21st Australasian Fluid Mechanics Conference
Adelaide, Australia
10-13 December 2018

# Assimilation of Lagrangian Drifter Data into a Hydrodynamic Model: Evaluation of Model Performance

**M. Khanarmuei[1], K. A. Suara[1], S. W. McCue[2] and R. J. Brown[1]**

[1]Environmental Fluid Mechanics Research Group
Queensland University of Technology, Queensland 4000, Australia

[2]School of Mathematical Sciences
Queensland University of Technology, Queensland 4000, Australia

### Abstract

Deployment of Lagrangian devices, such as drifters, can provide larger spatial coverage than Eulerian techniques. Although drifters have been widely used in the oceanography studies, they have been recently employed in shallow water applications. Since conventional hydrodynamic model variables are computed in Eulerian frame, assimilating of Lagrangian data into a hydrodynamic model to improve its accuracy is challenging. Using an ensemble Kalman filter is a practicable approach to this challenge. In this study, we investigate the assimilation of Lagrangian data into the Delft3D-FLOW, a hydrodynamic model, by using ensemble Kalman filter (EnKF) to improve its accuracy for estimating horizontal velocity fields for a shallow tidal estuary. Incorporating observations into the model was done through OpenDA which is a generic open source data assimilation tool. As different factors such as complexity of the flow field, interaction between flow field forcings, availability of the flow field and adequacy of the initial and boundary conditions can affect the performance of data assimilation, a sensitivity analysis was performed regarding ensemble size. To quantify the improvement of model estimations, model outputs with assimilation were compared against observations and model outputs without assimilation. Results revealed that assimilation of Lagrangian drifter data into Delft3D can successfully enhance its accuracy for estimating the horizontal velocity field for an estuary and performance of EnKF is dependent on ensemble size and observation density.

### Introduction

Large centres of worlds population are located around shallow water bodies, such as estuaries and coastal waters, and these waterways are among the most valuable natural resources due to their importance for economic development, providing required water for population and agriculture proposes, and ecological processes [1]. Population growth and human activities such as fishing, recreation, wastewater disposal and catchment development significantly impact on these shallow waterways and also increase the risk of pollution in these systems. Consequently, these activities make monitoring and managing estuaries a challenging area. Hydrodynamic modelling provides the basis of studying the problems related to the water systems, such as transport of pollutants and sediments. Although numerical models can provide larger spatial temporal coverage than field observations, model outputs are associated with uncertainties which can result from the imperfect knowledge about physical processes, discretization errors and uncertainties in initial and boundary conditions. On the other hand, observations are reliable, but they cannot generally provide a complete picture of the environmental processes especially those processes that vary spatially and temporally, such as flows in the shallow water bodies [2].

The numerical model outputs and observations are complementary. Therefore, the accuracy of the numerical models can be improved by assimilation of field observations. Data assimilation (DA) is an approach to incorporate observations into a numerical model in order to enhance its forecasts. In addition to forecasting, DA is used for estimating the model parameters and quantifying the model uncertainties [3]. Generally, DA is classified into two categories, variational and sequential methods. Variational schemes, such as three-dimensional variational (3D-Var) and four-dimensional variational (4D-Var), assimilate observed data into a hydrodynamic model via optimization. Optimization of the best model trajectory that fits the observation time series. On the other hand, sequential DA methods are techniques such as Kalman filter (KF), extended Kalman filter (EKF), ensemble Kalman filter (EnKF), particle filter (PF) and optimal interpolation. These schemes update model estimations every instance of time that new observed data is available [4].

In the last decades, many studies have successfully used different methods of DA in the fields of meteorology and oceanography. Compared to these two applications, there are less studies which have employed DA methods for estuarine systems. Drifters measure variables in Lagrangian frame, while model variables are computed in Eulerian frame. Therefore, assimilation of Lagrangian data poses some research challenges whilst providing potential opportunities for enhancement of the hydrodynamic model outputs by DA [5]. To this end, this study investigates on the assimilation of Lagrangian drifter data into the Delf3D-FLOW to improve its accuracy for estimating the horizontal velocity fields for micro-tidal estuary. In this study, EnKF, a sequential DA scheme is used to incorporate observations into the model. OpenDA, which is a generic open source data assimilation tool, is used to implement the EnKF.

### Study Region and Observed Data

Eprapah creek is a micro-tidal estuary in Queensland, Australia (Long. 153.293° E, Lat. 27.574° S). This estuary consists of relatively straight and meandering channels, which are characterized by irregular bathymetry. The estuarine zone is around 3.8 km. Eprapah Creek is characterized with a maximum depth between 3 and 4 m and it is sheltered from wind by mangroves [6]. This estuary is connected to the ocean and thus the main force in the creek is coming from the tide.

Observations consist of Lagrangian and Eulerian datasets, which were obtained from field experiment on $15-17^{th}$ of the July, 2015. The description of the field study, data quality control and analysis are provided in details in [7]. Datasets from acoustic Doppler velocimeters (ADV) are used as boundary conditions for the hydrodynamic model. Velocity data from four high–resolution drifters is used for the DA. These drifters were sampled at 10 Hz and their position accuracy is 2 cm [8]. Drifters were deployed in a relatively straight part of the estuary (Figure 1). Therefore this work considered only the fairly straight part of the Eprapah Creek to investigate the DA.

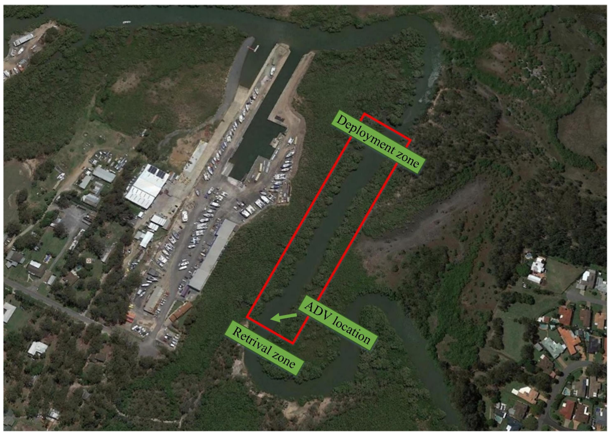


Figure 1: The study region, including the drifter deployment zone.

### Model Set up

In this study, the simulation is in 2D depth-average mode. This model solves the Navier Stokes equations under shallow water assumptions. As vertical accelerations are neglected for the vertical momentum equation, it uses the hydrostatic assumption-based pressure equation. The partial differential equations are solved by using a finite difference approach, using the boundary and initial conditions. In this study, the model used is Delft3D-FLOW to simulate the hydrodynamics of the estuary. This model has been successfully used in different applications, including oceanic, coastal and estuarine waters [9]. Delft3D-FLOW is a multi-dimensional hydrodynamic and transport simulation program. Using the flexible mesh, this model can be implemented for complex geometries, such as estuaries.

By doing a sensitivity analysis regarding mesh size, an unstructured mesh with the average resolution of 4 m (10 and 288 cells in width and length of the estuary, respectively) was used. This model was set up with two boundary conditions, flow discharge and water level at the upstream and downstream of the domain, respectively. The time series of the water level was manually obtained and flow discharge were extracted from ADV measurement. The bathymetry information for this region has been obtained from the Lidar data with a resolution of 5 m from the Geoscience Australia. Bathymetry and the time series of the water level are shown in Figure 2. The water level was converted to the Australian Height Datum (AHM). In addition, this figure indicates the period of the drifter deployment used in the DA. The friction coefficient was set as a uniform Manning of 0.023. The simulation time window is 48 hours and the time step is 30 seconds.

### Data Assimilation Algorithm

The traditional KF and EKF methods have several significant drawbacks that restricted their applications, such as weakness in dealing with the systems having a degree of nonlinearity. EnKF, which is a Monte Carlo approximation of the Kalman filter, was designed for non-linear models. It is the most popular sequential DA technique and it has been used widely in different applications such as meteorology and oceanography. The major idea for the EnKF is to propagate an ensemble (statistical samples) of the model state instead of using only one single state estimate and then calculating the covariance matrices from the ensemble, and finally, in the analysis step, calculate the Kalman gain by using the covariance matrices to update each ensemble member. The formulation of this method is presented in the following.

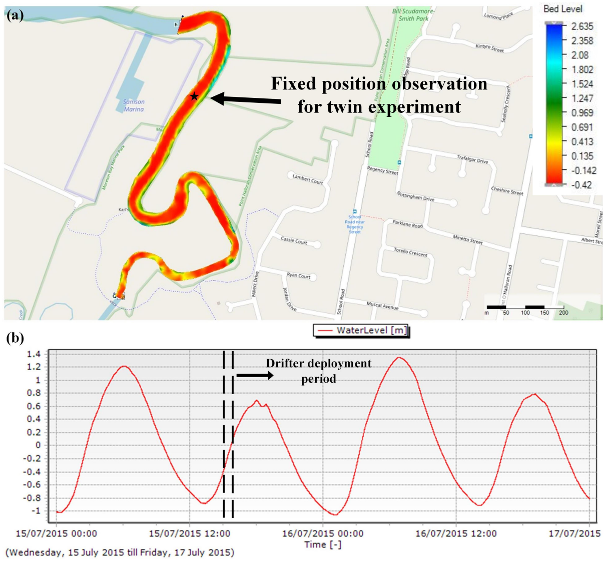


Figure 2: a: 5 m resolution bathymetry interpolation of the Eprapah Creek ($\star$ represents the location of the fixed position observation for twin experiment) and b: time series of the water level and drifter deployment period.

Similar to other sequential methods, EnKF is a step-by-step method which consists of alternating forecast and analysis steps. In the forecast step, an ensemble of the model forecast state with $m$ members is generated through propagating the model states in time as below:

$$x_i^f(t_k+1) = M(t_k)x_i^a(t_k) + \omega_i(t_k), \tag{1}$$

where $x_i^f$ is the ensemble forecast state, $x_i^a$ is the ensemble analysis state, $M$ is the model operator and $\omega_i$ is the model error. The ensemble forecast mean ($\bar{x}^f$) can be calculated from:

$$\bar{x}^f = \frac{1}{m}\sum_{i=1}^{m} x_i^f. \tag{2}$$

By using $[X_f]_i = \frac{x_i^f - \bar{x}^f}{\sqrt{m-1}}$, which $X_f$ is an $n \times m$ matrix with $m$ normalized perturbations, the forecast error covariance matrix would be approximated as $P^f = X_f X_f^T$. In the analysis step, each ensemble member ($x_i^a$ ) can be updated as follows

$$x_i^a(t_k) = x_i^f(t_k) + K(t_k)\left[y(t_k) - H(t_k)x_i^f(t_k) + \upsilon_i(t_k)\right], \tag{3}$$

where $K$ is the Kalman gain matrix, $y$ is the observations and $\upsilon_i$ is a realization of the observation error. The Kalman gain matrix can be expressed as:

$$K(t_k) = X_f(t_k)X_f^T(t_k)H^T(t_k)\left[H(t_k)X_f(t_k)X_f^T(t_k)H^T(t_k) + R(t_k)^T\right]^{-1}, \tag{4}$$

where $R$ is the observation error covariance matrix.

#### Implementation of EnKF via OpenDA

This study uses the OpenDA as a DA system to implement EnKF. To use the DA techniques available in OpenDA, only a coupling between OpenDA and the numerical model needs to be programmed. One of the most important parts of the EnKF is modelling of uncertainties that can be performed by imposing noise to the state vector for each ensemble member. In this study, it is assumed that the uncertainties in the model come from the uncertainties in the boundary conditions. Therefore, the noise is imposed to the boundaries by using the first order Auto Regressive process (AR(1)) to generate the replicates of the perturbed forcings. Moreover, for the observations, the uncorrelated errors with a standard deviation of 1.5% of the maximum drifter velocity (0.363 $ms^{-1}$) is assumed. This value is a representative error associated with drifter data which are provided in details in [8]. The EnKF assimilates the observations every 1 min for a 25 min as the assimilation window. A sensitivity analysis was performed regarding the ensemble size based on the accuracy of the assimilation outputs and the computational cost.

## Assimilation Results and Discussion

In this section, we qualitatively examine the results of the DA for two cases. The first case is a twin experiment which is a useful tool to evaluate the performance of the DA scheme for a simple or reduced-order model before applying to the realistic data. In this numerical experiment, velocity data from one drifter was used as a time series of the velocity in a fixed position in the domain. Indeed, we reduced the problem from Lagrangian to the Eulerian data assimilation. The location of this fixed position observation in the domain is demonstrated in Figure 2. To study the effect of the ensemble size ($n$), $n$ =20, 40, 50 and 100 were tested. Figure 3 presents the results for both model with and without assimilation and also the observation data. As it was mentioned, the assimilation window is 25 min. In Figure 3, it can be seen that the observations fluctuated within assimilation window. This reflects the spatio-temporal variability of the flow field in which the drifter was deployed. On the other hand, there is no significant change in the velocity of a fixed position obtained from the model without assimilation. Despite the difference between the physical representation of the observation and model forecast, the results showed that assimilated output using EnKF were weighted towards the observations. This demonstrates the effectiveness of assimilation scheme. Moreover, the accuracy of the model estimation were noticeably enhanced with the increase in the ensemble number.

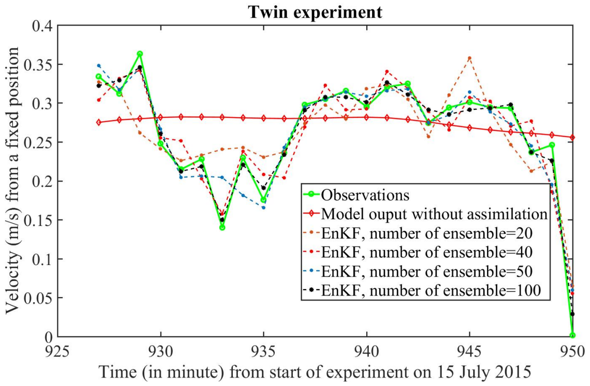


Figure 3: Twin experiment of EnKF.

In the second case, the DA was performed for a realistic situation. In fact, DA compares observations with the model outputs at the grids cells correspond to the drifter location at a certain time. At first, we examine the effectiveness of the dataset from a single drifter. Figure 4 qualitatively shows that the model outputs are improved after assimilation. Subsequently, all four drifters were used in DA process and the results are presented in Figure 5. It can be seen that DA successfully incorporated all four drifter data into the Delft3D-FLOW, as well. Although a wider area of the domain can be improved by using four drifters as the observations, it can be seen that data assimilation of the four drifter data requires larger ensemble sizes compare to the assimilation of one single drifter.

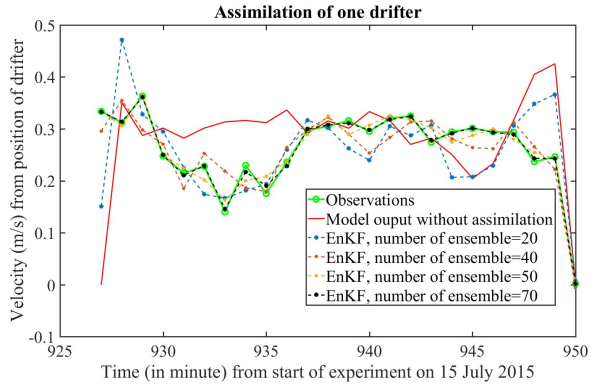


Figure 4: Results of EnKF by using velocity data from one single drifter.

As indicated in the results, the estimations of the model got more accurate by increasing the ensemble size. However, choosing a proper ensemble size is dependent on not only the improvement of the assimilation outputs but also the computational time. To quantitatively evaluate the improvement of the model, the root mean square error (RMSE) between the observations and assimilation outputs were calculated. Figure 6 shows that increasing the ensemble size causes a reduction in RMSE and also increasing computational time. This figure was extracted for the case in which all four drifters were used for DA. The computational time is the run time for the whole 48 hours simulation using a workstation with an Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz. Therefore, based on this analysis, 150 ensemble members is an appropriate ensemble size for this case since there is no significant reduction in RMSE for larger ensemble sizes.

## Conclusions

By using OpenDA, EnKF was implemented to improved the accuracy of a hydrodynamic model for estimating the horizontal velocity fields in a micro-tidal estuary. In this study, Lagrangian drifter data was used as the observations for the DA process. EnkF successfully incorporated Lagrangian observations into the model for both the twin experiment and the realistic situation. In the realistic case, EnKF was used for assimilating one single and four drifters as observations with different ensemble sizes. Results indicated that EnKF requires a larger ensemble size for assimilation of four drifter data in comparison to only one drifter data to reasonably improve the assimilation outputs. A sensitivity analysis was performed regarding the ensemble size based on the RMSE and computational time and it showed that 150 ensemble members is an appropriate ensemble size for this case. For the future works, authors aim to conduct a comprehensive sensitivity analysis regarding assimilation window, assimilation time step and examine the relationship between data density and effectiveness of assimilation process for Lagrangian drifter observations in shallow water estuaries.

## Acknowledgment


The authors would like to thank the Geoscience Australia for providing the bathymetry information. The project is supported through Australia Research Council Linkage grant LP1501072.

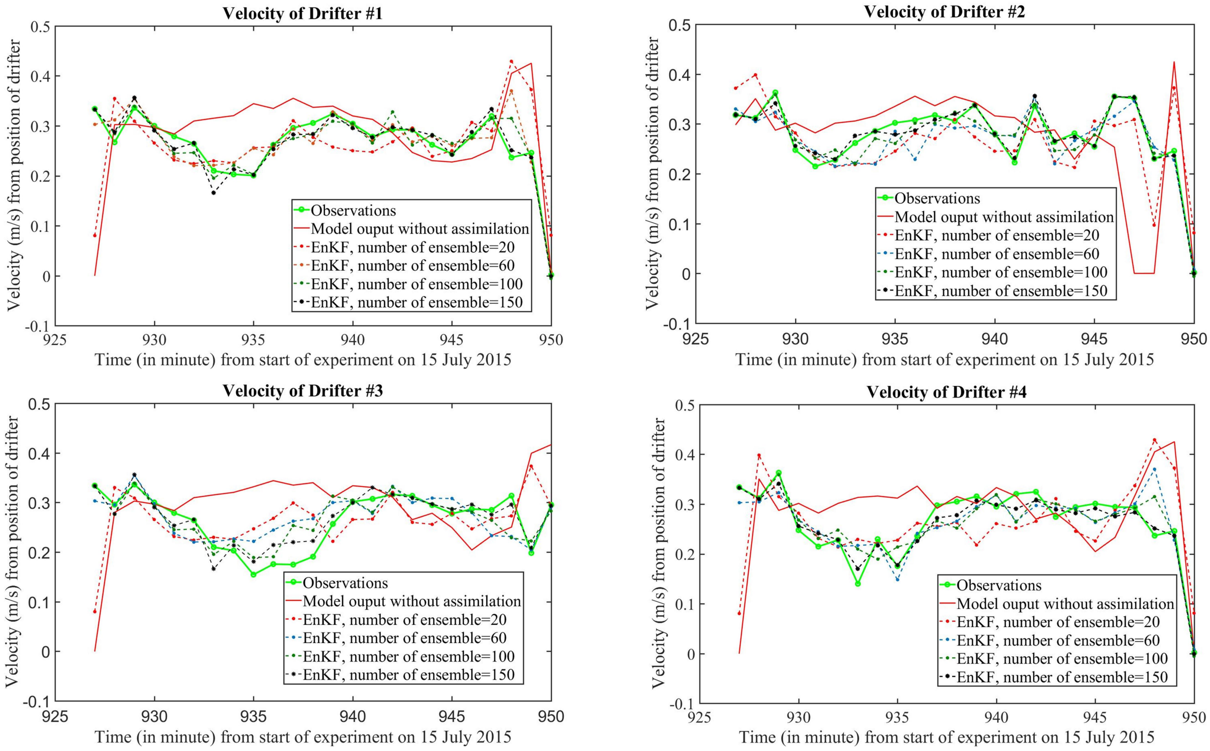


Figure 5: Results of EnKF by using velocity data from all four high-resolution drifters.

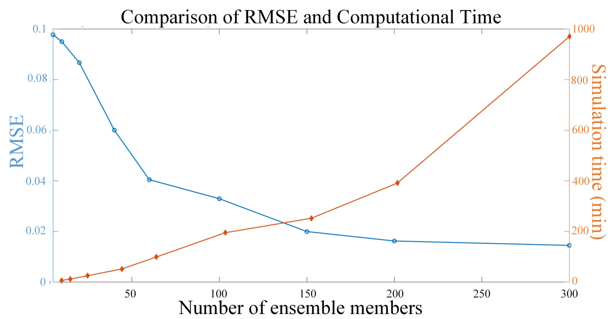


Figure 6: Comparison of RMSE and computational time of EnKF with different ensemble sizes for assimilation of all four drifters.

## References


[1] Costanza, R., d'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'neill, R.V., Paruelo, J., and others. The value of the world's ecosystem services and natural capital. *Nature*, **387**, 1997.

[2] Heemink, A.W., Hanea, R.G., Sumihar, J., Roest, M., Velzen, N. and Verlaan, M., Data assimilation algorithms for numerical models. *Advanced Computational Methods in Science and Engineering*, 2009, 107–142.

[3] Toye, H., Zhan, P., Gopalakrishnan, G., Kartadikaria, A., R., Huang, H., Knio, O. and Hoteit, I., Ensemble data assimilation in the Red Sea: sensitivity to ensemble selection and atmospheric forcing, *Ocean Dynamics*, **67**, 2017, 915–933.

[4] Bertino, L., Evensen, G. and Wackernagel, H., Combining geostatistics and Kalman filtering for data assimilation in an estuarine system, *Inverse problems*, **18**, 2002.

[5] Strub, I.S., Percelay, J., Tossavainen, O.P. and Bayen, A.M., Comparison of two data assimilation algorithms for shallow water flows. *Networks and Heterogeneous Media*, **4**, 2009, 409–430.

[6] Suara, K., Brown, R.J. and Borgas, M., Eddy diffusivity: a single dispersion analysis of high resolution drifters in a tidal shallow estuary. *Environmental Fluid Mechanics*, **16**, 2016, 923–943.

[7] Suara, K., Chanson, H., Borgas, M. and Brown, R.J., Relative dispersion of clustered drifters in a small micro-tidal estuary. *Estuarine, Coastal and Shelf Science*, **194**, 2017, 1–15.

[8] Suara, K., Wang, C., Feng, Y., Brown, R.J., Chanson, H. and Borgas, M., High-resolution GNSS-tracked drifter for studying surface dispersion in shallow water. *Journal of Atmospheric and Oceanic Technology*, **32**, 2015, 579–590.

[9] Garcia, M., Ramirez, I., Verlaan, M. and Castillo, J., Application of a three-dimensional hydrodynamic model for San Quintin Bay, BC, Mexico. Validation and calibration using OpenDA. *Journal of Computational and Applied Mathematics*,**273**, 2015, 428–437.