Journal of Marine Science and Technology; Vol. 18, No. 4; 2018: 358–367
DOI: 10.15625/1859-3097/18/4/12474
http://www.vjs.ac.vn/index.php/jmst

APPLICATION OF DATA ASSIMILATION METHOD FOR WAVE HEIGHT IN EASTERN VIETNAM SEA BY THE ENSEMBLE KALMAN FILTER

Nguyen Trung Thanh[1,*], Nguyen Minh Huan[2], Tran Quang Tien[3]

[1]*Northern Center for Planning and Investigation of Marine Resources-Environment, Hanoi, Vietnam*
[2]*Hanoi University of Science, VNU, Vietnam*
[3]*The National Centre for Hydro-Meteorological Forecasting*
[*]E-mail: thanhnt2212@gmail.com

Received: 26-4-2018; accepted: 17-7-2018

Abstract. Advances in ocean data assimilation have been improved with the power of computers over the past 30 years, but the theory of data assimilation has a long history with the probability, theory of mathematical simulation, and traditional numerical calculations. Therefore, the research and application of data assimilation technology in oceanography have been widely carried out in the world and have gotten considerable achievements. In Vietnam, the study of data assimilation has also been applied in the meteorology and has attracted the increasing interest in oceanography. This paper presents the results of the study to calculate the wave height data from satellites combined with measured wave height data at MSP1 station of Vietsovpetro using the Ensemble Kalman Filter (EnKF) method associated with SWAN wave model in the Eastern Vietnam Sea.

Keywords: SWAN, Calibration, SWAN-CI, EnKF, Bien Dong.

INTRODUCTION

Data assimilation technology in general and Ensemble Kalman Filter (EnKF) in particular have an important significance in increasing the accuracy of forecasting models. Nowadays, there are many different types of data assimilation schemes such as a Three-Dimensional Variational Data Assimilation (3DVar), Four-Dimensional Variational Data Assimilation (4DVar), Optimal Interpolation (IO) and so on. However, EnKF is assessed to have some outstanding advantages such as no necessity to develop tangential models, which allows EnKF to be easily applied to models without interfering with the core of the model. Furthermore, EnKF allows the creation of an initial noise field that changes over time [1]. The paper conducted a combination test of SWAN-CI model and EnKF method in the OpenDA software version 2017 developed by Martin Verlaan, Delft, the Netherlands [10] with SWAN model in the Eastern Vietnam Sea set on triangle grid. The data used in this research include direct measurements from satellites as tracks combined with data at the oil platform MSP1 of Vietsovpetro Company.

RESEARCH MATERIALS AND METHODS

Open DA software. OpenDA is a software developed for data assimilation, model stability analysis and model parameter calibration [10]. This is an open source software using Java programming language, supporting parallel computing, this software is developed independently of the models so users can easily add features. Users also apply the tools available in this software with any models.

WAN-CI - Automatic Parameter Calibration Method [7] is built into OpenDA to calibrate the parameters of SWAN model. The author studied experimental calculations with this method and published in [12].

This study focused on the application of a combination test of automatic parameter calibration method SWAN-CI and EnKF method for SWAN model based on data from satellites and data at the oil platform MSP1 of Vietsovpetro Company.

SWAN model. The SWAN wave model is a third-generation spectral wave model, is designed to obtain realistic wave estimates for any region of wind-generated surface gravity waves. The physics are governed by the wave action equation, Booij et al., [2].

$$\frac{\partial N}{\partial t}+\nabla_{\vec{x}}\cdot\left[\left(\vec{c}_g+\vec{U}\right)N\right]+\frac{\partial c_\sigma N}{\partial \sigma}+\frac{\partial c_\theta N}{\partial \theta}=\frac{S_{tot}}{\sigma} \quad (1)$$

Where: $N=N\left(\vec{x},t,\sigma,\theta\right)$ is the wave action density spectrum, as a function of the physical space and time $\vec{x}$ and t, σ and θ are the relative frequency and the wave direction. The first two terms on the left-hand side of balance equation are the local and advective rates of change of wave action density, where $\vec{c}_g$ is the wave group velocity, and $\vec{U}$ is the ambient current velocity. The third term, $\frac{\partial c_\sigma N}{\partial \sigma}$, represents the relative frequency shifting due to variations in depth and ambient current. The fourth term, $\frac{\partial c_\theta N}{\partial \theta}$, represents the depth-induced and current-induced refraction, where c_σ and c_θ are the propagation speeds in spectral space σ and θ. On the right-hand side, the term S_{tot} includes all of the source/sink terms and represents all the physical processes that generate, dissipate, and redistribute wave energy, Komen et al., [8].

The source/sink terms are defined via the energy density $E\left(\sigma,\theta\right)$. S_{tot} consists of six main parts:

$$S_{tot}=S_{in}+S_{nl3}+S_{nl4}+S_{ds,w}+S_{ds,b}+S_{ds,br} \quad (2)$$

Where: S_{in} denotes the wave growth by the wind, S_{nl3} and S_{nl4} represent nonlinear wave energy transfers due to triad and quadruplet wave-wave interactions, respectively, and $S_{ds,w}$, $S_{ds,b}$, and $S_{ds,br}$ are wave energy dissipations due to white-capping, bottom friction, and depth-induced wave breaking, respectively. Among those terms, the present research will focus on the contribution of wave growth by the wind input S_{in}. It is widely accepted that the transfer of wind energy to the waves may be conceived as the superposition of resonance, Phillips [13] and positive feedback mechanisms, Miles [11].

The SWAN model uses both finite difference and finite element diagrams to integrate the equation of wave balancing effects over time, space and spectrum. The paper uses SWAN model version 41.10 to calculate wave on the triangle grid in the Eastern Vietnam Sea.

EnKF Method. EnKF method is a data assimilation method developed by Evensen [4–6] to apply to nonlinear problems. This method combines Kalman filter method with Monte Carlo method to create a combination of states describing matrix error of covariance and matrix of covariance error.

Normally, the calculation process by EnKF method is divided into 2 steps as follows:

1) Prediction step: Assuming that at time step k there are q composite components to evaluate the state of the model x_k^f by using normal distribution errors. Considering the nonlinear discrete random model:

$$x_k^f=f\left(x_{k-1},\Phi_{k-1}\right)+\omega_{k-1} \quad (4)$$

Where: Φ_{k-1} is the control vector and ω_{k-1} is the average random error at time k-1. The observations corresponding to the state of the model can be described as follows:

$$y_k=Hx_k+\upsilon_k \quad (5)$$

Where: H relates to the model state x with observation y and and υ is the observed error.

If x and y have the same number, then H can become an identity matrix.

Assuming that at the time step k, there are q components of the forecasted model state, then:

$$x_k^f = \left[x_k^{f^1}, x_k^{f^2}, \ldots x_k^{f^q} \right] \quad (6)$$

2) Analysis step: Error of covariance or expected error of prediction model states at the analysis points and observed points can be determined by the ensemble components from the equation (6):

$$P_k^f H^T = \frac{1}{q}\sum_{i=1}^{q}\left(x_k^{f^i} - x_k^t\right)\left(Hx_k^{f^i} - Hx_k^t\right)^T \quad (7)$$

In which the true state x_k^t is unknown, using the average ensemble components to approximate the true state as follows:

$$x_k^t \approx \overline{x_k^f} = \frac{1}{q}\sum_{i=1}^{q} x_k^{f^i} \quad (8)$$

Equation (7) takes the form:

$$P_k^f H^T = \frac{1}{q-1}\sum_{i=1}^{q}\left(x_k^{f^i} - \overline{x_k^f}\right)(Hx_k^{f^i} - H\overline{x_k^f})^T \quad (9)$$

It should be noted that to make the estimate unequal, factor 1/(q-1) is used instead of 1/q.

$P_k H^T$ in equation (9) can be determined directly from the ensemble components and $HP_k H^T$ is the domain matrix for all model states to calculate the variance between the predicted values at the monitoring points. It should be noted that for all q ensemble components, the Kalman K component always has the same value as all the states of the reflected model instead of a specific value of each ensemble component, which reduces significant cost of calculation. In addition, the covariance error matrix does not need to be saved for the next analysis cycle, which also improves model efficiency by freeing up memory space to use for the next steps.

The following figure shows the steps to calculate and calibrate data in OpenDA:

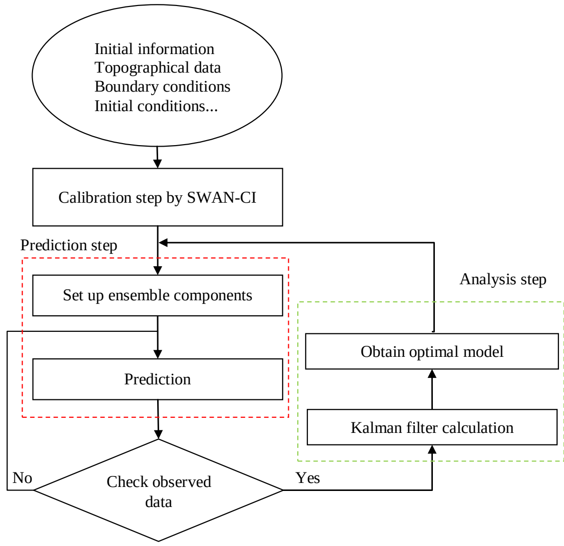

Fig. 1. Flow diagram of calculation steps in OpenDA

After the Kalman matrix K is obtained at the analysis step, the analysis is performed to update all the states of the model even for areas without direct observation in the calculation domain. The transition from the forecast step to the analysis step is done through checking observed data. If no observed data is available, then return to the forecast calculation in the next step, if there is observed data, go to the analysis step.

In order to well perform the data assimilation calculation in OpenDA, it is necessary to prepare the initial information such as topographical data, boundary conditions (wave height, direction and period, wind data), initial conditions, observed wave height in the study area. In addition to assimilating the data, the study also calculates to select the appropriate parameter set of the model for the study area with SWAN-CI, then uses this parameter set to serve the data assimilation.

Model setup and data. The domain is the entire Eastern Vietnam Sea stretching from 1.5^o to 25^{o}N (from 150,000 to 2,800,000 in VN2000 coordinate system, projection zone 105th) and 99^o to 121^{o}E (-150,000 to 2,200, 000 in VN2000 coordinate system, projection zone 105th). The grid is a triangle grid with length of each grid side about 10 km in nearshore areas and gradually offshore. The total number of grid cells is 12,858. The grid is created by ADCIRC software of SMS version 10.0 [9]. There are 3 boundaries: Taiwan Strait (1), Bashi Strait (2) and Malacca Strait (3). The bathymetric data were collected from NOAA [15] and the detailed topographic data in the nearshore area of the marine surveys in the study area.

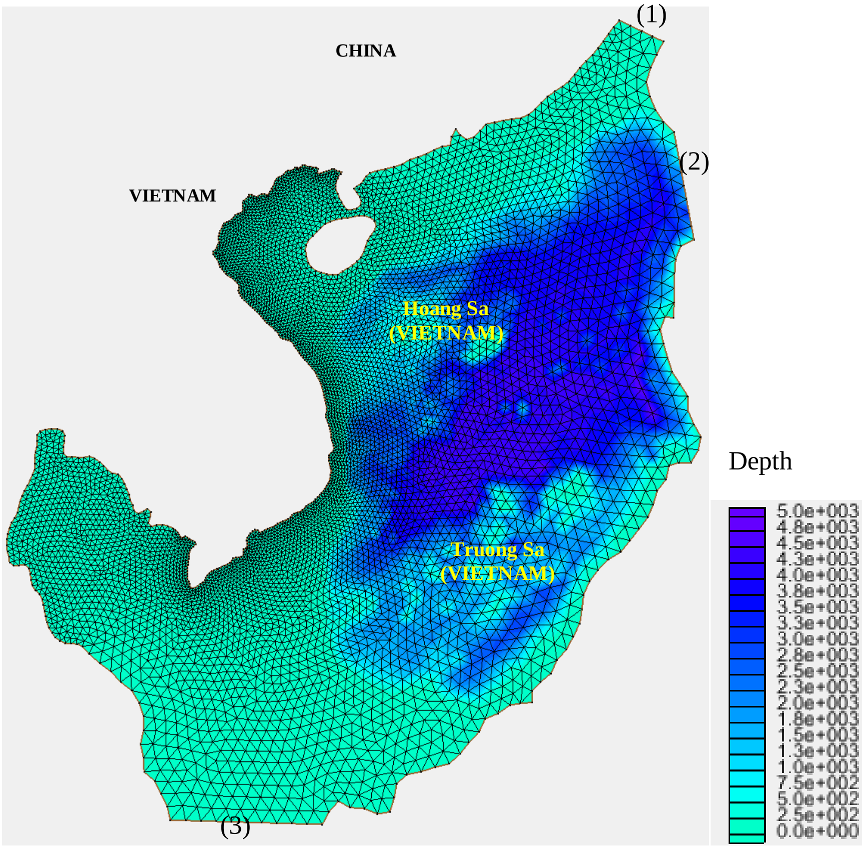

Fig. 2. Depth and grid

Wind data: The global reanalysis wind CFSv2 of National Center for Environmental Forecasting (NCEP) with a grid step of 0.2 × 0.2 degrees and a time of 1 hour from 1979 to present [14].

Wave data in boundaries (1-2-3): Wave height, direction and period are extracted from ECMWF's global re-analysis wave data with a grid of 0.25 × 0.25 degrees, time step 6 hours from 1979 to present [17].

Observed wave data are collected from 2 main sources:

Wave height data from satellites in the form of track routes are collected from Online Data Extraction Service [16], details of the operating time of the satellites are shown in the following table:

Table 1. List of satellites that receive wave height

No.	Satellite name	Activity time
1	ERS-2	1995–2003
2	Geosat - Follow (G2)	2000–2008
3	Jarson1	2002–2013
4	Envisat	2002–2012
5	Jarson2	2008–now
6	HY-2A	2014–now
7	Saral-Altica	2013–now
8	Jarson3	2016–now

With the development of the current satellite technology, the significant wave height (SWH) obtained from satellites has become more and more close to the observation data at buoy stations around the globe and has smaller errors 10% [3]. The paper chose to use the data directly from the above satellites corresponding to the calculation period as an important data source as well as input for calibration and data assimilation.

In order to perform model parameter calibration and data assimilation by satellite wave heights, it is necessary to compile these discrete data for equidistant time steps and for fixed spatial positions. This study uses time step 1 hour and 174 fixed points. Therefore, the wave height is interpolated as follows:

Select satellites and download track data files of each satellite in the simulation period.

Interpolate wave height data from tracks to 174 points (fig. 3): Choose all neighboring points of each point in 174 points with a radius of 0.5 degrees in space and 0.5 hour in time at every hour round to interpolate.

At each point in each time, no data will be assigned a value of -999 so the model does not include these values in the calibration process.

Real wave data measured at MSP1 station provided by Vietsovpetro Company. This data is assigned to point 175 with a 1 hour time step.

These data are processed and converted according to SWAN's input standard as well as OpenDA's to include in the calculation.

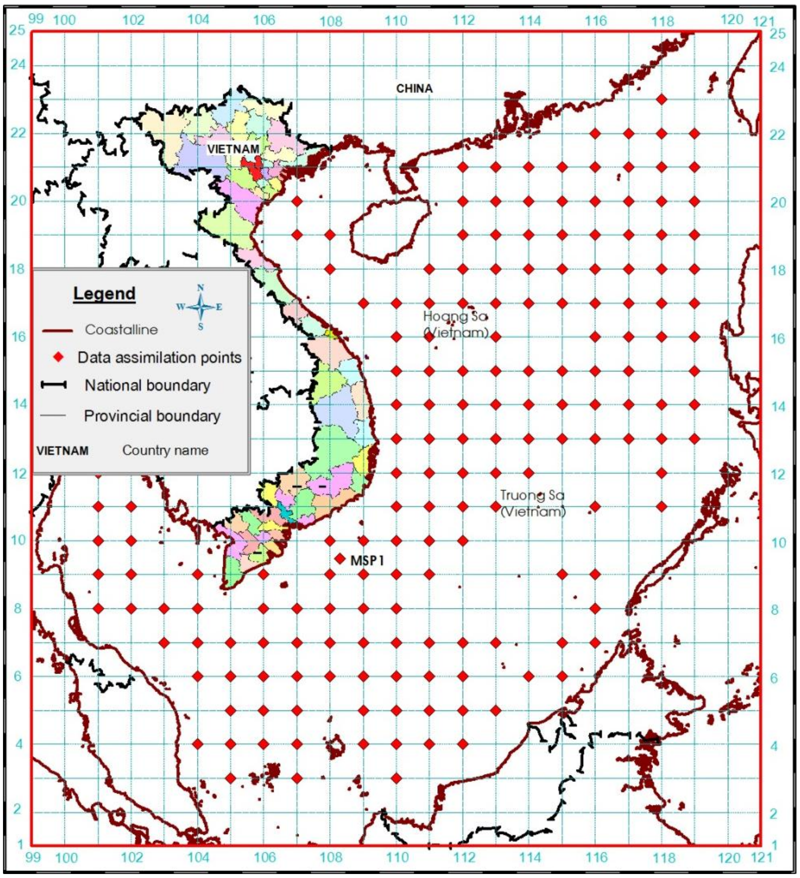

Fig. 3. Location of points involved in calibration and data assimilation

RESULTS AND DISCUSSION

Simulation in 3/2011. The calculation is made from 0:00 on March 6, 2011 to 23:00 on March 29, 2011. The total number of satellite data at the locations and the period is 852, the total data at MSP1 station is 552.

Before assimilating data, this study selected parameters of model with these data series. The calibration is done based on the evaluation of parameters through the GoF function (Goodness of Fitting) for 2 parameters CDS2 (coefficient for determining the rate of white capping dissipation) and POWST (power of steepness normalized with the wave steepness of a Pierson-Moskowitz spectrum). In particular, the GoF function is calculated based on the following formula:

$$GoF = \frac{1}{2}\frac{\sum_{i=1}^{N_i}\left[\omega_{H_s}^{i}\left(H_{s,obs}^{i} - H_{s,sim}^{i}\right)/\sigma_{H_{s,obs}}\right]^2}{\sum_{i=1}^{N_i}\omega_{H_s}^{i\;2}} \quad (10)$$

In which $H_{s,sim}^{i}$ and $H_{s,obs}^{i}$ are simulated and observed significant wave heights at *i*; $\omega_{H_s}^{i}$ is a weight function (for the present moment taken to be equal to 1); N_i is the number of spatial calibration data points; $\sigma_{H_{s,obs}}$ is the standard deviation of the observed parameter in the particular location.

The study used the conjugate gradient method, i.e. with the default parameters CDS2 and POWST in the model and the adjacent values of these two parameters to select the appropriate parameter values for the next step, until the optimal values is reached or the GoF function is minimum. The advantage of this method is that the rate of rapid convergence and the number of calibration time is less than the sequential method as done in the previous study [12]. The process of selecting parameters is shown in the following figures:

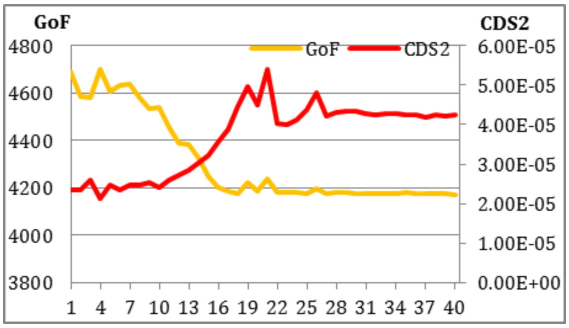

Fig. 4. Variation of the cost function GoF with CDS2 (coefficient for determining the rate of white capping dissipation)

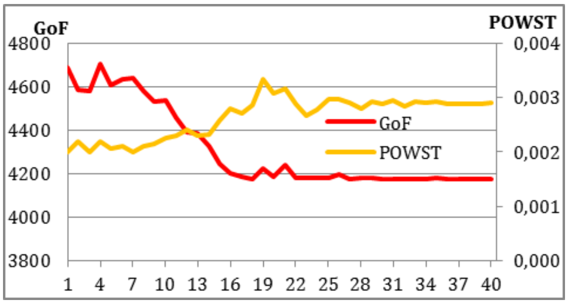

Fig. 5. Variation of the cost function GoF with POWST (power of steepness normalized with the wave steepness of a Pierson-Moskowitz spectrum)

The optimal values obtained after the adjustment of parameters CDS2 and POWST are 4.25 × 10^{-5} and 2.898 respectively. These parameters are used to calculate the data assimilation by EnKF method with the actual measured data series used for calibration. The number of ensemble components included in the calculation is 20, the results are shown as follows:

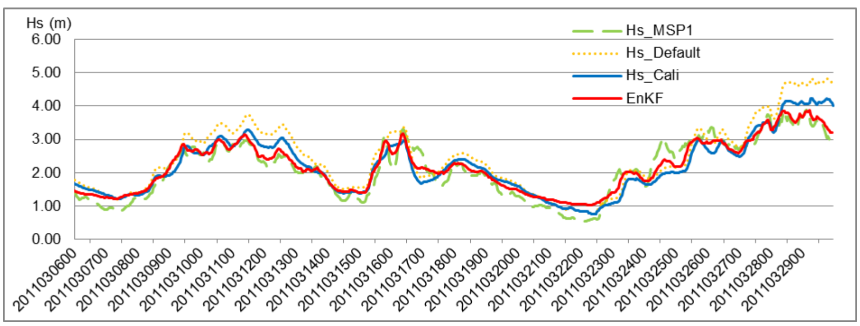

Fig. 6. Comparison of Hs in cases with data at MSP1 station in 3/2011

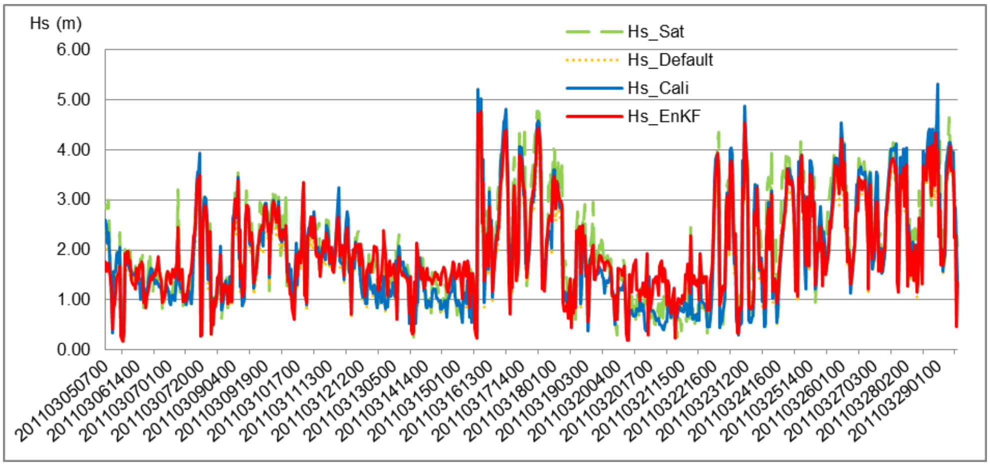

Fig. 7. Comparison of Hs in cases with satellite data in 3/2011

The calculation results are evaluated through Nash index as follows:

The results show that although the calibration process has significantly improved the forecasted results compared to the actual measurement, at some times when there is a big change of wave height value, it is still not close to reality (fig. 6–7). After data assimilation with the EnKF method, the results were more approachable than actual measurements even at times when there was a strong change in wave height. The change in results is shown more clearly through Nash index, this index has increased significantly from 0.57 to 0.78 after calibration and to 0.92 after data assimilation with number data at MSP1 station, from 0.75 to 0.88 after calibration and to 0.89 after data assimilation with satellite data (table 2).

Table 2. Comparison of Nash index in cases in 3/2011

No.	Data type	Nash index in default case	Nash index in calibration case	Nash index in EnKF case
1	MSP1	0.57	0.78	0.92
2	Satellite	0.75	0.88	0.89

Simulation in 10/2013. The calculation is made from 0:00 on October 6, 2011 to 23:00 on October 29, 2011. The total number of satellite data at the locations and the period is 570, the total data at MSP1 station is 552.

The study is done just like the above calculation, the GoF function is calculated based on the following formula.

The optimal values obtained after the adjustment of parameters CDS2 and POWST are 4.47 × 10^{-5} and 1.732 respectively. These parameters are used to calculate the data assimilation by EnKF method with the actual measured data series used for calibration. The number of ensemble components included in the calculation is 20, the results are shown as follows:

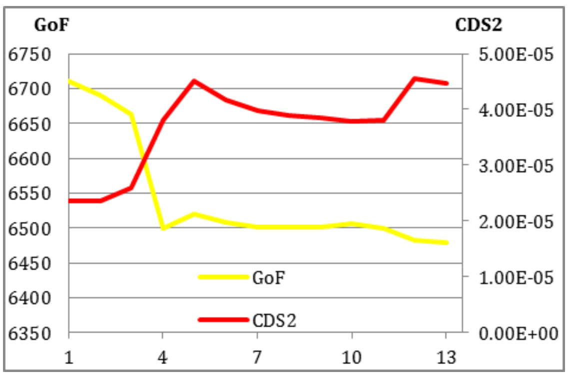

Fig. 8. Variation of the cost function GoF with CDS2 (coefficient for determining the rate of white capping dissipation)

The calculation results are evaluated through Nash index as follows:

The results in this study show that although the calibration process has significantly improved the forecasted results compared to the actual measurement, at some times when there is a big change of wave height value, it is still not close to reality (fig. 10, 11). After data assimilation with the EnKF method, the results were more approachable than actual measurements even at times when there was a strong change in wave height. The change in results is shown more clearly through Nash index, this index has increased significantly from 0.62 to 0.71 after calibration and to 0.87 after data assimilation with number data at MSP1 station, from 0.65 to 0.67 after calibration and to 0.77 after data assimilation with satellite data (table 3).

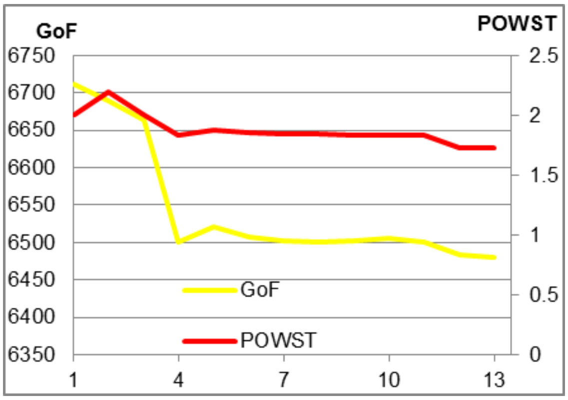

Fig. 9. Variation of the cost function GoF with POWST (power of steepness normalized with the wave steepness of a Pierson-Moskowitz spectrum)

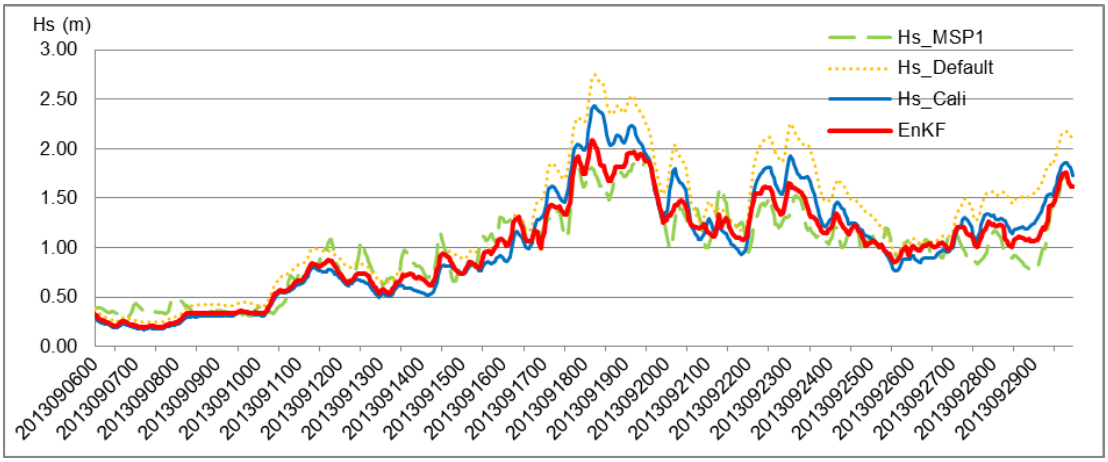

Fig. 10. Comparison of Hs in cases with data at MSP1 station in 10/2013

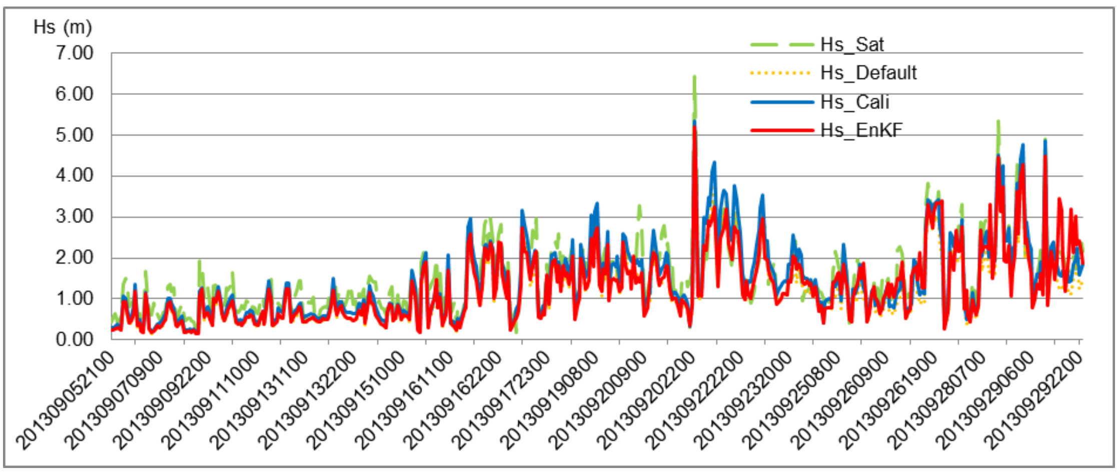

Fig. 11. Comparison of Hs in cases with satellite data in 10/2013

Table 3. Comparison of Nash index in cases in 10/2013

No.	Data type	Nash index in default case	Nash index in calibration case	Nash index in EnKF case
1	MSP1	0.62	0.71	0.87
2	Satellite	0.65	0.67	0.77

CONCLUSION

OpenDA software provides many useful tools for integrating different calibration methods and different data assimilation methods. Therefore, users can choose the method that conforms to their problem.

The calibration of the model by the conjugate gradient method has improved the calculation efficiency and partly improved the quality of the wave forecast calculation by comparing both data from satellites and data at MSP1 station.

After model calibration, this research uses the selected parameters to continue calculating the data assimilation by EnKF method combined with two data series in March 2011 and October 2013, it is shown that the calculation results are significantly improved compared to the calculations after model calibration.

REFERENCES

[1] Kieu Quoc Chanh, 2011. Overview of the Ensemble Kalman Filter and Its Application to the Weather Research and Forecasting (WRF). *Journal of Science, Vietnam National University, Hanoi,* (1S), 17–28.

[2] Booij, N. R. R. C., Ris, R. C., and Holthuijsen, L. H., 1999. A third- generation wave model for coastal regions: 1. Model description and validation. *Journal of geophysical research: Oceans*, **104**(C4), 7649–7666.

[3] Sølvsteen, C., and Hansen, C., 2006. Comparison of altimetry wave and wind data with model and buoy data. In *Proc. 15 years of progress in Radar Altimetry, ESA Symposium, Venice, Italy* (pp. 13–18).

[4] Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi- geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, **99**(C5), 10143–10162.

[5] Evensen, G., 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**(4), 343–367.

[6] Evensen, G., 2004. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, **54**(6), 539–560.

[7] Wenneker, I., Dhondia, J., Hummel, S., Verlaan, M., and Gerritsen, H., 2008. User manual calibration instrument SWAN. *Delft Hydraulics*.

[8] Komen, G., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., and Janssen, P. A. E. M., 1994. Dynamics and Modeling of Ocean Waves. *Cambridge Univ. Press, Cambridge, UK.*

[9] Laboratory, Environmental Modeling Research, 2009. SMS (Surface-water Modeling System).

[10] Martin Verlaan, 2017. Open DA User Documentation. *TU Delft*.

[11] Miles, J. W., 1957. On the generation of surface waves by shear flows. *Journal of Fluid Mechanics*, **3**(2), 185–204.

[12] Nguyen Trung Thanh, Nguyen Minh Huan, Tran Quang Tien, 2017. Application of automated calibration method to calibrate parameters in SWAN model using wave height data from satellite and MSP1 in Eastern Vietnam Sea. *Journal of Marine Science and Technology*, **17**(3), 271–278.

[13] Phillips, O. M., 1957. On the generation of waves by turbulent wind. *Journal of Fluid Mechanics*, **2**(5), 417–445.

[14] Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y. T., Chuang, H. Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Dool, H. V. D., Zhang, Q., Wang, W., Chen, M., and Becker, E., 2014. The NCEP climate

forecast system version 2. *Journal of Climate*, **27**(6), 2185–2208.
[15] http://ngdc.noaa.gov.
[16] http://odes.altimetry.cnes.fr.
[17] http://www.ecmwf.int/en/research/modelling-and-prediction/marine.