

Towards reliable hydrological model calibrations with river level measurements

J. Jian^a, D. Ryu^a, J. F. Costelloe^a and C.-H. Su^a

^aDepartment of Infrastructure Engineering, University of Melbourne, Victoria 3010, Australia
Email: jjian1@unimelb.edu.au

Abstract: Streamflow prediction based on hydrological models generally requires model calibration using observed discharge derived from a rating curve that requires accurate, frequent and sustained measurements of the flow rate. However, the remoteness of many catchments increases the difficulties and costs of measuring discharge data in these areas. Even at gauging stations with some discharge measurements, the resulting relationship can still have high uncertainties, particularly for larger magnitude flows where discharge measurements are absent.

In this study, methods to calibrate streamflow models using only the time series of stream level measurements, without converting it to discharge values, are explored. With an emerging remote sensing technique that estimates stream level using satellite altimeters, the methods can potentially enable streamflow model calibration over ungauged catchments. Two prospective calibration schemes, which are not reliant on observed discharge data, are examined: i) Spearman Rank Correlation (denoted by SRC) based calibration between observed stream level (h_{obs}) and modelled streamflow (Q_{est}); ii) Inverse Rating Curve (IRC) function based calibration using Nash-Sutcliffe Efficiency (NSE) between the estimated water level (h_{est}), which is calculated from the modelled streamflow using an inverse rating curve, and the observed water level data (h_{obs}). The new calibration schemes are applied to two catchments (604053 and G8110004 from the Bureau of Meteorology's Hydrologic Reference Stations) and the efficacy of the methods are examined by comparing the modelled discharges, which are calibrated by the new schemes, to the observed discharge. A control case that is conventionally calibrated against observed discharge is also applied in the efficiency analysis.

The results show that the new calibration schemes could properly predict the runoff events and timings, with the linear correlations between the modelled and observed discharges ranging between 0.675-0.889. However, compared to the control cases that show adequate prediction of the observed discharge with $R > 0.77$ and $NSE > 0.58$, the biases between the results and the observations remain large and are deemed unacceptable. The reasons for the biases are associated with the underlying principles of the calibration schemes. In particular, the SRC-based method is more sensitive to the rank of the data than to the differences between modelled and observed magnitudes. The IRC-based method has difficulty in reproducing the shape of the actual power function using only observed and modeled water levels. Of the two new calibration schemes, the IRC-based scheme has better performances than the SRC-based method, although the SRC-based method incurs less computation cost by calibrating fewer parameters.

The results give cause for optimism in devising a reliable calibration scheme with reduced reliance on large quantity of discharge measurements. Future work will focus on conjunctive use of river level data and small number of discharge measurements to overcome the aforementioned problem of biases and applications at more catchments and with satellite altimetry data.

Keywords: Hydrological modelling, model calibration, Spearman Rank correlation, inverse rating curve

1. INTRODUCTION

Hydrological modelling provides the detailed understanding of water cycles and hydrological systems, thereby supporting studies of climate change, water resource management, land use and infrastructure design (Lørup et al., 1998; Vörösmarty et al., 2001). Streamflow prediction is one of the most important applications of hydrological modelling. The modelling processes require calibration to optimize their performances (Gupta et al., 1998), and the observed stream discharge data are essential in the streamflow model calibration because streamflow discharge reflects the whole-of-catchment responses to meteorological forcing. However, discharge data are generally limited within a catchment, and for many catchments globally, even though automated data may be collected on stream level variations. This is because the one-to-one relationship between river levels and river discharge volume, namely the rating curve, must be established with accurate information of river bathymetry and requires numerous measurements across the entire discharge range to define the relationship (Dottori et al., 2009). This is not a trivial task in many remote basins and insufficient gauging data can lead to significant uncertainties in river discharge (Bjerklie et al., 2003; Smith et al., 1996).

Even where accurate discharge data are collected at gauging stations monitoring large catchments, many of significant rainfall-runoff events occur in the sparsely gauged, or ungauged, sub-catchments of these larger catchments, and so their contribution can be poorly constrained (Alsdorf et al., 2003). The remoteness of catchment and sub-catchment outlets and expense of instrumentation limit the number of gauged catchments. Since many monitoring projects are undertaken by individual institutions, the spatial and temporal coverage and the availability of the data are limited by the scopes and privacies of the individual projects (Vörösmarty et al., 2001). Also, the monitoring stations' capacity of supplying correct and timely discharge data is decreasing as well due to the decreasing number of monitoring stations (Vörösmarty et al., 2001).

To overcome this issue, alternative calibration methods for hydrological modelling that do not rely on observed discharge data are highly desirable. Satellite altimetry could be used as a potential source of stage data for estimating discharge for monitoring and calibrating models (Smith et al., 1996; Smith et al., 1995). However, the current use of the altimetry data in estimating discharge in ungauged locations is constrained by the need to convert altimeter water level to discharge with synthetic rating curves (Jung et al., 2010; Smith et al., 1996). This current practice still requires significant inputs (e.g., high-resolution river bathymetry) that are often unavailable in remote basins or subject to large uncertainties.

A method that directly uses river stage data in the hydrological model calibration would circumvent the need for the conversion to discharge and reduce uncertainty introduced by this conversion. In this way, satellite altimetry can provide river stage data at many inaccessible rivers and catchments. Motivated by this potential, this work aims to develop such methods to calibrate hydrological models using the river stage data and to assess its reliability. In particular, we propose two calibration schemes and apply it to a widely used conceptual lumped rainfall-runoff model, the Probability Distributed Model (PDM) (Moore, 2007) at two catchments with perennial and ephemeral flow regimes in southwest and north of Australia, respectively. Their performance and limitations can be distinguished by comparing against ground truth discharge data. This work therefore provides an important cornerstone for future research in using satellite altimetry data in model calibration.

2. STUDY CATCHMENTS

As this work is a precursor to using satellite altimetry data in model calibration, we focus on the calibration of the rainfall-runoff model PDM at two catchments suitable for use with the satellite data. In particular, their selection is based on the following criteria.

- The catchment's size is approximately 1000-5000 km² to satisfy the requirements of the lumped model, and the locations of the catchments are within the coverage of satellite altimetry.
- The study catchment should have a monitoring station that is close to the satellite altimetry track that can later provide discharge measurements for verification.
- The geographical condition of the catchments should not obstruct the satellite altimetry track; and the catchments should have the potential to provide altimeter data with sufficient quality.

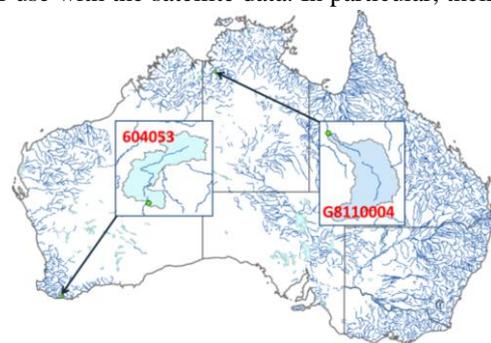


Figure 1. Distribution of the selected catchments.

Following the above conditions, Kent River at Styx Junction

(604053) in Western Australia, and East Baines River at VIC Highway (G8110004) in Northern Territory are selected. The details are shown in Table 1. The river stage and discharge data are obtained from Western Australia Government – Department of Water and Northern Territory Government – Department of Land Resource Management, separately.

Table 1. Site information.

Catchment ID	604053	G8110004
Description	Kent River at Styx Junction	East Baines River at VIC Highway
Catchment area (km ²)	1786	2443
State	WA	NT
Mean daily rainfall (mm/d)	2.058	2.39
Mean water level (m)	0.35	1.73
Mean daily discharge (mm/d)	0.097	0.441
Koppen climate zone classification	Temperate - moderately dry winter (warm summer)	Tropical - savanna
Flow duration type	Perennial	Ephemeral

3. METHOD

The PDM is used to model streamflow from 1 January 1990 to 31 December 2014 (25 years), and the Shuffled Complex Evolution algorithm (SCE-UA) (Duan et al., 1993) is applied to calibrate the model. To check the suitability of the model structure and benchmark the performances of the new calibration schemes (as follows), a control (conventional) case of using observed discharge data (Q_{obs}) in model calibration is made, depicted as Case 1 of Figure 2. The modelled discharge by the control case is represented as Q_{cr} .

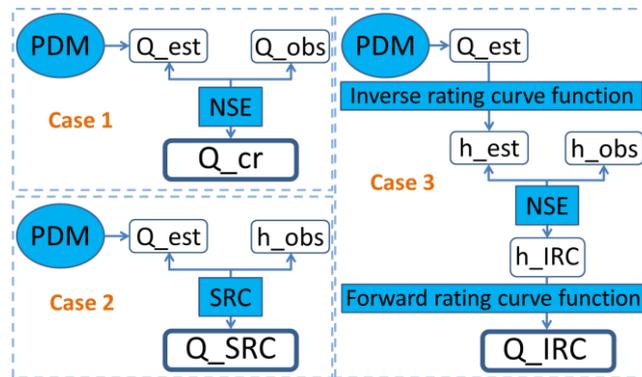


Figure 2. Illustration of control case (Step 1) and two new calibration schemes (Step 2 and Step 3). (Acronyms in the diagram: Q_{obs} - observed discharge; Q_{est} – estimated discharge, h_{obs} – observed water level; h_{est} – estimated water level, Q_{cr} – modelled discharge from the control case; Q_{SRC} – modelled discharge from the Spearman rank correlation based scheme; Q_{IRC} – modelled discharge from the inverse rating curve function based scheme; h_{IRC} – modelled water level from the inverse rating curve function based scheme.)

Two novel calibration schemes based only on river stage data (h_{obs}) are proposed:

- Spearman rank correlation based calibration scheme (denoted by SRC) (Case 2 of Figure 2)

The water level (h) and discharge (Q) are expected to share a monotonic relationship that can be measured by the Spearman Rank correlation coefficient, ρ

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}, \quad (1)$$

where $d_i = x_i - y_i$ is the difference between ranks and n is the sample size. ρ varies between -1 and 1, and a value of 1 or -1 means the paired data have a perfect monotonic relationship. The Spearman rank correlation was chosen due to its lower sensitivity to noise compared to an alternative correlation metric, Kendall-tau rank correlation (Glantz, 2012).

- Inverse rating curve function based calibration scheme (IRC) (Case 3 of Figure 2)

The relationship between water level h and discharge Q (Figure 3) can typically be expressed by a power function as:

$$Q = a_0 + a_1 h^{a_2} \quad (2)$$

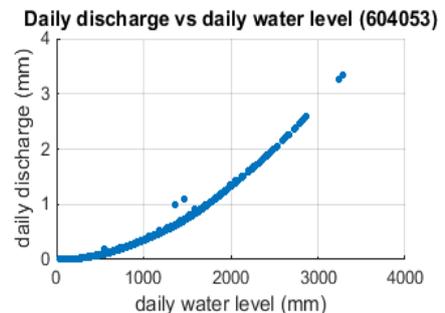


Figure 3. Typical relationship between water level and discharge (604053)

where a_0 , a_1 and a_2 are parameters requiring calibration. By inverting equation (2), the estimated water level data (h_{est}) can be obtained from the estimated discharge (Q_{est}):

$$h_{est} = \left(\frac{Q_{est} - a_0}{a_1} \right)^{\frac{1}{a_2}} \quad (3)$$

This calibration scheme aims to calibrate the estimated water level (h_{est}) against the observed water level data (h_{obs}) using Nash- Sutcliffe Efficiency (NSE) to achieve more accurate calibration results.

Several statistical metrics are used to evaluate the results. The first one is the correlation coefficient R , which is an indicator of the linear relationship between the observations and the estimated data (Moriassi et al., 2007). The second is the NSE which is commonly applied in hydrological modelling as a combined measure of correlation and multiplicative and additive bias between the observed data and estimated data (Gupta et al., 2009). We also compare the first and second statistical moments of the simulated streamflow to distinguish the presences of biases.

4. RESULTS AND DISCUSSIONS

The calibration performances for the two catchments are shown in Table 2. Control case performances (represented by R and NSE) are shown in columns 2 and 3. The modelled discharges from the two calibration schemes are compared to the observed discharge in columns 4 to 7. Similarly, the last four columns display the comparisons of the results from new schemes with that from the control case.

Table 2. Verification results.

Catchment ID	Control case performance		Estimated Q vs Observed Q				Estimated Q vs Q from control case			
	R of Q_{cr} vs Q_{obs}	NSE of Q_{cr} vs Q_{obs}	R of Q_{SRC} vs Q_{obs}	R of Q_{IRC} vs Q_{obs}	NSE of Q_{SRC} vs Q_{obs}	NSE of Q_{IRC} vs Q_{obs}	R of Q_{SRC} vs Q_{cr}	R of Q_{IRC} vs Q_{cr}	NSE of Q_{SRC} vs Q_{cr}	NSE of Q_{IRC} vs Q_{cr}
604053	0.899	0.782	0.853	0.889	-0.710	0.706	0.909	0.988	-2.463	0.629
G8110004	0.772	0.580	0.675	0.751	0.324	0.125	0.853	0.958	0.540	0.346

The control cases, which involved calibration against the observed discharge data, show that the PDM model was able to adequately predict the observed discharge with high $R \geq 0.77$ and $NSE \geq 0.58$, and thus both the study catchments are suitable for testing the new calibration schemes. The high values of correlation coefficients for the two study catchments mean that the PDM can simulate the trends of the streamflow; and the relatively high values of NSE indicate that the PDM has satisfactory skills in reproducing the overall behaviour of the observed streamflow time series. Some multiplicative and additive biases in the control case can be noted by examining the ratios of mean values and standard deviations between Q_{cr} and Q_{obs} (Table 3). The values represent the control cases' underestimation of the whole volume and variation of the observed discharge.

Table 3. Ratios of mean values and standard deviations between Q_{cr} and Q_{obs} . Biases correspond to deviation of the ratios from unity.

Catchment ID	$\frac{\text{mean}(Q_{cr})}{\text{mean}(Q_{obs})}$	$\frac{\text{stdev}(Q_{cr})}{\text{stdev}(Q_{obs})}$	$\frac{\text{mean}(Q_{SRC})}{\text{mean}(Q_{obs})}$	$\frac{\text{stdev}(Q_{SRC})}{\text{stdev}(Q_{obs})}$	$\frac{\text{mean}(Q_{IRC})}{\text{mean}(Q_{obs})}$	$\frac{\text{stdev}(Q_{IRC})}{\text{stdev}(Q_{obs})}$
604053	0.828	0.757	2.830	1.770	1.184	1.167
G8110004	0.577	0.849	0.295	0.364	1.128	1.443

4.1. Performance of the new calibration schemes

The calibration results from both the SRC- and IRC-based schemes are compared to the observed discharge data (Q_{obs}) and modelled discharge from the control case (Q_{cr}) to check whether the methods could simulate the streamflow (Table 2).

Both catchments show good linear relationships ($R \geq 0.68$) between the modelled and the observed discharge and even better relationships between the control case modelled discharge and the SRC and IRC modelled discharge. The hydrographs of the verification results (Figure 4) show that the modelled discharge simulates every runoff event and can properly fit their timings and trends. However, the results show considerable biases between the estimated and the observed data resulting in the low values of NSE in Table 2. In addition, ratios of mean values and standard deviations in Table 3 are considerably larger than the control case. This is in fact expected because, for the control case, the observational data obviously contain information on the true dynamic range of discharge. This information enables minimization of biases in simulated discharge via NSE-based optimization. However, when only river stage data were used in the SRC-based method, the simulated discharge will match the trend of the water level rather than its magnitude

because the information on the absolute range of discharge is absent. For the IRC-based scheme, the added analytical function may provide a slightly stronger relationship between discharge and water level. However erroneous estimates of the IRC parameters can also lead to biases.

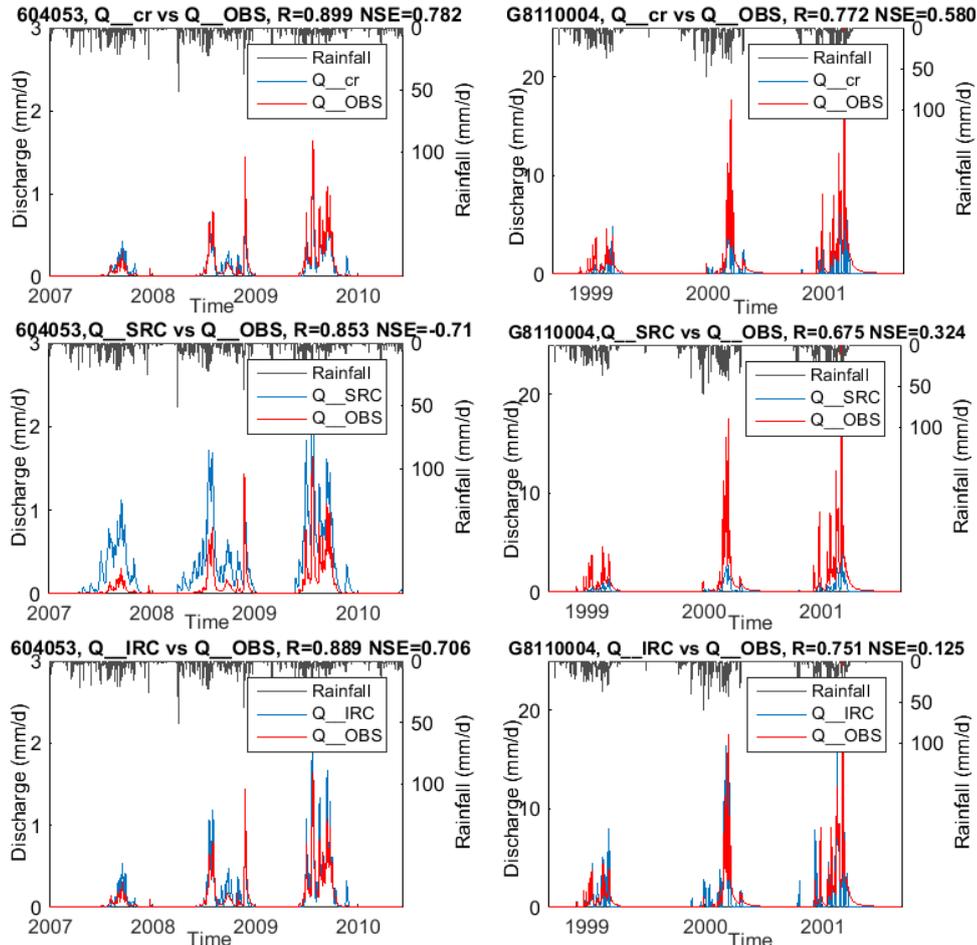


Figure 4. Observed and simulated streamflow hydrographs for both catchments. The figure displays three years’ verification results by control case, SRC-based and IRC-based schemes.

Table 4. Calibrated parameters from control case and the new schemes. Parameters which have more than -50% or +100% deviations from the control case are marked in red.

Parameter group	Parameter	604053			G8110004		
		Control case	SRC-based scheme	IRC-based scheme	Control case	SRC-based scheme	IRC-based scheme
Probability-distributed store	C_{max}	977.3	856.2	939.2	681.8	681.1	334.4
	b	0.801	1.476	1.365	1.478	0.467	1.614
	$C_{min}Ratio$	0.118	0.020	0.115	0.050	0.015	0.031
Evaporation function	b_e	20.6	13.4	164.4	316.5	87.68	387.4
	b_g	351.0	653.2	514.8	154.4	471.6	355.5
Recharge function	k_g	1506.6	1472.2	775.2	1799.1	1382.2	452.5
	$SRatio$	0.823	0.829	0.363	0.652	0.880	0.913
	k_b	3.26	27.63	10.42	14.45	59.92	37.03
Groundwater storage routing	k_1	3.296	7.822	2.986	0.048	2.030	0.466
	k_2Ratio	0.232	0.012	0.058	0.936	0.217	0.222
Inverse rating curve function	a_0	0*	-	-9.80e-3	0*	-	-5.47e-3
	a_1	6.04e-7*	-	4.18e-6	3.17e-4*	-	5.72e-4
	a_2	1.92*	-	1.739	1.32*	-	1.37

*derived from the observed rating curve

The calibrated model parameters from control cases and new calibration schemes are shown in Table 4. The similarity of the parameters is an important indicator of suitability of the new calibration schemes, while the dissimilarity does not necessarily indicate poor performances. The new calibration schemes, in comparison to

the control case, are less effective in simulating the parameters that represent the groundwater storage routing and surface routing. These mismatches of the new schemes contribute to the biases, which are significantly larger than the control case, between the estimated data and the observed data. Compared to the SRC-based scheme, the IRC-based scheme can generate calibrated parameters closer to the control case.

To mitigate the biases in the SRC- and IRC-based methods, there are two prospective approaches. First, in order to correct the biases in the modelled discharge relative to the observed data, an additional parameter, conceptually representing effective catchment area, could be used to rescale the calculated data. This parameter could be added in the calibration processes as a universally transferrable parameter based on a regionalized regression relationship between discharge and catchment area. Second, the simulated discharge data could be anchored to a small number of observed discharge values through a multi-objective optimization. Of course, these approaches will require additional *in situ* discharge measurements to supply information on absolute discharge volume. However it is important to note that the observational needs are arguably less extensive and costly than the current practice of defining an entire rating curve. In future work, we are extending the methods to more catchments and will attempt the bias-correction for all of those.

4.2. Spearman Rank correlation scheme vs inverse rating curve function scheme

The relative performance and benefits of using the SRC-based scheme and the IRC-based scheme are summarized in Table 5. Considering both the *R* and NSE metrics in Table 2, the latter has better performance in reproducing temporal variability of observed discharge and simulated discharge of the control case. The two calibration schemes have both pros and cons in the aspects of performance and calibration process.

Table 5. Comparison of SRC based scheme and inverse rating curve based scheme.

Items	SRC based scheme	Inverse rating curve based scheme
Calibration process	<ul style="list-style-type: none"> Fewer parameters (11 parameters) Less calculation time (about 1.5 hours*) Easy to obtain the global optima 	<ul style="list-style-type: none"> More parameters (14 parameters) More calculation time (about 2.5 hours*) Hard to obtain convergence and global optima
Calibration performance	<ul style="list-style-type: none"> Slightly lower performance: lower linear relationship with the observation, larger biases 	<ul style="list-style-type: none"> Better performance: higher linear relationship with the observation, smaller biases

*Based on the computation by Intel Core i7-3370@3.4GHz, RAM – 8.00GB

5. CONCLUSION

To overcome the challenges of lacking discharge data for modelling streamflow prediction, two schemes that use water level data in the hydrological model calibration are explored. The first scheme utilizes non-linear Spearman Rank correlation between modelled discharge and the observed water level. The second scheme adopts an inverse rating curve as a part of the streamflow model, making the model output comparable to the observed water level.

According to the results, both of these schemes provided reliable predictions of the temporal variability of runoff events as the modelled discharge shows strong linear correlation with the observation data (*R* values of modelled and observed discharge range from 0.675 to 0.889). The inverse rating curve based scheme exhibited better performance. However, the schemes have poor performances in fitting the observed discharge magnitudes. The reasons for the biases are likely to be inherent to the calibration schemes. The rank correlation method is more sensitive to the rank of the data than to the differences between modelled and observed magnitudes. The inverse rating curve function has large uncertainty in the calibration process, particularly in matching the shape of the power function using only observed and modeled water levels.

Future work will focus on improving the calibration results. It will investigate developing methods to correct the biases of the new method using catchment attributes or a small number of discharge observations. Moreover, more catchments with variable characteristics should be tested to uncover possible underlying meteorological and hydrological influences.

However, there are still some limitations in applying the altimeter water level data to the proposed calibration schemes. Firstly, the repeat frequency of some satellite data is large (i.e. 35-days repeat cycle for satellite ENVISAT) (Medina et al., 2008), thus it is hard to predict daily and weekly discharges. Secondly, as the size of footprint is usually large (~ a few hundred meters), these data are mainly suitable for large water bodies and may bring errors in the modelling of smaller rivers (Leon et al., 2006).

The application of water level data directly in the hydrological model calibration is proved promising in streamflow prediction. After the bias issue and the limitations of satellite altimeter data are addressed, future use of the altimeter water level data in discharge estimation in ungauged areas will be investigated.

REFERENCES

- Alsdorf, D., Lettenmaier, D., & Vörösmarty, C. (2003). The need for global, satellite-based observations of terrestrial surface waters. *Eos (0096-3941)*, 84(29), 269.
- Bjerklie, D. M., Lawrence Dingman, S., Vörösmarty, C. J., Bolster, C. H., & Congalton, R. G. (2003). Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, 278, 17-38. doi: 10.1016/S0022-1694(03)00129-X
- Dottori, F., Martina, M., & Todini, E. (2009). A dynamic rating curve approach to indirect discharge measurement. *Hydrology and Earth System Sciences*, 13(6), 847-863.
- Duan, Q., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, 76(3), 501-521.
- Getirana, A. C. (2010). Integrating spatial altimetry data into the automatic calibration of hydrological models. *Journal of Hydrology*, 387(3), 244-255.
- Glantz, S. A. (2012). The Spearman Rank Correlation Coefficient. *Primer of Biostatistics*.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2), 80-91.
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4), 751-763.
- Jung, H. C., Hamski, J., Durand, M., Alsdorf, D., Hossain, F., Lee, H., . . . Hoque, A. (2010). Characterization of complex fluvial systems using remote sensing of spatial and temporal water level variations in the Amazon, Congo, and Brahmaputra Rivers. *Earth Surface Processes and Landforms*, 35(3), 294-304.
- Leon, J. G., Calmant, S., Seyler, F., Bonnet, M.-P., Cauhopé, M., Frappart, F., . . . Fraizy, P. (2006). Rating curves and estimation of average water depth at the upper Negro River based on satellite altimeter data and modeled discharges. *Journal of Hydrology*, 328(3), 481-496.
- Lørup, J. K., Refsgaard, J. C., & Mazvimavi, D. (1998). Assessing the effect of land use change on catchment runoff by combined use of statistical tests and hydrological modelling: case studies from Zimbabwe. *Journal of Hydrology*, 205(3), 147-163.
- Medina, C. E., Gomez-Enri, J., Alonso, J. J., & Villares, P. (2008). Water level fluctuations derived from ENVISAT Radar Altimeter (RA-2) and in-situ measurements in a subtropical waterbody: Lake Izabal (Guatemala). *Remote Sensing of Environment*, 112(9), 3604-3617.
- Moore, R. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences Discussions*, 11(1), 483-499.
- Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. Asabe*, 50(3), 885-900.
- Refsgaard, J. C., van der Sluijs, J. P., Brown, J., & van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, 29(11), 1586-1597.
- Smith, L. C., Isacks, B. L., Bloom, A. L., & Murray, A. B. (1996). Estimation of discharge from three braided rivers using synthetic aperture radar satellite imagery: Potential application to ungaged basins. *Water Resources Research*, 32(7), 2021-2034.
- Smith, L. C., Isacks, B. L., Forster, R. R., Bloom, A. L., & Preuss, I. (1995). Estimation of discharge from braided glacial rivers using ERS 1 synthetic aperture radar: First results. *Water Resources Research*, 31(5), 1325-1329.
- Vörösmarty, C., Askew, A., Grabs, W., Barry, R., Birkett, C., Döll, P., . . . Kitaev, L. (2001). Global water data: A newly endangered species. *EOS, Transactions American Geophysical Union*, 82(5), 54-58.