Deriving Daily Discharges from Satellite Radar Altimetry and Ensemble Learning Regression in Poorly Gauged River Basins

Donghwan Kim[1], Hyongki Lee[1], Faisal Hossain, Susantha Jaysinghe[2], Senaka Basnayake[2], Farrukh Chishtie[2], Duong Du Bui[3]

[1]Department of Civil and Environmental Engineering, University of Houston, [2]Asian Disaster Preparedness Centre and SEVIR-Mekong, Bangkok, Thailand, [3]National Center for Water Resources Planning and Investigation, Hanoi, Vietnam, [4]Department of Environmental Technology and Management, National University of Civil Engineering, Hanoi, Vietnam, [5]Water Resources Research Center, K-Water Institute

2018 AGU Fall Meeting Abstract Number: GC31K-1384

Introduction

Hydraulic variables obtained from remotely sensed data have been successfully used to estimate river discharge (Q). However, most studies have used a rating curve based on a single hydraulic variable or the Manning equation (multiplicative method). *Kim et al.* (2019) have applied the ensemble learning regression method for estimating river discharge (ELQ), which provided more accurate results than the previous methods based on a single rating curve (depth-discharge relationship). However, ELQ has been tested only over the Congo River and obtained Q at sparsely distributed dates (35-day interval of Envisat altimetry data). In this study, we aimed at estimating daily Q from 2003 to 2005 (3-year) using ELQ and linearly interpolated altimetry-derived water levels over the lower Mekong River as a case study.

Motivation

Monitoring Q using remotely sensed data involves obtaining hydraulic variables, such as water levels, inundation areas, river widths, and surface water slopes, and Q has been estimated by (1) the multiplicative method, which is the Manning equation or by (2) a single rating curve. A key limitation in using the two representative methods of estimating Q is that those methods used a single rating curve or multiple hydraulic variables at a single location. In other words, a lack of information may arise when we obtain estimates of Q using a single rating curve or the Manning equation. In contrast, when we combine several base learners, additional information provides more information of Q variations in time. Thus, the ELQ process leads to accurate Q compared to the previous method based on a single rating curve.

Although ELQ is a novel method to estimate Q, there is a limitation that ELQ should use in-situ Q for training base learners and obtaining weights. Moreover, to date, the ELQ has been applied to only the Congo River basin. Therefore, to apply the ELQ to ungauged or poorly gauged rivers, further investigations are required on ELQ. Using satellite altimetry data and ELQ, this work proposes daily discharge estimation in poorly gauged river basins. Ungauged (or poorly gauged) basins consist of several categories: (1) genuinely ungauged basins, which have not been gauged; (2) decommissioned basins due to reductions of budgets for measuring Q; (3) temporarily discontinued basins due to instrumental failures or errors. We aim to estimate Q in the above cases (2) and (3).

Ensemble Learning Regression (ELQ)

Ensemble learning indicates a series of procedure to train several functions and combine their results based on an integration rule, either in classification or regression problem (*Brown*, 2011). Typically, the ensemble process consists of two parts: ensemble generation and ensemble integration (*Zhou*, 2015). Some studies added an ensemble pruning between the ensemble generation and integration. The conceptual design of the ensemble learning process is shown in Figure 1.

$$f_{ELQ} = w_1 f_{i1}(X_{i1}) + w_2 f_{i2}(X_{i2}) + \cdots + w_j f_{ij}(X_{ij})$$

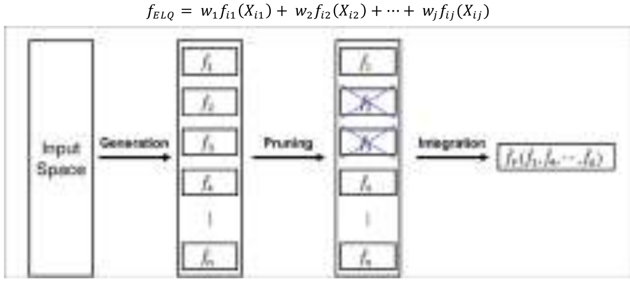

Figure 1. Conceptual design of the ensemble learning process (*Kim et al.*, 2019)

Study Area and Data

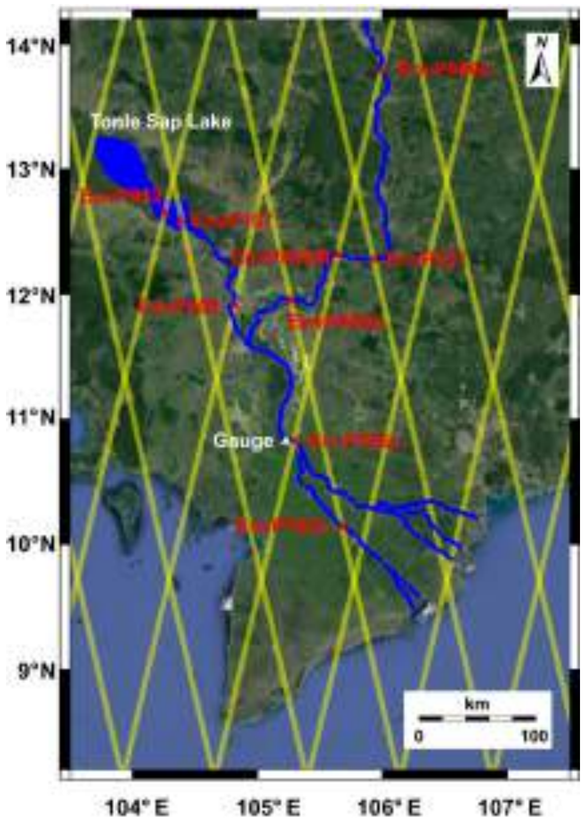

Figure 2. Map of the lower Mekong River with Envisat altimetry ground tracks. The white and red triangles indicate the in-situ gauge at Tan Chau (Latitude/Longitude: 10.801° N/105.248° E) and Envisat virtual stations, respectively. Yellow lines are Envisat altimetry ground tracks (map data ©2018 Google).

The study area of the lower Mekong River is shown in Figure 2. The Mekong River basin is the sixth largest in discharge (16,000 m³/s), and the river flows through six countries: China, Myanmar, Laos, Thailand, Cambodia, and Vietnam. In general, the water level of the river starts to rise in May, and it reaches a peak in October, and the low water season occurs in March and April. Figure 2 shows nine virtual stations over the lower Mekong River. The in-situ discharge data was provided by the Asian Disaster Preparedness Center (ADPC).

We used Envisat Radar Altimeter 2 (RA-2) Geophysical Data Record (GDR) of nine passes. We extracted ICE-1 retracked water levels from the nine passes. The instrument, ionosphere/troposphere, and geophysical corrections were applied.

Methods and Results

Firstly, altimetry-derived water levels from the nine virtual stations were linearly interpolated. Then candidate ensemble learners were generated based on a rating curve. Before the integrating process, several candidate ensemble learners (EnvP408, 866L, and 107) were pruned based on the Degree of compensation and performance of base learners. Finally, 15 combinations of ELQ results, which integrate two rating curves, were obtained among six base learners.

The ELQ-derived Q using two altimetry-derived water levels showed reduced RMSE of 711/763 m^3s^{-1} (RRMSE of 7.9/8.1%) on average for the training/validation datasets compared to the Q obtained using a single rating curve. It approximately corresponds to the mean annual discharges of Sacramento River (CA, USA).

Table 1. Error analysis of the best results of estimated Q using a rating curve and ELQ.

Method	RMSE (m^3s^{-1})	RRMSE (%)	R^2_{NS}
Rating curve (565U)	2348 / 2415	26.22 / 25.72	0.86 / 0.86
ELQ (952-866U)	1079 / 1089	12.05 / 11.60	0.97 / 0.97

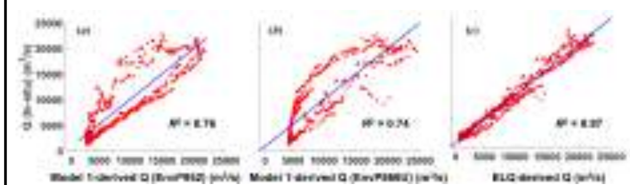

Figure 3. Performance comparison of (*a*) the Model 1-derived Q using EnvP952, (*b*) Model-derived Q using EnvP866U, and (*c*) ELQ-derived Q.

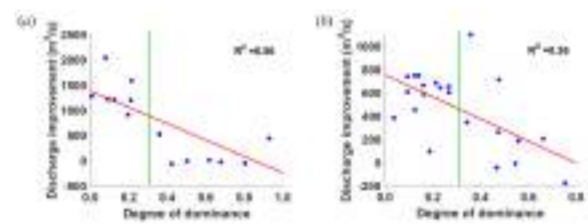

Figure 4. Degree of dominance (I_{DoD}) and Q improvements: (*a*) Mekong River and (*b*) Congo River. The Q improvement is calculated the RMSE of ($M_1 - f_{ELQ}$). The green vertical line indicates the boundary where I_{DoD} is 0.3.

Conclusions

(1) Our results indicate that ELQ outperforms the previous method based on a single rating curve over the Mekong River Basin

(2) We have developed an index termed I_{DoD} which represents the performance of ELQ. This finding might provide a clue of the weighting pattern in the ELQ process.

Acknowledgements

This study is supported by NASA Applied Sciences Program (NNX16AQ33G) for SWOT Science Team, SERVIR Program (NNX16AN35G), and GEO Program (80NSSC18K0423). Envisat altimetry data were provided by European Space Agency (ESA).