Optimal interpolation data assimilation of surface currents by utilizing pseudo measurement with Monte Carlo simulation

Lei Ren*, Stephen Nash, Michael Hartnett

[1]*Department of Civil Engineering, National Univerisity of Ireland, Galway, University Road, Galway, Ireland*

[2]*Ryan Institute, Galway, Ireland*

Received 15 May 2014, www.cmnt.lv

Abstract

Optimal Interpolation (OI) data assimilation is a technique to combine available observations with background states to improve prediction states. In this research, pseudo measurement of surface currents generated by adding noise with Monte Carlo simulation is used to update the background states with optimal interpolation. The core of Optimal Interpolation data assimilation is the definition of background error covariance, which determines to what extent the model background states will be corrected to match the observations. The background error covariance is computed before the data assimilation process. The model background errors are calculated from the mean over a short time interval ten minutes. A series of sensitivity tests with Optimal Interpolation are done by calculating Root Mean Square Error (RMSE) to decide the appropriate parameters. The improvement of Optimal Interpolation at reference points is measured in Taylor diagrams, and the surface current maps of test domain show the effectiveness of Optimal Interpolation.

Keywords: Optimal Interpolation, data assimilation, Monte Carlo, pseudo measurement, background error covariance

1 Introduction

Data assimilation is a technique to improve the modelling prediction ability by blending available measurement information with the background states. In general, there are two kinds of data assimilation algorithms: sequential and variational data assimilation (Robinson and Lermusiaux, 2000, Moore, Arango, Broquet, Powell, Weaver and Zavala-Garay, 2011, Ma, Zheng, Zhong and Zou, 2014). The analysis equation of the former is expressed by the linear combination between background and measurement states; the latter algorithm is generally derived from an objective function measuring the distance between observed states and background states (Zaron, 2009, Dong and Xue, 2012). In our current work, sequential Optimal Interpolation data assimilation scheme is undertaken to update the model background states.

For sequential Optimal Interpolation data assimilation schemes, innovation of calculating the background error covariance results in a variety of methods, such as Optimal Interpolation (Gu, Woo and Kim, 2011, Rienecker, 1991), Ensemble Optimal Interpolation (EnOI) (Oke, Brassington, Griffin and Schiller, 2010, Counillon and Bertino, 2009). Due to the inexpensiveness and flexibility of Optimal Interpolation data assimilation algorithm, it is becoming a popular data assimilation approach in oceanography (Counillon and Bertino, 2009, Oke, Brassington, Griffin and Schiller, 2008, Srinivasan, Chassignet, Bertino, Brankart, Brasseur, Chin, Counillon, Cummings, Mariano, Smedstad and Thacker, 2011). Optimal Interpolation and Ensemble Optimal Interpolation data assimilation had been applied in some operational oceanic hydrodynamic prediction systems (Oke, Brassington, Griffin and Schiller, 2010, Oke, Brassington, Griffin and Schiller, 2008, Carton and Giese, 2008). In this paper, Optimal Interpolation method is used to update the surface velocity components by using pseudo measurement generated with Monte Carlo simulation (Doucet, de Freitas and Gordon, 2001). In order to clearly analyse Optimal Interpolation data assimilation process, a small test area is defined as our data assimilation domain, the purpose of this work is to develop an Optimal Interpolation data assimilation system for coastal areas and assess the improvement of Optimal Interpolation data assimilation.

An outline of this paper is as follows: Section 2 and 3 describe the three dimensional numerical modelling and generation process of pseudo measurement. Section 4 presents Optimal Interpolation data assimilation schemes. Results of Optimal Interpolation data assimilation is presented in Section 5, followed by conclusions of Optimal Interpolation data assimilation in Section 6.

2 Numerical modelling

The Environmental Fluid Dynamics Code (EFDC) is applied to simulate the dynamic process of Galway Bay, which is located on west coast of Ireland. The numerical model EFDC solves the three-dimensional, vertically hydrostatic, free surface, turbulent averaged equations of

* ***Corresponding author*** e-mail: leirencomeon@gmail.com

motions for a variable density fluid. The module uses a sigma vertical coordinate and curvilinear, orthogonal horizontal coordinates. There are 380×241 grids in the rectangular simulation domain of model, the grid resolution is 150 metres, the physical domain is from (-9.71891E, 52.97371N) (left at the bottom) to (-8.87716E, 53.03773N) (right on the top), following picture shows the modelling area and data assimilation domain. The basic research area is Galway bay, a square domain with 961 wet cells is defined as our data assimilation domain, real dimension is a 4.65km×4.65km square area.

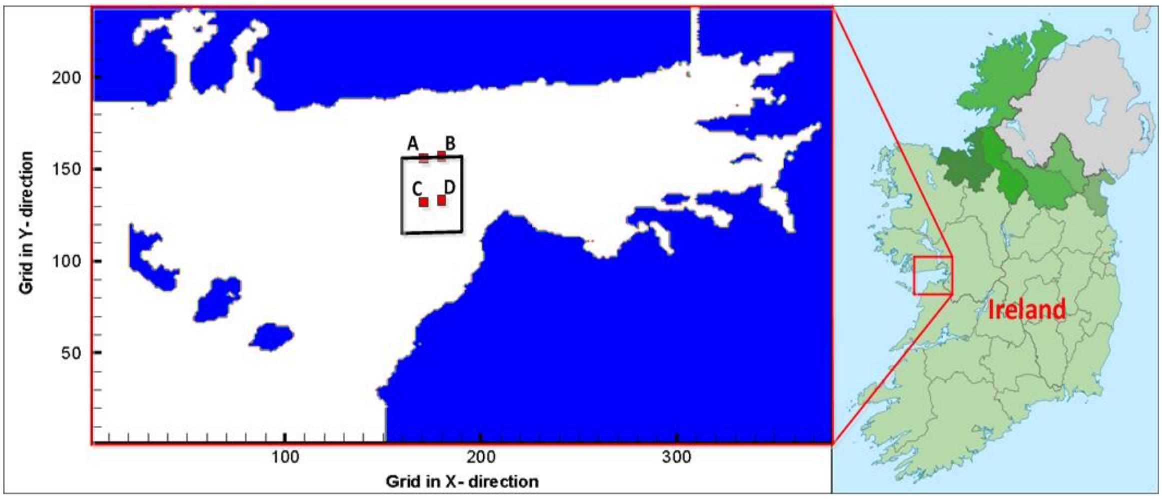

FIGURE 1 Research area and data assimilation test domain

The meteorological data (temperature, rain, solar radiation and relative humidity etc.) are obtained from the weather station located at National University of Ireland, Galway (http://weather.nuigalway.ie/). The river inflow of River Corrib was got from the Office of Public Works (http://www.opw.ie/hydro/). And the tidal information is generated from Oregon State University Tidal Prediction Software (OTPS), which provides tidal information on the western and southern open boundaries. In order to illustrate the Optimal Interpolation data assimilation process in detail, there are three main simulations performed: the Free run, which is initialised with no data assimilation for seven days (01/10/2011-07/10/2011), its output is applied to compute the background error covariance; the Assimilation run, which is initialised for seven days (14/10/2011-20/10/2011), but the surface velocity components are updated by utilising the pseudo measurement during the last four days; the Control run is the same as the Assimilation run, but with no data assimilation, which is regarded as the standard reference state, its output of surface velocity components are used to generate pseudo measurement by adding normal distribution noise.

3 Pseudo measurement

In order to update the background states in numerical model, pseudo measurement is generated by adding normal distribution noise to the output from a model run with Monte Carlo simulation, Monte Carlo simulation is undertaken to yield pseudo measurement based on the basic hydrodynamic trend(Doucet, de Freitas and Gordon, 2001). In order to clarify the difference between original model results and generated pseudo measurement, the noise is unbiased, the standard deviation of added noise is 20% of the maximum absolute difference of velocity components during model stable phase, the value of standard deviation are 4 cm/s and 3 cm/s for surface velocity components (u and v) separately. There are three obvious advantages by using this kind pseudo measurement: firstly, the generated pseudo measurement is based on the output from Control run, the general trend of generated pseudo measurement still follows the basic dynamical process; secondly, the value of noise can be controlled, which means that the sensitivity test of Optimal Interpolation data assimilation can be more accurately assessed; thirdly, the generation process of pseudo measurement is based on the background states field from the Control run and the noise field has the same structure of background states, so the yielded artificial measurement field matches well with the model state field.

4 Optimal interpolation data assimilation

Optimal Interpolation data assimilation combines observation states with model background states to obtain better prediction results, here, the background states are updated by utilising pseudo measurement.

The analysis equation of Optimal Interpolation is a linear combination of background states and measurement states, the optimal weight factor (Kalman gain) is derived by minimizing the analysis error covariance, the analysis equation could be expressed as follows (Kalnay, 2002):

$$x^a = x^b + K(y^0 - Hx^b)\,, \tag{1}$$

where x^a is the analysis state, x^b is the forecast or background state, K is the Kalman gain, H is the measurement operator, y^0 is the observation state.

The state in our research is surface velocity components *u* and *v*, which could be given as:

$$x = (u,v)^T . \quad (2)$$

The Kalman gain is obtained in the following formula by minimizing the analysis error covariance (Kalnay, 2002):

$$K = P^b H^T (HP^b H^T + R)^{-1} , \quad (3)$$

where P^b is the background error covariance, R is the measurement error covariance

The steps of Optimal Interpolation data assimilation updating surface velocity components with generated pseudo measurement by using Monte Carlo simulation is listed as follows:

- Run the free model from 01/10/2011 to 07/10/2011, calculate the background error variance, the error variance of measurement field is averaged on space using obtained background error variance, then calculate the Kalman gain with equation (3)
- Generate noise following normal distribution N (0, 16) and N (0, 9) for producing pseudo measurement, because surface velocity component are calculated separately in our numerical model, normal distribution noise N (0, 16) is for u component of surface velocity, normal distribution noise N (0, 9) is for v component of surface velocity.
- Control run is undertaken from 14/10/2011 to 20/10/2011 with no data assimilation, the time interval of its surface velocity components output in data assimilation domain is sixty minutes
- Add the generated noise to corresponding velocity components from Control run by using Monte Carlo simulation, the new dataset is the pseudo measurement field, then interpolate the pseudo measurement on time to every five minutes
- Update the surface velocity states every five minutes with the generated pseudo measurement by utilising Optimal Interpolation data assimilation scheme in the square data assimilation domain

Since the model takes about three days to be stable, the output of surface velocity components from the free run and data assimilation process are only taken during the last four simulation days.

4.1 BACKGROUND ERROR COVARIANCE (FORECAST ERROR COVARIANCE)

In Optimal Interpolation data assimilation scheme, the background error covariance is static, which means that the background error covariance is a constant matrix. It is calculated before the data assimilation process is performed (Robert, Blayo and Verron, 2005). According to the statistical relationship between the covariance and correlation coefficient, the background error covariance could be expressed as (Oke, 2002):

$$\operatorname{cov}(e_i, e_j) = \rho_{ij} \times \sigma_i \times \sigma_j , \quad (4)$$

where e_i, e_j are the errors at different locations, σ_i, σ_j are the standard deviation of errors at different locations, ρ_{ij} is the spatial correlation function, which is defined based on Gaussian function $\rho_{ij} = \alpha \exp(-\frac{d^2}{L^2})$, α is the scale factor, L is the correlation length, d is the spatial distance between two points. Our interest is to update the surface velocity components u and v. The appropriate parameters α and L are determined when the Root Mean Square Error reaches minimum.

According to equation (4), the background error covariance can be simplified as:

$$P^b = D^{\frac{1}{2}} C D^{\frac{1}{2}} , \quad (5)$$

D is the diagonal background state variance matrix describing the modelling error structure. The background state error e^b is computed from the difference over a short time interval ten minutes $(\Delta t = 10 \min s)$.

$$D = \operatorname{var}(e^b) = E[e^b e^{bT}] = E[(e^b - \overline{e^b})(e^b - \overline{e^b})^T] , \quad (6)$$

$$e^b = x_i^b - \overline{x_{\Delta t}^b} . \quad (7)$$

The overbar means the expected value.

e^b is the model error at different locations.

x_i^b is the background state at time step i .

C is the spatial correlation matrix. Every element is calculated from equation (4).

Since pseudo measurements are yielded from Control run by adding noise with Monte Carlo simulation, the pseudo measurement field has the same structure as background states of numerical modelling, the measurement operator H in equations (1) and (3) is an identity matrix.

In order to clearly show the improvement of sequential data assimilation schemes, in following Optimal Interpolation data assimilation, five minutes is chosen as the data assimilation interval.

4.2 ADJUSTMENT OF α AND L FOR OPTIMAL INTERPOLATION

For Optimal Interpolation data assimilation process, parameters α and L directly decide its effectiveness. Gu ((Gu, Woo and Kim, 2011)) had used different optimized values of the two parameters at different location to assimilate vertical current data to the unstructured grid ocean numerical model, the minimum and maximum of α was 1 and 4, the minimum and maximum value of correlation length L was 20 km and 100 km. Ragnoli (Ragnoli, Zhuk, Donncha, Suits and Hartnett, 2012) assimilated the High Frequency radar surface current data of Galway bay in numerical model, the scaling factor war chosen as 100 and correlation length was 0.3 km, their data assimilation area was the whole inner Galway bay. For different research area and using different types of measurement, the optimal parameters of scaling factor and correlation length are different. In our research, a variety of test cases with different values of these parameters are investigated, RMSE is employed to measure the degree of their match. For these tests, assimilation interval is five minutes. Firstly, the RMSE is calculated on space (961 grids) every five minutes, then it is averaged on time (see Table 1 and Table 2).

TABLE 1 RMSE of u component

a/L	0.15 km	0.30 km	0.45 km	0.75 km	1.05 km	1.50 km	3.00 km
1	0.0523	0.0508	0.0508	0.0538	0.0605	0.0769	0.1910
2	0.0550	0.0524	0.0535	0.0642	0.0846	0.1292	0.3770
3	0.0570	0.0541	0.0573	0.0754	0.1094	0.1834	0.5705

TABLE 2 RMSE of v component

a/L	0.15 km	0.30 km	0.45 km	0.75 km	1.05 km	1.50 km	3.00 km
1	0.0383	0.0373	0.0372	0.0393	0.0438	0.0556	0.1947
2	0.0404	0.0385	0.0392	0.0472	0.0630	0.0994	0.4089
3	0.0418	0.0398	0.0418	0.0561	0.0833	0.1466	0.6074

From the RMSE (u) and RMSE (v) in these cases, when $\alpha = 1.0$ and $L = 0.45km$, both of them are minimum, so these values are employed in our Optimal Interpolation data assimilation.

5 Results

The goal of Optimal Interpolation data assimilation is to enhance the modelling prediction capability referring to the measurement trajectory. Surface current maps at certain time steps are displayed and statistical comparison of surface velocity components time series for reference points is shown in Taylor diagrams. The surface current maps of assimilation model field are compared with control model field with no data assimilation process and pseudo measurement field at certain time steps. Here, only the data assimilation domain is displayed.

In Figure 2-4, the left panels show surface current map at t=4.0 days, the right panels show surface current map at t=6.0 days. For both surface current map at t=4.0 days and t=6.0 days, compared with results from original model in Figure 3, Figure 4 shows that the Optimal Interpolation data assimilation process absorbs useful information from pseudo measurement into numerical model, since the consideration of observation error in Optimal Interpolation data assimilation process, the pseudo measurement is not fully projected into the numerical model, the assimilation model just assimilated the basic trend of pseudo measurement into assimilation model. The pseudo measurement is generated by adding normal distribution noise into the results from Control model, the general trend in the data assimilation domain is chaotic. We use this way to test the sensitivity reflection of our Optimal Interpolation data assimilation process when this kind of pseudo measurement is used for update. The reason behind this is that observation data in real world is always noisy. Generally, Optimal Interpolation works well when chaotic pseudo measurement is used for assimilation, and compared with Control model, general trend of assimilation model in data assimilation domain is closer to pseudo measurement trajectory.

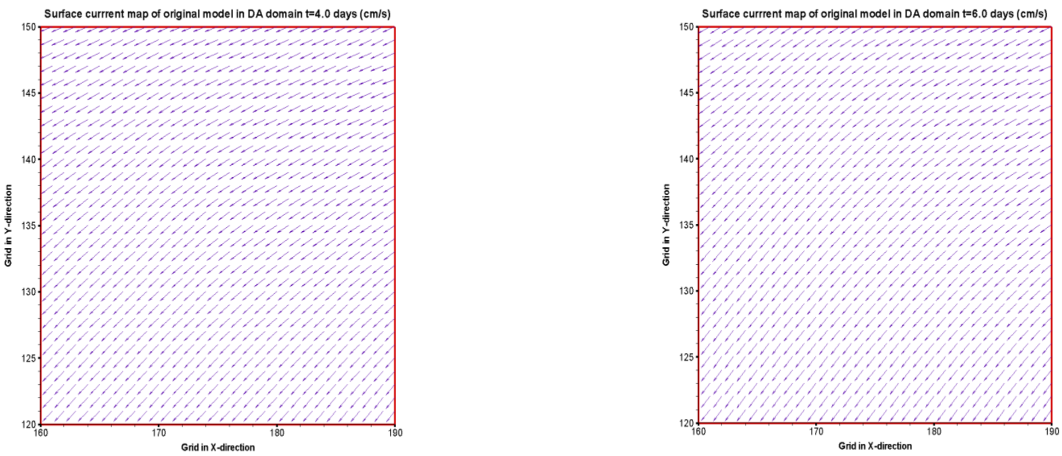

FIGURE 2 Surface current map with no data assimilation from Control model (t=4.0 days and t=6.0 days)

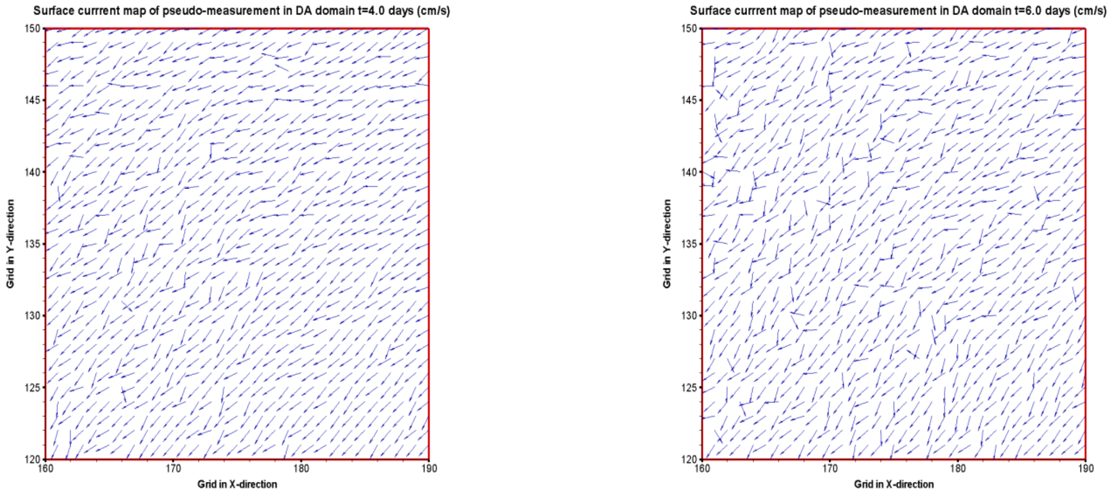

FIGURE 3 Surface current map of pseudo measurement (t=4.0 days and t=6.0 days**)**

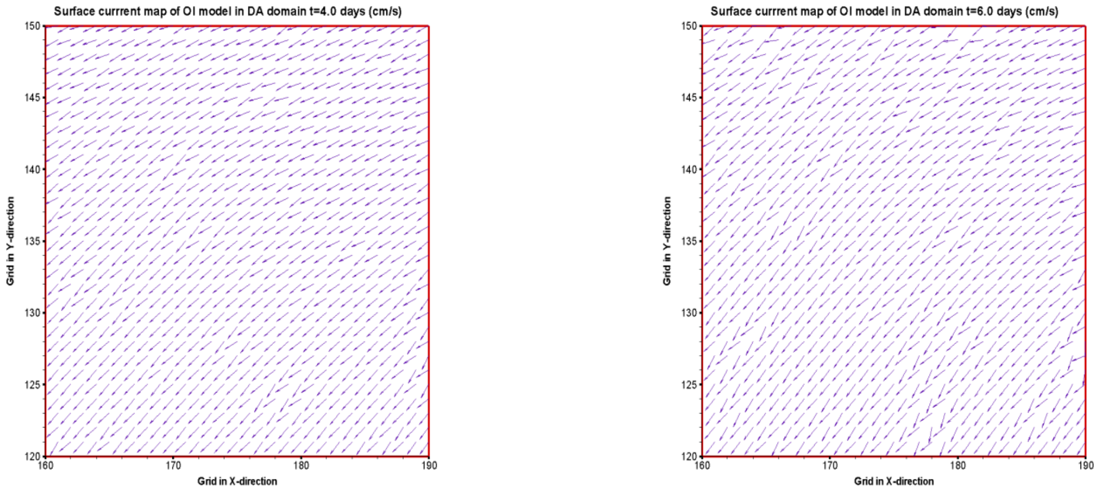

FIGURE 4 Surface current map with Optimal Interpolation data assimilation (t=4.0 days and t=6.0 days)

In order to further assess the effectiveness of Optimal Interpolation data assimilation, Taylor diagram (see Figure 5-12) of two inside reference point C, point D and two boundary reference (point A and point B) are displayed (see Figure 1). Data of these figures is from t=3.0 days to t=7.0 days. Taylor diagram is a graphical way to summarise degree of match between observation and reference models, their statistic (correlation, centred root-mean-square difference and standard deviation) could be concisely shown in terms of each model's position in the diagram (Taylor, 2001). In the following Taylor diagrams, the blue point means the results of data assimilation model, the red point stands for the results of control model with no data assimilation process from Control run, the hollow black dot on the horizontal axis is the measurement state. The centred root-mean-square difference between the modeming results and observed patterns is proportional to the distance to the point on the x-axis identified as measurement. The dotted contours indicate the RMS values. The dotted line from the origin to arch shows the correlation relationship between observation and modelling states. The standard deviation of the modelling results is proportional to the radial distance from the origin. Generally, the RMS values of assimilation states (blue point) is smaller than Control modelling results (red line), especially for point B, which means that the Optimal Interpolation data assimilation process makes the numerical modelling takes useful measurement information into the dynamic system. There is not obvious improvement of correlation relationship for point *A* and point *D*, but for point *B* and *v* component of point *C*, in other words, the assimilation process renders the model have a closer correlation relationship with measurement states. For the standard deviation, since the pseudo measurement is produced by adding normal distribution noise to the results from original modelling. Surface velocity components of four points time series are outputted at the exact data assimilation time step with five minutes assimilation interval. The chaotic pseudo measurement is not smooth comparing with results of the Control run, so the assimilation model could not show smaller standard deviation when noisy pseudo measurement is used. This also proves that the quality of measurement is of great importance for our data assimilation, although the measurement error is considered in the Kalman gain, check of measurement data is needed. From the below Taylor diagrams, the majority of the Taylor diagrams show that the Optimal Interpolation data assimilation process improves the modelling prediction ability, making the model states closer to observation states, which means that Optimal Interpolation data assimilation is an effective tool to enhance the numerical modelling by blending the available measurement data.

Since the surface u and v pseudo measurement are assimilated in the numerical modelling respectively, the impact of data assimilation on the direction of total velocity is also investigated. Taylor diagrams of these reference points direction (t=3.0~7.0 days) are shown as follows. Except for point A, the correlation relationship between measurement states and assimilation states is enhanced and the RMS values are reduced with assimilation.

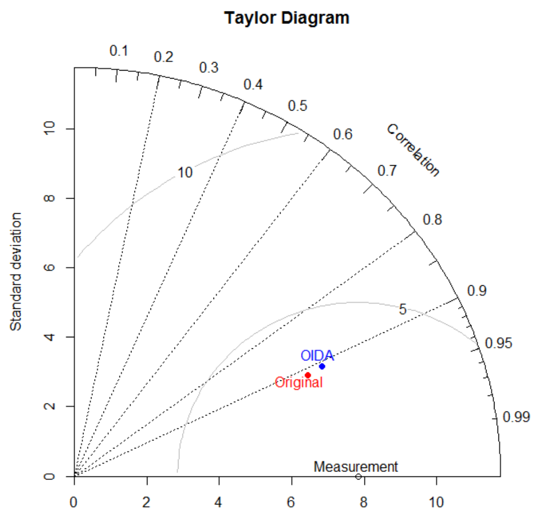

FIGURE 5 Taylor diagram of u component at point A

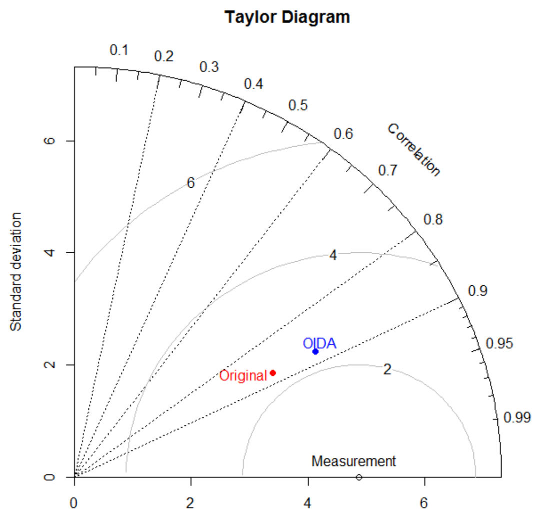

FIGURE 6 Taylor diagram of v component at point A

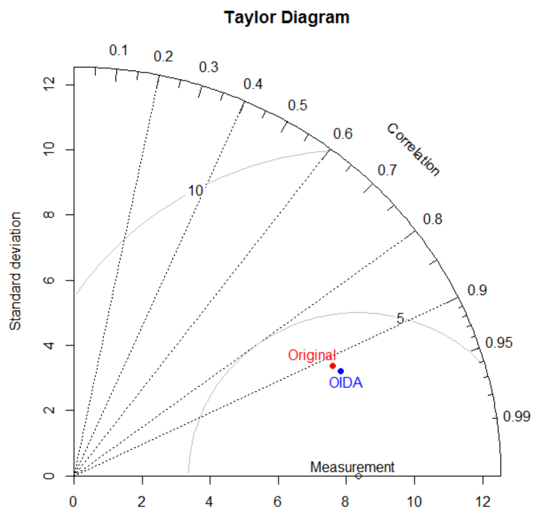

FIGURE 7 Taylor diagram of u component at point B

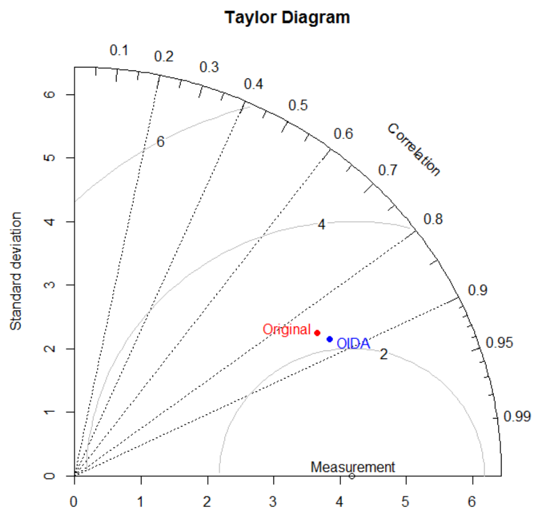

FIGURE 8 Taylor diagram of v component at point B

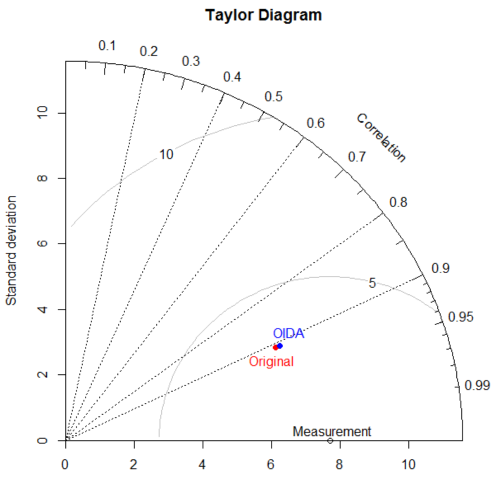

FIGURE 9 Taylor diagram of u component at point C

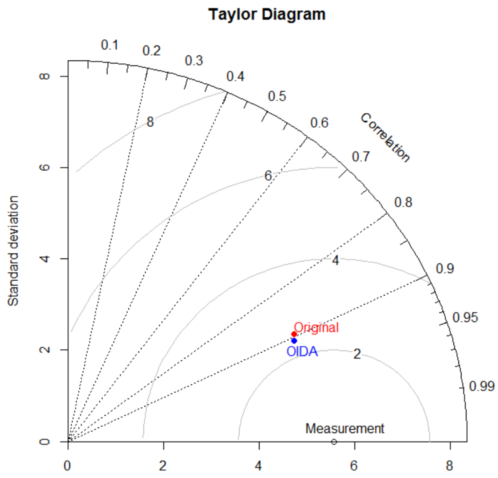

FIGURE 10 Taylor diagram of v component at point C

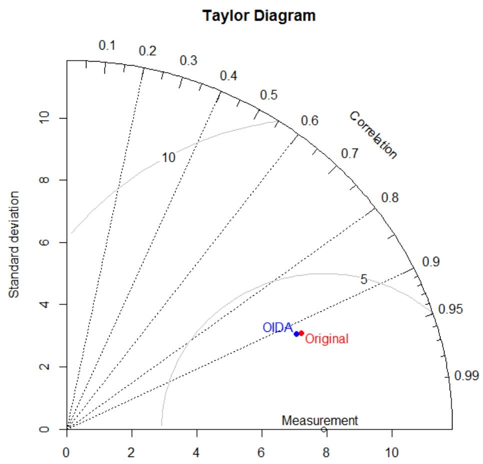

FIGURE 11 Taylor diagram of u component at point D

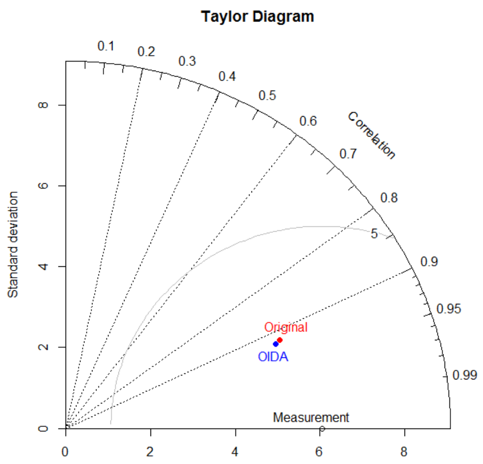

FIGURE 12 Taylor diagram of v component at point D

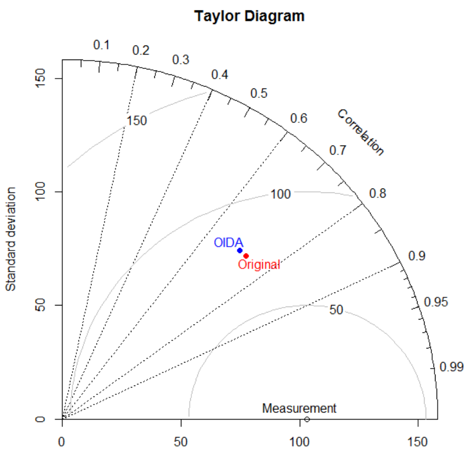

FIGURE 13 Direction Taylor diagram at point A

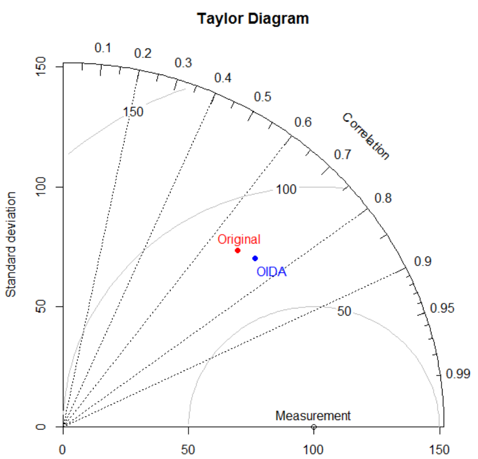

FIGURE 14 Direction Taylor diagram at point B

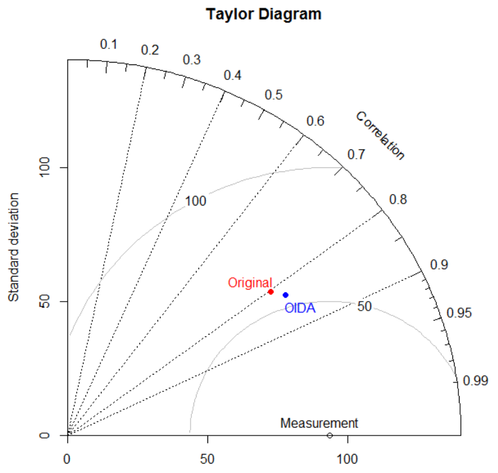

FIGURE 15 Direction Taylor diagram at point C

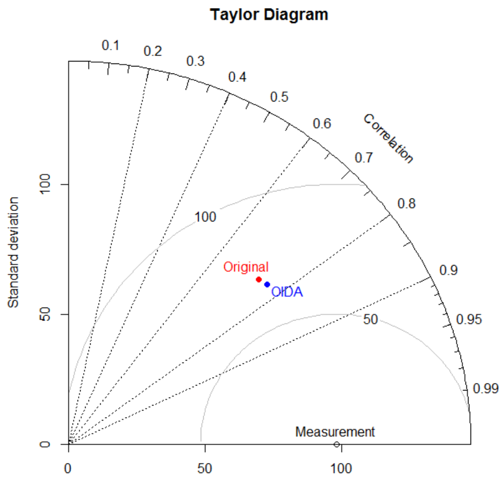

FIGURE 16 Direction Taylor diagram at point D

6 Conclusions

Taylor diagram at reference points and surface current maps in data assimilation domain showed that application of Optimal Interpolation to update model background states with pseudo measurement improves the modelling prediction ability in data assimilation domain, which means the method to calculate the background error covariance in our data assimilation system is meaningful. The improvement of Optimal Interpolation data assimilation is not obvious or the added analysis increment contaminates the background states in certain area or at few points (point A, v component). This is due to the modelling error covariance could not well stand for the development the modelling error, namely the background error covariance is stationary (Oke, Brassington, Griffin and Schiller, 2010, Counillon and Bertino, 2009, Oke, 2002). For further research, authors are trying to develop an operational real time forecasting surface current data assimilation system, real in situ measurement data will be used to update the background state in the following work.

The background error covariance was calculated from a free run, the model error was defined by subtracting the mean of background states over ten minutes from the model states. The time series improvement of surface velocity components at four inside reference points during the last four simulation days is displayed in Taylor diagram, the shown statistical values in Taylor diagrams depict that the model states are closer to the measurement trajectory when the Optimal Interpolation data assimilation is applied. The surface current map at certain time steps describe the Optimal Interpolation data assimilation process has assimilated the useful information from measurement into the model. When comparing with the original model that contained no data assimilation process, the general regional tends to follow the measurement trend after assimilation, which proves that the method used to compute the background error covariance is reasonable and Optimal Interpolation data assimilation works when pseudo measurement is used to update the model states. Optimal Interpolation data assimilation scheme of updating with pseudo measurement does improve the model prediction ability, which proves that this new way of computing background error covariance is efficient.

Acknowledgements

We acknowledge the funding from College of Engineering and Informatics, National University of Ireland, Galway (NUIG) and China Scholarship Council (CSC). This research was carried out as a part of the EnergyMare project, which is funded by the European Regional Development Fund (ERDF) through the Atlantic Area Transnational Programme (INTERREG IV). We also would like to thank Informatics Research Unit for Sustainable Engineering (IRUSE) providing the weather data.

References

[1] Robinson A R, Lermusiaux P F J 2000 Overview of data assimilation Harvard Reports in Physical/Interdisciplinary (Ocean Sciences); The Division of Engineering and Applied Sciences, Harvard University: Cambridge, Massachusetts (NUMBER 62).
[2] Moore A M, Arango H G, Broquet G, Powell B S, Weaver A T, Zavala-Garay J 2011 The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems *Progress In Oceanography* **91**(1) 34-49
[3] Ma Z, Zheng Y, Zhong S, Zou X 2014 Study of maneuvering target tracking algorithm based on Kalman filter and ANFIS *Computer Modelling & New Technologies* **18**(1) 31-7
[4] Zaron E D 2009 *Introduction to Ocean Data Assimilation*
[5] Dong J, Xue M 2012 Assimilation of radial velocity and reflectivity data from coastal WSR-88D radars using an ensemble Kalman

filter for the analysis and forecast of landfalling hurricane Ike *Q J Roy Meteor Soc* **139**(671) 467-87

[6] Gu B, Woo S, Kim S 2011 Application of Optimal Interpolation to Vertical Current Data in the Ieo Island of Korea Using Unstructured Grid Ocean Numerical Model (FVCOM) *OCEANS* 2012 - YeosuYeosu, 1-8

[7] Rienecker M M 1991 Ocean Data Assimilation Using Optimal Interpolation With a Quasi-Geostrophic Model *Journal of Geophysical Research* **96** 93-103

[8] Oke P R, Brassington G B, Griffin D A, Schiller A 2010 Ocean data assimilation a case for ensemble optimal interpolation *Australian Meteorological and Oceanographic Journal* **59** 67-76

[9] Counillon F, Bertino L 2009 Ensemble Optimal Interpolation: multivariate properties in the Gulf of Mexico *Tellus A* **61**(2) 296-308

[10] Oke P R, Brassington G B, Griffin D A, Schiller A 2008 The Bluelink ocean data assimilation system (BODAS) *Ocean Modelling* **21**(1-2) 46-70

[11] Srinivasan A, Chassignet E P, Bertino L, Brankart J M, Brasseur P, Chin T M, Counillon F, Cummings J A, Mariano A J, Smedstad O M, Thacker W C 2011 A comparison of sequential assimilation schemes for ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM): Twin experiments with static forecast error covariances *Ocean Modelling* **37** 85-111

[12] Carton J A, Giese B S 2008 A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA) *Monthly Weather Review* **136**(8) 2999-3017

[13] Doucet A, de Freitas N, Gordon N 2001 *Sequential Monte Carlo methods in practice* Springer

[14] Kalnay E 2002 *Atmospheric Modeling, Data Assimilation and Predictability* Cambridge University Press.

[15] Robert C l, Blayo E, Verron J 2006 Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean *Proc., Modelling, observing, and forecasting sea level, ocean tides and ocean circulation; reviews and recent progress Special issue in honour and in memory of Christian Le Provost, Springer* 624-33

[16] Oke P R 2002 Assimilation of surface velocity data into a primitive equation coastal ocean model *Journal of Geophysical Research* **107**(C9)

[17] Ragnoli E, Zhuk S, Donncha F O, Suits F, Hartnett M 2012 An Optimal Interpolation Scheme for Assimilation of HF Radar Current Data into a Numerical Ocean Model *Hampton Road* **VA** 1 - 5

[18] Taylor K E 2001 Summarizing multiple aspects of model performance in a single diagram *Journal of Geophysical Research* **106**(D7) 7183.

Authors

Lei Ren, born on June 10, 1985, Hubei province, China

Current position, grades: PhD student
University studies: Master degree was earned in major of harbour, coastal and offshore engineering, Hohai University in 2011. He began his PhD research from October, 2011 in National University of Ireland, Galway
Scientific interest: Hydrodynamic modelling, data assimilation

Stephen Nash, born on June 4, 1976, Ireland

Current position, grades: lecturer
University studies: Ph.D. degree was earned in major of civil engineering, National University of Ireland, Galway in 2010
Scientific interest: physical and computer modelling of oceanic, estuarine and freshwater systems

Michael Hartnett, born on August 21, 1959, Birmingham, England

Current position, grades: professor
University studies: Ph.D. degree was earned in major of civil engineering, Trinity College Dublin in 2000.
Scientific interest: marine modelling: tidal energy; water quality; climate change; coastal flooding
Publications: 39 journal publications and 61 conference publications