A method to calibrate daily rainfall-runoff models to monthly streamflow data

Julien Lerat [a] , **Dmitri Kavetski** [b] , **David McInerney** [b] and **Mark Thyer** [b]

[a] CSIRO Land & Water, [b] School of Civil, Environmental and Mining Engineering, University of Adelaide, South Australia
Email: julien.lerat@csiro.au

Abstract: Conceptual hydrological models that predict streamflow at daily time steps are widely used in water forecasting, water resources planning and operations. Typically, these models are calibrated using daily observed streamflow data. However, in practice there are several common circumstances where observed data is not available at a daily resolution. In particular, a key practical application motivating this work is the operation of large dams. In large dam operation, inflows are typically estimated from (known) dam outflow and water levels by methods such as inverse pool routing. These methods are generally stable at monthly time steps but can become unstable at shorter time steps (e.g. daily) due to the uncertainty in water level data and the storage/volume relationships. Development of robust approaches for calibrating daily rainfall-runoff models to monthly streamflow data is hence of major practical interest.

This study describes an empirical analysis of 508 Australian catchments over two evaluation periods using the GR4J daily hydrological model. It compares the performance attained after calibrating the model to daily streamflow data versus monthly data. Multiple performance metrics are used: fit of the daily and monthly flow duration curve, daily pattern metrics (NSE, correlation, peak timing errors), and long-term bias.

A key finding is that monthly calibration schemes reach comparable performance to daily calibrations for performance metrics such as the fit of daily flow duration curves in low, medium and high flows, with a majority of sites and periods reaching similar or better metric values. This finding holds despite monthly calibration having no access to daily streamflow data. On the other hand, and as expected, monthly calibration performs generally worse than daily calibration for metrics related to daily patterns of the hydrograph, such as NSE, correlation and peak time errors.

However, monthly calibration can degrade the accuracy of estimated timing of flood peaks and, more generally, daily hydrograph patterns. Our results indicate that this limitation can be alleviated by estimating those parameter(s) controlling hydrograph timing using regionalisation from nearby catchments.

Overall, the empirical findings indicate that monthly calibration is a viable alternative to daily calibration when no daily flow data is available. Importantly, the findings hold over a large sample of Australian catchments. Future work on extending the findings to broader classes of rainfall-runoff models is recommended.

Keywords: *Streamflow prediction, rainfall-runoff models, model calibration, monthly streamflow*

1. INTRODUCTION

Conceptual hydrological models that predict catchment streamflow are widely used in water forecasting, water resources planning and operations. These models generally require calibration to observed data (streamflow time series and climate inputs) at the catchment of interest. Typically, the calibration data has the same temporal resolution (time step) as the intended model predictions. However, observed data are not always available with such resolution. For example, in many situations of operational interest, streamflow predictions are required at a daily time step (for given climate forcings) but observed streamflow data is available at a monthly time step only.

A key practical application motivating this work is the operation of large dams for which inflows are typically estimated from (known) dam outflows and water levels by methods such as inverse pool routing (D'Oria et al., 2012). These methods are generally stable at monthly time steps but can become unstable at shorter time steps (e.g. daily) due to uncertainty in water level data and the storage/volume relationships (Deng et al., 2015).

Relatively few studies have evaluated the daily performance of the same model calibrated at monthly vs daily time steps. For example, Sudheer et al. (2007) and Adla et al. (2019) reported that monthly calibration produced unrealistic daily simulations. However, their analysis was limited to two catchments and a single hydrological model (SWAT). Calibration to data at aggregated temporal scales are expected to lose timing information, and therefore impact on model parameters associated with quick flow processes (e.g., Kavetski et al., 2011). Potentially, limitations in parameter identifiability could be overcome by parameter regionalisation; e.g., see Rojas-Serna et al. (2016), which however focused solely on streamflow predictions the daily scale.

Recent work by Lerat et al. (2020) investigated daily versus monthly calibration of a daily rainfall-runoff model (GR4J) over a large sample of catchments, considering a range of objective functions and a wide range of performance metrics for both daily and monthly predictions. Compared to Lerat et al. (2020), this paper provides a simplified set of recommendations for modellers wishing to calibrate daily rainfall-runoff models when streamflow data are available at monthly time step only. This paper focusses on the performance of the model assessed against daily streamflow data (after monthly calibration), which is of most relevance to the operation of dams and related applications. More specifically, we consider the following aims:

- Compare daily streamflow predictions obtained from a daily-step model calibrated against monthly streamflow data to daily predictions obtained from the same daily-step model calibrated against daily streamflow data ("idealised" setup traditionally used in the literature).
- Mitigate the reduced information about flow timing using a parameter regionalisation strategy.

This study employs the hydrological model GR4J, which is widely used in streamflow prediction applications in Australia and worldwide (Perrin et al., 2003; Woldemeskel et al., 2018).

2. THEORY

Consider a hydrological model that simulates daily streamflow as a function of forcing data and parameters, (as well as initial conditions, which in this work we consider to be fixed). This daily model can be calibrated at the daily or monthly scale, as described next.

2.1. Daily calibration

In daily-scale calibration, the model parameter set $\boldsymbol{\theta}$ is estimated using an objective function $F(\boldsymbol{Q}, \boldsymbol{Q}^{\theta})$, formulated in terms of daily flows,

$$\boldsymbol{\theta}^{(\text{opt})} = \text{argmin}_{\boldsymbol{\theta}} \{ F(\boldsymbol{Q}, \boldsymbol{Q}^{\theta}) \} \tag{1}$$

where $\boldsymbol{Q}$ is the time series of observed daily streamflow and $\boldsymbol{Q}^{\theta}$ is the corresponding time series of daily flow simulated by the model with parameter set $\boldsymbol{\theta}$.

2.2. Monthly calibration

For the monthly-scale calibration, model parameters are estimated using a "monthly scale" objective function, formulated in terms of monthly flows

$$\boldsymbol{\theta}^{(\text{opt})} = \text{argmin}_{\boldsymbol{\theta}} \{ F(\boldsymbol{q}, \boldsymbol{q}^{\theta}) \} \tag{2}$$

where $\boldsymbol{q}$ are the time series of observed monthly streamflow and $\boldsymbol{q}^{\theta}$ are the corresponding time series of monthly streamflow computed by aggregating daily flows.

In order to support a meaningful evaluation of results, here the monthly observed streamflow is obtained by aggregating daily observed streamflow. In most practical applications of interest, monthly time series will be available directly – and indeed daily time series will not be available.

2.3. Objective function and optimisation

The specification of the objective function F can have considerable impact on model calibration. In this work the objective function is the sum of squared errors (SSE),

$$F(\boldsymbol{Y}, \boldsymbol{Y}^\theta) = \Sigma_j \left[\; z(Y_j^\theta; \lambda, A) - z(Y_j ; \lambda, A) \; \right] \tag{3}$$

Like any objective function, the SSE can be computed for either daily or monthly time series, hence the general notation Y is used for the time series. The function z denotes the Box-Cox transformation, with power parameter $\lambda = 0.2$ and offset $A = 0.01$ which follows the recommendations of McInerney et al. (2017). The Box-Cox transformation tends to stabilise errors with respect to the streamflow magnitude, which in turns is useful for parameter estimation. A wider range of objective functions were explored by Lerat et al. (2020).

The optimisation of the SSE is carried out using a hybrid method in two stages: 1. initial seeding of hydrological model parameter sets according to available prior knowledge, and 2. downhill simplex method applied to the best candidates. In our experience this approach improves global convergence (Lerat et al., 2020).

2.4. Combination of calibration and regionalisation

Previous work shows that parameters associated with quick flow processes are the most sensitive to the time scale of calibration data and can be poorly identifiable from aggregated data (e.g., Kavetski et al., 2011). This concern is particularly pertinent for monthly calibration of a daily model. Hence, quick flow parameters are estimated using a kriging regionalisation model (Merz & Blöschl, 2004). The regionalisation approach is detailed in Appendix A of Lerat et al. (2020).

Specifically, when regionalisation is used, the set of model parameters is partitioned in two, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_c, \boldsymbol{\theta}_r\}$, where $\boldsymbol{\theta}_c$ is the subset of parameters estimated by calibration and $\boldsymbol{\theta}_r$ is the subset of parameters estimated by regionalisation. As noted above, the subset $\boldsymbol{\theta}_r$ includes parameters known *a priori* to control timing of the hydrograph. For commonly used models, including the GR4J model used in this work, the parameters controlling quick flow (and hence hydrograph timing) are well known; see section 3.2.

To avoid artificial skill being introduced, parameter regionalisation is implemented using a "leave-one-out" validation approach (Lerat et al., 2020). Regionalised parameters are estimated using a kriging model calibrated to all catchments, except for the target catchment.

3. METHODS

3.1. Study Catchments

This study employs 508 catchments, covering all climate regions across Australia, with locations shown in Figure 1 (Lerat et al., 2020).

3.2. Hydrological Model

This study employs GR4J, a conceptual rainfall-runoff model which simulates daily catchment-scale streamflow time series from daily rainfall and PET inputs (Perrin et al., 2003). The GR4J model has a total of four free parameters that require estimation. The parameter θ_4 is known to control the timing of the hydrograph. Therefore, the regionalisation procedure described in the earlier section was applied to estimate this parameter.
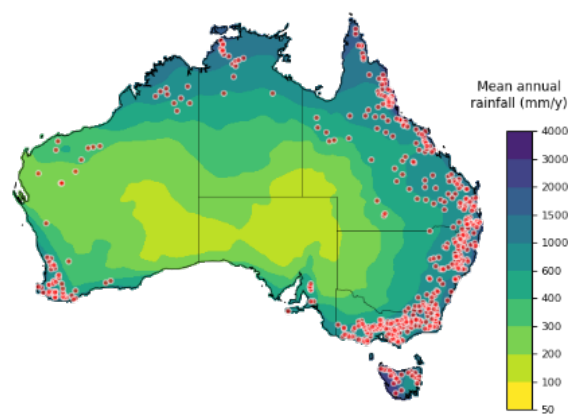
Figure 1. Locations of the 508 case study catchments. Figure adapted from Lerat et al. (2020).

3.3. Calibration schemes

Three calibration schemes are considered in this paper: a daily calibration as detailed in section 2.2, a monthly calibration as detailed in section 2.3 where all four parameters are calibrated, and a monthly calibration where three parameters are calibrated, and θ_4 is regionalised. This last scheme is referred to as "monthly- θ_4".

3.4. Performance Metrics for Daily Streamflow predictions

In this study we focus on the following five performance metrics computed from daily streamflow predictions.

Bias in long-term average streamflow: the following simple metric is used.

$$Bias = \left| 1 - \frac{mean(\boldsymbol{Y}^\theta)}{mean(\boldsymbol{Y})} \right| \tag{4}$$

where mean($\boldsymbol{Y}$) denotes the mean of vector $\boldsymbol{Y}$.

Flow duration curve (FDC) error: the mean absolute relative error is defined as

$$FDCfit_{p_1-p_2} = \frac{1}{N_p} \sum_{k=1}^{N_p} \left| 1 - \frac{Pct(\boldsymbol{Y}^\theta, p_k)}{Pct(\boldsymbol{Y}, p_k)} \right| \tag{5}$$

where $Pct(\boldsymbol{Y}, p_k)$ is the p'th percentile of series $\boldsymbol{Y}$, p_1 and p_2 are the lowest and highest percentiles considered, N_p is the number of percentile interpolation points, and $p_k = p_1 + (p_2 - p_1)(k - 1) / N_p$ is the linearly interpolated percentile. The FDC fit metric is similar to the FDC diagnostic metric used by Yilmaz et al. (2008), but is restricted to a portion of the curve. We apply this metric to three ranges of percentiles: low flows defined by the range of 0% to 10% (i.e. $p_1 = 0$ and $p_2 = 10$), medium flows in the range of 30% to 70% (i.e. $p_1 = 30$ and $p_1 = 70$), and high flows defined 90% to 100% (i.e. $p_1 = 90$ and $p_2 = 100$).

Pattern matching: Spearman rank correlation is used to compare hydrograph patterns,

$$Corr = \frac{\sum_j \left(R_j^\theta - mean(\boldsymbol{R}^\theta) \right) \left(R_j - mean(\boldsymbol{R}) \right)}{sdev(\boldsymbol{R}^\theta)\, sdev(\boldsymbol{R})} \tag{5}$$

where R_j and R_j^θ are the respective ranks of observed and simulated streamflow values at day j.

Timing of daily peaks: the "error in timing" metric introduced by Ficchì et al. (2016) is used,

$$T_{err} = \frac{1}{N_a} \sum_{a=1}^{N_a} \left| t_a - \underset{t \in [t_a-5, t_a+5]}{argmax} \left(Q_t^\theta \right) \right| \tag{6}$$

where N_a is the number of years simulated, t_a is the day at which the observed peak flows of water year a occurs. The simulated peak time is identified within a window of 5 days before and after t_a to minimise the risk of selecting a peak from another flood event.

Overall accuracy: the Nash-Sutcliffe efficiency (NSE) is used

$$NSE = 1 - \frac{\sum_j \left(Y_j^\theta - Y_j \right)^2}{\sum_j \left(mean(\boldsymbol{Y}) - Y_j \right)^2} \tag{7}$$

All these metrics are intended to assess daily flows. A broader study that also includes an assessment of monthly streamflow predictions is provided by Lerat et al. (2020).

3.5. Cross-validation

We implemented a temporal split-sample validation procedure, with available data split into two periods: 1975-1992 and 1992-2015. Calibration is performed on each period. The model is then run on the other period to compute performance (validation) metrics. For a given catchment, this procedure yields two sets of validation periods, two corresponding parameter sets, and two sets of validation metrics.

3.6. Performance differences

A monthly calibration scheme (e.g. monthly calibration) is assessed by considering the proportion of sites and periods where its performance metrics is classified as "worse", "similar" or "better" in comparison to daily calibration. These proportions are computed over the 508 catchments and 2 validation periods, i.e., over a total of 1016 sites-period combinations. Metric values are considered "similar" if they differ by less than a "practical significance" threshold Δ. The choice of Δ is subjective and context-specific. In this work, we set it to 0.05 for all metrics, based on previous experience of the authors.

4. RESULTS

Figure 2 shows the proportion of sites/periods where the performance of monthly calibration is "worse than", "similar to", and "better than", daily calibration.
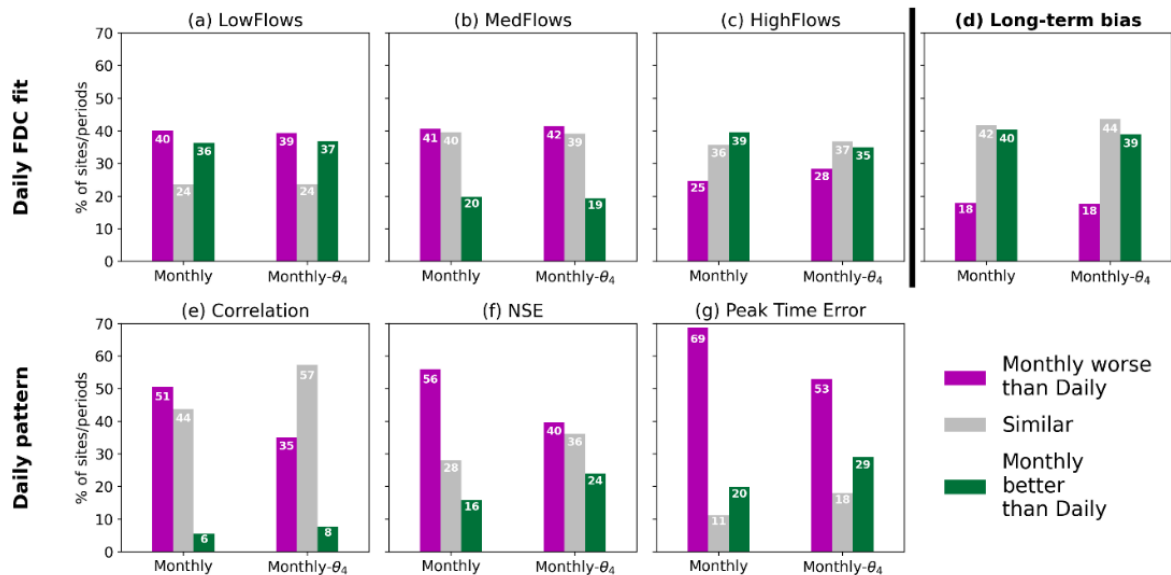
Figure 2. Proportion of sites/periods where monthly calibration is respectively "worse that", "similar to", and "better than" daily calibration. The numbers within each bar indicate the percentage of sites/period in the corresponding category. Figure adapted from Lerat et al. (2020).

4.1. Monthly versus Daily Calibration

Monthly calibration with no parameter regionalisation (labelled "monthly") performs similarly or better than daily calibration for daily FDC fit metrics in a majority of sites and periods (Figure 2). More specifically, Figures 2a, 2b and 2c indicate that performance of monthly calibration without parameter regionalisation (bars on the left hand side of the plots in Figure 2) is *similar* to daily in 24%, 40% and 36% of sites-periods for low, medium and high flow FDC fit, respectively. Interestingly, monthly calibration *outperforms* daily calibration in 36%, 20% and 39% of sites-periods, for the same metrics, respectively. Overall, monthly calibration achieves similar or better FDC fit than daily calibration in 60-75% of the sites-periods. A more detailed analysis of these results is presented by Lerat et al. (2020): nearly all summary statistics related to FDC fit on low and high flows are equivalent or better for monthly calibration compared to daily.

On the other hand, as expected, monthly calibration performs worse for most sites/periods in terms of daily pattern metrics: Figure 2e, 2f and 2g suggest that daily calibration reaches better performance than monthly in 51%, 56% and 69% of sites/periods for correlation, NSE and peak time error, respectively.

4.2. Impact of regionalised parameters on monthly calibration performance

Use of regionalisation to estimate the quick flow parameter, θ_4, substantially improves the daily pattern metrics of monthly calibration (Figure 2). For example, for daily correlation, Figure 2e shows that monthly calibration is similar to, or better than, daily calibration at 49% of sites when no regionalisation is used ("monthly" scheme) and at 65% of sites when θ_4 is regionalised ("monthly- θ_4" scheme). Similar findings are seen for the NSE metric, where this proportion increases from 44 to 60% (Figure 2f), and for the peak time error metric, where the proportion increases from 31 to 47% (Figure 2g).

In addition, monthly calibration using regionalised θ_4 achieves similar performance to daily calibration for the other performance metrics, namely daily FDC fit and long-term bias (Figure 2).

5. DISCUSSION

5.1. How does monthly calibration compare to daily calibration?

When a rainfall-runoff model operating at a daily time step is calibrated to streamflow data aggregated to a longer time scale such as monthly, it is reasonable to expect at least some loss of performance, especially in the timing of hydrographs (e.g., Adla et al., 2019). However, the empirical results in this work indicate that monthly calibration can be comparable with daily calibration, and indeed achieve similar performance to daily calibration in terms of daily pattern metrics and potentially even better performance in terms of the FDC fit metrics (especially for high flows). This is perhaps surprising, because monthly calibration does not access the daily streamflow data and is discussed briefly as follows.

First, note that the FDC contains no timing information and hence can be reproduced by a monthly calibration scheme. Second, the more comprehensive study by Lerat et al. (2020) suggests that a higher skew in daily rainfall may play a role when monthly calibration outperforms daily calibration. More specifically, daily calibration struggles to match hydrograph peaks in catchments that experience very intense rainfall events, where rainfall timing may be poorly measured. This constraint does not impact monthly calibration significantly due to the aggregated nature of the streamflow data used in the objective function.

Finally, performance improvements achieved by monthly calibration could also be interpreted by considering the estimated values of the GR4J parameters. Lerat et al. (2020) reports that monthly calibration tends to yield smaller values of the GR4J storage parameters θ_1 and θ_3. As a result, GR4J responds faster to rainfall and generates higher peak flows – which in turn could lead to a better match of large observed flood peaks compared to a daily calibration scheme (where the estimated storage parameters have larger values).

5.2. Mitigating loss of timing information in monthly calibration

An important finding of this work relates to the benefits of using regionalisation rather than calibration, to estimate timing (quick flow) parameters such as θ_4, when undertaking monthly calibration. Estimating θ_4 using regionalisation reduces timing errors significantly compared to the "standard" approach where all parameters are calibrated, without compromising other performance metrics. As such, regionalisation helps mitigate the loss of timing information inherent when aggregating daily time series to monthly vales. issing information about quick flow parameters is "transferred" from similar catchments as part of regionalisation. Note that the increased data requirements imposed by the regionalisation approach: it requires daily streamflow data to be available at nearby catchments so that daily calibration can be carried out there. This assumption appears reasonable at least for dam inflow estimation applications, because gauging stations are generally installed close to the dam to support reservoir design and operation.

6. CONCLUSIONS AND RECOMMENDATIONS

The ability to successfully calibrate a daily rainfall-runoff model to monthly streamflow data represents an important finding with benefits for many practical modelling applications. We carried out an empirical analysis over 508 Australian catchments and two evaluation periods with the following key conclusions:

1. **Daily rainfall-runoff models calibrated to monthly streamflow data perform similarly or better than models calibrated to daily streamflow data, except for metrics related to daily patterns**

   Monthly calibration schemes reach comparable performance to daily calibrations for performance metrics such as the fit of daily flow duration curves in low, medium and high flows, with a majority of sites and periods reaching similar or better metric values. This finding holds despite monthly calibration having no access to daily streamflow data. On the other hand, and as expected, monthly calibration performs generally worse than daily calibration for metrics related to daily pattern of the hydrograph, such as NSE, correlation and peak time errors.

2. **Daily patterns of monthly calibrated models can be improved significantly by using regionalised values for timing parameters**

> Monthly calibration can degrade the accuracy of the estimated timing of flood peaks and, more generally, daily hydrograph patterns. Our results indicate that this limitation can be alleviated by estimating parameter(s) controlling hydrograph timing using regionalisation from nearby catchments.

Overall, the empirical findings indicate monthly calibration is a viable alternative to daily calibration when no daily flow data is available. Importantly, these findings hold over a large sample of Australian catchments. Future work on extending the findings to broader classes of rainfall-runoff models is recommended.

ACKNOWLEDGMENTS

REFERENCES

Adla, S., Tripathi, S., & Disse, M. (2019). Can we calibrate a daily time-step hydrological model using monthly time-step discharge data? *Water (Switzerland)*, *11*(9). https://doi.org/10.3390/w11091750

D'Oria, M., Mignosa, P., & Tanda, M. G. (2012). Reverse level pool routing: Comparison between a deterministic and a stochastic approach. *Journal of Hydrology*, *470–471*, 28–35. https://doi.org/10.1016/j.jhydrol.2012.07.045

Deng, C., Liu, P., Guo, S., Wang, H., & Wang, D. (2015). Estimation of nonfluctuating reservoir inflow from water level observations using methods based on flow continuity. *Journal of Hydrology*, *529*(1198), 1198–1210. https://doi.org/10.1016/j.jhydrol.2015.09.037

Ficchì, A., Perrin, C., & Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. *Journal of Hydrology*, *538*(454), 454–470. https://doi.org/10.1016/j.jhydrol.2016.04.016

Kavetski, D., Fenicia, F., & Clark, M. P. (2011). Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment. *Water Resources Research*, *47*(5), 5501. https://doi.org/10.1029/2010WR009525

Lerat, J., Thyer, M., McInerney, D., Kavetski, D., Woldemeskel, F., Pickett-Heaps, C., Shin, D., & Feikema, P. (2020). A robust approach for calibrating a daily rainfall-runoff model to monthly streamflow data. *Journal of Hydrology*, *591*. https://doi.org/10.1016/j.jhydrol.2020.125129

Merz, R., & Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, *287*(1–4), 95–123. https://doi.org/10.1016/j.jhydrol.2003.09.028

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, *279*(1–4), 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7

Rojas-Serna, C., Lebecherel, L., Perrin, C., Andréassian, V., & Oudin, L. (2016). How should a rainfall-runoff model be parameterized in an almost ungauged catchment? A methodology tested on 609 catchments. *Water Resources Research*, *52*(6), 4765–4784. https://doi.org/10.1002/2015WR018549

Sudheer, K. P., Chaubey, I., Garg, V., & Migliaccio, K. W. (2007). Impact of time-scale of the calibration objective function on the performance of watershed models. *Hydrological Processes*, *21*(25), 3409–3419. https://doi.org/10.1002/hyp.6555

Woldemeskel, F., McInerney, D., Lerat, J., Thyer, M., Kavetski, D., Shin, D., Tuteja, N., & Kuczera, G. (2018). Evaluating post-processing approaches for monthly and seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, *22*(12), 6257–6278. https://doi.org/10.5194/hess-22-6257-2018

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9). https://doi.org/10.1029/2007WR006716