceptualizations and the associated conceptualized river cross section. The three dotted lines in each panel indicates downstream stage heights generated by similar upstream stage height and rainfall depth under each model conceptualization. Left panel: The relationship between downstream stage height and upstream stage height and rainfall is linear, and additive, which can be conceptualized by a rectangular river cross section. There is no interaction term. h_{a0}, h_{b0} and h_{c0} are the corresponding increments of stage height for consecutive units of upstream rainfall received on the previous day r_a, r_b and r_c respectively. In this case $h_{a0} = h_{b0} = h_{c0}$; Middle panel: The relationship between the log downstream stage height and log upstream stage height and rainfall is linear and includes an interaction term, to allow for the impact of upstream rainfall on downstream stage height to depend on the upstream stage height (Equation 1 with $k = 1$). This can be conceptualized by a trapezoidal river cross section. Intuitively this model allows for the river to widen as the height increases. The dashed partial rectangle provides reference to the river cross section given in the left panel; Right panel: The relationship between log downstream stage height and log upstream stage height is linear in the absence of other factors. The impact of rainfall on downstream stage height varies with upstream stage height, and this variation itself changes (Equation 1 with $k = 2$). Intuitively this model allows for the river to widen as stage height increases but at different rates. This model can be conceptualized by multiple trapezoidal river cross sections, such as may be found with overbank flows. The dashed trapezoid provides reference to the river cross section given in the middle panel. h_{a1}, h_{b1} and h_{c2} are the corresponding increments of stage height for consecutive units of upstream rainfall received on the previous day r_a, r_b and r_c respectively. In this case $h_{a1} > h_{b1} > h_{c2}$ and $h_{c1} > h_{c2}$.

$$
\begin{aligned}
p(\mathbf{h}_2|\mathbf{h}_1, \mathbf{r}, \boldsymbol{\theta}) &:= \prod_{t=2}^{T+1} \sum_{k=1}^{2} p_{tk}\left(h_{t,2}|\mathbf{x}_t, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \sigma^2_{\mu_k}, \sigma^2_{\eta_k}, \gamma_t = k\right) \Pr\left(\gamma_t = k|\boldsymbol{\alpha}, \mathbf{x}_t\right) \\
&:= \prod_{t=2}^{T+1} \sum_{k=1}^{2} p_{tk}\left(h_{t,2}|\mathbf{x}_t, \boldsymbol{\beta}_k, \boldsymbol{\phi}_k, \sigma^2_{\mu_k}, \sigma^2_{\eta_k}, \gamma_t = k\right) \pi_{tk}(\mathbf{x}_t; \boldsymbol{\alpha}),
\end{aligned} \tag{1}
$$

where $\mathbf{x}_t$ is the t-th row of X, γ is an indicator variable with $\gamma_t = k$ if observation $h_{t,2}$ is generated from component k, for $k = 1, 2$, $p_{tk}(\cdot)$ are the mixture components and $\pi_{tk}(\cdot) := \Pr(\gamma_t = k|\cdot)$ are the weights attached to these components, $\boldsymbol{\beta}_k := (\beta_{0k}, \beta_{1k}, \beta_{2k}, \beta_{3k})'$ and $\boldsymbol{\phi}_k := (\phi_{0k}, \phi_{1k}, \phi_{2k}, \phi_{3k})'$. Equation 1 assumes that, conditional on lagged upstream stage height, rainfall, the indicator variable, γ, and the parameters, $\boldsymbol{\theta}$, the downstream stage heights are independent of each other. When $k = 1$, Equation 1 reduce to the single component model while $k = 2$ represents the two component mixture model.

The two most commonly used distributions for representing hydrological variables are gamma and log-normal distribution (Charbeneau, 1978; Langat & Koech, 2019; Singh, 1998) with the gamma distribution preferred to represent stage (stream) height (Moramarco et al., 2008; Tayfur & Moramarco, 2022) because it provides a better approximation of the empirical distribution for daily stage height than the log-normal distribution. In this study, we assume the mixture components, $p_{tk}(\cdot)$, are gamma density functions, each with mean μ_{tk} and standard deviation σ_{tk} given by Equations 2 and 3 respectively. As explained in Section 2.2, we propose a mixture distribution to allow for the dependency between upstream conditions and downstream stage height to vary over time in response to changes in upstream catchment conditions (rainfall and stage height).

$$
\mu_{tk} := \exp\left(\mathbf{x}_t\boldsymbol{\beta}_k + \epsilon_{t,\mu_k}\right), \text{ where } \epsilon_{t,\mu_k} \sim \mathcal{N}\left(0, \sigma^2_{\mu_k}\right) \tag{2}
$$

$$
\sigma_{tk} := \exp\left(\mathbf{x}_t\boldsymbol{\phi}_k + \epsilon_{t,\eta_k}\right), \text{ where } \epsilon_{t,\eta_k} \sim \mathcal{N}\left(0, \sigma^2_{\eta_k}\right) \tag{3}
$$

$\boldsymbol{\beta}_k := (\beta_{0k}, \beta_{1k}, \beta_{2k}, \beta_{3k})'$, and $\boldsymbol{\phi}_k := (\phi_{0k}, \phi_{1k}, \phi_{2k}, \phi_{3k})'$, are the vectors of regression coefficients for μ_{tk} and σ_{tk}, for $k = 1, 2$, in Equations 2 and 3 respectively. The notation $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2.

The mixture weights are also functions of lagged upstream stage height and rainfall. Specifically;

$$\begin{aligned} \pi_{t1} &:= \frac{\exp(-\mathbf{x}_t\boldsymbol{\alpha})}{1 + \exp(-\mathbf{x}_t\boldsymbol{\alpha})} \\ \pi_{t2} &:= 1 - \pi_{t1}, \end{aligned} \tag{4}$$

where $\boldsymbol{\alpha} := (\alpha_0, \alpha_1, \alpha_2, \alpha_3)'$ are the mixture weight parameters.

2.5. Priors

We place multivariate-normal priors on $\boldsymbol{\beta}_{1:2}$ and $\boldsymbol{\phi}_{1:2}$

$$\begin{aligned} \boldsymbol{\beta}_k &\sim \mathcal{MVN}(0, c\boldsymbol{I}_4) \\ \boldsymbol{\phi}_k &\sim \mathcal{MVN}(0, c\boldsymbol{I}_4), \end{aligned}$$

where the notation $\mathcal{MVN}(\boldsymbol{\nu}, \Sigma)$ is the multivariate-normal distribution with mean $\boldsymbol{\nu}$, variance/covariance matrix Σ. c is a scalar constant and $\boldsymbol{I}_M$ denotes the $M \times M$ identity matrix. We set $c = T$, the number of data points, to allow the prior to be sufficiently wide, w.r.t the likelihood, reflecting limited knowledge on the relevant time varying dependencies between upstream rainfall and stage height with downstream stage height in this study. If useful hydrological information is available, other forms of priors could be used to reflect this prior knowledge about the hydrological system in the catchment.

The prior for $\boldsymbol{\alpha}$ is also multivariate-normal with $c = T$ similar to the prior on $\boldsymbol{\beta}_{1:2}$ and $\boldsymbol{\phi}_{1:2}$,

$$\boldsymbol{\alpha} \sim \mathcal{MVN}(0, c\boldsymbol{I}_4).$$

We place inverse-gamma priors on $\boldsymbol{\sigma}_{\mu_{1:2}}$ and $\boldsymbol{\sigma}_{\eta_{1:2}}$.

$$\begin{aligned} \sigma_{\mu_k} &\sim \mathcal{IG}(a, b) \\ \sigma_{\eta_k} &\sim \mathcal{IG}(a, b), \end{aligned}$$

where the notation $\mathcal{IG}(\tau, \upsilon)$ denotes an inverse-gamma distribution with shape τ and scale υ. We set $a = 1$ and $b = 2$ for the inverse-gamma priors to reflect our prior belief in smaller values of the standard deviation of the error terms (in Equations 2 and 3) while admitting larger values are also possible (Kruschke, 2015), with support only on non-zero, positive values.

2.6. Performance Metrics

We use the Watanabe-Akaike/Widely Applicable Information Criteria (WAIC) (Watanabe, 2010), a probabilistic performance metric, to estimate the expected log pointwise predictive density (*elppd*) given by Equation 5. $\hat{p}_{\text{waic}}$ given by Equation 7 is the effective number of parameters in the model (Vehtari et al., 2017),

$$\widehat{elppd}_{\text{waic}} := \widehat{lppd} - \hat{p}_{\text{waic}} \tag{5}$$

$$\widehat{lppd} := \sum_{t=2}^{T+1} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(h_{t,2}|\boldsymbol{\theta}^s)\right) \tag{6}$$

$$\hat{p}_{\text{waic}} := \sum_{t=2}^{T+1} V_{s=1}^{S}\left(\log p(h_{t,2}|\boldsymbol{\theta}^s)\right) \tag{7}$$

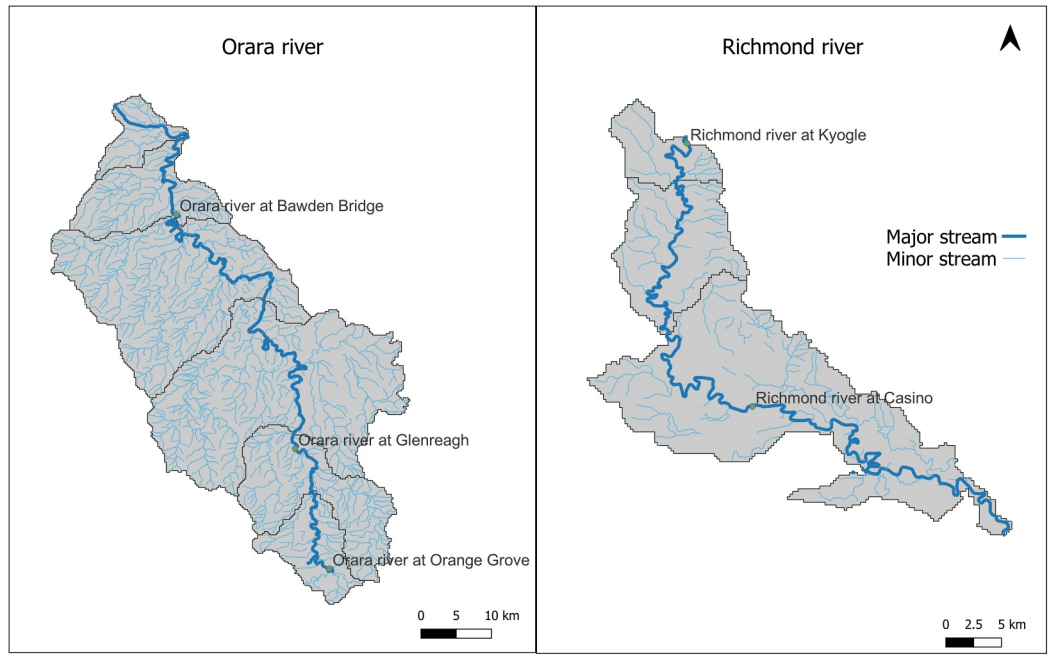

Figure 2. Station locations: On left the Orara river and on right the Richmond river.

The quantity, $\widehat{lppd}$ in Equation 6 denotes the log pointwise predictive density, and $V_{s=1}^{S}$ in Equation 7 represents the sample variance, for the posterior draws $s = (1, \ldots, S)$. The ˆsymbol denotes that these quantities are computed estimates using the posterior draws.

Since the proposed BHME is a probabilistic approach, the most suitable metric to compare the model performance is the WAIC. However, we also compute three other measures of predictive performance; mean squared error (*MSE*), the Nash-Sutcliffe efficiency (*NSE*) (Nash & Sutcliffe, 1970) and the correlation coefficient (R) between the observed and predicted downstream stage heights.

3. Data

The proposed BHME model was applied to data from three stations in Australia (Figure 2) given in Table 1.

The above stations were selected for the following reasons:

1. The availability of two or more stations along the main river channel without branching (bifurcation) between the stations.
2. The availability of a continuous record of stage height data with no missing values for at least a period of 3 years (with the exception of Nana Glen station which had four missing data records for rainfall. These were removed from the data set).

The model was estimated using 3 years of daily data (1096 data points) and the predictive performance of the model was assessed using one year of daily data (365 data points), following the model estimation period.

Table 1
Streamflow Gauging Stations Used in the Study

River	Station	Training period	Upstream gauging station	Closest upstream rainfall station	River length between the two stations
Orara	Bawden	2008–2010	Glenreagh	Nana Glen	≈ 80 km
Orara	Glenreagh	2008–2010	Orange Grove	Nana Glen	≈ 30 km
Richmond	Casino	2012–2014	Kyogle	Kyogle	≈ 60 km

4. Inference

The joint and marginal posterior distributions of the model parameters, $\boldsymbol{\theta}$, are approximated by Markov Chain Monte Carlo (MCMC) (Gelfand & Adrian, 1990). More specifically, we approximate the posterior, $p(\boldsymbol{\theta}|\mathbf{H},\mathbf{r})$, by a set of S samples, $\boldsymbol{\theta}^{[1]},\ldots,\boldsymbol{\theta}^{[S]}$, drawn using the No-U-Turn Sampler (NUTS) (Homan & Gelman, 2014). NUTS is a variation of Hamiltonian Monte Carlo (HMC) (Duane et al., 1987) where the hyper-parameters are decided dynamically. These samples are then used to estimate features of the posterior distribution of interest, for example, the posterior expected value of $\boldsymbol{\theta}$ which is approximated by,

$$\mathbb{E}(\boldsymbol{\theta}|\mathbf{H},\mathbf{r}) = \int \ldots \int p(\boldsymbol{\theta}|\mathbf{H},\mathbf{r}) \cdot \boldsymbol{\theta} d\boldsymbol{\theta}$$
$$\approx \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\theta}^{[s]}.$$

The NUTS sampling was implemented using Numpyro (Bingham et al., 2019; Phan et al., 2019), a probabilistic programming package in python which uses JAX (Bradbury et al., 2018) for automatic differentiation. Three chains were initiated with different starting points for each run to draw 10,000 samples following 2000 warmup samples. The BHME model estimation took on average 95 s to complete a single chain while using a single CPU. The convergence was assessed using trace plots, a selection of which is provided in the Supporting Information S1 section. The relevant data along with the sampling implementation is available online at: https://github.com/Rajitha-Athukorala/BHME_streamflow.git.

5. Simulation Study

In this section we demonstrate the performance of the model via simulation. We do this by generating two sets of 50 realizations from Equation 1. The first set of realizations are generated from Equation 1 with $k = 1$, which corresponds to generating data from a single component model. The second set of realizations are generated from Equation 1 with $k = 2$, which corresponds to a two component mixture model. We refer to these models as M_1 and M_2 respectively.

For all realizations, we obtain two estimates of the model parameters and predictive densities. The first estimate assumes that the data are generated from a single component mixture model and the second estimate assumes that data are generated from a two component mixture model. We refer to these estimates as $\mathcal{E}_1$ and $\mathcal{E}_2$ respectively.

Using this notation, the quantity $\mathcal{E}_1|M_1$ is the estimate we obtain if we assumes the data are generated from a single component mixture and the data were indeed generated from a single component mixture. Whereas the quantity $\mathcal{E}_2|M_1$ is the estimate we obtain if we assume the data are generated from a two component mixture model, but the data were generated from a single component mixture model. The quantities $\mathcal{E}_1|M_2$, and $\mathcal{E}_2|M_2$ are similarly defined for data generated from a two component mixture model.

This comparison is important because it tests the robustness of the proposed technique to model misspecification. For example, we know that if the data are generated from M_2 then $\mathcal{E}_2$ should give better estimates of parameters and predictive densities than $\mathcal{E}_1$ - allowing for complexity will give better results when the underlying truth is complex. However we also wish to establish the cost of allowing for complexity. For example, if the data are generated from the simpler model M_1, will allowing for complexity, that is using the estimate $\mathcal{E}_2$, lead to over-fitting and if so by how much?

To establish answers to these issues, Section 5.2 presents results for $\mathcal{E}_1|M_1$ and $\mathcal{E}_2|M_1$, that is the estimates when the data are generated from the single component mixture model. Section 5.3 presents results for the estimates when the data are generated from a two component mixture model, that is $\mathcal{E}_1|M_2$ and $\mathcal{E}_2|M_2$.

We now describe the simulation settings.

5.1. Simulated Data Generation and Inference

We generate 50 realizations from each of the models M_1 and M_2 and Table 2 lists the parameter values of both models that were used to generate the simulated data. To construct the design matrix, $\mathbf{X}$, for these simulations, we use the actual lagged upstream stage height and rainfall from Bawden station in Orara river, Australia.

Table 2
Values of Parameters Used to Generate Stage Height Data for the Simulation Study

Parameter	M_1	M_2: Component 1	M_2: Component 2
β_0	−0.81	−0.80	−0.03
β_1	3.85	3.65	3.02
β_2	0.71	−0.28	−2.87
β_3	−1.29	−1.57	2.24
ϕ_0	−2.54	−3.26	−0.59
ϕ_1	5.13	3.64	2.14
ϕ_2	7.29	6.44	0.57
ϕ_3	−9.00	15.05	−0.32
σ_μ	0.04	0.04	0.17
σ_σ	0.11	0.12	0.37
α_0	NA	6.47	–
α_1	NA	−11.84	–
α_2	NA	−22.35	–
α_3	NA	5.56	–

The parameter values used to generate data from M_2 were assigned close to the estimated posterior mean of the parameters for the M_2 model on real data from Bawden station in Orara river. To simulate data from M_2, we first compute the probability of downstream stage height on day t to come from one of two gamma distributions by introducing an auxiliary variable, $\gamma_t \sim \mathcal{B}(\pi_{1,t}(\mathbf{x}_t;\boldsymbol{\alpha}))$. If $\gamma_t = 1$ then $h_{t,2}$ is generated from component 1 and if $\gamma_t = 2$, then $h_{t,2}$ is generated from component 2. The notation $\mathcal{B}(\pi)$ denotes a Bernouilli distribution with probability of success π.

To simulate data from M_1, parameter values close to the first component of the M_2 model data simulation was used except for β_2. β_2 corresponds to the regression coefficient for rainfall on μ_t, and was increased to mimic a true stage height data set.

Each estimate $\mathcal{E}_k$, for $k = 1,2$, contains the estimates of the posterior distributions of the parameters, $p_k(\boldsymbol{\theta}|\mathbf{H},\mathbf{r})$ as well as estimates of the predictive densities $p_k(\mathbf{h}^*|\mathbf{H},\mathbf{r})$ for $k = 1,2$. To quantify the accuracy of these estimates we defined the following quantities.

Let the Root Mean Square Error (RMSE) between the downstream actual stage height, simulated from model M_j, at time t, for realization l, $h_{t,2,j,l}$ and the predicted mean of downstream stage height, of estimate $\mathcal{E}_k$, $\hat{h}^*_{t,2,j,k,l}$, be defined as

$$RMSE_{jkl} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(h_{t,2,j,l} - \hat{h}^*_{t,2,j,k,l}\right)^2},$$

for $l = 1,\ldots,50$, j and $k = 1,2$. Let

$$\hat{\boldsymbol{\theta}}_{j,k,l} := \left(\hat{\boldsymbol{\beta}}_{j,k,l}, \hat{\boldsymbol{\phi}}_{j,k,l}, \hat{\boldsymbol{\sigma}}_{\mu_{j,k,l}}, \hat{\boldsymbol{\sigma}}_{\eta_{j,k,l}}, \hat{\boldsymbol{\alpha}}_{j,k,l}\right),$$

be the estimated posterior means of the parameters for estimate $\mathcal{E}_k$ and realization l, when the data are generated from model M_j. Note that for $k = 1$, there is no $\hat{\boldsymbol{\alpha}}_{j,1,l}$. The parameters in $\boldsymbol{\theta}$ are described in Section 2.4, however note that these parameters now have the additional subscripts j, k and l to identify the model, estimate and realization respectively.

For example, if $j = 1$, and $k = 1$ then $\hat{\boldsymbol{\beta}}_{1,1,l} = \left(\hat{\beta}_{0,1,1,l}, \hat{\beta}_{1,1,1,l}, \ldots, \hat{\beta}_{3,1,1,r}\right)$. If $j = 2$ and $k = 1$, then $\hat{\boldsymbol{\beta}}_{2,1,l} = \left(\hat{\beta}_{0,2,1,l}, \hat{\beta}_{1,2,1,l}, \ldots, \hat{\beta}_{3,2,1,l}\right)$, where $\hat{\beta}_{p,j,k,l}$ is the regression coefficient for the p^{th} covariate, $p = 0,1,2,3$. If $j = 2$ and $k = 2$, then $\hat{\boldsymbol{\beta}}_{2,2,l} = \left(\tilde{\beta}_{0,2,2,l,1}, \tilde{\beta}_{0,2,2,l,2}, \ldots, \tilde{\beta}_{3,2,2,1,1}, \tilde{\beta}_{3,2,2,l,2}\right)$ where $\tilde{\beta}_{p,j,k,l,m}$ is the regression co-efficient for the p^{th} covariate, and the m^{th} component, $m = 1,2$

5.2. Results for Data Generated From M_1

This section presents the results of the accuracy of the estimates of $\mathcal{E}_1$ and $\mathcal{E}_2$ when data are generated from M_1.

Let $\hat{\boldsymbol{\Theta}}_{j,k} := \left(\hat{\boldsymbol{\theta}}_{j,k,1}, \ldots, \hat{\boldsymbol{\theta}}_{j,k,50}\right)$. Figure 3, displays the density estimates of the posterior means of the elements of $\hat{\boldsymbol{\Theta}}_{1,k}$ for the 50 realizations described in Section 5.1, for $k = 1,2$. Orange plots correspond to estimates from $k = 1$, the estimate which assumes the data were generated from a single component. The blue plots correspond to estimates from $k = 2$, the estimate which assumes the data were generated from a two component mixture, for the first component of that mixture. The red plots also correspond to estimates from $k = 2$, but for the second component in that mixture. Dark blue dashed lines are the true parameter values used to simulate the data.

Three features are apparent from Figure 3. The first is that the distribution of estimated parameter posterior means are centered over the true parameter values for both estimators. The second is that distribution of estimated parameter posterior means obtained from $k = 2$ are slightly more variable than those obtained from $k = 1$. This is

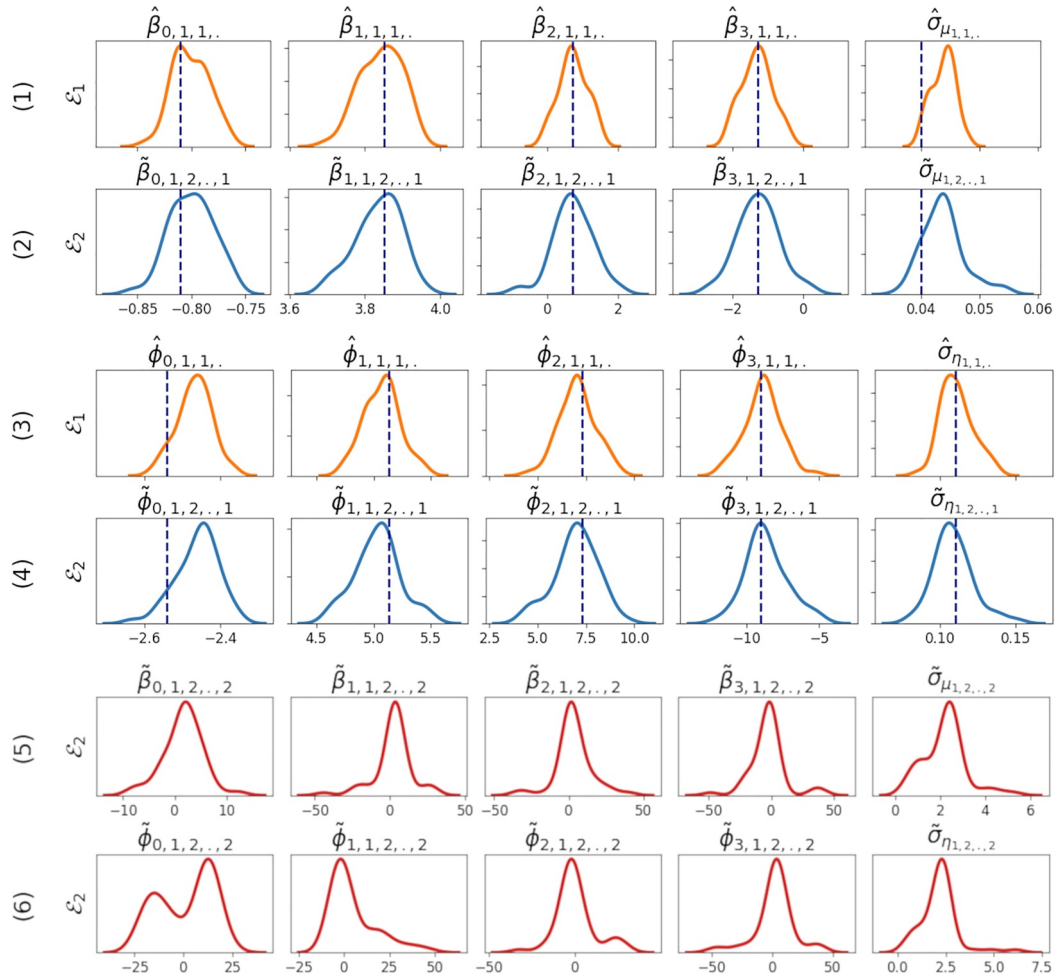

Figure 3. Plots of density estimates of the posterior means of parameters $\hat{\boldsymbol{\theta}}_{1,k,l}$ using $\mathcal{E}_1$ and $\mathcal{E}_2$ for data generated from M_1. These density estimates are plotted using the seaborn package (Waskom, 2021) which uses a Gaussian kernel to generate a smooth density estimate from discrete bin data Top two rows (1–2): Density estimates of the posterior means constructed from the $\hat{\boldsymbol{\beta}}_{1,k,l}$'s, (first four columns) and the $\hat{\boldsymbol{\sigma}}_{\mu_{1,k,l}}$'s (final column). Row 1 (orange) is for estimate $\mathcal{E}_1$ while row 2 (blue) is for estimate $\mathcal{E}_2$, component 1 Middle two rows (3–4): Density estimates of the posterior means constructed from the $\hat{\boldsymbol{\phi}}_{1,k,l}$'s, (first four columns) and the $\hat{\boldsymbol{\sigma}}_{\eta_{1,k,l}}$'s (final column), Row three (orange) is for estimate $\mathcal{E}_1$ while row 4 (blue), is for estimate $\mathcal{E}_2$, component 1 Bottom two rows (5–6): Row 5 (red) is the density estimates of the posterior means constructed from the $\hat{\boldsymbol{\beta}}_{1,2,l}$'s, (first four columns) and the $\hat{\boldsymbol{\sigma}}_{\mu_{1,2,l}}$'s (final column) using $\mathcal{E}_2$ for component 2, while the second row 5 (red) is the density estimates of the posterior means constructed from the $\hat{\boldsymbol{\phi}}_{1,2,l}$'s, (first four columns) and the $\hat{\boldsymbol{\sigma}}_{\eta_{1,2,l}}$'s (final column) using $\mathcal{E}_2$ for component 2 The dashed dark blue lines indicate the true parameter value used to generate the data from M_1 as given in Table 2.

to be expected given that $\mathcal{E}_2$ has more than double the number of parameters than $\mathcal{E}_1$, and hence there will be more uncertainty in the estimates of those parameters. The third feature is that the distribution of parameters corresponding to the second component in the mixture estimate $\mathcal{E}_2$ are essentially the prior distribution of those parameters, because data were rarely assigned to that component. Indeed the maximum mixture weight attached for the second component of $\mathcal{E}_2$ for the whole simulation time period was 0.006.

The ability of $\mathcal{E}_2$ to discern that the data came from a single data generating process, and not a mixture, by assigning a probability of approximately zero to the second component in the mixture explains why $\mathcal{E}_2$ does not overfit. Box plots of $RMSE_{1kr}$ in Figure 4, reinforce these points; on average the estimator $\mathcal{E}_2$ performs as well as $\mathcal{E}_1$, however there is more variance in the performance of $\mathcal{E}_2$ than in $\mathcal{E}_1$.

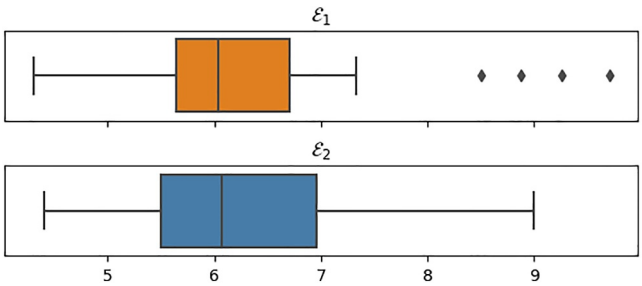

Figure 4. Boxplot of $RMSE_{1kl}$; Top: $k = 1$; Bottom: $k = 2$, The $\mathcal{E}_2$ performs as good as $\mathcal{E}_1$, even when the data is generated from M_1. However, the variability in the performance is increased which is to be expected with the additional number of parameters in M_2 than in M_1.

5.3. Results for Data Generated From M_2

This section presents the results obtained using the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ when data are generated from M_2, with Figure 5 analogous to Figure 3. The density estimates of the elements of $\boldsymbol{\Theta}_{2,2}$ in Figure 5 (rows 2,3,5,6 and 7), shows the ability of the mixture estimation technique to recover the true parameter estimates given in Table 2.

Figure 6 shows how the estimated posterior mean weights of each component of the mixture model changes over time for a slice of 50 days in the simulation period along with the corresponding covariate values.

The better predictive performance of the $\mathcal{E}_2$ compared to $\mathcal{E}_1$ is highlighted by the lower values of the $RMSE_{2kl}$'s in Figure 7.

5.4. Predictive Performance of $\mathcal{E}_1$ and $\mathcal{E}_2$ Using Simulated Data

Predictions from $\mathcal{E}_1$ and $\mathcal{E}_2$, for a single random realization of the data generated from M_1 and M_2, are given in Figure 8. The top row in Figure 8 indicates that when data are generated from M_1 the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ produce similar results. This is not surprising because the $\mathcal{E}_2$ attaches negligible weight to the second component (top right panel).

The bottom row in Figure 8 shows the better performance of $\mathcal{E}_2$ over $\mathcal{E}_1$ when the data are generated from M_2. The $\mathcal{E}_1$ prediction captures the timing of the peaks but over estimates the magnitude of the peak heights as shown in the bottom left panel of Figure 8. These points in time where the $\mathcal{E}_1$ over predicts are correctly estimated by the $\mathcal{E}_2$ (Figure 8: Bottom right panel) by switching between the model components (as seen by the transition of the dominant component from blue to red).

The over estimation of peak stage heights from the $\mathcal{E}_1$ is because the single component mixture assumes that the joint impact between upstream stage height and rainfall on downstream stage height is constant over time (and therefore the slope of the river bank is constant but $<\infty$) whereas the $\mathcal{E}_2$ estimate has the flexibility to account for the varying joint impact between upstream stage height and rainfall on downstream stage height (and therefore the slope of the river is not constant and can vary between 0 and ∞). The impact of rainfall on downstream stage height varies with upstream stage height, and this variation itself changes in the $\mathcal{E}_2$ estimate by switching between the components as seen in Figure 8, bottom right panel and thus avoids over estimating the peak stage heights. These observations are in agreement with a typical river cross section (see Figure 13 for the surveyed cross section for Bawden and Casino station) and how the stage height increments vary with the upstream condition as explained in Section 2.3, Figure 1. Models lacking the flexibility to account for varying joint impact of upstream stage height and rainfall on downstream stage height such as $\mathcal{E}_1$, will results in parameter estimates that predominantly reflect dependencies for the more frequently occurring lower flow regimes.

In the next Section we compare the performance of the simple and complex predictive models (M_1 and M_2) along with the original model proposed by Ossandon et al. (2021) (referred to as M_0) on real data.

6. Results for Real Catchments

We compute several metrics (Section 2.6) to compare the performance of the proposed BHME model (M_2) with its single component counterpart (M_1) and a baseline single component model originally proposed by Ossandon et al. (2021) (M_0)

- M_0: The model setup proposed by Ossandon et al. (2021).
- M_1: The single component model (when $k = 1$ in Equation 1).
- M_2: The two component BHME model (when $k = 2$ in Equation 1).

The BHME model outperforms the single component mixture model (M_1) and the baseline single component model originally proposed by Ossandon et al. (2021) (M_0) at all 3 stations when the predictive accuracy is assessed for out of sample predictions (Table 3). The performance improvement from M_0 to M_1 is significant, particularly due to the changes introduced in Section 2.2 bullet points 1–3. It is also worth noting that linear

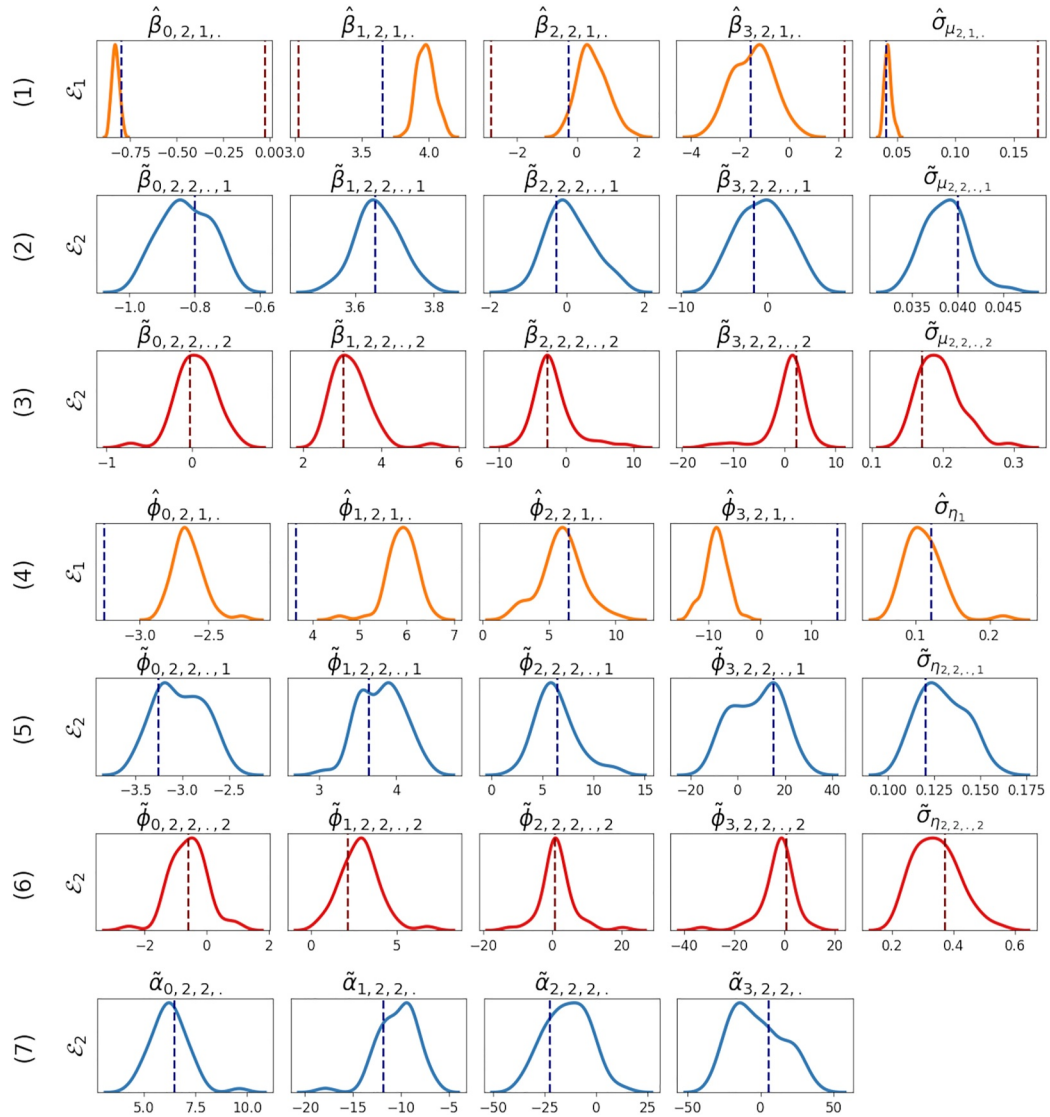

Figure 5. Plots of density estimates of the posterior means of parameters $\hat{\boldsymbol{\theta}}_{2,k,l}$ using $\mathcal{E}_1$ and $\mathcal{E}_2$ for data generated from M_2Top three rows (1–3): Density estimates of the posterior means constructed from the $\hat{\boldsymbol{\beta}}_{2,k,l}$'s, (first four columns) and the $\hat{\boldsymbol{\sigma}}_{\mu_{2,k,l}}$'s (final column). Row 1 is for estimate $\mathcal{E}_1$ while rows 2 and 3 are for estimate $\mathcal{E}_2$, components 1 and 2 respectively. Rows 4–6: Density estimates of the posterior means constructed from the $\hat{\boldsymbol{\phi}}_{2,k,l}$'s, (first four columns) and the $\hat{\boldsymbol{\sigma}}_{\eta_{2,k,l}}$'s (final column). Row 4 is for estimate $\mathcal{E}_1$ and rows 5 and 6 are for estimate $\mathcal{E}_2$, components 1 and 2 respectively. Last row (7): Density estimates of the posterior means constructed from the $\hat{\boldsymbol{\alpha}}_{2,2,l}$'s Given in orange color are the density estimates of posterior means of the parameters using $\mathcal{E}_1$ Given in blue and red colors are the density estimates of posterior means of the parameters using $\mathcal{E}_2$ for components 1 and 2 respectively. The dashed dark blue and red lines indicate the true parameter values used to generate the data as given in Table 2 from M_2, component 1 and 2 respectively.

correlation as expressed by R as a performance metric is less sensitive to the differences between models than the other performance metrics, possibly due to the impact of the peak stage heights on this statistic (Barber et al., 2020).

Accurate predictions of high flow events is particularly important when modeling stage height with relevance to flood events. The performance of each model was assessed for the three stations for stage heights above the 0.75 quantile and given in Table 4.

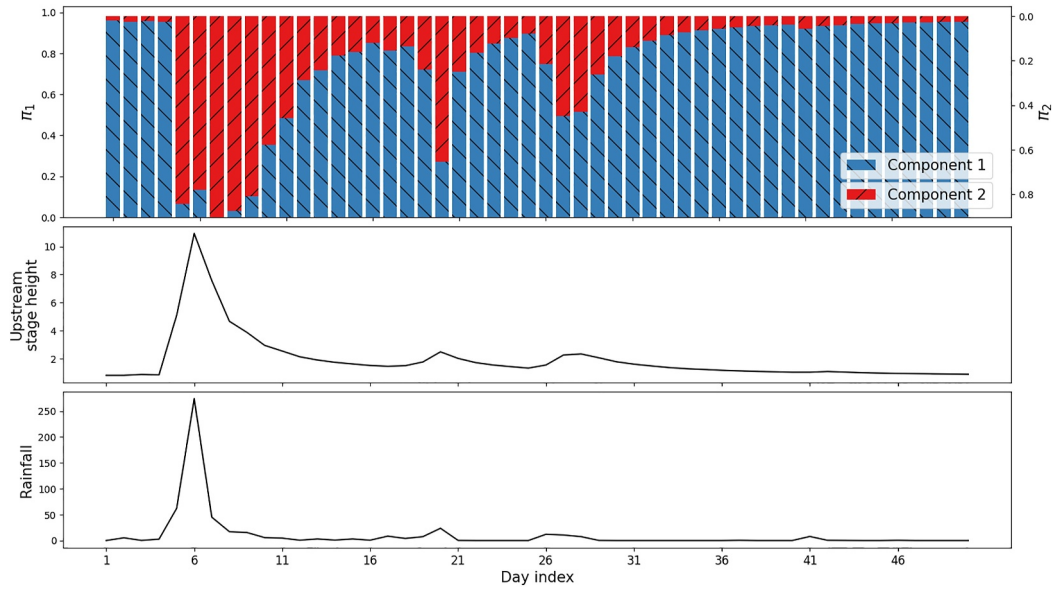

Figure 6. Top panel: Posterior means of mixture weights for component 1 and 2 in the estimate $\mathcal{E}_2$ given in blue and red respectively for a slice of 50 days Middle panel: The corresponding upstream stage height Bottom panel: The corresponding upstream rainfall The mixture weight for component 2 (π_2) increases with higher values of upstream stage height and rainfall. In the recession limbs of peak events, the model shifts gradually from component 2 dominant to component 1 dominant.

6.1. Predictions From the BHME Model

The posterior predictive mean of downstream stage height from the BHME model for the three stations considered along with the 95% credible interval is given in Figure 9. The preference of the BHME component 1 to predict low stage height events and component 2 for higher stage height events is evident. While in general, higher levels of uncertainty in predictions are attributed to higher stage height events, the widening of the credible interval during and around peak stage height events (e.g., the recession limbs of the peak stage height events for the Glenreagh station) denote instances where the model struggles to correctly identify the relevant mixture component during the transition from low/high to high/low events.

The separate components represent different parts of the downstream stage height distribution as can be seen in Figure 10 with low to mid values of stage heights primarily captured by the component 1 and high stage height events primarily captured by component 2. However, there is some blending between the different mixtures for the higher and lower stage heights, potentially indicating that the boundary between "high" and "low" flow processes is not clear cut. This separation of the stage height distribution comes from the changing dependency between downstream stage height and the upstream stage height and rainfall depth, modeled by the mixture weights as explained in Section 2.3.

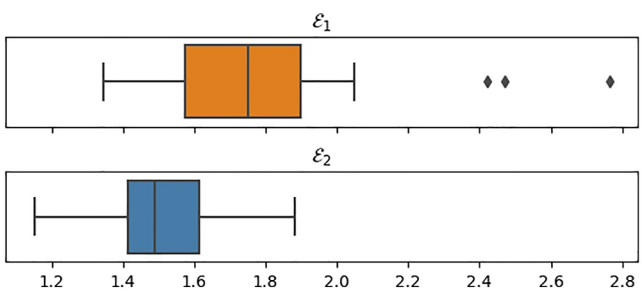

Figure 7. Boxplot of $RMSE_{2kr}$; Top: $k = 1$; Bottom: $k = 2$ The $\mathcal{E}_2$ performs significantly better than $\mathcal{E}_1$ which is to be expected as the data is generated from M_2.

Figure 11 presents how the BHME assigns weights to each component depending on the current catchment condition as predicted by the upstream stage height and rainfall depths. Linking back to the conceptualization explained in Section 2.3, the BHME assigns different weights to each component depending on the upstream conditions. For example, even if the upstream stage height is moderate (and hence the downstream stage height is also likely to be moderate), if the rainfall is high, component two is favored. Figure 12 shows how each of the components in the mixture predicts the downstream stage heights. Accounting for the varying river bank slopes, the mixture weights are assigned to reflect the varying impact of upstream rainfall on downstream stage height depending on the prevailing upstream stage height. The two contrasting gradients representing the difference in the rate of change of downstream stage height with upstream stage height and rainfall for the two components (Component 1 which is dominant in low flow conditions

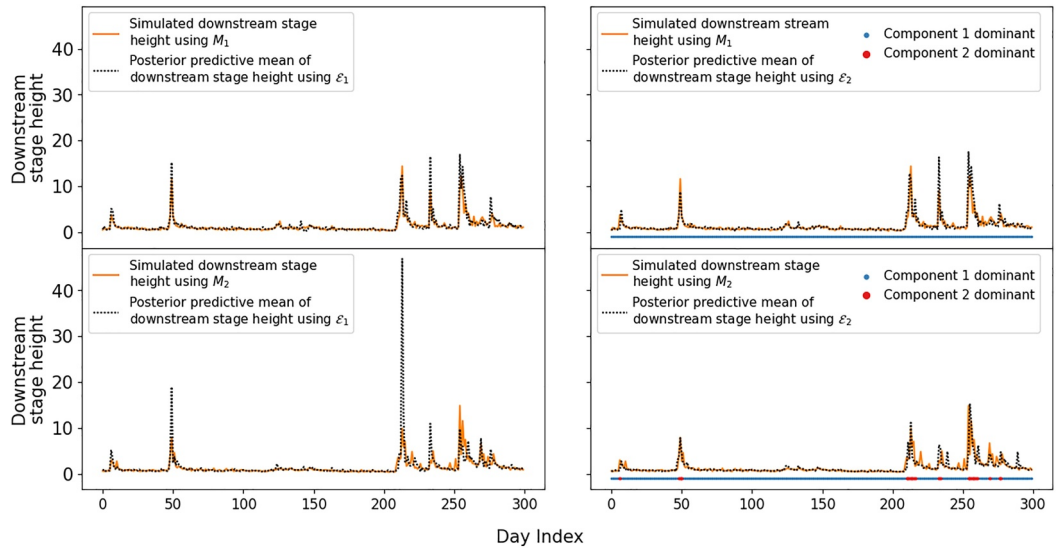

Figure 8. Posterior predictive mean stage height using $\mathcal{E}_1$ and $\mathcal{E}_2$ for a single realization of simulated data. Top row: Data generated from M_1. Bottom row: Data generated from M_2. Left column: Posterior predictive mean using $\mathcal{E}_1$. Right column: Posterior predictive mean using $\mathcal{E}_2$. The horizontal intervals denoted by the blue and red markers indicate the dominant component, where dominant is defined as $\pi_{1,t} > 0.5$, (see Equation 4). The dominance of component 1 is indicated by a blue marker while that of component 2 is indicated by a red marker.

has a higher rate of change of downstream stage height compared to that of component 2 which is dominant in high flow conditions) are characteristic of a typical river cross section as explained in the conceptualization of the model in Section 2.3.

The surveyed river cross sections by WaterNSW (https://realtimedata.waternsw.com.au/water.stm) for the Glenreagh and Casino stations are given in Figure 13. A partial trapezoid in dashed line is plotted as a reference to the conceptualized river cross section by a single component model as explained in Section 2.3. The difference between the surveyed river cross section and the single component model conceptualized cross section further validates the forth bullet point given in Section 2.2 about the potential for model misspecification in a single component mixture model in estimating the underlying true stage height distribution. The result of this model misspecification is that the peak stage heights are over estimated as seen in the simulation study (Bottom left panel in Figure 8).

Table 5 presents the regression parameters for μ_t from both the single component model (M_1) and the two component mixture model (M_2) for the Casino station. It also includes the covariate values for a single high stage height day and the corresponding predictions from each model. The observed downstream stage height for this particular day is 10.41. The overprediction of peak stage heights by the single component model is clear (Prediction by $M_1 = 29.03$), highlighting the necessity of the two-component mixture model (Prediction by $M_2 = 10.11$) to accurately estimate the true stage height distribution.

Table 3
Model Performance Metrics for Out of Sample Predictions of Downstream Stage Heights, $\mathbf{h}_2$: Models With Lower Values of MSE, and Higher Values of $\widehat{elppd}_{waic}$, NSE and R Are Preferred

	Bawden			Glenreagh			Casino		
Metric	M_0	M_1	M_2	M_0	M_1	M_2	M_0	M_1	M_2
$\widehat{elppd}_{\text{waic}}$	−2120.2	96.3	**293.2**	−1844.1	32.2	**351.9**	−1037.8	1920.9	**2055.0**
MSE	7.768	0.583	**0.448**	3.419	0.228	**0.184**	0.226	0.215	**0.089**
NSE	−1.250	0.831	**0.870**	−2.984	0.734	**0.786**	0.708	0.721	**0.885**
R	0.902	0.917	**0.933**	0.835	0.898	**0.902**	0.906	0.867	**0.942**

Note. Given in bold is the best value obtained out of the three models for each performance metric.

Table 4
Model Performance Metrics for Out of Sample Predictions of Downstream Stage Heights Above the 75th Quantile for Each Station

	Bawden			Glenreagh			Casino		
Metric	M_0	M_1	M_2	M_0	M_1	M_2	M_0	M_1	M_2
MSE	7.12	2.65	**1.69**	2.24	0.81	**0.63**	0.73	0.82	**0.36**
NSE	−3.12	0.58	**0.79**	−3.10	0.72	**0.73**	0.65	0.50	**0.80**

Note. Given in bold is the best value obtained out of the three models for each performance metric.

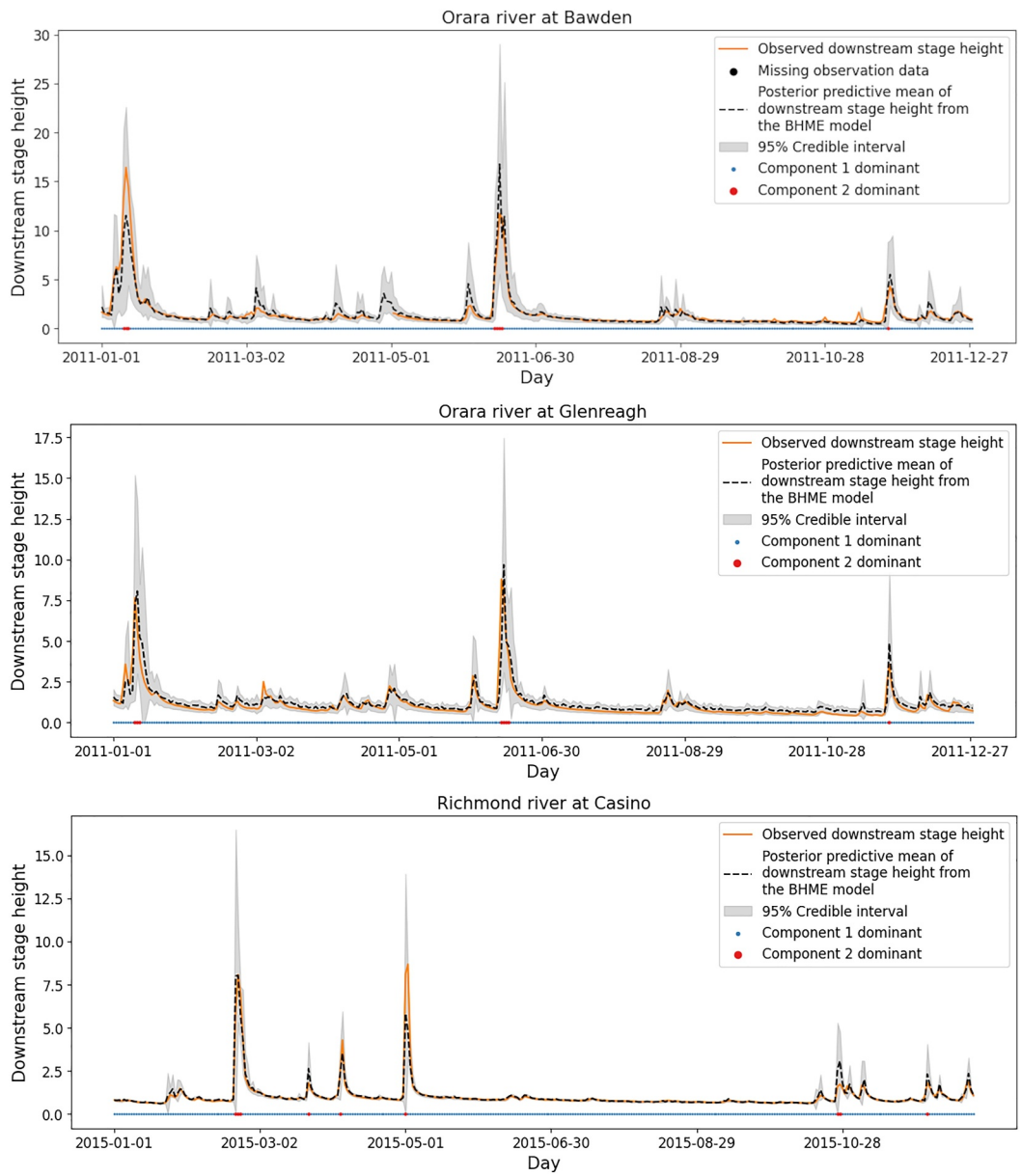

Figure 9. Out of sample posterior predictive means for downstream stage height from the BHME model for three stations Top panel: Orara river at Bawden, Middle panel: Orara river at Glenreagh, Bottom panel: Richmond river at Casino, The dominance of each component, component 1 in blue - for low stage height events and component 2 in red - for high stage events showcase the ability of the BHME to provide reliable prediction for both flow regimes.

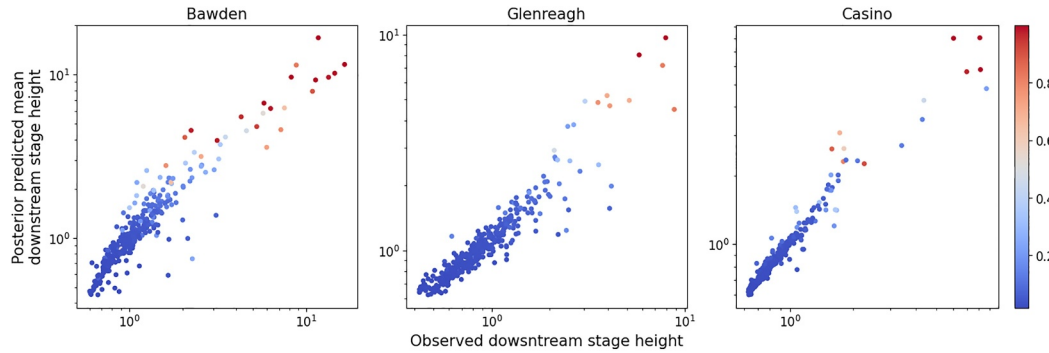

Figure 10. Posterior predicted mean downstream stage height versus observed downstream stage height in meters. Both axes are in log scale. The color scale represents the posterior mean of the weight assigned to component 2. It begins with blue, denoting zero weight allocated to component 2 (thus, component 1 is dominant), and transitions to red, indicating the maximum weight of one assigned to component 2 (thus, component 2 is dominant).

7. Discussion

The proposed Bayesian Hierarchical Mixture of Experts (BHME) model provides 1-day ahead forecasts of stage height, and is currently tested for Australian catchments. The motivation behind the proposed method is that catchment dynamics (which we define to be the varying dependencies between upstream conditions and downstream stage height) change over time (Saft et al., 2016) and therefore models which predict downstream stage height need to account for these changes and quantify uncertainty surrounding these predictions.

The BHME model's ability to accurately capture the timing of sudden peaks and estimate the magnitude of peak stage heights, as illustrated in Figure 9, is crucial in the context of Australian rivers. This flexibility of the our BHME to accommodate changes in catchment dynamics is particularly important because, as this paper demonstrates, a single component mixture model is insufficient to adequately represent the stage height distribution in a catchment. Single component hydrological models often fall short in capturing the complex streamflow response in semi-arid systems (Jiang et al., 2015). Similarly, this paper demonstrates that a single component mixture model is also insufficient to adequately represent the stage height distribution in a catchment. In particular a single component model fails to adequately represent less frequent and more extreme characteristics of the stage height distribution, as evidenced from the results of the simulation study, presented in Section 5, where it was found that the single component mixture model estimate, $\mathcal{E}_1$, over estimates the less frequent peak stream heights. This finding accords with other hydrological studies in literature (Buzacott et al., 2019; Coron et al., 2012; Shin & Kim, 2021). In contrast the mixture model, M_2, allows for more flexibility compared to M_1, and therefore is more robust to model misspecification in capturing the stage height distribution typical for semi-

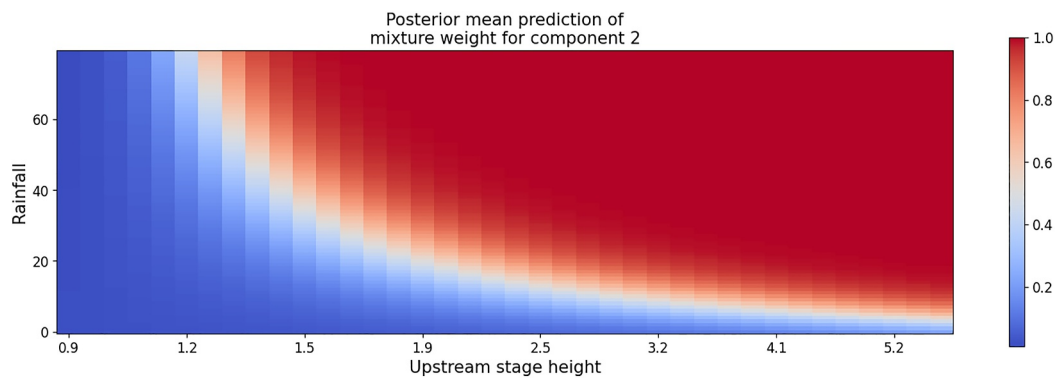

Figure 11. Posterior mean prediction of mixture weights for component 2 with upstream stage height and rainfall for Richmond river at Casino station. The *x* axis is in log scale. The color scale represents the weight assigned to component 2. It begins with blue, denoting zero weight allocated to component 2 (thus, component 1 is dominant), and transitions to red, indicating the maximum weight of one assigned to component 2 (thus, component 2 is dominant).

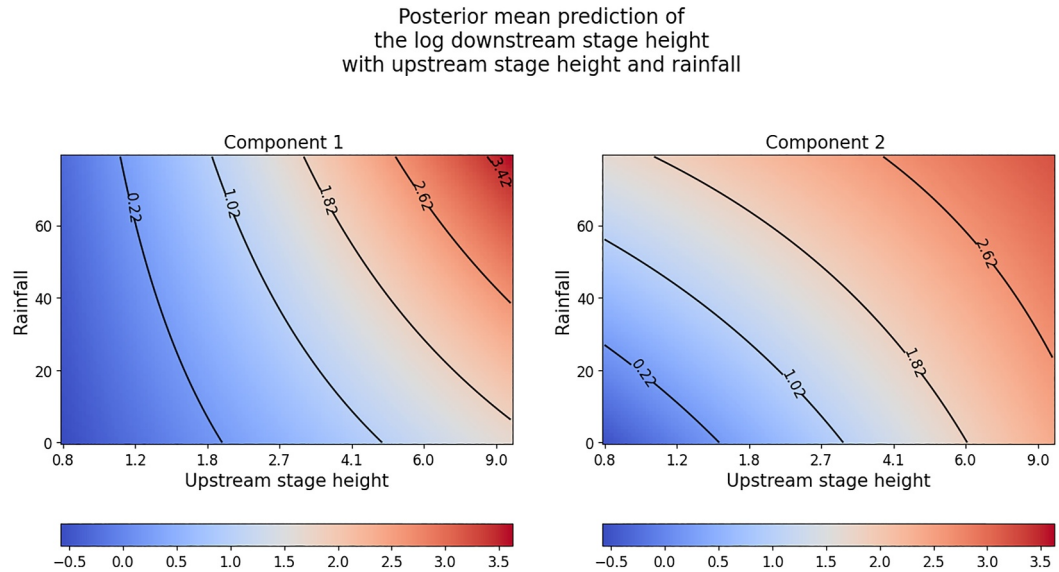

Figure 12. Posterior mean prediction of the log of mean downstream stage height from the two components for Richmond river at Casino, with upstream stage height and rainfall. The *x* axes are in log scale. The color indicates the log mean downstream stage height starting with blue for low values and transitions to red for high values. Contours drawn indicate similar log mean downstream stage heights Left: Component 1; Right: Component 2.

arid systems. Figure 14 shows how M_2 utilize this flexibility by assigning weights to each mixture component, capturing the changes is catchment dynamics over time for the Bawden station.

The specification of the mixture weights in the BHME model to depend on upstream conditions is important in the context of generating catchment specific hydrological knowledge through data. For example, the conditions of upstream stage height and rainfall which result in a high mixture weight for component two, which represents high stage height events (see Figure 11). Such insight from the mixture weights, indicating the preference for each component based on upstream catchment conditions, can potentially enhance the understanding of flood behavior specific to the catchment.

Furthermore, the mixtures may suggest different flow paths or processes within the catchment that generate peak flows or recessions. This presents an opportunity for further research, such as comparing with hydrogeochemical tracer data (Tunaley et al., 2017) to identify the catchment's specific hydrological signatures.The proposed BHME was developed with a focus on capturing the rapid changes in stage height commonly seen in Australia. For different regions with unique hydrological characteristics-like snowmelt-driven catchments or tropical catchments, the use of suitable covariates to inform the model of these processes would be required. The relevant covariates to be selected should be guided by the hydrological process knowledge about the catchment.

7.1. Probabilistic Forecasts for When "The Creeks Run Dry or Ten Feet High"

The proposed BHME model provides probabilistic stage height forecasts along with the posterior distributions of all the parameters. This is a key advantage when applied to water resource management and flood forecasting as both these applications benefit from uncertainty quantification for decision making (Herrera et al., 2022; Krzysztofowicz, 1999).

Single gamma distributions lack the flexibility to represent different parts of the daily stage height distribution, which is captured through the mixture representation as shown in Figure 15.

7.2. Limitations and Possible Extensions

We implement the proposed BHME model using two stations, upstream and downstream without bifurcation (a single stream separating into two or more streams) or confluence (opposite of bifurcation, two or more streams combine to form a single stream) between them. A possible extension of the proposed BHME, to account for confluence rivers would be to use the upstream stage height data from the merging tributaries in the design matrix.

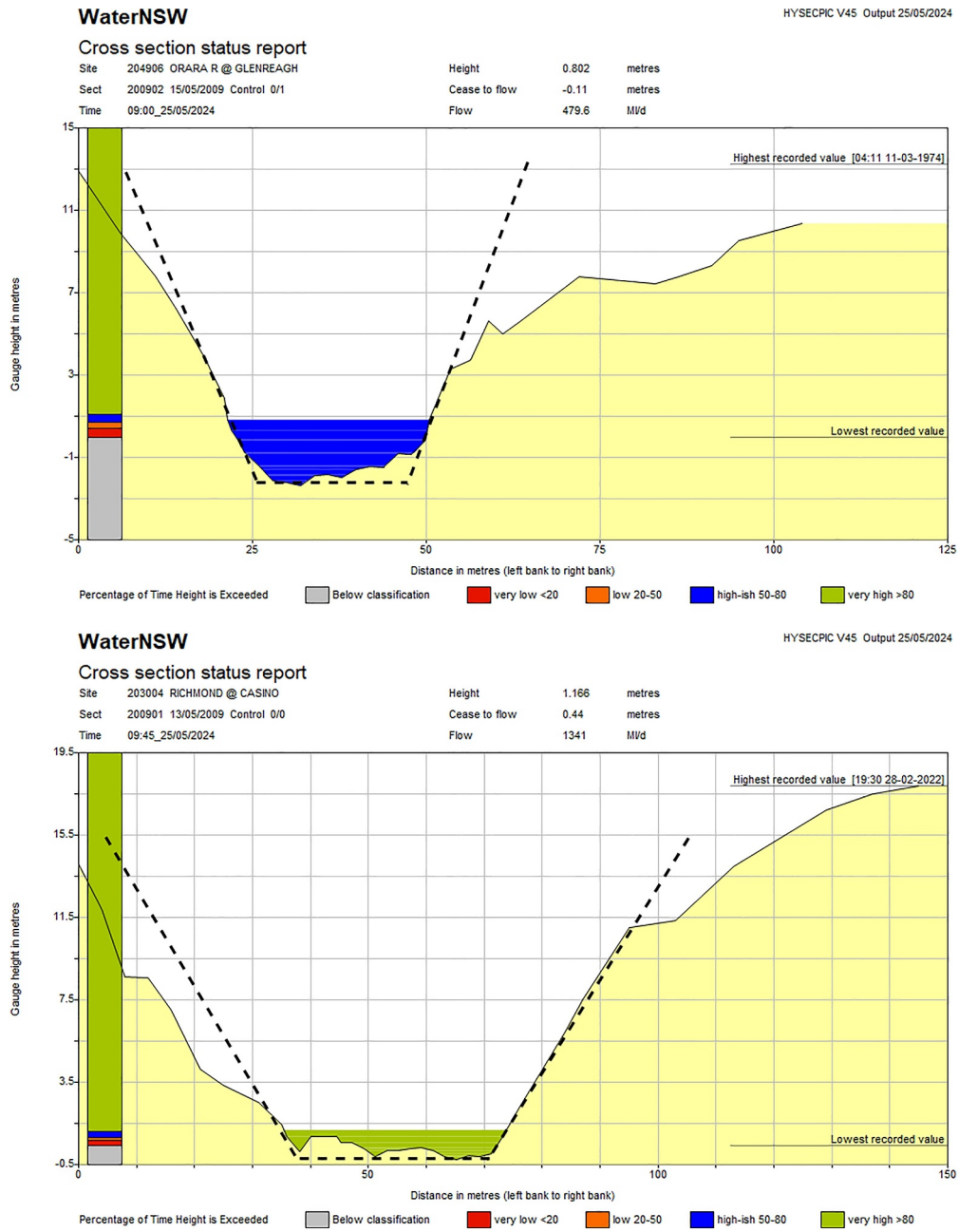

Figure 13. Cross sections of the gauging stations provided by WaterNSW (https://realtimedata.waternsw.com.au/water.stm). The dashed partial trapezoid provides reference to the conceptualization of the river cross section by a single component mixture (M_1) discussed in Section 2.3 Top: Orara river at Glenreagh station. Bottom: Richmond river at Casino station.

It is also important to note the time lag (time taken for the upstream conditions to impact downstream stage height) of the covariates used in this study. We demonstrate the methodology with a one day lag as the maximum cross correlation between the covariates and the downstream stage height was achieved with a 1 day lag (see Supporting Information S1 section). However, longer lags can easily be incorporated. The decision on what lags to be included for specific catchments will depend on the catchment properties and the distance to the closest upstream station, or can depend on the length of the forecast that is required. When expert knowledge on the suitable lag time is not available, a variable selection method (Villani et al., 2012) or a model selection method (Ghonchepour et al., 2021) based on information criteria using different time lags can be tested.

Table 5
Regression Parameters for μ_t From Both the Single Component Model (M_1) and the Two Component Mixture Model (M_2) for the Casino Station

Parameter	M_1	M_2 component 1	M_2 component 2	Min-Max scaled covariate
β_0 (intercept)	−0.68	−0.64	−0.47	1.00
β_1 ($\log(\mathbf{h}_1)$)	2.24	2.24	2.89	0.91
β_2 ($\mathbf{r}$)	0.50	0.43	2.06	0.43
β_3 ($\log(\mathbf{h}_1) \odot \mathbf{r}$)	3.59	2.16	−1.72	0.50
Prediction	29.03	10.11		Observed stage height = 10.41

Note. The covariate values for a single high streamflow day and the corresponding predictions from both models are given in the last row of the table.

The decision on the number of mixtures can be based on existing knowledge about the catchment specific flow path signatures, or using tracer data (Tunaley et al., 2017). The other approach is to identify the number of mixture components from the data (one example is to use an infinite mixture of experts (Ferguson, 1973; Neal, 1992)) when the focus is on the predictive skill. However, it is worth noting that the gain in predictive performance using such methods will be at the expense of interpretability.

Australian river systems have 2 main features in common. The first feature is the rapid change in stage height which we account for with the proposed BHME. The second feature is that about 70% of the rivers are non-perennial (i.e., rivers have no flow for at least some part of the year) (Shanafield et al., 2024). The proposed BHME model currently cannot account for zero stage height observations as the gamma density used in each component does not have support over zero. However, a possible extension of the current model is to add a separate distribution to represent the zero heights (ex: a point mass at zero) (Bertolacci et al., 2019) which extends the current mixture to three components instead of two.

8. Conclusion

Given the limited applications of Bayesian data driven models in hydrology, this paper extend this field by proposing a Bayesian Hierarchical Mixture of Experts model (BHME) with two mixture components to model stage height.

Compared to the Bayesian Hierarchical Network model proposed by Ossandon et al. (2021), the proposed BHME:

1. Models the measured stage height variable instead of streamflow which is estimated using a rating curve.

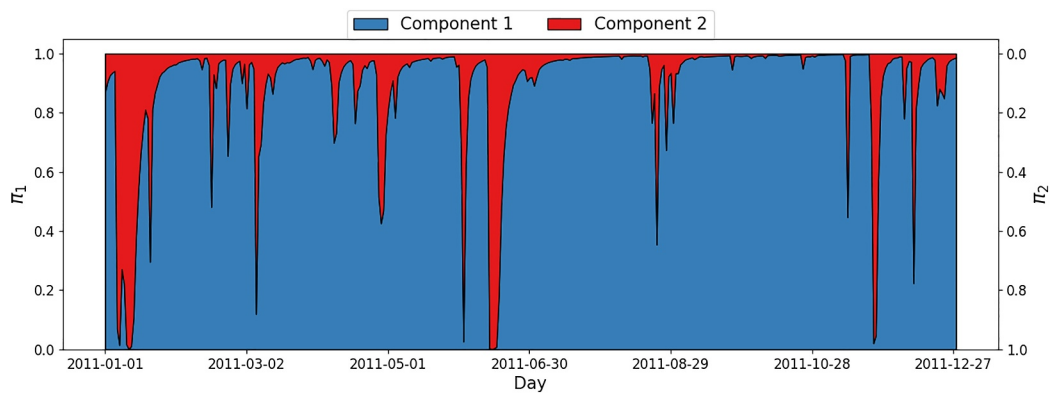

Figure 14. Posterior mean of mixture weights for Bawden station: The mixture weight for component 1 is given in blue while red represents the weight attached to component 2. The ability of the BHME to assign weights for each component as informed by the time varying covariates is presented.

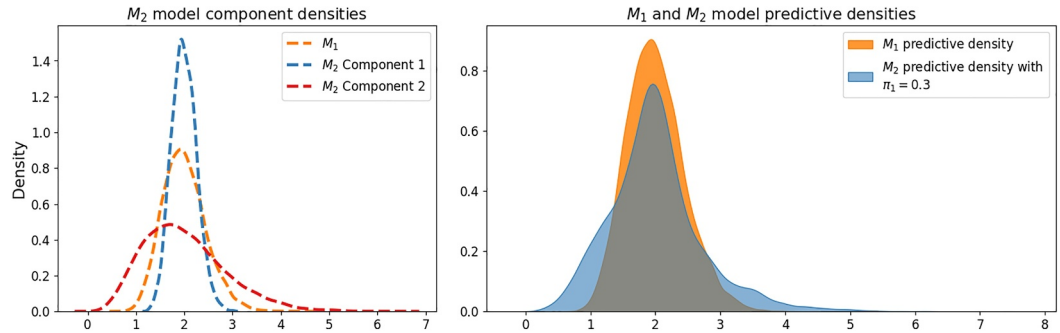

Figure 15. The predictive densities for a single day for Bawden station from the BHME model (M_2) and the single component model (M_1) are given in orange and blue respectively, on the right hand side plot. The individual component densities for M_2 which compose the mixture density as well as the single component density of M_1 are given on the left hand side plot. The weight attached to component 1 of M_2 on this day was 0.3.

2. Models the downstream stage height using a mixture of gamma distributions to offer greater flexibility in estimating the true underlying distribution and to mitigate the potential for model misspecification, which is more likely with a single component model.
3. Assumes the means and the standard deviations of the component gamma distributions are not deterministic functions of upstream stage height and rainfall but rather a noisy signal of these covariates.
4. Impose the positivity constraint on downstream stage height by assuming a power relationship between downstream stage height with upstream stage height and rainfall.
5. Includes an interaction term between the log upstream stage height and rainfall to account for the fact that, for a given volume of rainfall, both the river's height and width will increase, reducing the impact of rainfall on up/downstream stage height as the up/downstream stage height increases.

The BHME model outperforms the single component model and the earlier model proposed by Ossandon et al. (2021) across all the performance metrics evaluated in this study. In addition to providing better fits to the data, the BHME model generates useful catchment specific information that can help understand sudden changes and the variability in stage height commonly seen in Australia.

Data Availability Statement

The data along with the code implementation to reproduce the results presented in this paper can be accessed through Athukorala et al. (2025). Software and data–https://doi.org/10.5281/zenodo.16516280.

Acknowledgments
This research was supported by the Australian Government through the Australian Research Council's Industrial Transformation Research Program funding scheme through project IC190100031. Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

References

Allenby, G., Rossi, P., & McCulloch, R. (2005). Hierarchical Bayes models: A practitioners guide [Journal Article]. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.655541

Athukorala, R., Afshar, H., Cripps, S., & Vervoort, R. W. (2025). "Rajitha-athukorala/bhme_streamflow:" Where the creeks run dry or ten feet high: A probabilistic approach to stage height forecasting in Australia (BHME) [Dataset and Software]. *Zenodo*. https://doi.org/10.5281/zenodo.16516280

Barber, C., Lamontagne, J. R., & Vogel, R. M. (2020). Improved estimators of correlation and R2 for skewed hydrologic data [Journal Article]. *Hydrological Sciences Journal*, *65*(1), 87–101. https://doi.org/10.1080/02626667.2019.1686639

Bates, B. C., & Campbell, E. P. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling [Journal Article]. *Water Resources Research*, *37*(4), 937–947. https://doi.org/10.1029/2000wr900363

Bergström, S. (1992). *The HBV model-its structure and applications* (Report No. 02831104 (ISSN)). SMHI. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:smhi:diva-2672(2016-07-08T15:29:30.392+02:00

Bertolacci, M., Cripps, E., Rosen, O., Lau, J. W., & Cripps, S. (2019). Climate inference on daily rainfall across the Australian continent, 1876-2015 [Journal Article]. *Annals of Applied Statistics*, *13*(2), 683–712. https://doi.org/10.1214/18-AOAS1218

Beven, K. (2006). A manifesto for the equifinality thesis [Journal Article]. *Journal of Hydrology*, *320*(1), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., et al. (2019). Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, *20*(28), 1–28:6. Retrieved from http://jmlr.org/papers/v20/18-403.html

Boughton, W. (2004). The Australian water balance model [Journal Article]. *Environmental Modelling & Software*, *19*(10), 943–956. https://doi.org/10.1016/j.envsoft.2003.10.007

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., et al. (2018). JAX: Composable transformations of Python+NumPy programs. Retrieved from http://github.com/google/jax

Burrell, M., Petrovic, A. J., Ali, Nicholls, D., Ching, M., & Ooi, X. (2021). *General purpose water accounting report 2019-20: Namoi catchment (Report)*. NSW Department of Planning, Industry and Environment.

Buzacott, A. J. V., Tran, B., van Ogtrop, F. F., & Vervoort, R. W. (2019). Conceptual models and calibration performance–investigating catchment bias. *Water*, *11*(11), 2424. https://doi.org/10.3390/w11112424

Charbeneau, R. J. (1978). Comparison of the two- and three-parameter log normal distributions used in streamflow synthesis [Journal Article]. *Water Resources Research*, *14*(1), 149–150. https://doi.org/10.1029/wr014i001p00149

Chiew, F., McMahon, T., Zhou, S.-L., & Piechota, T. (2000). *Streamflow variability, seasonal forecasting and water resources systems [Book Section]* (pp. 409–428). Springer. https://doi.org/10.1007/978-94-015-9351-9_25

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept [Journal Article]. *Water Resources Research*, *51*(4), 2498–2514. https://doi.org/10.1002/2015wr017198

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models [Journal Article]. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2007WR006735

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments [Journal Article]. *Water Resources Research*, *48*(5). https://doi.org/10.1029/2011wr011721

Cripps, S., & Durrant-Whyte, H. (2023). Uncertainty: Nothing is more certain. *Environmetrics*, *34*(1), e2745. https://doi.org/10.1002/env.2745

Dal Molin, M., Kavetski, D., & Fenicia, F. (2021). Superflexpy 1.3.0: An open-source python framework for building, testing, and improving conceptual hydrological models [Journal Article]. *Geoscientific Model Development*, *14*(11), 7047–7072. https://doi.org/10.5194/gmd-14-7047-2021

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo [Journal Article]. *Physics Letters B*, *195*(2), 216–222. https://doi.org/10.1016/0370-2693(87)91197-X

Evin, G., Kavetski, D., Thyer, M., & Kuczera, G. (2013). Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration [Journal Article]. *Water Resources Research*, *49*(7), 4518–4524. https://doi.org/10.1002/wrcr.20284

Fasbender, D., & Ouarda, T. B. M. J. (2010). Spatial Bayesian model for statistical downscaling of AOGCM to minimum and maximum daily temperatures [Journal Article]. *Journal of Climate*, *23*(19), 5222–5242. https://doi.org/10.1175/2010jcli3415.1

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2008). Understanding catchment behavior through stepwise model concept improvement [Journal Article]. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2006wr005563

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. [Journal Article]. *Annals of Statistics*, *1*(2), 209–230. https://doi.org/10.1214/aos/1176342360

Gelfand, A. E., & Adrian, F. M. S. (1990). Sampling-based approaches to calculating marginal densities [Journal Article]. *Journal of the American Statistical Association*, *85*(410), 398–409. https://doi.org/10.2307/2289776

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images [Journal Article]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596

Ghonchepour, D., Sadoddin, A., Bahremand, A., Croke, B., Jakeman, A., & Salmanmahiny, A. (2021). A methodological framework for the hydrological model selection process in water resource management projects [Journal Article]. *Natural Resource Modeling*, *34*(3), e12326. https://doi.org/10.1111/nrm.12326

Green, E. J., Finley, A. O., & Strawderman, W. E. (2020). Introduction [Book Section]. In E. J. Green, A. O. Finley, & W. E. Strawderman (Eds.), *Introduction to Bayesian methods in ecology and natural resources* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-030-60750-0_1

Herrera, P. A., Marazuela, M. A., & Hofmann, T. (2022). Parameter estimation and uncertainty analysis in hydrological modeling [Journal Article]. *WIREs Water*, *9*(1), e1569. https://doi.org/10.1002/wat2.1569

Holsclaw, T., Greene, A. M., Robertson, A. W., & Smyth, P. (2017). Bayesian nonhomogeneous Markov models via polya-gamma data augmentation with applications to rainfall modeling [Journal Article]. *Annals of Applied Statistics*, *11*(1), 393–426. https://doi.org/10.1214/16-aoas1009

Homan, M. D., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo [Journal Article]. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Hrachowitz, M., & Clark, M. P. (2017). Hess opinions: The complementary merits of competing modelling philosophies in hydrology [Journal Article]. *Hydrology and Earth System Sciences*, *21*(8), 3953–3973. https://doi.org/10.5194/hess-21-3953-2017

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts [Journal Article]. *Neural Computation*, *3*(1), 79–87. https://doi.org/10.1162/neco.1991.3.1.79

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model [Journal Article]. *Water Resources Research*, *29*(8), 2637–2649. https://doi.org/10.1029/93wr00877

Jeremiah, E., Marshall, L., Sisson, S. A., & Sharma, A. (2013). Specifying a hierarchical mixture of experts for hydrologic modeling: Gating function variable selection [Journal Article]. *Water Resources Research*, *49*(5), 2926–2939. https://doi.org/10.1002/wrcr.20150

Jiang, Y., Liu, C., Li, X., Liu, L., & Wang, H. (2015). Rainfall-runoff modeling, parameter estimation and sensitivity analysis in a semiarid catchment [Journal Article]. *Environmental Modelling & Software*, *67*, 72–88. https://doi.org/10.1016/j.envsoft.2015.01.008

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods or graphical models [Journal Article]. *Machine Learning*, *37*(2), 183–233. https://doi.org/10.1023/a:1007665907178

Jordan, M. I., & Jacobs, R. A. (1993). *Hierarchical mixtures of experts and the EM algorithm [Conference Proceedings]*. IEEE. https://doi.org/10.1109/ijcnn.1993.716791

Kruschke, J. K. (2015). Chapter 6: Metric-predicted variable on one Ortwo groups [Book Section]. In J. K. Kruschke (Ed.), *Doing Bayesian data analysis* (2nd ed., pp. 454–456). Academic Press. https://doi.org/10.1016/B978-0-12-405888-0.00001-5

Krzysztofowicz, R. (1999). Bayesian theory of probabilistic forecasting via deterministic hydrologic model [Journal Article]. *Water Resources Research*, *35*(9), 2739–2750. https://doi.org/10.1029/1999wr900099

Kuczera, G., Kavetski, D., Franks, S., & Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters [Journal Article]. *Journal of Hydrology*, *331*(1), 161–177. https://doi.org/10.1016/j.jhydrol.2006.05.010

Langat, K., & Koech. (2019). Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow [Journal Article]. *Water*, *11*(4), 734. https://doi.org/10.3390/w11040734

Li, D., Marshall, L., Liang, Z., Sharma, A., & Zhou, Y. (2021). Characterizing distributed hydrological model residual errors using a probabilistic long short-term memory network [Journal Article]. *Journal of Hydrology*, *603*, 126888. https://doi.org/10.1016/j.jhydrol.2021.126888

Li, M., Wang, Q. J., Bennett, J. C., & Robertson, D. E. (2016). Error reduction and representation in stages (ERRIS) in hydrological modelling for ensemble streamflow forecasting [Journal Article]. *Hydrology and Earth System Sciences*, *20*(9), 3561–3579. https://doi.org/10.5194/hess-20-3561-2016

Marshall, L., Nott, D., & Sharma, A. (2007). Towards dynamic catchment modelling: A Bayesian hierarchical mixtures of experts framework [Journal Article]. *Hydrological Processes*, *21*(7), 847–861. https://doi.org/10.1002/hyp.6294

Masoudnia, S., & Ebrahimpour, R. (2014). Mixture of experts: A literature survey [Journal Article]. *Artificial Intelligence Review*, *42*(2), 275–293. https://doi.org/10.1007/s10462-012-9338-y

McInerney, D., Thyer, M., Kavetski, D., Bennett, B., Lerat, J., Gibbs, M., & Kuczera, G. (2018). A simplified approach to produce probabilistic hydrological model predictions [Journal Article]. *Environmental Modelling & Software*, *109*, 306–314. https://doi.org/10.1016/j.envsoft.2018.07.001

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors [Journal Article]. *Water Resources Research*, *53*(3), 2199–2239. https://doi.org/10.1002/2016WR019168

McMahon, T. A., & Peel, M. C. (2019). Uncertainty in stage-discharge rating curves: Application to Australian hydrologic reference stations data [Journal Article]. *Hydrological Sciences Journal*, *64*(3), 255–275. https://doi.org/10.1080/02626667.2019.1577555

Mehrnegar, N., Jones, O., Singer, M. B., Schumacher, M., Bates, P., & Forootan, E. (2020). Comparing global hydrological models and combining them with grace by dynamic model data averaging (DMDA) [Journal Article]. *Advances in Water Resources*, *138*, 103528. https://doi.org/10.1016/j.advwatres.2020.103528

Moges, E., Demissie, Y., Larsen, L., & Yassin, F. (2020). Review: Sources of hydrological model uncertainties and advances in their analysis [Journal Article]. *Water*, *13*(1), 28. https://doi.org/10.3390/w13010028

Moges, E., Demissie, Y., & Li, H. (2016). Hierarchical mixture of experts and diagnostic modeling approach to reduce hydrologic model structural uncertainty [Journal Article]. *Water Resources Research*, *52*(4), 2551–2570. https://doi.org/10.1002/2015wr018266

Mohammadi, B., Moazenzadeh, R., Christian, K., & Duan, Z. (2021). Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models [Journal Article]. *Environmental Science and Pollution Research*, *28*(46), 65752–65768. https://doi.org/10.1007/s11356-021-15563-1

Moramarco, T., Pandolfo, C., & Singh, V. P. (2008). Accuracy of kinematic wave approximation for flood routing. II. Unsteady analysis [Journal Article]. *Journal of Hydrologic Engineering*, *13*(11), 1089–1096. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:11(1089)

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A discussion of principles [Journal Article]. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Neal, R. M. (1992). *Bayesian mixture modeling [Book Section]* (pp. 197–211). Springer. https://doi.org/10.1007/978-94-017-2219-3_14

Ossandon, A., Rajagopalan, B., Lall, U., Nanditha, J. S., & Mishra, V. (2021). A Bayesian hierarchical network model for daily streamflow ensemble forecasting [Journal Article]. *Water Resources Research*, *57*(9), e2021WR029920. https://doi.org/10.1029/2021wr029920

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation [Journal Article]. *Journal of Hydrology*, *279*(1), 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7

Phan, D., Pradhan, N., & Jankowiak, M. (2019). Compostable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv*. https://doi.org/10.48550/ARXIV.1912.11554

Prieto, C., Kavetski, D., Le Vine, N., Á1lvarez, C., & Medina, R. (2021). Identification of dominant hydrological mechanisms using Bayesian inference, multiple statistical hypothesis testing, and flexible models [Journal Article]. *Water Resources Research*, *57*(8), e2020WR028338. https://doi.org/10.1029/2020wr028338

Prieto, C., Le Vine, N., Kavetski, D., Fenicia, F., Scheidegger, A., & Vitolo, C. (2022). An exploration of Bayesian identification of dominant hydrological mechanisms in ungauged catchments [Journal Article]. *Water Resources Research*, *58*(3), e2021WR030705. https://doi.org/10.1029/2021wr030705

Rastogi, N. K., Rajagopalan, B., & Ossandón, Ã. (2025). Bayesian hierarchical network model for forecasting daily river stage for rainfed river networks [Journal Article]. *Journal of Hydrology*, *654*, 132894. https://doi.org/10.1016/j.jhydrol.2025.132894

Ravindranath, A., Devineni, N., Lall, U., Cook, E. R., Pederson, G., Martin, J., & Woodhouse, C. (2019). Streamflow reconstruction in the upper Missouri river basin using a novel Bayesian network model [Journal Article]. *Water Resources Research*, *55*(9), 7694–7716. https://doi.org/10.1029/2019wr024901

Resilience, A. I. f. D. (2022). Major incidents report 2021-22 (Report). Retrieved from https://disasterphilanthropy.org/disasters/2022-australian-flooding/

Saft, M., Peel, M. C., Western, A. W., & Zhang, L. (2016). Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics [Journal Article]. *Water Resources Research*, *52*(12), 9290–9305. https://doi.org/10.1002/2016wr019525

Shanafield, M., Blanchette, M., Daly, E., Wells, N., Burrows, R. M., Korbel, K., et al. (2024). Australian non-perennial rivers: Global lessons and research opportunities [Journal Article]. *Journal of Hydrology*, *634*, 130939. https://doi.org/10.1016/j.jhydrol.2024.130939

Shao, Q., Lerat, J., Podger, G., & Dutta, D. (2014). Uncertainty estimation with bias-correction for flow series based on rating curve [Journal Article]. *Journal of Hydrology*, *510*, 137–152. https://doi.org/10.1016/j.jhydrol.2013.12.025

Shin, M.-J., & Kim, C.-S. (2021). Component combination test to investigate improvement of the IHACRES and GR4J rainfall-runoff models. *Water*, *13*(15), 2126. https://doi.org/10.3390/w13152126

Singh, V. P. (1998). *Gamma distribution [Book Section]* (pp. 202–230). Springer. https://doi.org/10.1007/978-94-017-1431-0_13

Smith, T., Marshall, L., & Sharma, A. (2015). Modeling residual hydrologic errors with Bayesian inference [Journal Article]. *Journal of Hydrology*, *528*, 29–37. https://doi.org/10.1016/j.jhydrol.2015.05.051

Tayfur, G., & Moramarco, T. (2022). Kinematic reverse flood routing in natural rivers using stage data [Journal Article]. *Applied Water Science*, *12*(8), 182. https://doi.org/10.1007/s13201-022-01707-2

Thiemann, M., Trosset, M., Gupta, H., & Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrological models [Journal Article]. *Water Resources Research*, *7*, 21–35.

Todorović, A., Grabs, T., & Teutschbein, C. (2024). Improving performance of bucket-type hydrological models in high latitudes with multi-model combination methods: Can we wring water from a stone? [Journal Article]. *Journal of Hydrology*, *632*, 130829. https://doi.org/10.1016/j.jhydrol.2024.130829

Tomkins, K. M. (2014). Uncertainty in streamflow rating curves: Methods, controls and consequences [Journal Article]. *Hydrological Processes*, *28*(3), 464–481. https://doi.org/10.1002/hyp.9567

Tunaley, C., Tetzlaff, D., Birkel, C., & Soulsby, C. (2017). Using high-resolution isotope data and alternative calibration strategies for a tracer-aided runoff model in a nested catchment [Journal Article]. *Hydrological Processes*, *31*(22), 3962–3978. https://doi.org/10.1002/hyp.11313

Ulzega, S., & Albert, C. (2022). Bayesian parameter inference in hydrological modelling using a Hamiltonian Monte Carlo approach with a stochastic rain model [Unpublished Work]. https://doi.org/10.5194/egusphere-2022-857

Van Kempen, G., Van Der Wiel, K., & Melsen, L. A. (2021). The impact of hydrological model structure on the simulation of extreme runoff events [Journal Article]. *Natural Hazards and Earth System Sciences*, *21*(3), 961–976. https://doi.org/10.5194/nhess-21-961-2021

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC [Journal Article]. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Víctor, P., Fuentes-Aguilera, P., & Enrique, M. (2018). Identifying advantages and drawbacks of two hydrological models based on a sensitivity analysis: A study in two Chilean watersheds [Journal Article]. *Hydrological Sciences Journal*, *63*(12), 1831–1843. https://doi.org/10.1080/02626667.2018.1538593

Villani, M., Kohn, R., & Nott, D. J. (2012). Generalized smooth finite mixtures [Journal Article]. *Journal of Econometrics*, *171*(2), 121–133. https://doi.org/10.1016/j.jeconom.2012.06.012

Wagener, T., & Wheater, H. S. (2006). Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty [Journal Article]. *Journal of Hydrology*, *320*(1–2), 132–154. https://doi.org/10.1016/j.jhydrol.2005.07.015

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory [Journal Article]. *ArXiv, abs/1004.2316*.

Westerberg, I. K., & Karlsen, R. H. (2024). Sharing perceptual models of uncertainty: On the use of soft information about discharge data [Journal Article]. *Hydrological Processes*, *38*(5), e15145. https://doi.org/10.1002/hyp.15145

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments [Journal Article]. *Water Resources Research*, *52*(3), 1847–1865. https://doi.org/10.1002/2015wr017635

Xiong, L., Shamseldin, A. Y., & O'Connor, K. M. (2001). A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system [Journal Article]. *Journal of Hydrology*, *245*(1), 196–217. https://doi.org/10.1016/S0022-1694(01)00349-3