26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Bayesian Inference for Model Construction in Hydrology

R.K. Niven [a], **N. Taghavi** [a,b], **L. Cordier** [c], **A. Mohammad-Djafari** [d], **M. Abel** [e], **M. Quade** [e]

[a] *School of Engineering and Technology, The University of New South Wales, Canberra ACT 2600, Australia.*
[b] *Sydney Water, Parramatta NSW 2150, Australia.*
[c] *Institut Pprime / CNRS, Poitiers, France.*
[d] *CentraleSupélec, Gif-sur-Yvette, France.*
[e] *Ambrosys GmbH, Potsdam, Germany.*
Email: r.niven@unsw.edu.au

Abstract:

Many phenomena in hydrological and environmental systems can be represented by the equation:

$$\mathbf{y} = f(\mathbf{x}) \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{y} \in \mathbb{R}^m$ is dependent variable vector and f is the system model, with $n, m \in \mathbb{N}$. Examples include: (i) Dynamical systems $\dot{\mathbf{x}} = f(\mathbf{x})$, where $\dot{\mathbf{x}}$ are derivatives of the state vectors $\mathbf{x}$, both of which can be functions of space and/or time; (ii) Groundwater pollution systems $\mathbf{C} = f(\mathbf{H})$, where $\mathbf{C}$ are pollutant concentrations and $\mathbf{H}$ are hydrological variables, both of which can be functions of space and/or time; and (iii) Rainfall-runoff systems $\mathbf{Q} = f(\mathbf{P})$, where $\mathbf{Q}$ is the runoff and $\mathbf{P}$ is the precipitation, which can be measured at a fixed location as a function of time, or over a catchment as functions of space and time. Commonly, f is unknown and must be identified from the data, referred to as an *inverse problem*.

Eq. (1) is generally observed in discrete form, based on time-series or spatial data [1]:

$$\mathbf{Y} = \mathbf{\Theta}(\mathbf{X})\mathbf{\Xi} \tag{2}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the state and dependent variable matrices, arranged as time or spatial series, $\mathbf{\Theta}(\mathbf{X})$ is a library matrix composed of functions of $\mathbf{X}$, and $\mathbf{\Xi}$ is a matrix of model coefficients. The inverse problem is how to infer $\mathbf{\Xi}$ from the time-series or spatial data. A wide assortment of heuristic methods have been developed to determine $\mathbf{\Xi}$, especially for hydrological systems. One group of methods involves regularization, based on minimization of the objective function [2]:

$$J(\mathbf{\Xi}) = ||\mathbf{Y} - \mathbf{\Theta}(\mathbf{X})\mathbf{\Xi}||_\beta^\alpha + \psi||\mathbf{\Xi}||_\delta^\gamma \tag{3}$$

where $||\mathbf{a}||_p = (\sum_i |a_i|^p)^{1/p}$ is the L_p norm of vector $\mathbf{a}$ with $p \in \mathbb{R}$, $\psi \in \mathbb{R}$ is the regularization coefficient and $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ are constants. Most regression or optimisation methods fit into the framework of (3). However, such methods suffer from many disadvantages, including (i) seemingly *ad hoc* choices of the regularisation term and its coefficient; (ii) a limited ability to rank models; and (iii) the lack of a deeper mathematical foundation, for example to enable uncertainty quantification.

We propose a Bayesian framework for solution of the inverse problem (2), based on Bayes' rule [2]. Bayesian inference provides a common platform for the analysis of all uncertainties, including inverse model identification, model ranking, hypothesis testing, estimation, interpolation, extrapolation, classification, experimental design or any other analysis of data [3]. It also offers many methodological advantages, including the selection of models or model parameters, the ability to rank models, the quantification of model uncertainties, and the estimation of unknown (nuisance) parameters. If desired, the user also has access to the full Bayesian apparatus for estimation of the posterior posterior probability density function (pdf) over the model space.

In the present study, we consider the maximum *a posteriori* (MAP) point estimator, which seeks the maximum posterior pdf $p(\mathbf{\Xi}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{\Xi})p(\mathbf{\Xi})$, where $p(\mathbf{Y}|\mathbf{\Xi})$ and $p(\mathbf{\Xi})$ are the likelihood and prior pdfs. This method provides a Bayesian generalization of Tikhonov regularization [4], in which the likelihood corresponds to the residual term and the prior corresponds to the regularization term. The Bayesian framework is demonstrated by several applications, including (ii) several dynamical systems with added noise [3], and (ii) the analysis of groundwater vulnerability to nitrate pollution in the Burdekin Basin, Queensland, Australia [5]. These

applications demonstrate the advantages of the Bayesian framework for the solution of inverse problems in hydrology, especially the ability to select and rank models, and (for dynamical systems) the ability to quantify the uncertainties of the estimated model coefficients.

References

[1] Brunton, S.L., Proctor, J.L. & Kutz, J.N., Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *PNAS* **113**(15), 3932-3937 (2016).

[2] Bayes, T., An essay towards solving a problem in the doctrine of chance, *Phil. Trans. Royal Soc. London* **53**, 370-418 (1763).

[3] Niven, R.K., Mohammad-Djafari, A., Cordier, L., Abel, M. & Quade, M., Dynamical system identification, model selection and model uncertainty quantification by Bayesian inference, Chaos, **34**: 083140 (2024).

[4] Tikhonov, A.N., Solution of incorrectly formulated problems and the regularization method, *Doklady Akademii Nauk SSSR* **151**, 501-504 (1963).

[5] Taghavi, N., Niven, R.K., Kramer, M. & Paull, D.J., Enhancing groundwater vulnerability assessment through Bayesian inference, Journal of Hydrology, **653**: 132781 (2025).

Keywords: *Bayesian inference, inverse problem, model identification, model ranking, dynamical system, groundwater vulnerability*

