Sensitivity Tests of Direct Insertion Data Assimilation with Pseudo Measurements

Lei Ren, Michael Hartnett, and Stephen Nash

Abstract—Data Assimilation (DA) was a technique to improve the modelling forecasting ability. The measured data was blended with model background states in the DA process. Whether the model was robust to synthesize the measured data needs to be studied. In our research, Direct Insertion (DI) data assimilation was applied to update the model background states. In order to test the sensitivity of Direct Insertion data assimilation, five kinds of data assimilation model with different time interval were utilized. The influence of Direct Insertion data assimilation on surface layer was also studied. A small test area was chosen for data assimilation. Constant pseudo measurements were used to update the surface velocity components in data assimilation domain.

Index Terms—Data assimilation, direct insertion, sensitivity test, pseudo measurement.

I. INTRODUCTION

Data assimilation combines available measured data with model background states to obtain improved states [1]-[4]. Generally, there are two kinds of data assimilation schemes, sequential and variational data assimilation [5]-[7]. The formal combines the measurement data into the numerical model [8]-[11]. The latter projects the difference between measurement and model background states at every calculation steps [2], [12].

Direct Insertion is a simple sequential data assimilation scheme. This is used to reset the initial conditions of surface current at data assimilation steps with constant pseudo measurements [10], [13]-[17]. The main aims of this study are to test the sensitivity of data assimilation time interval, check the stability of numerical modeling and explore the influence of Direct Insertion data assimilation on surface layer.

II. NUMERICAL MODELLING

The Environmental Fluid Dynamics Code (EFDC) simulates the dynamic process of Galway Bay in Ireland. The EFDC model was developed at the Virginia Institute of Marine Science and is currently supported by the U.S. Environmental Protection Agency (EPA) [18]. The numerical model EFDC solves the three-dimensional, vertically hydrostatic, free surface, turbulent averaged equations of motions for a variable density fluid. The module uses a sigma vertical coordinate and curvilinear, orthogonal horizontal coordinates. The model has been applied to a

Manuscript received April 19, 2014; revised July 10, 2014.

variety of modelling studies. EFDC could provide the capability of internally linking four major modules: hydrodynamics, water quality, sediment transport, toxics [19], [20]. In our research, there are 380×241 grids in the 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). Fig. 1 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.



Fig. 1. Data assimilation domain and Galway bay research area.

The meteorological data (temperature, rain, solar radiation and relative humidity etc.) is obtained from the weather station located at National University of Ireland, Galway. The river inflow of River Corrib was got from the Office of Public Works. 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. The model is initialized for seven days from 14/10/2011 to 20/10/2011. In order to clearly show the results from EFDC model, *t*=0.0 means 14/10/2011 00:00.

III. DIRECT INSERTION DATA ASSIMILATION

A. Sensitivity Test of Data Assimilation Interval

Insertion data assimilation pseudo Direct with measurement is to check the stability of EFDC and to influence of data assimilation on surface layer. Direct insertion data assimilation is that the surface velocity components are substituted by pseudo measurement at certain time steps. Regarding the updating, there is an implicit assumption under the Direct Insertion, "the measurement" represents the true state very well. The Direct Insertion is to render the dynamical process of model follow the general trend of the measurement trajectory. In our current research, constant values of surface velocity components u=v=25 cm/s are taken as our measured data in data assimilation area. Since these data is not from the real in-situ observation, it is called pseudo measurement. There are three reasons to apply this kind of pseudo measurement

The authors are with the Ryan Institute, National University of Ireland, Galway (e-mail: leirencomeon@gmail.com, michael.hartnett@nuigalway.ie, stephen.nash@nuigalway.ie).

data into our data assimilation system: firstly, the measurement interval is available to control. We can use different DA time intervals to test the sensitivity of sequential data assimilation. Secondly, we can produce the data for all the grid points in our data assimilation domain, which ignores the influence resulting from data deficiency. Thirdly, since we set the measurement as constant for all the data, there is an obvious difference between model background states and measurement, the difference changes in time. Since in reality, the in-situ measured data is always noisy. The obviously changing difference can better show this real trend.

For the Direct Insertion, the assimilation equation could be expressed as follows:

$$X_{t} = \begin{cases} measurement & t = t_{1} \\ mod \ el \cdot resolution & t = t_{2} \end{cases}$$

where X_{i} means the state vector

 t_1 is the assimilation time step

 t_2 is the time step with no assimilation

At the data assimilation steps, the initial conditions of surface current in data assimilation domain is reset by using the known measured values. For the rest steps, the model calculates the dynamic process.

In order to test the reaction of model with different assimilation time interval, five kinds of Direct Insertion data assimilation model are run. Table I shows the Direct Insertion data assimilation models with different updating interval. The numerical model EFDC needs few days to reach stable state, the Direct Insertion data assimilation starts to work from the end of the third day till the end of run.

TABLE I: DIRECT INSERTION MODELS			
Model	DI	Updating interval (minutes)	Measurement (cm/s)
Original model	No		
DI-C01	Yes	5	u=v=25
DI-C02	Yes	10	<i>u</i> = <i>v</i> =25
DI-C03	Yes	15	<i>u</i> = <i>v</i> =25
DI-C04	Yes	20	u=v=25
DI-C05	Yes	25	<i>u</i> = <i>v</i> =25

The following figures display the total surface velocity time series at point D (see Fig. 1). The red line means the original model or free run. The blue line is the results from Direct Insertion data assimilation models. Since model needs few days to reach stable states. The comparison happens during the last four days.





The output time interval is three minutes in all models for comparison. From Fig. 2-Fig. 6, it is clean that the shorter data assimilation time interval, the stronger influence. When the data assimilation interval is ten or twenty minutes, the influence of Direct Insertion data assimilation at point D is not obvious. This may be due to the output time interval of modelling results is three minutes. The outputted data is not at the exact Direct Insertion time step. There is a delay weaken influence of Direct Insertion data assimilation. Since the extreme constant values are used to update the model background states, the model is still stable after Direct Insertion data assimilation for four days. This proves that the EFDC model is robust to develop a data assimilation system. The constant values 25 cm/s are taken as pseudo measurement for both surface velocity components. The trend of Direct Insertion data assimilation model is closing to the pseudo measurement trajectory in Fig. 2. The closing trend is more obvious when the Direct Insertion data assimilation interval is smaller.

B. Influence of Direct Insertion on Surface Layer

The Direct Insertion data assimilation process works in the test domain, every certain time steps. That makes the initial conditions of surface velocity components reset with constant values at every data assimilation step. The difference distribution of surface velocity components between Direct Insertion model and original model with no data assimilation is shown.

In order to show the significant influence of Direct Insertion, the following Fig. 7-Fig. 10 are plotted at ebb and flood tide with Direct Insertion data assimilation interval five minutes.



Fig. 7. Difference distribution of surface u component flood tide (cm/s).



Fig. 8. Difference distribution of surface v component flood tide (cm/s).



Fig. 9. Difference distribution of surface *u* component ebb tide (cm/s).



Fig. 10. Difference distribution of surface v component ebb tide (cm/s).

The above figures show that the Direct Insertion data assimilation affects the surface velocity components of the outside area. The main difference appears at the adjacent area. The extreme values of difference appear in the data assimilation domain.

IV. DISCUSSION

The performance of Direct Insertion data assimilation with different assimilation interval by using constant pseudo measurement shows that the EFDC is a robust tool to build up coastal surface currents data assimilation system. The shorter data assimilation interval is, the stronger trend of the model background states close to the measurement trajectory. Moreover, the application of Direct Insertion data assimilation in a small test area has an impact on the initial conditions of outside points, the influence degree decreases with distance. The significant difference between Direct Insertion data assimilation model and original model with no data assimilation appears in data assimilation domain. From the perspective of energy balance, Direct Insertion data assimilation with constant pseudo measurement to update the surface current breaks the internal balance of model. The increased or decreased energy from the psudeo measurement is balanced by the model itself. In other word, the increased is extend to the outside area, the decreased is absorbed from the adjacent area. Our future work is to apply more advanced sequential data assimilation schemes, such as Optimal Interpolation (OI), Ensemble Optimal Interpolation (EnOI) into EFDC model by blending real time in-situ measurement data from high frequency Coastal Ocean Dynamics Applications Radar (CODAR) system [21].

V. CONCLUSIONS

Direct Insertion is a simple sequential data assimilation scheme, in this study, extreme constant values are used as pseudo measurement states to update the surface velocity components. We focus on the sensitivity test of different data assimilation interval and the influence of data assimilation on surface layer. The main results could be listed as follows:

- The EFDC model is robust to frequently combine extreme constant measurement states with the model background states. This is of great importance when we use in-situ measurement data to update our background model states, since there are always noises in the real measured dataset.
- 2) The model is sensitive to the data assimilation interval, the shorter Direct Insertion data assimilation interval, the stronger influence on the model background states. When the Direct Insertion data assimilation time interval is five minutes, the analysis states following the pseudo measurement trajectory is obvious.
- 3) The Direct Insertion data assimilation has an impact on the surface area outside of data assimilation domain, the main difference comes out in the data assimilation domain, the influence degree of Direct Insertion data assimilation decreases with the distance to data assimilation domain.

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) for providing us the atmospheric data.

REFERENCES

- [1] P. R. Oke *et al.*, "The Bluelink ocean data assimilation system (BODAS)," *Ocean Modelling*, vol. 21, pp. 46-70, 2008.
- [2] G. Broquet *et al.*, "Application of 4D-variational data assimilation to the California Current System," *Dynamics of Atmospheres and Oceans*, vol. 48, pp. 69-92, 2009.
- [3] C. I. Robert *et al.*, "Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean," *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*, Toulouse, France, 2005, pp. 624-633.
- [4] D. R. Jackson et al., "Use of Canadian quick covariances in the met office data assimilation system," *Quarterly Journal of the Royal Meteorological Society*, vol. 134, pp. 1567-1582, 2008.
- [5] A. R. Robinson and P. F. J. Lermusiaux, "Overview of data assimilation," *Harvard Reports in Physical/Interdisciplinary (Ocean Sciences); The Division of Engineering and Applied Sciences*, Harvard University: Cambridge, Massachusetts, 2000.
- [6] B. S. Powell *et al.*, "4DVAR data assimilation in the Intra-Americas sea with the regional ocean modeling system (ROMS)," *Ocean Modelling*, vol. 25, pp. 173-188, 2008.
 [7] A. M. Moore *et al.*, "The regional ocean modeling system (ROMS)
- [7] A. M. Moore *et al.*, "The regional ocean modeling system (ROMS) 4-dimensional variational data assimilation systems," *Progress in Oceanography*, vol. 91, pp. 34-49, 2011.
- [8] M. R. Turner *et al.*, "Ensemble member generation for sequential data assimilation," *Remote Sensing of Environment*, vol. 112, pp. 1421-1433, 2008.
- [9] X. Li and X. J. Han, "An evaluation of the nonlinear/non-gaussian filters for the sequential data assimilation," *Remote Sensing of Environment*, pp. 1434-1449, 2007.
- [10] Y. Han et al., "A new sequential data assimilation method," Science in China Series E Technological Sciences, vol. 52, pp. 1027-1038, 2009.
- [11] J. Dong and M. Xue, "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," *Quarterly Journal of the Royal Meteorological Society*, vol. 139, pp. 467-487, January 2013.
- [12] J. L. Steward *et al.*, "Impact of non-smooth observation operators on variational and sequential data assimilation for a limited-area shallow-water equation model," *Quarterly Journal- Royal Meteorological Society*, vol. 138, pp. 323-339, 2012.
- [13] G. Gopalakrishnan, "Surface current observations uning high frequency radar and its assimilation into the New York harbor

observing and prediction system," Doctor of Philosophy, Stevens Institute of Technology, 2008.

- [14] E. D. Zaron, "Introduction to ocean data assimilation," in *Operational Oceanography in the 21st Century*, 2009, ch. 13, pp. 321-350.
- [15] K. Ide *et al.*, "Unified notation for data assimilation operational, sequential and variational," *Journal of the Meteorological Society of Japan*, 1997.
- [16] A. V. Penenko, "Sequential variational data assimilation [6160-83]," *Atmospheric and Ocean Optics/Atmospheric Physics*, Tomsk, Russia, 2005, p. 61602D.
- [17] L. J. Walstad and D. J. McGillicuddy, "Data assimilation for coastal observing systems," *Oceanography*, vol. 13, pp. 47-53, 2000.
- [18] J. M. Hamrick, A Three-Dimensional Environmental Fluid Dynamics Computer Code: Theroretical and Computational Aspects, Virginia Institute of Marine Science, School of Marine Science, The College of William and Mary, 1992.
- [19] R. Zou *et al.*, "An integrated hydrodynamic and water quality modeling system to support nutrient TMDL development for Wissahickon Creek," *Journal of Environmental Engineering*, vol. 132, pp. 555-566, 2006.
- [20] K. R. Jin and Z. G. Ji, "Case study: Modeling of sediment transport and wind-wave impact in Lake Okeechobee," *Journal of Hydraulic Engineering*, vol. 130, pp. 1055-1067, 2004.
- [21] L. Ren *et al.*, "Empirical Orthogonal Function analysis of High Frequency radar surface currents in Galway Bay," *3rd IAHR Europe Congress*, Porto, Portugal, 2014.



Lei Ren was born in Hubei province, China. He received the master degree in harbor, coastal and offshore engineering, College of Harbor, Coastal and Offshore Engineering, Hohai University in 2011. He is studying in water engineering as a PhD student in National University of Ireland, Galway from October, 2011.



Michael Hartnett was born in Birmingham, England. He received Ph.D. degree in civil engineering, Trinity College Dublin in 2000. He is a professor in College of Engineering and Informatics, National University of Ireland, Galway. He has published 39 journal papers and 61 conference papers. His scientific interest is marine modelling: tidal energy; water quality; climate change; coastal flooding.



Stephen Nash received Ph.D. degree in civil engineering, National University of Ireland, Galway, in 2010. He is a lecturer in College of Engineering and Informatics, National University of Ireland, Galway. His scientific interests are physical and computer modelling of oceanic, estuarine and freshwater systems.