

Hydrological utilization of satellite precipitation estimates in a data-scarce lake region

Tengfei Hu, Jingqiao Mao, Peipei Zhang, Diandian Xu, Weiyu Chen and Huichao Dai

ABSTRACT

In developing regions, accurate rain gauge measurements and satellite precipitation estimates that effectively capture rainfall spatial variability are promising sources of rainfall information. In this study, the latest Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) research product, 3B42V7, was validated against ground measurements in the region surrounding the Dongting Lake in China. In the subsequent model-based evaluation and comparison, the two precipitation datasets were separately included as the inputs for data-driven predictive models of the daily Dongting Lake level. The results show that (i) the daily 3B42V7 agrees well with the gauge measurements (correlation coefficient: 0.64–0.73); (ii) 3B42V7 underestimates the frequency of low-intensity (0–30 mm/day) rainfall and the contribution of low-intensity rainfall to the total rainfall volume, but slightly overestimates those of more intense rainfall; (iii) the lake level models driven by rainfall data from the two sources have similar performance, highlighting the potential of using 3B42V7 in data-driven modeling and prediction of hydrological variables in data-scarce regions; and (iv) the inclusion of rainfall as the model input helps achieve a balance between underestimation and overestimation of the lake levels in terms of both magnitude and quantity.

Key words | hydrological utility, lake level prediction, precipitation product, rain gauge measurement, support vector regression

Tengfei Hu
Jingqiao Mao (corresponding author)
Peipei Zhang
Diandian Xu
Weiyu Chen
Huichao Dai
College of Water Conservancy and Hydropower
Engineering,
Hohai University,
Nanjing 210098,
China
E-mail: maojq@hhu.edu.cn

INTRODUCTION

Reliable rainfall information is a prerequisite for effective regional water resources management (Kanakoudis *et al.* 2016; Kanakoudis *et al.* 2017). In developing regions, conventional rain gauge networks can provide relatively accurate point-based rainfall data (Mitra *et al.* 2009; Luo *et al.* 2016). Given that increasing the network density for more detailed monitoring is often infeasible in these regions, estimation of rainfall fields can be achieved by either spatial interpolation methods or remote sensing techniques (Kurtzman *et al.* 2009).

Methods that interpolate rain gauge measurements over the region of interest have long been studied (e.g., Wu *et al.* 2016). When applied to data-scarce regions,

these methods usually incorporate additional land surface information (e.g., elevation) to account for the scarcity of rain gauges. Even so, their practical applications suffer from several drawbacks: (i) the scarcity of rain gauges leads to rough estimates of the actual rainfall fields (Collischonn *et al.* 2008); and (ii) the performance of the interpolation method being used is difficult to evaluate because of the lack of a representative subnetwork of rain gauges for validation (Wagner *et al.* 2012).

The development of measurement techniques, especially remote sensing, has improved the availability of fine-resolution precipitation products since the 1980s (Singh &

Woolhiser 2002). Among the advanced measurement techniques, ground-based radars generally have non-continuous coverage and an uneven distribution across the world (Collischonn *et al.* 2008). Satellite-based techniques, in contrast, have reached a good level of maturity in recent years (Kidd & Levizzani 2011). For example, they can provide gridded precipitation retrievals that have near-global coverage.

In general, both gauge measurements, which are relatively accurate, and satellite estimates, which well describe rainfall spatial variability, are promising rainfall data sources for data-scarce regions. It is thus interesting to investigate their relative applicability and usefulness in such places. To compare the two rainfall data sources, a commonly used method is to validate the satellite estimates at different spatiotemporal scales against the gauge measurements that are considered the ground truth (e.g., Prakash *et al.* 2015). Another possible way is to evaluate and compare the two precipitation datasets based on their respective skills in driving hydrological models (e.g., Li *et al.* 2012). However, point-based gauge measurements cannot be evaluated by conceptual and physically based hydrological models directly, as these models are forced by rainfall fields. Some sort of interpolation operation is needed to obtain the spatial distribution of rainfall intensity from the point measurements, which may mask the original skill of the gauge data. Data-driven models, in contrast, can readily incorporate each of the two sources of rainfall data as the model input (e.g., Campolo *et al.* 2003; Akhtar *et al.* 2009).

The primary objective of this paper is to determine the better source of rainfall data from gauge measurements and satellite estimates for a data-scarce region through a two-step process. Validating the latter source against the former is conducted first, followed by the data-driven prediction of water level fluctuations in a large lake in the study region. The two precipitation datasets are separately included as the lake level model inputs and are then evaluated and compared based on the accuracy and reliability of the corresponding models. Furthermore, special attention is paid to the potential benefit of including rainfall as the additional input of the lake level model, which could be driven purely by hydrological data.

STUDY AREA AND DATA

Study area

The study area (27.7°–30.5° N, 110.8°–114.3° E) lies in the northeast part of the Dongting Lake Basin in the Yangtze River floodplain (Figure 1(a)). It covers an area of approximately 58,600 km², accounting for 20% of the basin. This area, together with its four inflowing rivers (i.e., the Xiang River, Zi River, Yuan River and Li River), drains via the Dongting Lake into the Yangtze River (Figure 1(b)). The spatial extent of the study area is determined by removing the contributing areas of hydrological stations #1 – #4 on the four rivers (i.e., stations Xiangtan, Taojiang, Taoyuan and Shimen) from the Dongting Lake Basin.

The Dongting Lake (~2,600 km² in the wet season) is the second largest freshwater lake in China. It is of great importance to its surrounding area for water supply, irrigation and fisheries (Mao *et al.* 2016). Apart from the four major tributaries, the Dongting Lake also receives water from the Yangtze River (Lai *et al.* 2014).

The temperature in this area seldom drops below 0 °C (probability less than 1% based on the data from 2009 to 2012); therefore, the terms ‘rainfall’ and ‘precipitation’ are not distinguished from each other in this paper.

Data acquisition

Rain gauge measurements

Nine rain gauges, numbered 1 to 9, are present within or near the study area (Figure 1(c)). Daily precipitation data at these sites from 2009 to 2012 have been collected from the National Meteorological Information Center (NMIC) of the China Meteorological Administration (CMA).

Satellite precipitation estimates

The satellite precipitation product used in this paper is the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) research product (<http://trmm.gsfc.nasa.gov/>). TMPA combines many satellite measurements and provides precipitation products at 0.25° × 0.25° with a 3-hour resolution at a quasi-global

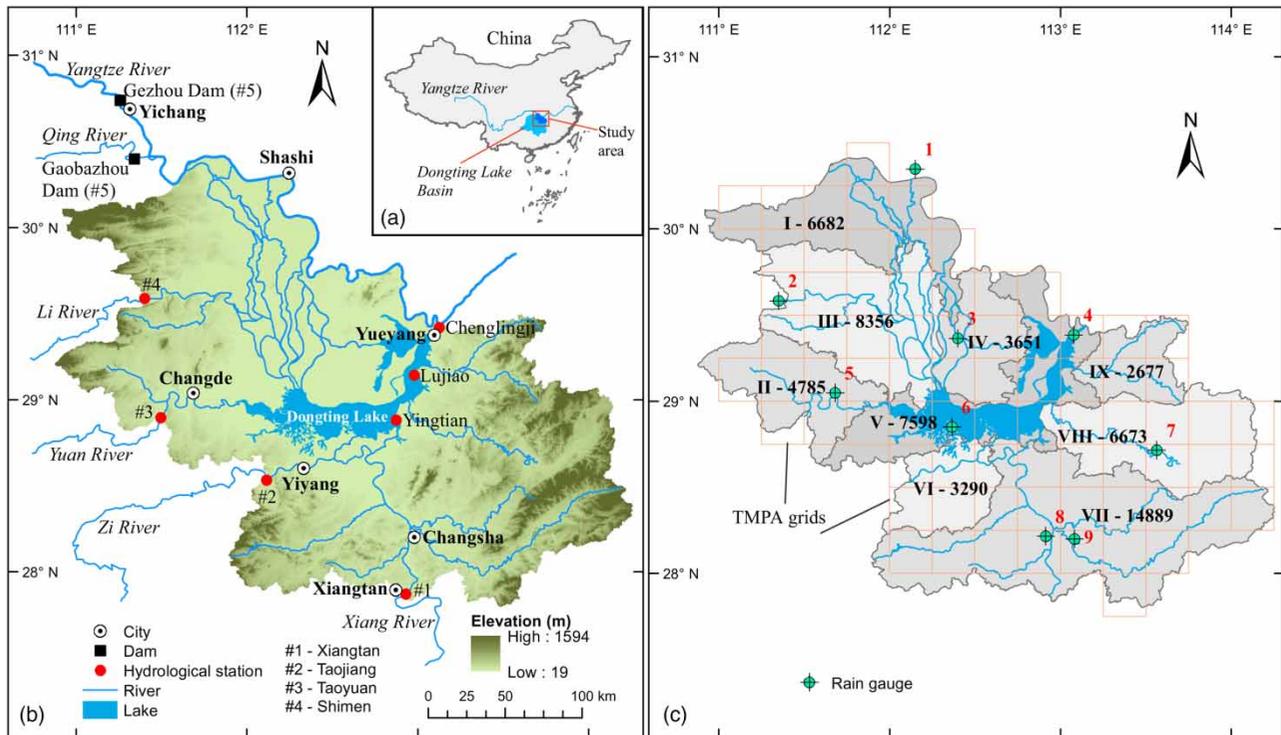


Figure 1 | Map of the study area, showing its (a) location, (b) topography and river-lake system, and (c) rain gauge distribution, delineation of subbasins (the area of each subbasin is given in km²), and the 0.25° × 0.25° TMPA grids.

scale (Huffman *et al.* 2007; Yan *et al.* 2017). The latest TMPA research product, 3B42V7, was released in 2012, and the historical measurements since 1998 were retrospectively processed.

Figure 1(c) shows 77 TMPA grid boxes with centroids falling within the study area. Daily precipitation estimates were generated for these boxes for the period from 2009 to 2012 by accumulating the 3-hourly 3B42V7.

Hydrological data

The water levels of the Dongting Lake and the discharges of the relevant rivers from 2009 to 2012 have been collected for the lake level modeling. The Dongting Lake levels were measured at the three lake stations shown in Figure 1(b) (i.e., Chenglingji, Lujiao and Yingtian) at 8:00 every day. The mean daily discharges of the four major tributaries to the lake were obtained at river stations #1 – #4. The daily outflow of the Gaobazhou Dam on the Qing River was added to the daily release of the Gezhou Dam on the

Yangtze River to jointly reflect the Yangtze River discharge (at a ‘virtual station’, #5).

METHODS

Problem formulation

The water level of the Dongting Lake can be considered a function of the time-lagged discharges of the major lake tributaries and the Yangtze River, as well as the antecedent lake levels:

$$y_t = f(D_{t-m_1}^1, D_{t-m_1-1}^1, \dots, D_{t-n_1}^1, \dots, D_{t-m_5}^5, D_{t-m_5-1}^5, \dots, D_{t-n_5}^5, L_{t-m_6}, L_{t-m_6-1}, \dots, L_{t-n_6}) \quad (1)$$

where y_t is the water level at a lake station measured on day t ; m_i and n_i ($i = 1, \dots, 6$) are the minimum and maximum time lags (in days), respectively; D_{t-j}^i ($i = 1, \dots, 5; j = m_i, \dots, n_i$) is the daily discharge at river station # i with a time lag of j day(s);

and L_{t-j} ($j= m_6, \dots, n_6$) is the water level at the same lake station measured j day(s) before day t .

When the rainfall data of the study area are taken into consideration, the Dongting Lake level can be expressed as follows:

$$y_t = f(D_{t-m_1}^1, D_{t-m_1-1}^1, \dots, D_{t-n_1}^1, \dots, D_{t-m_5}^5, D_{t-m_5-1}^5, \dots, D_{t-n_5}^5, L_{t-m_6}, L_{t-m_6-1}, \dots, L_{t-n_6}, \mathbf{R}) \tag{2}$$

where \mathbf{R} represents the input of daily rainfall.

Rainfall input $\mathbf{R} = (R_{t-p_1}^1, R_{t-p_1-1}^1, \dots, R_{t-q_1}^1, \dots, R_{t-p_N}^N, R_{t-p_N-1}^N, \dots, R_{t-q_N}^N)$, where p_i and q_i ($i = 1, \dots, N$) are the minimum and maximum time lags (in days), respectively; R_{t-j}^i ($i = 1, \dots, N; j = p_i, \dots, q_i$) is the daily rainfall volume on day $t-j$ either measured at the i th rain gauge (for gauge measurements) or averaged over the i th subbasin (for satellite estimates).

Figure 1(c) shows the delineation of the drainage areas of the Dongting Lake tributaries. The study area was divided into nine subbasins numbered I to IX, with the area of each subbasin given in km^2 . The daily rainfall volumes (on day $t-j$) of the grid boxes with centroids falling within subbasin i were then averaged to obtain R_{t-j}^i .

Support vector regression (SVR)

In this study, the regression function of Equation (1) or Equation (2) was estimated with support vector regression (SVR). SVR is based on the statistical learning theory by Vapnik (1998) and uses structural risk minimization (SRM), rather than empirical risk minimization (ERM) that is used in most conventional artificial neural networks (ANNs) (Gunn 1998). SVR has been successfully applied in modeling environmental and water resources variables (Maier et al. 2010).

SVR can be viewed as a universal approximator. Consider the dataset $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l)\}$, where $\mathbf{x}_i \in R^n$ is the input vector and $y_i \in R^1$ is the target output ($i = 1, \dots, l$). The input-output relationship of dataset D can be approximated with the following function:

$$f(\mathbf{x}) = \boldsymbol{\omega}^T \phi(\mathbf{x}) + b \tag{3}$$

where $\boldsymbol{\omega}$ is the weight vector, $\phi(\mathbf{x})$ is an embedding map that projects \mathbf{x} into a higher dimensional feature space where linear regression can be performed, and b is the bias.

Based on the ϵ -insensitive loss function, the optimal regression function is given by minimizing the regularized risk function (Vapnik 1998):

$$\begin{aligned} \min_{\boldsymbol{\omega}, b, \xi, \xi^*} & \frac{\boldsymbol{\omega}^T \boldsymbol{\omega}}{2} + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t. } & \boldsymbol{\omega}^T \phi(\mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i \\ & y_i - \boldsymbol{\omega}^T \phi(\mathbf{x}_i) - b \leq \epsilon + \xi_i^* \quad i = 1, \dots, l \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{4}$$

where C is the regularization parameter ($C > 0$) and the slack variables ξ_i and ξ_i^* are the lower and upper excess deviations, respectively.

Normally, the Lagrange multiplier technique is used to transform Equation (4) to the dual problem, which is easier to solve:

$$\begin{aligned} \max_{\alpha, \alpha^*} & \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) - \epsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i) \\ & (\alpha_i^* - \alpha_j) K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t. } & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq C \quad i = 1, \dots, l \end{aligned} \tag{5}$$

where α_i^* and α_i are Lagrange multipliers and $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is the kernel function.

By solving Equation (5), the regression function is as follows:

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(\mathbf{x}_i, \mathbf{x}) + b. \tag{6}$$

Among the commonly used kernels, the Gaussian radial basis function has attracted significant attention. This function is also adopted in this study in the following form:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{7}$$

where $\gamma > 0$ is the kernel parameter that determines the width of the kernel.

Input variable selection (IVS) and SVR parameter optimization

This study employed a synchronized search to determine the optimal time lags of the input variables and SVR parameters (i.e., ϵ , C and γ). The search was implemented in a performance-oriented manner using a genetic algorithm (GA), and SVR was incorporated to assess each individual (a combination of variable time lags and SVR parameters). Furthermore, the 5-fold cross validation was used in the model training to avoid overfitting.

Model comparison

Two Dongting Lake level models (Models 1 and 2) driven by gauge measurements and satellite estimates, respectively, were developed. The better source of rainfall data was determined by comparing the performance and reliability of the two models. In addition, Model 0, which was not driven by any rainfall input (Equation (1)), was also developed as the benchmark model. Models 1 and 2 were compared with the benchmark model to explore the benefit of including rainfall as the additional model input.

Evaluation metrics

Multiple statistical metrics were used to characterize the differences between gauge measurements and TMPA 3B42V7, including Pearson correlation coefficient (CC), relative bias (RB), root mean square error (RMSE) and

mean absolute error (MAE). Moreover, the performance of different lake level models was characterized via the RMSE and MAE.

RESULTS AND DISCUSSION

Satellite precipitation validation

TMPA 3B42V7 over the period from January 2009 to December 2012 was validated against the gauge measurements. Rainfall data at gauge 1 were not considered in the validation because this gauge is not covered by any selected TMPA grid box (see Figure 1(c)).

According to Table 1, the daily precipitation rates of 3B42V7 show an acceptable level of agreement with the rates of gauge measurements, with CC values ranging between 0.64 and 0.73 for the eight gauges. 3B42V7 slightly overestimates rainfall at two of the eight gauges, with RB values of 8.23% at gauge 4 and 6.76% at gauge 6, while the systematic bias is not obvious at the other six gauges. Compared with the daily validation, the 3B42V7 estimates agree better with the gauge data in terms of CC, RMSE and MAE when validated on a monthly basis. For example, the range of CC values increases to 0.87–0.95 for the monthly validation. Moreover, the monthly RMSE values are less than 1.60 mm/day, much smaller than the corresponding daily values (≥ 8.50 mm/day). Similarly, the maximum monthly MAE is 1.01 mm/day, which is lower than the minimum daily value of 2.92 mm/day.

Table 1 | Validation results of precipitation rates of TMPA 3B42V7

Location	Daily				Monthly			
	CC	RB (%)	RMSE (mm/day)	MAE (mm/day)	CC	RB (%)	RMSE (mm/day)	MAE (mm/day)
Gauge 2	0.72	1.74	9.01	2.92	0.88	1.75	1.57	1.01
Gauge 3	0.64	1.21	9.22	3.01	0.88	1.44	1.23	0.84
Gauge 4	0.73	8.23	8.53	3.32	0.95	8.29	0.97	0.69
Gauge 5	0.73	4.00	8.50	3.01	0.92	4.18	1.03	0.77
Gauge 6	0.67	6.76	8.73	3.48	0.87	6.93	1.37	0.85
Gauge 7	0.65	3.80	9.50	3.35	0.91	3.89	1.23	0.95
Gauge 8	0.68	-1.84	9.42	3.42	0.95	-1.89	0.88	0.67
Gauge 9	0.69	0.69	9.00	3.43	0.92	0.68	1.13	0.79

Figure 2 shows the occurrence frequencies of rainfall events with different intensities over time for both gauge measurements and TMPA 3B42V7, and the relative contributions of these events to the total rainfall volume. According to the left two bars indicating non-rain frequencies, 3B42V7 has difficulty in detecting all rainfall events. More specifically, 3B42V7 underestimates the frequency of low-intensity (0–30 mm/day) rainfall events (by ~11%), especially those with rates below 5 mm/day. However, the number of medium- to high-intensity (>30 mm/day) rainfall events is overestimated by the TMPA product. In keeping with such a trend, the gauge measurements have a larger proportion of low-intensity rainfall volume than 3B42V7, and 3B42V7 has a larger proportion of more intense rainfall volume than the gauge measurements. The above findings agree well with the results that Wang *et al.* (2016) obtained in the Xiang River Basin.

Dongting Lake level prediction

In this study, the years 2010 and 2012 (wetter) were selected as the training period, while the years 2009 and 2011 (drier) were selected as the testing period.

Model performance

Models 0–2 were independently built and calibrated for stations Chenglingji, Lujiao and Yingtian, respectively. Figure 3 shows the RMSEs and MAEs of Models 0–2 at

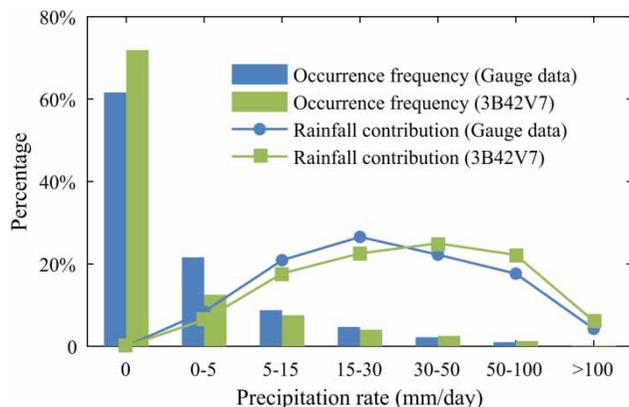


Figure 2 | Occurrence frequencies of rainfall events with different intensities over time for gauge measurements and TMPA 3B42V7, and the relative contributions of these events to the total rainfall volume.

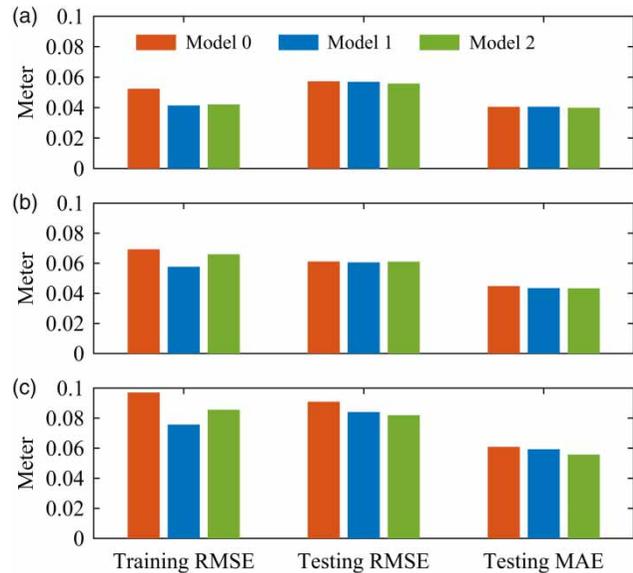


Figure 3 | Performance of the lake level models at (a) Chenglingji, (b) Lujiao and (c) Yingtian.

the three lake stations. In the model training phase, Model 1 (driven by gauge measurements) outperforms Model 2 (driven by 3B42V7 estimates), especially at Lujiao and Yingtian. However, the performance differences between them become marginal in terms of both RMSE and MAE in the testing phase. Models fed with the additional rainfall input (Models 1 and 2) clearly have smaller RMSEs than the benchmark model (Model 0) in the model training phase at all sites. In the testing phase, however, the benefit of incorporating rainfall data in the lake level prediction is not evident (except at Yingtian).

Prediction errors and bias

The prediction errors of Models 0–2 at the three lake sites are plotted in Figure 4. Model 0 tends to produce serious underestimates of the observed lake levels at all sites. This finding most likely arises from the fact that the effects of rainfall are ignored in the lake level modeling and prediction.

In terms of reducing the underestimation of the Dongting Lake levels, Model 1 (gauge measurements) has a marginal advantage over Model 2 (3B42V7 estimates), especially at Lujiao and Yingtian. The most serious underestimates of the observed lake levels produced by Model 1 are -0.43 m and -0.42 m at Lujiao and Yingtian, respectively, slightly

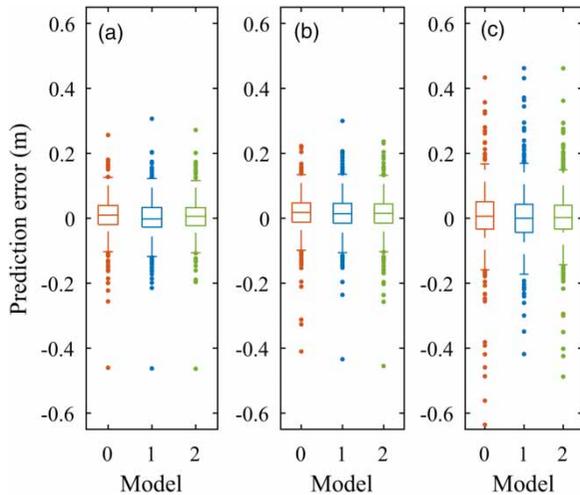


Figure 4 | Box plots of lake level prediction errors (predicted minus observed) in the testing period at (a) Chenglingji, (b) Lujiao and (c) Yingtian.

better than the corresponding values of Model 2 (-0.45 m and -0.49 m, respectively).

Figure 4 also shows that, once the precipitation data are included as the model input, the resulting models (Models 1 and 2) achieve a balance between underestimation and overestimation of the lake levels in terms of magnitude.

Figure 5 shows the proportions of overestimates and underestimates produced by Models 0-2 in the testing period. Model 0 yields more overestimates than underestimates at the three sites. This pattern can be partly

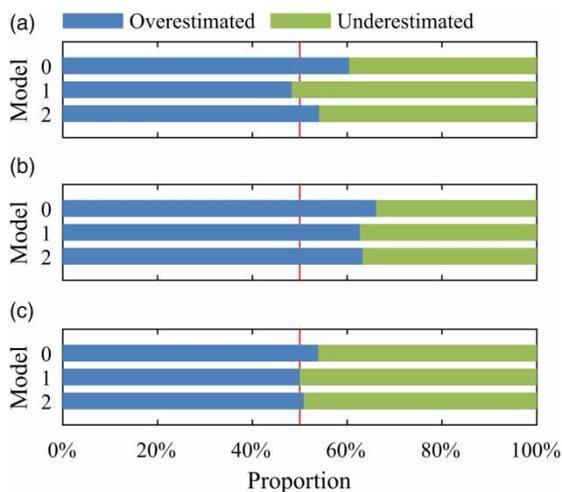


Figure 5 | Proportions of overestimated and underestimated lake levels in the testing period at (a) Chenglingji, (b) Lujiao and (c) Yingtian. The red line in each subplot indicates that 50% of the lake levels are overestimated. Please refer to the online version of this paper to see this figure in color: <http://dx.doi.org/10.2166/ws.2017.223>.

attributed to the performance-oriented model calibration with rainfall ignored in the lake level modeling. Rises in the Dongting Lake level caused by rainfall in the study area lead to underestimation of the lake levels. Therefore, the calibrated model increases the predicted lake levels to reduce the underestimation. As the RMSE is sensitive to large errors, the increases tend to cater for serious underestimates arising from extreme rainfall, which finally results in occasional large underestimates and overall slight overestimates.

Model 1 (gauge measurements) outperforms Model 2 (3B42V7 estimates) in terms of balancing the numbers of overestimated and underestimated lake levels. However, at all three stations, the advantage of Model 1 is very slight.

Figure 5 also shows that incorporating rainfall in the study area helps correct the prediction bias of the lake level models. The bias correction is most evident at station Chenglingji, where the proportion of overestimated lake levels decreases from 60.42% for Model 0 to 48.31% and 54.08% for Models 1 and 2, respectively.

CONCLUSIONS

This study has evaluated and compared the precipitation data from a sparse rain gauge network and the latest TMPA research product, 3B42V7, over the area surrounding the second largest freshwater lake in China, the Dongting Lake. Based on the results and discussion, the following conclusions have been reached. Daily 3B42V7 estimates agree satisfactorily with the gauge measurements, and the two datasets are much closer when compared on a monthly basis. 3B42V7 underestimates the frequency of low-intensity (0–30 mm/day) rainfall and its contribution to the total rainfall volume, but overestimates those of more intense rainfall. The two precipitation datasets perform similarly in terms of driving the lake level models, indicating the potential of using 3B42V7 as an alternative to gauge measurements in data-driven modeling and prediction of hydrological variables in data-scarce regions. The inclusion of rainfall as the additional input only slightly improves the lake level model. Therefore, the original model without rainfall may be preferable, given that its training process is relatively simple and merely requires hydrological data. However,

the benefit of taking rainfall into account is still noticeable and interesting. Such information helps strike a balance between underestimation and overestimation of the lake levels in terms of both magnitude and quantity.

This study was carried out before the Global Precipitation Measurement (GPM) mission products became available. Future research efforts can thus be directed toward the evaluation and comparison of gauge and GPM precipitation data.

ACKNOWLEDGEMENTS

This research is funded by the National Key Research and Development Program of China (2016YFC0402204), the National Natural Science Foundation of China (51379059) and the Fundamental Research Funds for the Central Universities (2015B34014 and 2015B15314). The support from the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) is also acknowledged.

REFERENCES

- Akhtar, M. K., Corzo, G. A., van Andel, S. J. & Jonoski, A. 2009 River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges river basin. *Hydrology and Earth System Sciences* **13** (9), 1607–1618.
- Campolo, M., Soldati, A. & Andreussi, P. 2003 Artificial neural network approach to flood forecasting in the River Arno. *Hydrological Sciences Journal* **48** (3), 381–398.
- Collischonn, B., Collischonn, W. & Tucci, C. E. M. 2008 Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology* **360** (1–4), 207–216.
- Gunn, S. 1998 *Support Vector Machines for Classification and Regression*. University of Southampton, UK.
- Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. & Stocker, E. F. 2007 The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology* **8** (1), 38–55.
- Kanakoudis, V., Tsitsifli, S., Papadopoulou, A., Curk, B. C. & Karleusa, B. 2016 Estimating the water resources vulnerability index in the Adriatic Sea Region. *Procedia Engineering* **162** (Supplement C), 476–485.
- Kanakoudis, V., Tsitsifli, S., Papadopoulou, A., Curk, B. C. & Karleusa, B. 2017 Water resources vulnerability assessment in the Adriatic Sea region: the case of Corfu island. *Environmental Science and Pollution Research* **24** (25), 20173–20186.
- Kidd, C. & Levizzani, V. 2011 Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences* **15** (4), 1109–1116.
- Kurtzman, D., Navon, S. & Morin, E. 2009 Improving interpolation of daily precipitation for hydrologic modelling: spatial patterns of preferred interpolators. *Hydrological Processes* **23** (23), 3281–3291.
- Lai, X. J., Liang, Q. H., Jiang, J. H. & Huang, Q. 2014 Impoundment effects of the Three-Gorges-Dam on flow regimes in two China's largest freshwater lakes. *Water Resources Management* **28** (14), 5111–5124.
- Li, X. H., Zhang, Q. & Xu, C. Y. 2012 Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. *Journal of Hydrology* **426**, 28–38.
- Luo, J. N., Lu, W. X., Ji, Y. F. & Ye, D. J. 2016 A comparison of three prediction models for predicting monthly precipitation in Liaoyuan city, China. *Water Science and Technology-Water Supply* **16** (3), 845–854.
- Maier, H. R., Jain, A., Dandy, G. C. & Sudheer, K. P. 2010 Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling & Software* **25** (8), 891–909.
- Mao, J. Q., Zhang, P. P., Dai, L. Q., Dai, H. C. & Hu, T. F. 2016 Optimal operation of a multi-reservoir system for environmental water demand of a river-connected lake. *Hydrology Research* **47** (S1), 206–224.
- Mitra, A. K., Bohra, A. K., Rajeevan, M. N. & Krishnamurti, T. N. 2009 Daily Indian precipitation analysis formed from a merge of rain-gauge data with the TRMM TMPA satellite-derived rainfall estimates. *Journal of the Meteorological Society of Japan* **87A**, 265–279.
- Prakash, S., Mitra, A. K., Momin, I. M., Pai, D. S., Rajagopal, E. N. & Basu, S. 2015 Comparison of TMPA-3b42 versions 6 and 7 precipitation products with gauge-based data over India for the southwest monsoon period. *Journal of Hydrometeorology* **16** (1), 346–362.
- Singh, V. P. & Woolhiser, D. A. 2002 Mathematical modeling of watershed hydrology. *Journal of Hydrologic Engineering* **7** (4), 270–292.
- Vapnik, V. 1998 *Statistical Learning Theory*. John Wiley & Sons, New York.
- Wagner, P. D., Fiener, P., Wilken, F., Kumar, S. & Schneider, K. 2012 Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *Journal of Hydrology* **464**, 388–400.

- Wang, J. L., Chen, H., Xu, C. Y., Li, X. Q. & Zeng, Q. 2016 [Evaluation of accuracy and streamflow simulation of TRMM satellite precipitation data](#). *Journal of Water Resources Research* **5**, 434–445 (in Chinese).
- Wu, L., Liu, X. & Ma, X. Y. 2016 [Spatio-temporal evolutions of precipitation in the Yellow River basin of China from 1981 to 2013](#). *Water Science and Technology-Water Supply* **16** (5), 1441–1450.
- Yan, J. B., Jia, S. F., Lv, A. F., Mahmood, R. & Zhu, W. B. 2017 [Analysis of the spatio-temporal variability of terrestrial water storage in the Great Artesian Basin, Australia](#). *Water Science and Technology-Water Supply* **17** (2), 324–341.

First received 19 July 2017; accepted in revised form 31 October 2017. Available online 13 November 2017