Uncertainty estimation for river system models

Shaun S H Kim [a], **Justin D Hughes** [a], **Julien Lerat** [a], **Dom Regan-Beasley** [b]

[a] CSIRO, Environment, Canberra, ACT, Australia, [b] NSW Department of Climate Change, Energy, the Environment and Water, Parramatta, NSW, Australia
Email: shaun.kim@csiro.au

Abstract: The fundamental design and scale of river system models presents a unique challenge for uncertainty estimation. River system models are complicated networks of nodes and links that represent rainfall runoff, routing, bifurcations, diversions and transmission losses. Observed streamflow data in Australian basins are commonly obtained from many gauging stations distributed across the river system. Thus, river system models essentially consist of a combination of many models, often different types (e.g., rainfall runoff, routing), and contain numerous observed streamflow sites.

Calibrations are usually achieved via a reach-by-reach approach, a stepwise process which breaks the system into manageable sections delineated by the streamflow gauging sites. This avoids the high-dimensionality issues that may arise when using a system calibration approach (Hughes et al., 2016). Regardless of calibration approach, due to the size and scale of many Australian basins, common parameter uncertainty analysis algorithms, such as Markov Chain Monte Carlo (Gelman et al., 1995), are impractical to implement.

The current study attempts to develop a simple yet rigorous method to account for uncertainty in river system models. The method is designed for straightforward application to existing calibrated river system models as a post-processing step. The method statistically characterises and simulates streamflow errors and is called the Conditional Discrete Non-Parametric (CDNP) approach. As the name implies, the approach requires no estimation of parameters and utilises discrete mathematics to conveniently derive conditional probabilities. The final aim of CDNP is to produce many viable error time series for different points in a river system. CDNP only requires simulation data and residual error time series. CDNP errors can be generated within calibrated river system models at different locations, allowing the production of simulation ensembles. CDNP would be especially valuable when applied through platforms like eWater Source.

The CDNP functions including help documentation are publicly available as an R package at https://github.com/shaunkim079/CDNP. The CDNP functions first divide the range of residual errors, ε, into n bins between certain intervals. Furthermore, the previous day errors, ε_{t-1}, are coupled with current day simulated streamflow values, q_t, and classified into k groups using a k-means clustering algorithm (Hartigan and Wong, 1979). Conditioning on ε_{t-1} accounts for error autocorrelation, while q_t addresses error heteroscedasticity. Grouping variables into bins and clusters allows computation of $P(b_i|c_j)$ for each residual bin, b_i, and each cluster, c_j. This provides a lookup table of probabilities for each bin and cluster combination.

For error simulation at $t = 1$, $\varepsilon_{t=0}$ and $q_{t=1}$ is assumed given. The cluster these belong to is determined by the closest cluster centroid by Euclidean distance. Then, a residual error bin is randomly sampled but weighted based on the lookup table probabilities for the given cluster. The final simulated error is randomly sampled from the original residual error values that reside in the selected bin. This process starts again with $\varepsilon_{t=1}$ and $q_{t=2}$ for the subsequent time step.

Results from the Border Rivers and Gwydir catchments show that the method provides a simple and robust means of estimating uncertainty without the strict requirement of recalibration. Furthermore, median daily streamflow predictions from CDNP perform better than deterministic simulations in a range of fit metrics including KGE and bias.

REFERENCES

Gelman A, Carlin JB, Stern HS & Rubin DB (1995) Bayesian data analysis. Chapman and Hall/CRC.

Hartigan JA & Wong MA (1979) Algorithm AS 136: A K-means clustering algorithm. Applied Statistics 28, 100-108. doi:10.2307/2346830.

Hughes JD, Kim SSH, Dutta D & Vaze J (2016) Optimization of a multiple gauge, regulated river-system model. A system approach. Hydrological Processes, 30(12), 1955-1967.

Keywords: *Uncertainty estimation, river system modelling, CDNP, post-process*