Bayesian merging of numerical modeling and remote sensing for saltwater intrusion quantification in the Vietnamese Mekong Delta

Anh Phuong Tran[1], Pham Van Chien[2], Phong V.V. Le[3], Thanh Thuy Nguyen[2], Manh Cuong Tran[1], Hai V. Pham[4]

[1]Water Resources Institute, 8 Phao Dai Lang, Dong Da, Hanoi, Vietnam

[2]Thuyloi University, 175 Tay Son, Dong Da, Ha Noi, Vietnam

[3]Faculty of Hydrology Meteorology and Oceanography, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam

[4]INTERA INC, 9600 Great Hills Trl Ste 300W, Austin, TX 78759, USA

Corresponding author: Anh Phuong Tran (phuongtran.monre@gmail.com)

Abstract: Saltwater intrusion has become one of the most concerning issues in the Vietnamese Mekong Delta (VMD) due to its increasing impacts on agriculture and food security of Vietnam. Reliable estimation of salinity plays a crucial role to mitigate the impacts of saltwater intrusion. This study developed a hybrid technique that merges satellite imagery with numerical simulations to improve the estimation of salinity in the VMD. The salinity derived from Landsat images and by numerical simulations was fused using the Bayesian inference technique. The results indicate that our technique significantly reduces the uncertainties and improves the accuracy of salinity estimates. The Nash-Sutcliffe coefficient is 0.73, which is much higher than that of numerical simulation

(0.69) and Landsat estimation (0.67). The correlation coefficient between the merged and measured salinity is relatively high (0.75). The variance of the ensemble salinity errors (2.57 ppt^2) is lower than that of Landsat estimation (3.65 ppt^2) and numerical simulations (8.69 ppt^2). The proposed approach in this study shows a great potential to combine multiple data sources of a variable of interest to improve its accuracy and reliability wherever these data are available.

1. Introduction

The Mekong river with a basin area of 795,000 km^2 and an annual volumetric flow of about 473 billion m^3 is one of the largest rivers in the world. The river plays a crucial role in water supply for agricultural production and is an important waterway for transportation and business in Southeast Asia. The Vietnamese Mekong Delta (VMD) region is located in the downstream of the Mekong river and the southern end of Vietnam. The VMD has a great potential in agriculture, fisheries, and mangrove and is considered as the largest "rice bowl" of Vietnam, contributing up to 50% of the rice crop, 65% of aquaculture, and 70% of fruit production. Hence, the region plays a key role in ensuring the food security of Vietnam. However, situated in the downstream of the Mekong River, the VMD is highly vulnerable to the socio-economic activities at the upstream and in the region itself such as the development of hydro-electrical dams (Pearse-Smith et al., 2012; Kuenzer et al., 2013), excessive groundwater extraction (Minderhoud et al., 2017) and sand mining (Brunier et al., 2014; Eslami et al., 2019). In addition to climate change and sea level rise (Toan et al., 2014), these anthropogenic activities have caused severe problems for the VMD, namely, land subsidence (e.g., Erban et al., 2014), river bank erosion (e.g.,

Lam-Dao et al., 2011; Brunier et al., 2014), land loss (e.g., Schmitt et al., 2017) and drought and saltwater intrusion (e.g., CGIAR, 2016). Among the abovementioned problems, saltwater intrusion is considered as one of the most severe issues, especially under the direct compound impacts of sea level rise, land subsidence, and reduction of upstream sediment.

Saltwater intrusion has caused severe losses for the VMD in recent years. Kotera et al. (2008) showed that rice cropping intensities decreased with increasing salinity levels. Based on the interview approach, Bergqvist et al. (2012) found that the most profound consequences of saltwater intrusion are the reduction of the crops and yields and fruit production. They also showed that saltwater intrusion also causes a shortage of drinking water supply for local people. There have been several mitigation measures to reduce the impacts of saline water. For example, many regions in the VMD converted from salinity-affected rice monoculture areas to integrated rice-fish or rice-shrimp aquaculture. However, this measure has negative impacts on the livelihoods of the farmers and the ecosystem (CGIAR, 2016). Another measure is to operate the sluice gates to take water from the river system to the irrigation channels when salinity is predicted at an acceptable level for drinking or irrigation. Nevertheless, these measures require quantitative information on salinity and saltwater water intrusion to better inform the decision-makers when each measure should be implemented.

Salinity is often determined by in situ measurement of electrical conductivity (EC). However, because this approach is time-consuming and labor-intensive, it cannot provide salinity with a high spatio-temporal resolution over a large spatial area. Recently, there have been increasing studies that use satellite imagery to estimate salinity for relatively

large area coverage and finer spatial resolution compared to the EC measurement. The theory underpinning this application of remote sensing is the negative correlation between salinity and the optical absorption coefficient of colored dissolved organic matter (CDOM), which can be derived from remote sensing. For example, Palacios et al. (2009) developed a linear multivariate relationship between salinity versus temperature and CDOM absorption (a_{CDOM}) for the Columbia River plume. After that, the salinity can be obtained from a_{CDOM} and temperature derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) images. Molleri et al. (2010) used the absorption coefficient for dissolved and detrital material from Sea-viewing Wide Field-of-view Sensor (SeaWiFS) to investigate the spatiotemporal variations of salinity in the Amazon River plume. Bai et al. (2013) estimated salinity in the Changjiang River from satellite- derived CDOM and investigated its inter-annual variation. Keith et al. (2016) used the Medium Resolution Imaging Spectrometer (MERIS) and the International Space Station (ISS) Hyperspectral Imager to estimate $a_{CDOM,}$ which were then used to determine the surface salinities in New England, Middle Atlantic, and Gulf Coast Estuaries. In the VMD, Nguyen et al. (2018) constructed several data-driven models between the reflectances from the Landsat-8 Operational Land Imager images and salinity levels measured in situ. A total of 103 observed samples were divided into 50% training and 50% testing datasets using Multiple Linear Regression, Decision Trees, and Random Forest algorithms. The study found that the random forest approach provided the best-fitted model. Recently, Tran et al. (2019) developed a new Landsat-based satellite salinity index to investigate the annual salinity variations and the relationship between these variations with drought.

Salinity can be also estimated by solving mathematical equations that simulate the dynamics of saline water in the river system using analytical or numerical solutions. For example, Nguyen and Savenije (2006) developed a predictive analytical approach for the quantification of salinity intrusion in multi-channel estuaries based on the tidal water level and discharge from the upstream, and then applied this approach to estimate the salinity in the VMD. Nguyen et al. (2010) also applied the abovementioned analytical approach for investigating the impacts of the upstream development scenarios on drought and salinity intrusion in the VMD. Duong et al. (2018) combined 1D and 2D hydrodynamic models to predict future flows, water level, and salinity intrusion in the Hau River—one main river branch in the Mekong Delta. Vu et al. (2018) applied a 1D model to assess the impacts of sea level rise on seawater intrusion in the VMD. The study showed that a combination of changes in temperature, rainfall and sea level rise significantly altered saltwater intrusion. Eslami et al. (2019) showed that under the impacts of decreased sediment from the upstream and sand mining in the downstream, tidal amplification and saltwater intrusion in the VMD increased by 0.2–0.5 PPT (parts per thousand).

The above studies showed that although mathematical models and remote-sensing techniques have been widely employed to study saltwater intrusion in the Mekong river delta, each for these methods has its advantages and limitations. Numerical models can provide salinity with a high spatio-temporal resolution, but it is usually suffered from errors related to uncertainties of the inputs, model structures and parameterization. Moreover, for a complex and dense channel network of the VMD, saltwater intrusion modeling is challenging and suffered from large numerical errors. Meanwhile, remote sensing

techniques also have large errors caused by the non-uniqueness of the relationship between the reflectance and salinity as well as measurement errors of remote sensing.

This study aims to present a hybrid approach that combines numerical modeling simulation and remote sensing estimation to improve the accuracy of salinity estimates in the VMD. Salinity simulated by a numerical model is merged with that obtained from Landsat images using the Bayesian inference technique. This merging approach not only increases the accuracy of salinity estimates but also reduces its uncertainties.

2. Methods and materials

2.1. Methods

2.1.1. Estimation of salinity by numerical modelling

In this study, the 1-D hydrodynamic (HD) and advection-dispersion (AD) modules of the MIKE 11 package developed by the DHI Water and Environment (www.dhigroup.com) were used to simulate saltwater intrusion in the VMD and estimate the salinity along with the river system. The HD module simulates the water hydrodynamics in the channels by solving the 1-D Saint-Venant equations. The AD module simulates the dynamical process of salinity by numerically solving the advection-dispersion equation. By combining the HD and AD modules, we can simulate the dynamics of salinity in the river system.

2.1.2. Estimation of salinity from remote sensing

Salinity was estimated from Landsat-8 images using the linear multivariate regression method. Accordingly, the relationship between the salinity (S) in channels and Lansat-8 reflectances was formulated as below:

$$S = a_0 + a_1 d + a_2 B_2 + a_3 B_3 + a_4 B_4 + a_5 B_7 \quad (1)$$

in which d is the distance from the sea to the location that we estimate the salinity; B_2, B_3, B_4, B_7 are the Landsat-8 bands. Nguyen et al. (2018) showed that these Landsat bands were the most sensitive to the river water salinity. However, unlike Nguyen et al. (2018), in this study, we used the distance from the sea instead of the latitude and longitude coordinates to represent the geographical location where salinity is estimated. Coefficients a_k (k=1,…,5) are the coefficients of the linear regression equation. These coefficients were obtained by the least square method using historical measurements of salinity at 10 locations (Figure 1) and Landsat-8 bands at these locations as well as their distances from the sea.

2.1.3. Bayesian inference

We applied the Bayesian inference technique to combine salinity in the VMD obtained by numerical model and remote sensing, which is expected to improve the salinity estimate. Bayesian inference is a merging technique based on Bayes' theorem which constructs the probability density distribution (*pdf)* of a variable of interest, which is salinity in this study, from multiple sources of information. According, assuming that the salinity (s) has a prior *pdf* $p(s)$, which was formed from the modeling simulations. If the salinity derived from remote sensing is available, it can be used to construct the posterior *pdf* using the Bayesian's theorem as below:

$$p(s|S = \boldsymbol{RS}) = \frac{p(s)p(S=\boldsymbol{y}|\theta)}{p(S=\boldsymbol{RS})} \quad (2)$$

in which $p(\theta|S = \boldsymbol{RS})$ is the posterior (conditional) *pdf* of salinity given the salinity derived from remote sensing $\boldsymbol{RS}$ dataset**;** $p(s)$ is the prior *pdf* representing the prior distribution of salinity based on modeling simulations; $p(S = \boldsymbol{RS}|s)$ is the likelihood

function, which was constructed from the salinity derived from the remote sensing. When doing inference, the normalizing constant of Bayes' theorem $p(S = \boldsymbol{RS})$ was not considered. As a result, Equation (2) can be rewritten as:

$$p(s|S = \boldsymbol{RS}) \propto p(s)p(S = \boldsymbol{RS}|s) \quad (3)$$

As shown in Equation (3), in order to estimate the posterior *pdf* $p(s|S = \boldsymbol{RS})$ of salinity, we need to determine the prior *pdf* and the likelihood function. The detailed process of construction of the posterior *pdf* of salinity was presented as below:

- *Construction of prior pdf*: The prior *pdf* was assumed to follow a normal distribution as below:

$$p(s) = \frac{1}{\sqrt{2\pi\sigma_0^2}} exp\left\{-\frac{(s-\mu_{0t})^2}{2\sigma_0^2}\right\} \quad (4)$$

in which μ_{0t} is the mean and σ_0^2 is the variance, which were computed as:

$$\mu_{0t} = \beta S_t + \alpha \quad (5)$$

$$\sigma_0^2 = \frac{\sum_{t=1}^{n}(O_t-\beta S_t-\alpha)^2}{n-1} \quad (6)$$

in which O_t and S_t are the measured and modeled salinity at time *t*, respectively; *n* is the length of measured and modeled salinity dataset; α is the y-intercept coefficient and β is the slope coefficient of the best-fitted linear regression relationship between measured and modeled salinity:

$$O = \alpha + \beta S \quad (7)$$

Using the least-square method, α and β can be computed from the of measured and simulated salinity dataset as:

$$\alpha = \bar{O} - \beta\bar{S} \quad (8)$$

$$\beta = \frac{\sum_{t=1}^{n}(O_t-\bar{O})(S_t-\bar{S})}{\sum_{t=1}^{n}(S_t-\bar{S})^2} \quad (9)$$

in which $\bar{O}$ and $\bar{S}$ represent the average values of measured and modeled salinity, respectively. It is noting that while the mean of the prior *pdf* changes with time *t*, the variance was fixed and computed from the whole measured and modelled dataset.

- *Construction of likelihood function*: The likelihood function was established from salinity derived from remote sensing under the normal distribution assumption as:

$$p(Y = y|\theta) = \frac{1}{\sqrt{2\pi\sigma_{rs}^2}} exp\left\{-\frac{(RS_t-\theta)^2}{2\sigma_{rs}^2}\right\} \quad (10)$$

in which RS_t is the salinity data derived remote sensing at time *t*; σ_{rs}^2 is the error variance of salinity estimated by remote sensing. This variance was estimated based on the Landsat-derived and measured salinity as below:

$$\sigma_{rs}^2 = \frac{\sum_{t=1}^{n}(O_t-RS_t)^2}{n-1} \quad (11)$$

- *Construction of posterior distribution*: Insert Equation (4) and Equation (10) into (3), the posterior *pdf* becomes:

$$p(s|S = \boldsymbol{RS}) = \frac{1}{\sqrt{2\pi\sigma_s^2}} exp\left\{-\frac{(s-\mu_{st})^2}{2\sigma_s^2}\right\} \propto \frac{1}{\sqrt{\sigma_0^2\sigma_{rs}^2}} exp\left\{-\frac{(s-\mu_{0t})^2}{2\sigma_0^2} - \frac{(s-RS_t)^2}{2\sigma_{rs}^2}\right\} \quad (12)$$

Because both of the prior *pdf* and likelihood function are normally distributed, the posterior *pdf* is a normal distribution. The mean (μ_{st}) and variance (σ_s^2) of this posterior *pdf* were calculated as:

$$\mu_{st} = \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma_{rs}^2}}\mu_{0t} + \frac{\frac{1}{\sigma_{RS}^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma_{rs}^2}}RS_t \quad (13)$$

$$\sigma_s^2 = \frac{\sigma_0^2\sigma_{rs}^2}{\sigma_0^2+\sigma_{rs}^2} \quad (14)$$

It is worth noting that the mean of the posterior *pdf* is a weighted average of the numerical estimation μ_{0t} and the remote sensing estimation RS_t in which the weighting factors are inversely proportional to the variances of salinity calculations using numerical models and remote sensing. This implies that if one estimation is more reliable (smaller variance) than the other, then it will be assigned a larger weight. The variance of the posterior *pdf* (σ_{st}^2) is always smaller than those of both prior *pdf* and likelihood function, which indicates that the merging helps to increases the reliability of salinity estimation.

2.1.4. Evaluation metrics

In order to evaluate the accuracy of calibration and validation results of different methods, the following criteria were used in this study:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - M_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O_i})^2} \quad (15)$$

$$r = \frac{\sum_{i=1}^{N}(O_i - \overline{O_i})(M_i - \overline{M_i})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O_i})^2}\sqrt{\sum_{i=1}^{N}(M_i - \overline{M_i})^2}} \quad (16)$$

$$BIAS = \frac{\left|\sum_{i=1}^{N}|O_i| - \sum_{i=1}^{N}|M_i|\right|}{\sum_{i=1}^{N}|O_i|} \times 100\% \quad (17)$$

In which *NSE*, *r* and *BIAS* are the Nash-Sutcliffe, Pearson correlation and bias coefficients, respectively. O_i and M_i denote the measured and estimated values at time i^{th}; N is the length of the dataset.

2.2. Materials

2.2.1. Study area

The study area is the VMD with an area of 39,400 km^2 and a population of around 17 million people. With a flat topography, the VMD has an average elevation of 0.7-1.2 m above the sea level, except several high hills and mountains in the northern plain of An Giang province. Along the Cambodian border, the elevation ranges from 2.0 to 4.0 m, then gradually reduce to 1.0-1.5 m at the plain center, and only 0, 3-0.7 m at the coastal area. The river network in the VMD is extremely complicated with two main rivers (Tien and Hau rivers) and multiple channels that are interconnected (Figure 1).

The VMD has a tropical monsoon climate in which the dry season usually occurs from June to November, while the wet season begins in December and ends in May. During the wet season, the delta is inundated due to a large amount of flow from the upstream and high in-situ precipitation. In the dry season, due to lack of in-situ rainfall and low flow from the upstream, saltwater intrudes into the rivers and channels, which strongly influences agricultural production, especially in drought years. In the context of climate change and land subsidence, the problem of saltwater intrusion has become more severe, which threats the sustainable development of the VMD and the food security of Vietnam. Hence, finding solutions to cope with saltwater intrusion is an urgent demand in the VMD.

2.2.2. Data availability

Observed data: Water level data at Tonle Sap and discharge at Kratie station were used as the upper boundary conditions for the HD model that simulates the spatio-temporal variation of water level and discharge in the VMD river system. Water level data at Tan

234 Chau, Chau Doc, My Tho, Can Tho were used for calibrating and validating the HD model.
235 As for the AD model, which simulates the saltwater intrusion, salinity was set to zero at
236 the upstream boundaries and 33 ppt (equal to sea salinity) at the downstream boundaries.
237 The salinity data were collected at two stations in the Hau river and eight stations in the
238 Tien river (Figure 1) for calibrating and validating the AD model. All data collected during
239 the dry seasons from 2013 to 2016 with an hourly time step.

240 *Remote sensing data:* Landsat-8 images were used to estimate the salinity. To match with
241 the numerical model simulation time, we collected Landsat images during the dry seasons
242 from 2013 to 2016 in the whole VND. The Landsat-8 images has a revisit span of 16 days
243 and spatial resolution of 30 m. Only images with cloud cover less than 10% were collected
244 and processed for salinity derivation. These datasets were obtained from the Google Earth
245 Engine platform (https://developers.google.com/earth-engine/datasets/catalog/landsat),
246 which allows to quickly retrieve and process Landsat datasets for the region of interest
247 comparing to other traditional methods.

248 **3. Results**

249 **3.1. Calibration and validation of MIKE 11 model**

250 Figure 2 shows the schematic river network that was constructed in MIKE 11 model to
251 simulate the saltwater intrusion. In this river network, the HD boundary conditions are the
252 hourly time-series of water level at the Tonle Sap Lake and discharge at the Kratie station
253 (Figure 2). The AD boundary conditions at these locations are zero salinity. There are
254 multiple downstream boundaries at the downstream ends of the Mekong river system. The
255 HD boundary conditions at these locations are the hourly tide water level and the AD
256 boundary conditions were set at a salinity of 33 ppt, which is the salinity of seawater in the

257 VMD. Both HD and AD modules were calibrated and validated using water level and
258 salinity at different locations along with the river system as shown in Figure 1. In this study,
259 parameters including the Manning roughness coefficient of the HD module and the
260 dispersion coefficient of the AD module were obtained using the trial-and-error method.

261 Water level and salinity data in the 2013 and 2014 dry seasons were used for model
262 calibration and those in the 2015 and 2016 were used for model validation. The calibration
263 and validation processes are presented as below. Firstly, we calibrated and validated the
264 HD module using the water level data at the Tan Chau, Chau Doc, My Thuan, My Tho and
265 Can Tho stations. Then, we calibrated and validated the AD module using the salinity data
266 at the Can Tho, Hoa Binh and Dai Ngai stations.

267 Table 1 shows the criteria that evaluate the accuracy of modeled water level at the five
268 stations Tan Chau, Chau Doc, My Thuan, My Tho and Can Tho, whereas Figure 3
269 compares measured and simulated water level at Tan Chau and My Tho stations. There is
270 a good agreement between modeling and measurements in both calibration and validation
271 stages. For calibration, the Nash-Sutcliffe coefficient ranges from 0.6 at the Tan Chau
272 station to 0.96 at the My Tho station. The modeled and modeled water levels are highly
273 correlated with correlation coefficients greater than 0.83. The bias errors of all stations are
274 lower than 3%. For validation, the agreement between measured and modeled water level
275 is lower than that for calibration. The *RSE* coefficients are lowest at the Tan Chau (0.62)
276 and highest at the My Tho station (0.92). The correlation coefficients of all stations are
277 greater than or equal to 0.81. The bias ranges from 1.1% at the Chau Doc to 3.2% at the
278 Can Tho station. Comparing to the Can Tho, My Thuan and My Tho stations, the water

279 level simulations at Tan Chau and Chau Doc are less accurate. This could be attributed to
280 the lack of detailed channel cross-sections of the river system in the Cambodia territory.

281

282 As for the calibration and validation results of the AD module, Table 2 presents the Nash-
283 Sutcliffe, Pearson correlation and bias coefficients at 10 stations in the Tien and Hau rivers.
284 Figure 4 compared the simulated and measured salinity at the Hoa Binh and Dai Ngai
285 stations in 2014 (calibration) and 2016 (validation) for illustration. For the calibration
286 period, the correlation between modeling and measurement at both stations is relatively
287 good (correlation coefficient ranges from 0.75 at the Dai Ngai to 0.9 at the Hoa Binh
288 station). However, the absolute difference between measurement and modeling is relatively
289 high, especially at the Dai Ngai station. The *NSE* and bias coefficients at this station are
290 0.44 and 17%, respectively. The validation was obtained with similar accuracy as the
291 calibration. The *NSE* values range from 0.63 at the Hoa Binh down to 0.4 at the Dai Ngai
292 station and their correlation coefficients are 0.83 and 0.74, respectively. The bias criterion,
293 which is greater than 8% at all stations, is relatively high. The differences between
294 modelling and measurement mainly comes from the errors of both hydrodynamics and
295 advection-diffusion modeling. In addition, due to the lack of salinity measurements at the
296 downstream boundaries, we used the salinity at the sea for the downstream boundary
297 conditions, which contributes errors to the simulation. As a result, it is necessary to find a
298 solution to improve the numerical modeling estimate of salinity.

299 **3.2. Salinity estimation from Landsat data**

300 Figure 5 compared the Landsat-estimated and measured salinity at 10 locations in the VMD
301 (see Figure 1 for the locations of these stations). The figure indicates that there is a

relatively good agreement between measurement and Landsat estimation with an *NSE* of 0.79, a correlation coefficient of 0.89 and a bias of 0.1%. This indicates that the regression Equation (3) can be used to estimated salinity from Landsat measurements. However, similar to the numerical modeling, the salinity estimated by Landsat images is also suffered from errors caused by, e.g., the sensitivity of Landsat bands with the salinity, the nonuniqueness of the relationship between salinity and Landsat reflectance as well as the quality of Landsat products. Hence, a combination of salinity derived from Landsat and numerical modeling is expected to improve the accuracy of salinity estimate and reduce its uncertainties.

3.3. Bayesian inference of salinity

In this section, the salinity estimated by numerical modeling and Landsat images were fused using the Bayesian inference presented in Section 2.1. This section estimates the mean μ_{st} (Equations (13)) and variance σ_s^2 (Equations (14)) of the posterior distribution of salinity (Equation (12)).

We firstly calculated the mean and variance of the prior *pdf* from numerical modeling simulations. Figure 6 compares the measured and numerically-modeled salinity at all stations in the validation period. The *NSE* and correlation coefficients for all stations are 0.46 and 0.79, respectively. The best-fitted linear regression equation between the measurement and modelling is $O = 1.679 + 0.876S$, which indicates that $\alpha = 1.679$ and $\beta = 0.876$ (Equation 7). Based on the values of α and β, the variance of the posterior *pdf* (σ_0^2) was calculated as σ_0^2 =3.65 using Equation (6). The mean of the prior *pdf* was written as:

$$\mu_{0t} = 0.876S_t + 1.679$$

Next, we estimated the variance of Landsat-derived salinity. This parameter was computed using Equation (11) from measured and Landsat-derived salinity as $\sigma_{RS}^2 = 8.69$. Finally, we performed data fusion of salinity estimated by modelling and remote sensing to obtain the ensemble salinity and its variance using Equations (13) and (14). Figure 7 compared the merged salinity with that obtained from numerical modeling and Landsat images at 10 stations in the Tien and Hau rivers. The results indicate that the merged salinity better agrees with the measured salinity than that derived from the numerical model and Landsat images. While the *NSE* coefficients corresponding to the Landsat estimation and numerical simulation are, respectively, 0.67 and 0.69, this criterion is 0.73 after merging. The merged and measured salinity also shows a better correlation with a correlation coefficient of 0.75. The bias after merging is approximately equal to 1. As for the estimation uncertainty, the variance of ensemble salinity significantly reduces from 3.65 ppt^2 for numerical model and 8.69 ppt^2 for Landsat estimation to 2.57 ppt^2 (Figure 8). The proposed method based on Bayesian merging improved the accuracy and reliability of salinity estimation.

4. Conclusion

In this study, we presented a Bayesian merging approach to combine remote sensing images with numerical model simulations to improve the salinity estimation in the VMD. Firstly, numerical simulations based on MIKE11 were performed to simulate the saltwater intrusion along the Mekong river system. Next, Landsat-8 images and measured salinity at 10 stations in Tien and Hau rivers were processed to derive a remote sensing-based regression equation for salinity estimation. Finally, the salinity estimated by both Landsat-8 images and numerical simulation at the same time and location was combined to provide

ensemble salinity estimations. All three evaluation metrics (*NSE*, correlation coefficient and bias) indicated that the merging approach significantly improved the accuracy of salinity estimates. The variance of ensemble salinity (2.57 ppt^2) is also significantly lower than that of numerical model (3.65 ppt^2) and Landsat estimation (8.69 ppt^2).

Though applied for only two salinity datasets in this study, the merging approach can be extended to multiple datasets to further improve the salinity estimation. In addition, due to the coarse temporal resolution of Landsat data owing to a long revisit span (16 days) and the impact of cloud cover, there are only a few Landsat images in each dry season, which is insufficient to evaluate the temporal variation of saltwater intrusion. In addition, the Landsat images used in this study have a spatial resolution of 30 m. It is suitable for large river branches in the Mekong. However, for smaller rivers, it requires finer spatial resolution images. As a result, our future research will concentrate on using other remote sensing products to improve both spatial and temporal resolution of salinity estimates.

Acknowledgment

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 105.06-2017.320.

References

1. Bai, Y., Pan, D., Cai, W. J., He, X., Wang, D., Tao, B., & Zhu, Q. (2013). Remote sensing of salinity from satellite- derived CDOM in the Changjiang River dominated East China Sea. *Journal of Geophysical Research: Oceans*, *118*(1), 227-243.

2. Bergqvist, A., Eitrem Holmgren, K., & Rylander, P. (2012). Impacts of saline water intrusion on the daily lives in the Mekong Delta Viet Nam. Swedish University of Agricultural Sciences, Department of Soil and Environment, Uppsala, Sweden.

3. Brunier, G., Anthony, E. J., Goichot, M., Provansal, M., & Dussouillez, P. (2014). Recent morphological changes in the Mekong and Bassac river channels, Mekong delta: The marked impact of river-bed mining and implications for delta destabilisation. *Geomorphology*, *224*, 177-191.

4. CGIAR Research Centers in Southeast Asia (2016). The drought and salinity intrusion in the Mekong River Delta of Vietnam. Assessment Report.

5. Erban, L. E., Gorelick, S. M., & Zebker, H. A. (2014). Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environmental Research Letters*, *9*(8), 084010.

6. Eslami, S., Hoekstra, P., Trung, N. N., Kantoush, S. A., Van Binh, D., Quang, T. T., & van der Vegt, M. (2019). Tidal amplification and salt intrusion in the Mekong Delta driven by anthropogenic sediment starvation. *Scientific Reports*, *9*(1), 1-10.

7. Minderhoud, P. S. J., Erkens, G., Pham, V. H., Bui, V. T., Erban, L., Kooi, H., & Stouthamer, E. (2017). Impacts of 25 years of groundwater extraction on subsidence in the Mekong delta, Vietnam. *Environmental Research Letters*, *12*(6), 064006.

8. Lam-Dao, N., Pham-Bach, V., Nguyen-Thanh, M., Pham-Thi, M. T., & Hoang-Phi, P. (2011). Change detection of land use and riverbank in Mekong Delta, Vietnam using time series remotely sensed data. *Journal of Resources and Ecology*, *2*(4), 370-374.

9. Keith, D. J., Lunetta, R. S., & Schaeffer, B. A. (2016). Optical models for remote sensing of colored dissolved organic matter absorption and salinity in New England, Middle Atlantic and gulf coast Estuaries USA. *Remote Sensing*, *8*(4), 283.

10. Kotera, A., Sakamoto, T., Nguyen, D. K., & Yokozawa, M. (2008). Regional consequences of seawater intrusion on rice productivity and land use in coastal area of the Mekong River Delta. *Japan Agricultural Research Quarterly: JARQ*, *42*(4), 267-274.

11. Kuenzer, C., Campbell, I., Roch, M., Leinenkugel, P., Tuan, V. Q., & Dech, S. (2013). Understanding the impact of hydropower developments in the context of upstream–downstream relations in the Mekong river basin. *Sustainability science*, *8*(4), 565-584.

12. Molleri, G. S., Novo, E. M. D. M., & Kampel, M. (2010). Space-time variability of the Amazon River plume based on satellite ocean color. *Continental Shelf Research*, *30*(3-4), 342-352.

13. Palacios, S. L., Peterson, T. D., & Kudela, R. M. (2009). Development of synthetic salinity from remote sensing for the Columbia River plume. *Journal of Geophysical Research: Oceans*, *114*(C2).

14. Pearse-Smith, S. W. D. (2012). The impact of continued Mekong Basin hydropower development on local livelihoods. *Consilience*, (7), 73-86.

15. Schmitt, R. J. P., Rubin, Z., & Kondolf, G. M. (2017). Losing ground-scenarios of land loss as consequence of shifting sediment budgets in the Mekong Delta. *Geomorphology, 294*, 58-69.

16. Toan, T. Q. (2014). Climate change and sea level rise in the Mekong delta: Flood, tidal inundation, salinity intrusion, and irrigation adaptation methods. In *Coastal Disasters and Climate Change in Vietnam* (pp. 199-218). Elsevier.

17. Tran, T. V., Tran, D. X., Myint, S. W., Huang, C. Y., Pham, H. V., Luu, T. H., & Vo, T. M. (2019). Examining spatiotemporal salinity dynamics in the Mekong River Delta using Landsat time series imagery and a spatial regression approach. *Science of the total environment*, *687*, 1087-1097.

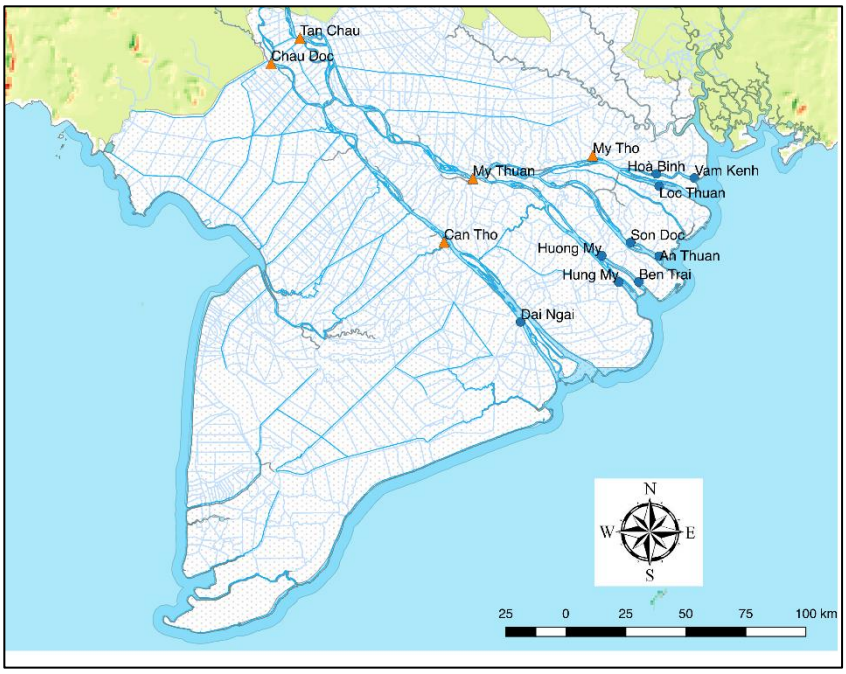
18. Vu, D. T., Yamada, T., & Ishidaira, H. (2018). Assessing the impact of sea level rise due to climate change on seawater intrusion in Mekong Delta, Vietnam. *Water Science and Technology*, *77*(6), 1632-1639.

Table 1. Criteria that evaluate the agreement between measured and modeled water levels during calibration and validation process at the My Tho, Can Tho, Tan Chau, Chau Doc stations

Station	Calibration			Validation		
	NSE	***r***	***BIAS (%)***	***NSE***	***r***	***BIAS (%)***
My Tho	0.96	0.98	1.3	0.92	0.85	1.7
Tan Chau	0.60	0.83	2.4	0.62	0.81	2.7
Chau Doc	0.68	0.87	0.8	0.65	0.82	1.1
My Thuan	0.85	0.89	1.7	0.87	0.83	2.1
Can Tho	0.93	0.97	2.9	0.89	0.92	3.2

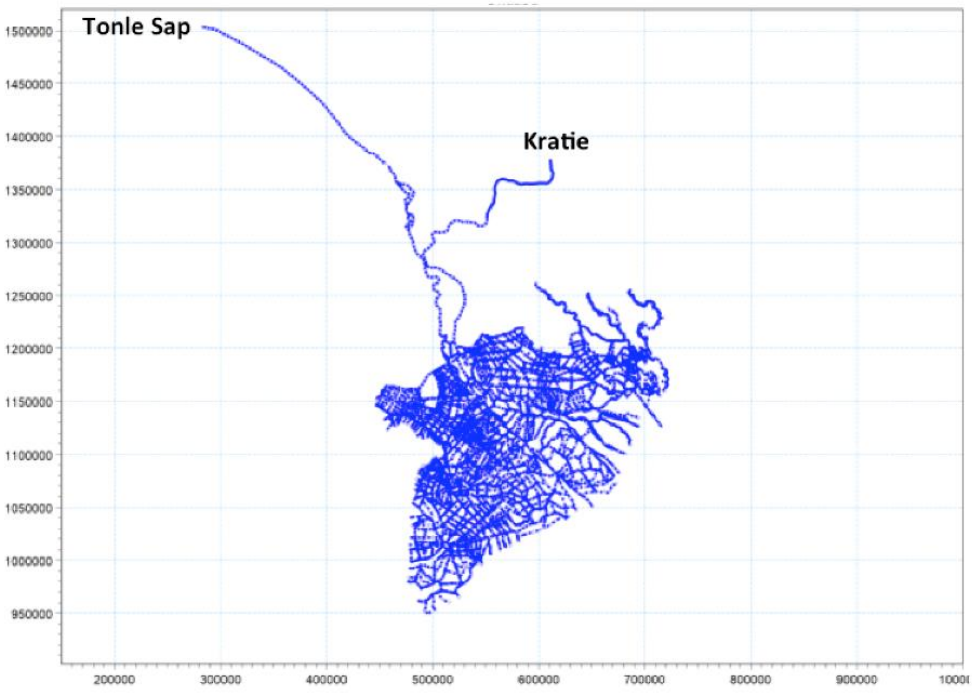
Table 2. Criteria that evaluate the agreement between measured and modeled salinity during model calibration and validation at the Hoa Binh and Dai Ngai stations

	Calibration			Validation		
Station	***RSE***	***r***	***BIAS (%)***	***RSE***	***r***	***BIAS (%)***
Can Tho	0.77	0.85	13.9	0.59	0.82	10.1
Dai Ngai	0.44	0.75	17.4	0.40	0.74	17.9
Huong My	0.55	0.77	15.3	0.45	0.78	13.6
Hung My	0.53	0.79	14.5	0.48	0.79	11.2
Ben Trai	0.55	0.77	15.3	0.45	0.78	13.6
Son Doc	0.53	0.79	14.5	0.48	0.79	11.2
An Thuan	0.63	0.81	14.2	0.58	0.81	13.2
Hoa Binh	0.77	0.91	13.4	0.63	0.83	7.9
Loc Thuan	0.60	0.75	15.7	0.52	0.72	16.2
Vam Kenh	0.75	0.88	13.9	0.55	0.79	14.5



433

434 **Figure 1. The study area of VMD with the river network. Locations of water level**
435 **(triangle shape) and salinity (circle shape) stations for calibrating and validating**
436 **numerical model are also presented**
437



438

439 **Figure 2. Schematic river network simulation in MIKE 11**

440

441

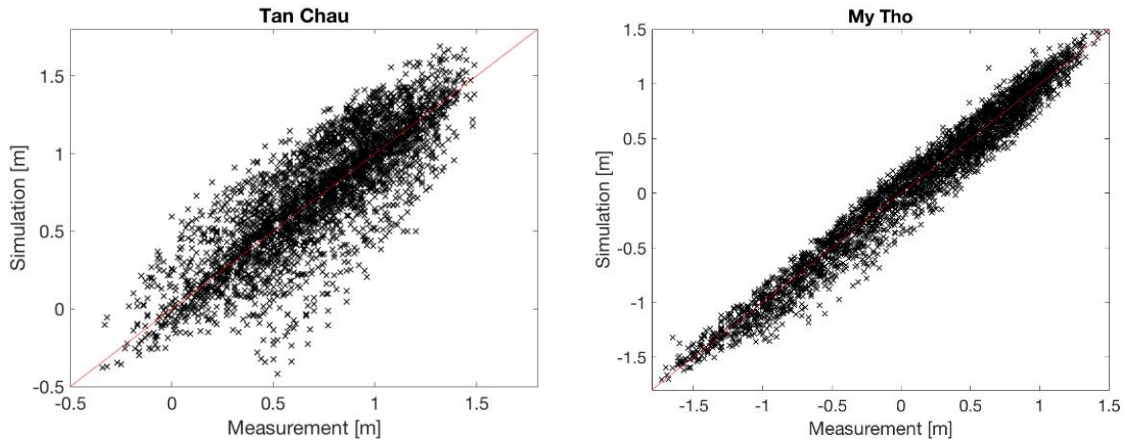
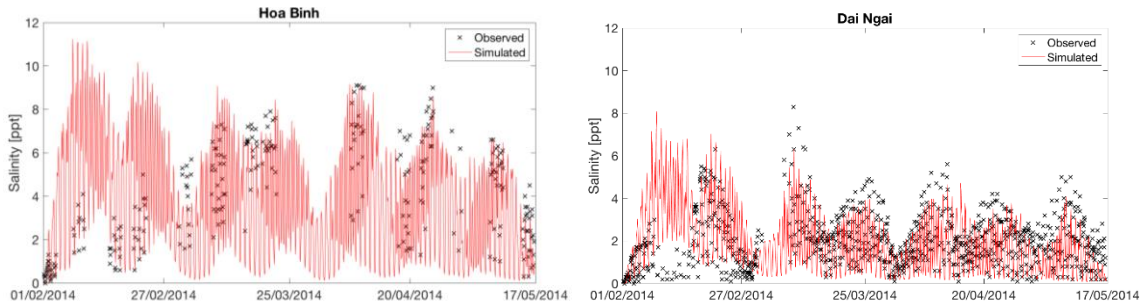
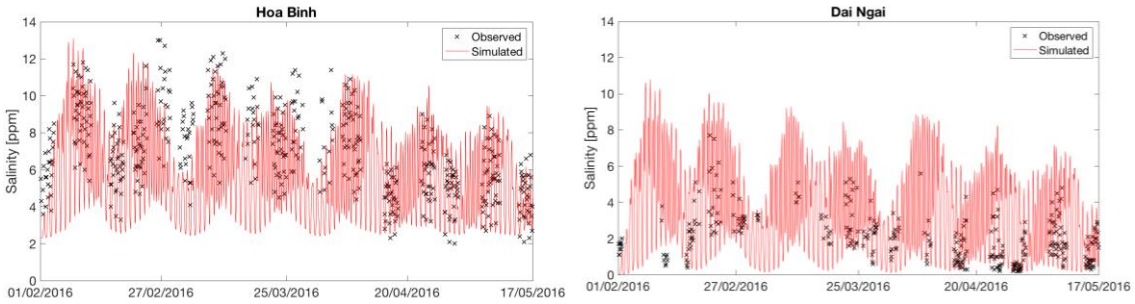


Figure 3. Comparison of measured and modeled water level at the Tan Chau, Chau Doc, My Tho and Can Tho stations. The red-solid line represents a perfect match between the simulation and measurement



a) Calibration



b) Validation

Figure 4. Comparison of measured and modeled salinity at the Can Tho and Dai Ngai (at Hau river) and Huong My, Hung My, Ben Trai, Son Doc, An Thuan, Hoa Binh, Loc Thuan and Vam Kenh (at Tien river) stations in 2014 and 2016

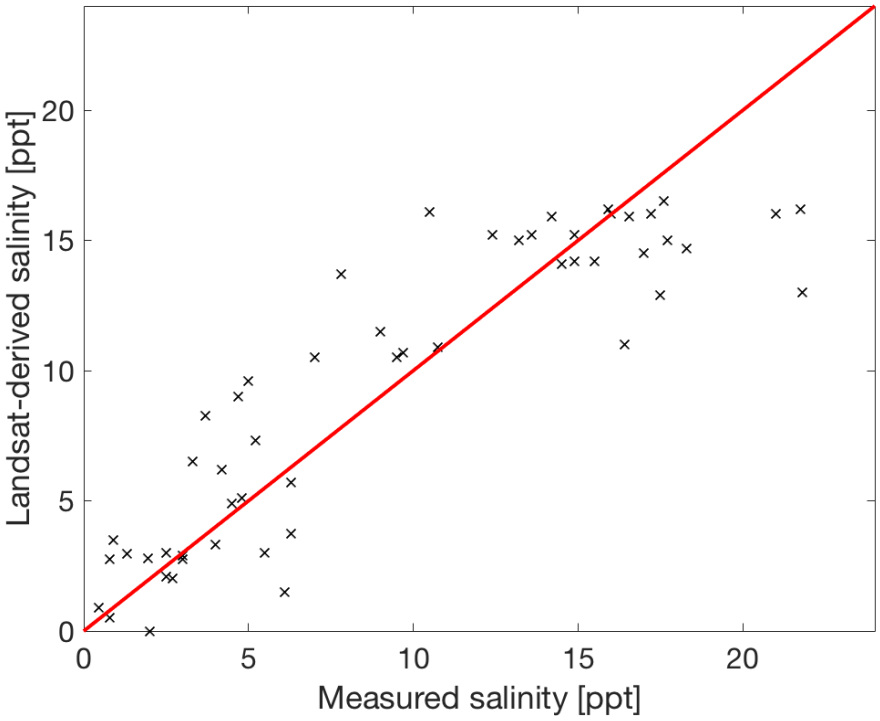


Figure 5. Comparison between Landsat-derived and measured salinity at different locations in the VMD

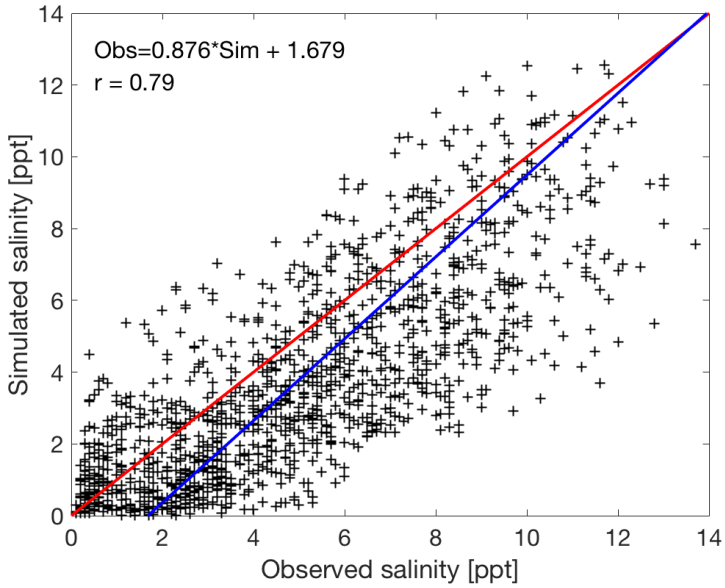
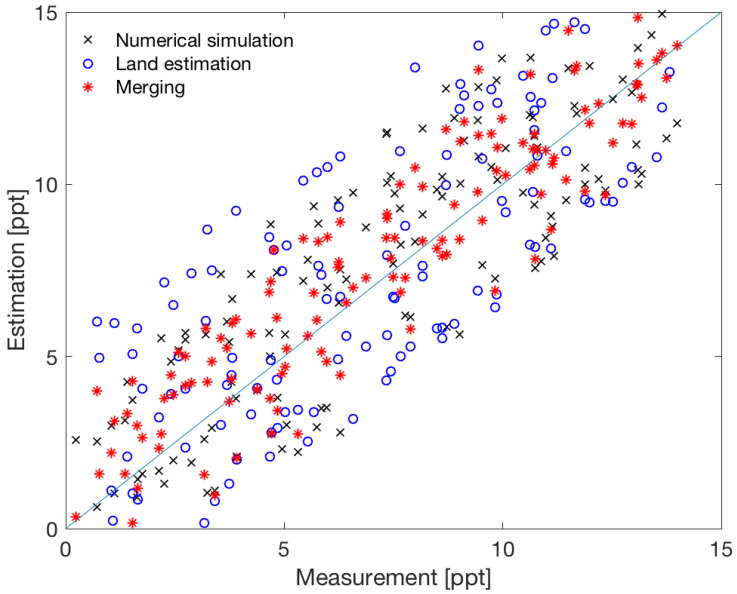


Figure 6. Modeled and measured salinity collected at 10 stations in the Tien and Hau rivers. The red line represents a perfect match between the modeled and measured salinity. The blue solid line denotes the best-fitted linear regression equation between the modeled and measured salinity, which are shown at the upper-left corner of the figure



461

462 **Figure 7: Comparison of salinity estimated by modeling simulations, Landsat**
463 **derivation and merging approach**

464
465
466

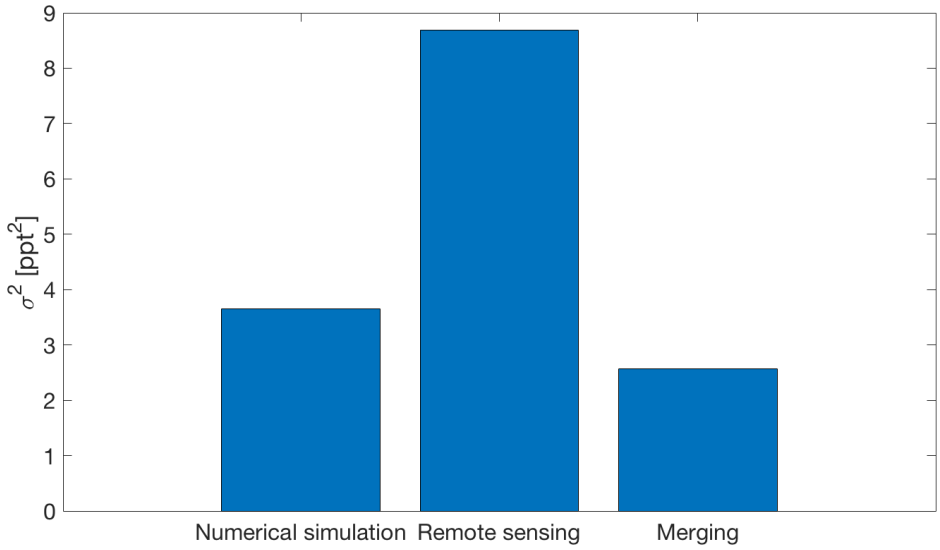


Figure 8: Comparison of the variance of salinity estimated by numerical simulation, remote sensing and merging of numerical simulation and remote sensing