Dmitri Kavetski, Julien Lerat, David McInerney and Mark Thyer

Can a daily rainfall-runoff model be successfully calibrated to monthly streamflow data?

Proceedings of the Hydrology and Water Resources Symposium (HWRS 2021), 2021, pp.806-817

ISBN number for HWRS 2021 is 978-1-925627-53-4

Can a daily rainfall-runoff model be successfully calibrated to monthly streamflow data?

Dmitri Kavetski[1], Julien Lerat[2], David McInerney[1] and Mark Thyer[1]

[1] School of Civil, Environmental and Mining Engineering, University of Adelaide, South Australia

[2] Land and Water, CSIRO, Canberra, ACT

Corresponding author email: dmitri.kavetski@adelaide.edu.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 several important practical applications, including the prediction of inflows into large dams, only monthly streamflow estimates are often available for model calibration. Development of robust approaches for calibrating daily rainfall-runoff models to monthly streamflow data is hence of major practical interest. This study assess the calibration of a daily hydrological model (GR4J) to monthly streamflow data and compares the resulting performance to the performance attained after calibration to daily streamflow data. Multiple performance metrics are used: fit of the daily and monthly flow duration curve, daily and monthly pattern metrics, and long-term bias. The analysis is carried for 508 Australian catchments and two evaluation periods. It is found that monthly calibration performs similar or better than daily calibration in a majority of sites and periods in terms of bias and fit of the daily flow duration curve. However, performance of monthly calibration is worse than daily calibration for daily pattern metrics such as Nash-Sutcliffe efficiency in a majority of sites and periods. This performance loss can be reduced significantly by using regionalised values for the flow-timing parameter of GR4J. Similar results are obtained for other pattern metrics. Overall, our findings suggest that monthly calibration of rainfall-runoff models to daily-rainfall/monthly-streamflow is a viable alternative to daily calibration when no daily streamflow data are available.

INTRODUCTION

Conceptual hydrological models that predict streamflow at daily time steps are widely used in water forecasting, water resources planning and operations (Pagano et al., 2010). In order to make predictions at a given catchment, these models require calibration to observed data (streamflow time series and climate inputs). 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 due to the prohibitive cost of data acquisition or limited historical records. 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 the monthly time step only.

Notable examples where daily streamflow predictions are needed despite streamflow calibration data being monthly include the calibration of basin scale models (Welsh et al., 2013), where a daily time step model is used to simulate catchment processes, including inflows, dam operations and water management rules (Steinfeld et al., 2015). In some locations available daily streamflow records might be of limited quality, compromising the model performance at those locations. In other locations daily streamflow data may not be available altogether, with only monthly data being available, e.g. from digitised paper archives (Fry, 2014). Note that these data resolution limitations are less pronounced for climate data such as rainfall, which can be obtained from global datasets, e.g. satellite rainfall.

A key practical application motivating this work is the operation of large dams. Dam inflows are typically estimated from (known) dam outflow and water levels by methods such as inverse pool routing (e.g., D'Oria et al., 2012). These methods are generally stable at a monthly time step, but can become unstable at shorter time steps (e.g. daily-weekly) due to uncertainty in water level data and the storage/volume relationship (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 performed worse than daily for simulating both daily and monthly flows, producing unrealistic daily simulations. However, their conclusions are based on results from two catchments and a single hydrological model (SWAT) and objective function (Nash-Sutcliffe efficiency). Broader studies using "mismatched" time steps produced more encouraging conclusions. For example, Wang et al. (2011) compared the monthly WAPABA model (monthly climate inputs and calibrated to monthly streamflow) against two daily models (AWBM and SIMHYD) over 331 catchments in Australia. They concluded that the monthly model achieved similar performance to the daily models in terms of monthly correlation, Nash-Sutcliffe efficiency, and long-term bias. However, no evaluation of the underlying daily-step performance was undertaken.

In general, calibration to data at an aggregated scale can be expected to lose timing information, impacting on model parameters associated with quick flow processes (e.g., Kavetski et al., 2011). Of interest here are studies where inference of quick flow parameters was improved through parameter regionalisation; e.g., see Rojas-Serna et al. (2016) and Pool et al., (2019), however those studies focused solely on the daily scale. Another pertinent study is Bennett et al. (2016), where an hourly model was calibrated to daily data, as such also representing calibration to aggregated data.

The recent work by Lerat et al. (2020) investigated the impact of daily versus monthly calibration of a daily rainfall-runoff model (GR4J) on streamflow predictive performance and parameter estimates over a large sample of catchments, using a wide range of performance metrics computed at monthly and daily time steps. In this paper, we focus on daily scale performance, which is of most relevance to the practical applications listed earlier, and provide a simplified set of recommendations.

This study has the following aims:

● Compare daily streamflow predictions obtained from monthly calibration of a daily-step model (practical setup faced by dam managers) to daily predictions obtained from a daily calibration of the same daily-step model ("idealised" setup).

● Use a parameter regionalisation strategy to mitigate the reduced information about flow timing.

The study uses the hydrological model GR4J, which is commonly employed in streamflow prediction applications in Australia and worldwide (Perrin et al., 2003; Woldemeskel et al., 2018).

THEORY

General Model Setup

Consider a hydrological model H that simulates daily streamflow on day t as a function of forcing data up to day t, $\mathbf{X}_{1:t}$, parameters $\boldsymbol{\theta}$ and initial conditions S_0,

$$Q_t^{\boldsymbol{\theta}} = H(\boldsymbol{\theta}; \mathbf{X}_{1:t}, \mathbf{S}_0) \tag{1}$$

This daily model can be calibrated at the daily or monthly scale, as described next.

Daily Calibration

In daily-scale calibration, parameters are estimated using an objective function F, formulated in terms of daily flows,

$$\boldsymbol{\theta}^{\text{opt}} = \arg\min_{\boldsymbol{\theta}} F(\mathbf{Q}^{\text{obs}}, \mathbf{Q}^{\boldsymbol{\theta}}) \tag{2}$$

where $\mathbf{Q}^{\text{obs}}$ are the time series of observed daily streamflow and $\mathbf{Q}^{\boldsymbol{\theta}}$ are the corresponding time series of daily flow computed using equation (1).

Monthly Calibration

In monthly-scale calibration, parameters are estimated using a "monthly scale" objective function F, formulated in terms of monthly flows

$$\boldsymbol{\theta}^{\text{opt}} = \arg\min_{\boldsymbol{\theta}} F(\mathbf{q}^{\text{obs}}, \mathbf{q}^{\boldsymbol{\theta}}) \tag{3}$$

where $\mathbf{q}^{\text{obs}}$ are the time series of observed monthly streamflow and $\mathbf{q}^{\boldsymbol{\theta}}$ are the corresponding time series of monthly streamflow computed by aggregating daily flows computed using equation (1). Note that in order to undertake the comparisons within this study, the monthly observed streamflow is obtained by aggregating observed daily streamflow, though in practical applications of interest here monthly time series will be available directly (and indeed daily time series will not be available).

Objective Function and Optimisation

In this work the objective function F in both equations (2) and equation (3) is the sum of squared errors (SSE), which can be computed for either daily or monthly time series.

The sum-of-squared errors function (SSE) has the form

$$F(\mathbf{y}^{\text{obs}}, \mathbf{y}^{\boldsymbol{\theta}}) = \left(z(\mathbf{y}^{\text{obs}}) - z(\mathbf{y}^{\boldsymbol{\theta}})\right)^2 \tag{4}$$

where z denotes the Box-Cox transform, here used with power parameter $\lambda = 0.2$ following recommendations of McInerney et al. (2017).

$$z(y) = (y^{\lambda} - 1) / \lambda \tag{5}$$

The use of the Box-Cox transformation in the objective function tends to stabilise the errors with respect to the streamflow magnitude.

The optimisation of the SSE is carried out using the downhill simplex method combined with an initial seeding scheme designed according to available prior knowledge of hydrological model parameters. This approach improves the global convergence of the optimisation (Lerat, 2018; Lerat el al., 2020).

It is emphasised that in this work the only difference between a daily calibration (using equation (2)) and monthly calibration (using equation (3)) is the time aggregation of the streamflow data. All other elements of the schemes are identical, including the daily climate inputs **X**, the model daily simulation time step, the mathematical form of the objective function F, and the optimisation algorithm.

Combination of Calibration and Regionalisation

Previous work has indicated that calibration of a daily model to monthly streamflow data results in poor identifiability of parameters associated with quick flow processes (Kavetski et al., 2011). Hence, these parameters are estimated using a regionalisation model, in order to benefit from additional conditioning information. More specifically, the regionalisation model is constructed by applying kriging to quick flow parameter values estimated in nearby catchments with available daily streamflow data (Merz and Blöschl, 2004). The regionalisation approach is detailed in Appendix A of Lerat el al. (2020).

When regionalisation is used, the model parameter vector is split into two partitions, $\boldsymbol{\theta} = (\boldsymbol{\theta}_c, \boldsymbol{\theta}_r)$, where $\boldsymbol{\theta}_c$ is the subset of parameters estimated by calibration and $\boldsymbol{\theta}_r$ is the subset of parameters estimated by regionalisation. The subset $\boldsymbol{\theta}_r$ of hydrological model parameters is taken as the subset of parameters know a priori to control the timing of the hydrograph. For common models, including the GR4J model used in this study, such parameters are well known (see latter section).

To avoid artificial skill being introduced, parameter regionalisation is implemented using a leave-one-out approach, where regionalised parameters are estimated using a kriging model calibrated to all catchments except the target catchment. Consequently, a separate kriging model is developed for each catchment and for each calibration period.

Note that the kriging-based regionalisation approach used in this study does not take into account catchment descriptors. A simple alternative using linear regression against catchment descriptors was also tested, but did not provide better performance (results not reported).

CASE STUDY DESCRIPTION

Case Study Catchments

The empirical case study employs 508 catchments in Australia, with locations shown in Figure 1; see Lerat et al. (2020) for detailed characteristics. Daily rainfall time series were extracted from the Bureau of Meteorology AWAP archive (Jones et al., 2009). Daily time series of potential evapotranspiration were estimated using the Penman equation (1948) within the Bureau of Meteorology landscape model AWRA (Frost et al., 2018). Daily streamflow data were extracted from the Water Data Online service (Bureau of Meteorology, 2019). As such, the catchment dataset covers all main Australian climate regions, lending robustness to the study conclusions. A wide range of hydro-climate regimes is included, with (daily) rainfall skewness in the range 2.4-16.4, catchment aridity index (ratio of mean annual rainfall and mean annual potential evapotranspiration) in the range 0.09-2.57, and runoff coefficient in the range 0.01-0.92.

Three representative catchments with distinct hydroclimatic characteristics are selected to illustrate the findings of this study in more detail: Macalister River at Glencairn in Victoria (catchment 225219), Dumaresq River at Farnbro in Queensland (catchment 416310) and Scott River at Brennans Ford in Western Australia (catchment 609002); see Lerat et al. (2020) for details.

Hydrological Model

The conceptual rainfall-runoff model GR4J is used in equation (1) to simulate daily catchment-scale streamflow time series from rainfall and PET inputs (Perrin et al., 2003). GR4J has a parsimonious model structure with four parameters to represent interception, infiltration and percolation, and is routinely used in research and operations in Australia and worldwide (e.g., Woldemeskel et al., 2018).

The GR4J parameter θ_4 is known to control the timing of the hydrograph. Therefore the regionalisation procedure described in the earlier section was applied to this parameter.

Calibration Schemes

Three calibration schemes are considered: daily, monthly and monthly with regionalised GR4J parameter θ_4 (denoted "monthly-θ_4").

Performance Metrics for Daily Streamflow Predictions

Five performance metrics computed from daily streamflow predictions are used. See Lerat et al. (2020) for a broader study that also includes an assessment of monthly streamflow predictions.

Bias in long-term average streamflow is computed as

$$\mathrm{Bias}(\mathbf{q}^{\mathrm{obs}},\mathbf{q}^{\boldsymbol{\theta}}) = \left|1 - \frac{\mathrm{mean}\,\mathbf{q}^{\mathrm{obs}}}{\mathrm{mean}\,\mathbf{q}^{\boldsymbol{\theta}}}\right| \tag{6}$$

where mean $\mathbf{y}$ denotes the mean of vector $\mathbf{y}$.

Flow duration curve (FDC) error, expressed as the mean absolute relative error as follows:

$$e_{\mathrm{FDC}}(\mathbf{q}^{\mathrm{obs}},\mathbf{q}^{\boldsymbol{\theta}}) = \frac{1}{N_P}\sum_{k=1}^{N_P}\left|1 - \frac{\mathrm{percentile}(\mathbf{q}^{\boldsymbol{\theta}}, p_k)}{\mathrm{percentile}(\mathbf{q}^{\mathrm{obs}}, p_k)}\right| \tag{7}$$

where $\mathrm{percentile}(\mathbf{y}, p)$ is the p'th percentile of series $\mathbf{y}$, p_1 and p_2 are the lowest and highest percentile considered, N_P is the number of percentile interpolation points, and $p_k = p_1 + (p_2 - p_1)(k-1)/(N_p - 1)$ is the linearly interpolated percentile. In this work $N_P = 50$.

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: 0% to 10% (i.e. $p_1 = 0$ and $p_2 = 10$) to quantify FDC fit in low flows, 30% to 70% (i.e. $p_1 = 30$ and $p_2 = 70$) for medium flows, and 90% to 100% (i.e. $p_1 = 90$ and $p_2 = 100$) for high flows.

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

$$\mathrm{spear}(\mathbf{q}^{\mathrm{obs}},\mathbf{q}^{\boldsymbol{\theta}}) = \frac{\sum_{t=1}^{T}(\mathrm{rank}\,q_t^{\mathrm{obs}} - \mathrm{rank}\,q_t^{\boldsymbol{\theta}})(\mathrm{rank}\,q_t^{\mathrm{obs}} - \mathrm{rank}\,q_t^{\boldsymbol{\theta}})}{\mathrm{sdev}\,\mathrm{rank}\,\mathbf{q}^{\mathrm{obs}} \times \mathrm{sdev}\,\mathrm{rank}\,\mathbf{q}^{\mathrm{obs}}} \tag{8}$$

where $\mathrm{rank}\,q_t^{\mathrm{obs}}$ and $\mathrm{rank}\,q_t^{\boldsymbol{\theta}}$ are the, respectively, ranks of the observed and simulated streamflow values at day t, and T is the total number of (daily) time steps.

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

$$e_{\mathrm{timing}}(\mathbf{q}^{\mathrm{obs}},\mathbf{q}^{\boldsymbol{\theta}}) = \frac{1}{N_A}\sum_{a=1}^{N_A}\left|t_a - \underset{t\in[t_a-5,t_a+5]}{\arg\max}\, q_t^{\mathrm{obs}}\right| \tag{9}$$

where N_A is the number of years simulated, t_a is the day at which the observed peak flows of water year a takes place. 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. The water year (rather than calendar year), is used, with start and end dates defined, respectively, as 5 months before and 5 months after the maximum mean inter-annual monthly flow. This definition avoids splitting a flood event across two years.

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

$$F(\mathbf{q}^{\mathrm{obs}},\mathbf{q}^{\boldsymbol{\theta}}) = 1 - \frac{\mathrm{var}(\mathbf{q}^{\mathrm{obs}} - \mathbf{q}^{\boldsymbol{\theta}})}{\mathrm{var}(\mathbf{q}^{\mathrm{obs}})} \tag{10}$$

The metrics described above can be categorised into three groups: long-term bias, FDC fit metrics and pattern metrics.

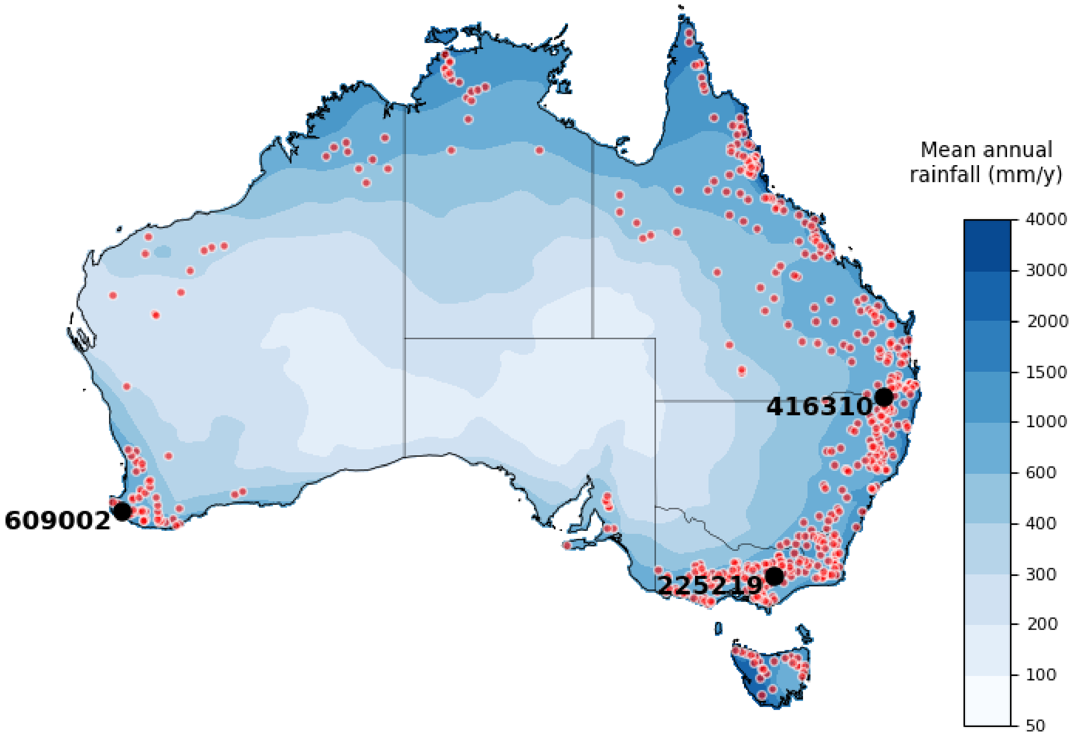

Figure 1. Locations of the 508 case study catchments and the 3 selected catchments. The red dots indicate the catchment outlets. Figure adapted from Lerat et al. (2020).

Cross-validation

A temporal split-sample validation procedure is implemented, 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 metrics. This procedure yields two sets of validation periods – and hence two validation metrics – for each catchment.

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/periods 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.

RESULTS

Figure 2 shows the proportion of sites/periods where the performance of monthly calibration is, respectively, "worse than", "similar to", and "better than" daily calibration. Two monthly calibration configurations are shown: (1) “Monthly”, where all parameters are calibrated (no regionalisation), and (2) “Monthly-θ4”, where GR4J parameter θ_4 is regionalised as described in a preceding section. Note that all performance metrics are computed from daily streamflow simulations.

Monthly vs Daily Calibration

Monthly calibration with no parameter regionalisation (labelled "monthly" in Figure 2) performs similarly or better than daily calibration for daily FDC fit metrics in a majority of sites and periods. Figure 2a, 2b and 2c reveals that monthly calibration performance is similar to daily in 24%, 40% and 36% of sites/periods for low, medium and high flow FDC fit, respectively. Moreover, 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 (Lerat et al., 2020) confirms these findings: nearly all summary statistics of FDC fit on low and high flows are lower (i.e. better) for monthly calibration compared to daily. All statistics on medium flows are higher (i.e. worse) for monthly compared to daily calibration, but the difference remains below 0.06 percentage points except for the maximum value (Lerat et al., 2020).

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

Impact of Regionalised Parameters on Monthly Calibration Performance

We now consider the benefits of parameter regionalization, by comparing monthly calibration coupled with parameter regionalisation ("monthly-θ4" in Figure 2) to the performance of daily calibration.

Using regionalised estimates of θ_4 improves daily pattern metrics of monthly calibration significantly: for daily correlation, Figure 2d shows that the proportion of sites/periods where monthly calibration is similar to, or better than daily calibration increases from 49% when no regionalisation is used to 65% with regionalised θ_4. This proportion increases from 44 to 60% for NSE (Figure 2e), and from 31 to 47% for peak time error (Figure 2f).

Figure 2 also shows that using regionalised θ_4 in the monthly calibration achieves similar performance to daily calibration for the other performance metrics, namely daily FDC fit and long-term bias.

Hydrographs for Representative Catchments

Figure 3 shows the daily observed and simulated streamflow in the three representative catchments. Hydrographs for the two largest floods are shown, using the period of 1995-2015 for validation, and the period of 1975-1992 for calibration.

The hydrographs shown in Figure 3 highlight the advantages of using a regionalised estimate of θ_4 to better reproduce flood peak timing when using a monthly calibration scheme. When all parameters are calibrated ("monthly" in Figure 3), the daily hydrographs obtained using the monthly calibration scheme are systematically lagged behind with respect to the (daily) observed hydrographs. In contrast, fixing parameter θ_4 to regionalised values ("monthly-θ_4" in Figure 3) greatly improves flood simulations, as can be seen in the majority of hydrographs.

DISCUSSION

How Does Monthly Calibration Compare to Daily Calibration?

Previous research has indicated that, when a rainfall-runoff model that operates at one time scale is calibrated to streamflow data aggregated to a larger time scale, there is typically some loss of performance, especially in the timing of hydrographs (Adla et al., 2019).

However, the results suggest that a carefully constructed monthly calibration can often (though not always) achieve comparable performance to daily calibration in terms of daily pattern metrics and potentially even *better* performance for FDC fit metrics, especially for high flows. Results in representative catchments in Figure 3 illustrate these findings using hydrographs. This finding is perhaps surprising because monthly calibration does not access the daily streamflow data – though the FDC contains no timing information and hence it can be reproduced by a monthly calibration scheme. 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 with 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.

The performance improvements achieved by monthly calibration for certain metrics could also be interpreted by considering the changes in GR4J parameter values. Inspection of GR4J model parameters in Lerat et al. (2020) suggests that monthly calibration tends to yield smaller values of the two 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 storage parameter are estimated with larger values).

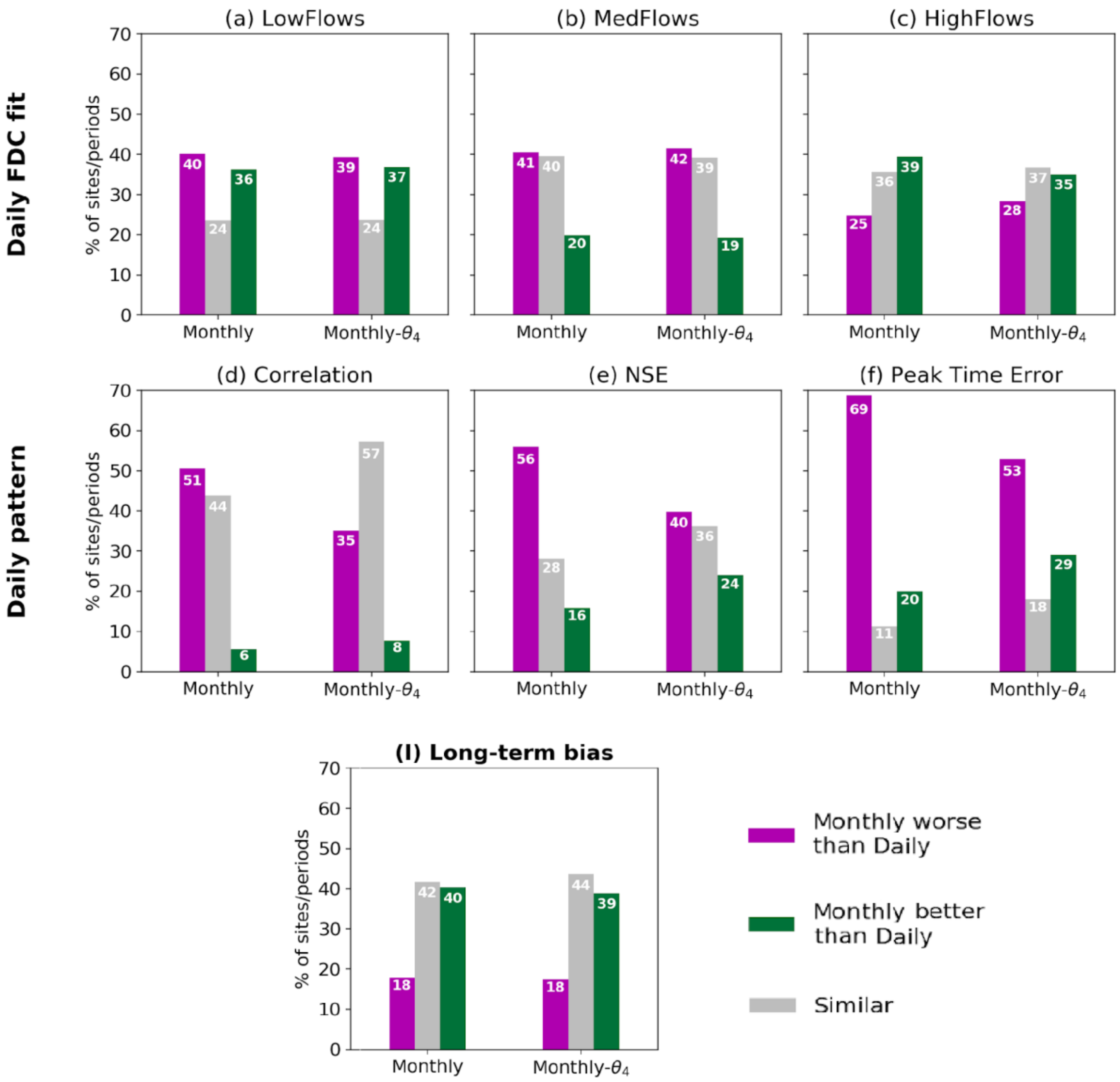

Figure 2. Model performance following monthly vs daily calibration. Proportion of sites/periods where monthly calibration is respectively "worse than", "similar to", and "better than" daily calibration. Proportions are computed from 508 catchments and two validation periods (1016 combinations). The labels "Monthly" and "Monthly-θ4" refer, respectively, to regular monthly calibration and to monthly calibration where the GR4J parameter θ_4 is fixed a priori to a regionalised value (i.e. three parameters calibrated). The numbers within each bar indicate the performance proportions (for the three categories listed earlier) and add up to 100%. Figure adapted with modifications from Lerat et al. (2020).

Mitigating Loss of Timing Information in Monthly Calibration

The empirical results also lend credence to using regionalisation rather than calibration to estimate timing (quick flow) parameters such as θ_4. This approach assumes that daily streamflow data are available at nearby catchments for a daily calibration to be implemented there. In the case of simulating dam inflows, this assumption appears reasonable as gauging stations are generally installed close to the dam site to support reservoir design and operation. The approach of estimating θ_4 using regionalisation reduces timing errors significantly compared to the "standard" approach where all parameters are calibrated, without compromising other performance metrics. It is of course possible that regionalisation leads to inaccurate estimates of θ_4 compared to daily calibration. However, our results suggest that these inaccuracies do not compromise monthly calibration performance.

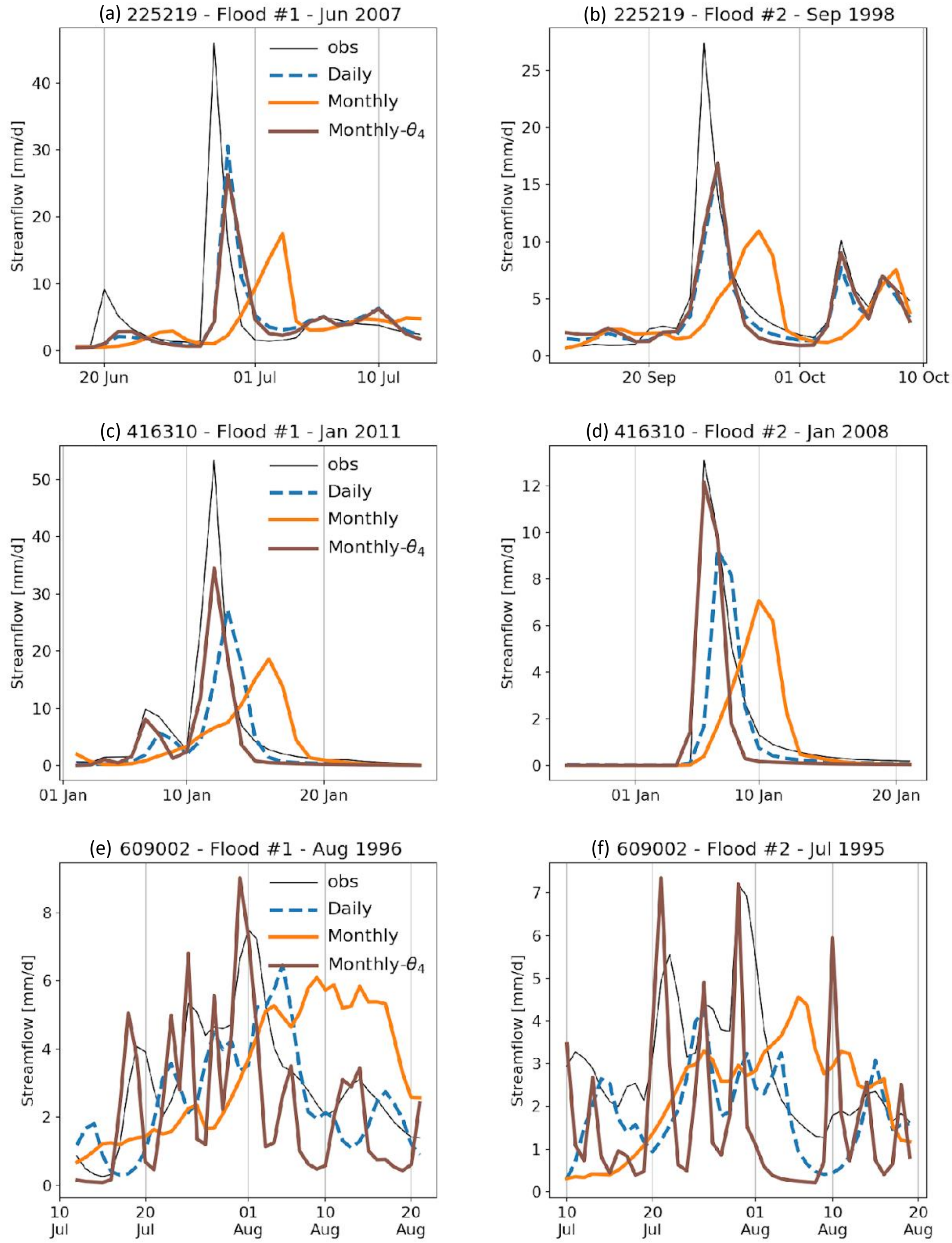

Figure 3. Streamflow time series for three representative catchments in the validation period 1992-2015. Simulations corresponding to three calibration schemes are shown: daily, monthly and monthly with regionalised θ_4 parameter (monthly-θ_4). Figure adapted with modifications from Lerat et al. (2020).

Limitations and Future Work

Several follow up research directions will be pursued in future studies.

First, it is of interest to verify if our findings regarding the value of monthly calibration for the parsimonious model GR4J hold for other rainfall-runoff models, particularly more complex models with a larger number of parameters, and for spatially resolved models.

Second, it is of interest to better elucidate the reasons why monthly calibration can produce better daily performance metrics than daily calibration. Separating timing errors from amplitude errors (for example using the spectral methods proposed by Ehret et al., 2011, or Liu et al., 2011), and relating specific hydrograph features to specific parameters, could be pivotal to progress our understanding.

Third, the investigations could be extended to probabilistic prediction (e.g. McInerney et al., 2018; and others), which can be expected to be easier at the monthly scale compared to the daily scale, due to reduced autocorrelation in the errors (Huard and Maillot, 2008, Evin et al., 2013).

Fourth, further analysis is required to ensure that monthly calibration remains a valid alternative during extremely dry or wet periods, for example in the context of changing climate.

Finally, the regionalisation methods could be enhanced with a better representation of local catchment characteristics and (potentially) a more sophisticated geostatistical model (Bárdossy and Li, 2008).

CONCLUSIONS AND RECOMMENDATIONS

The ability to successfully calibrate a daily rainfall-runoff model to monthly streamflow data with (relatively) little loss of performance – and indeed with some gains in performance – represents an important finding with benefits for many practical modelling applications.

Empirical analysis over 508 Australian catchments and two evaluation periods indicated the following key conclusions:

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

Monthly calibration schemes reach comparable performance to daily calibrations for a wide range of 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 is arguably surprising because monthly calibration does not access daily streamflow data. On the other hand, and arguably 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

When using monthly calibration schemes, the timing of flood peaks and, more generally, daily patterns, are not reproduce accurately. This limitation can be alleviated by estimating the GR4J parameter controlling hydrograph timing using regionalisation from nearby catchments.

Overall, the 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 models is recommended.

ACKNOWLEDGMENTS

The results presented in this work were obtained with the support of the iSSF4 and soil moisture forecasting projects funded by the Bureau of Meteorology. The authors acknowledge the support provided by the AWAP (Jones et al., 2009), AWRA (Frost et al., 2018) and Water Data Online (BoM, 2019) services in supplying high quality datasets for the case study catchments. The second author (JL) was employed at the Bureau of Meteorology at the time the project was carried out. The contributions of Fitsum Woldemeskel, Christopher Pickett-Heaps, Daeyhok Shin and Paul Feikema as co-authors of the full publication Lerat el al. (2020) are gratefully acknowledged.

REFERENCES

Adla, S., Tripathi, S., and Disse, M. (2019). Can we calibrate a daily time-step hydrological model using monthly time-step discharge data? *Water*, 11(9):1750.

Bárdossy, A. and Li, J. (2008). Geostatistical interpolation using copulas. *Water Resour. Res.*, 44(7).

Bennett, J. C., Robertson, D. E., Ward, P. G., Hapuarachchi, H. P., and Wang, Q. (2016). Calibrating hourly rainfall-runoff models with daily forcings for streamflow forecasting applications in mesoscale catchments. *Env. Modelling and Software*, 76:20–36.

Bureau of Meteorology (2019). *Water Data Online*. at URL http://www.bom.gov.au/waterdata. accessed on 2019-07-16.

Deng, C., Liu, P., Guo, S., Wang, H., and Wang, D. (2015). Estimation of nonfluctuating reservoir inflow from water level observations using methods based on flow continuity. *J. Hydrol.*, 529:1198– 1210.

D'Oria, M., Mignosa, P., and Tanda, M. G. (2012). Reverse level pool routing: Comparison between a deterministic and a stochastic approach. *J. Hydrol.*, 470:28–35.

Ehret, U. and Zehe, E. (2011). Series distance–an intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events. *Hydrology and Earth System Sciences*, 15(3):877–896.

Evin, G., Kavetski, D., Thyer, M., and Kuczera, G. (2013). Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. *Water Resour. Res.*, 49(7):4518–4524.

Ficchì, A., Perrin, C., and Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. *J. Hydrol.*, 538:454–470.

Frost, A., Ramchurn, A., and Smith, A. (2018). *The Australian Landscape Water Balance model (AWRA-L v6)*. Technical Description of the Australian Water Resources Assessment Landscape model version 6. Technical report, Bureau of Meteorology.

Fry, M. (2014). *Guidelines for hydrological data rescue*. Technical Report No. 1146, World Meteorological Organisation.

Huard, D. and Mailhot, A. (2008). Calibration of hydrological model GR2M using bayesian uncertainty analysis. *Water Resour. Res.*, 44(2).

Jones, D. A., Wang, W., and Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal*, 58(4):233.

Kavetski, D., Fenicia, F., and 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 Resour. Res.*, 47(5), W05501.

Lerat, J. (2018). Incorporating prior knowledge in the calibration of hydrological models for water resources forecasting. In Anderssen, R. S., Broadbridge, P., Fukumoto, Y., Kajiwara, K., Simpson, M., and Turner, I., editors, *Agriculture as a Metaphor for Creativity in All Human Endeavors*, pages 61–72, Singapore. Springer Singapore.

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, doi:10.1016/j.jhydrol.2020.125129

Liu, Y., Brown, J., Demargne, J., and Seo, D.-J. (2011). A wavelet-based approach to assessing timing errors in hydrologic predictions. *J. Hydrol.*, 397(3-4):210–224.

McInerney, D., Thyer, M., Kavetski, D., Bennett, B., Lerat, J., Gibbs, M., and Kuczera, G. (2018). A simplified approach to produce probabilistic hydrological model predictions. *Env. Modelling and Software*, 109:306–314.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., and Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resour. Res.*, 53(3):2199–2239.

Merz, R. and Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrolology*, 287(1-4): 95–123.

Pagano, T., Hapuarachchi, P., and Wang, Q. (2010). *Continuous rainfall-runoff model comparison and short-term daily streamflow forecast skill evaluation*. Technical Report EP103545, CSRIO.

Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 193(1032):120–145.

Perrin, C., Michel, C., and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *J. Hydrol*., 279(1-4):275–289.

Pool, S., Viviroli, D., and Seibert, J. (2017). Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? *J. Hydrol*., 554:613–622.

Rojas-Serna, C., Lebécherel, L., Perrin, C., Andréassian, V., and Oudin, L. (2016). How should a rainfall-runoff model be parameterized in an almost ungauged catchment? a methodology tested on 609 catchments. *Water Resour. Res*., 52(6):4765–4784.

Steinfeld, C., Kingsford, R., Webster, E., and Sharma, A. (2015). A simulation tool for managing environmental flows in regulated rivers. *Env. Modelling and Software*, 73:117–132.

Sudheer, K., Chaubey, I., Garg, V., and Migliaccio, K. W. (2007). Impact of time-scale of the calibration objective function on the performance of watershed models. *Hydrol. Processes*, 21(25):3409– 3419.

Wang, Q., Pagano, T., Zhou, S., Hapuarachchi, H., Zhang, L., and Robertson, D. (2011). Monthly versus daily water balance models in simulating monthly runoff. *J. Hydrol*., 404(3-4):166–175.

Welsh, W. D., Vaze, J., Dutta, D., Rassam, D., Rahman, J. M., Jolly, I. D., Wallbrink, P., Podger, G. M., Bethune, M., Hardy, M. J., et al. (2013). An integrated modelling framework for regulated river systems. *Env. Modelling and Software*, 39:81–102.

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

Yilmaz, K. K., Gupta, H. V., and Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NSW distributed hydrologic model. *Water Resour. Res*., 44(9).

BIOGRAPHY

Dmitri Kavetski is a Professor of Civil and Environmental Engineering at the University of Adelaide. His research work focuses on hydrological modelling, where he has contributed a range of new techniques for model development, calibration and prediction. A particular emphasis has been on understanding, quantifying, and where possible reducing, the sources of uncertainty in streamflow predictions at the catchment scale. Dmitri's broader research interests include probabilistic methods, numerical modelling and optimization. Dmitri also delivers regular short courses and workshops on hydrological modelling to Australian and international audiences.