Improved sea level anomaly prediction through combination of data relationship analysis and genetic programming in Singapore Regional Waters

Alamsyah Kurniawan a,*, Seng Keat Ooi a, Vladan Babovic b

a Singapore-Delft Water Alliance, National University of Singapore, DW1-02-05, 2 Engineering Drive 2, Singapore 117577, Singapore
b NUSDeltares, National University of Singapore, E1-08-24, 1 Engineering Drive 2, Singapore 117576, Singapore

A R T I C L E   I N F O

Article history:
Received 27 March 2014
Received in revised form 26 May 2014
Accepted 3 July 2014
Available online 16 July 2014

Keywords:
Data model integration
Average mutual information
Error forecasting
Tide-surge interaction

A B S T R A C T

With recent advances in measurement and information technology, there is an abundance of data available for analysis and modelling of hydrodynamic systems. Spatial and temporal data coverage, better quality and reliability of data modelling and data driven techniques have resulted in more favourable acceptance by the hydrodynamic community. The data mining tools and techniques are being applied in variety of hydro-informatics applications ranging from data mining for pattern discovery to data driven models and numerical model error correction. The present study explores the feasibility of applying mutual information theory by evaluating the amount of information contained in observed and prediction errors of non-tidal barotropic numerical modelling (i.e. assuming that the hydrodynamic model, available at this point, is best representation of the physics in the domain of interest) by relating them to variables that reflect the state at which the predictions are made such as input data, state variables and model output. In addition, the present study explores the possibility of employing ‘genetic programming’ (GP) as an offline data driven modelling tool to capture the sea level anomaly (SLA) dynamics and then using them for updating the numerical model prediction in real time applications. These results suggest that combination of data relationship analysis and GP models helps to improve the forecasting ability by providing information of significant predictive parameters. It is found that GP based SLA prediction error forecast model can provide significant improvement when applied as data assimilation schemes for updating the SLA prediction obtained from primary hydrodynamic models.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The understanding of the generating mechanisms and the ability to generate forecasts of tidal and non-tidal flow phenomena (also referred to sea level anomaly) and their forcing mechanisms for the highly complex Singapore Regional Waters are of both scientific and economic importance. Given its geographically constricted location, the island of Singapore, part of the Sunda Shelf, experiences a direct impact of nonlinear dynamical interactions between the South China Sea, Andaman Sea and Java Sea. The complex shallow water hydrodynamics generated due to multiple ocean currents moving into and out of its region, combined with short term meteorological effects leads to a very high variability of the sea water level around Singapore’s coast. In such narrow straits separating larger water bodies, it is often observed that the water currents and levels deviate significantly from their regular tidal behaviour. These deviations or residual components are generally not accounted for during ocean weather forecasting and hence seriously affect coastal planning and navigation in the region. Hence, analysis and accurate prediction of these sea level anomaly (SLA) and current anomaly becomes an important part of oceanographic modelling, especially in of such shallow water zones. A major step in their analysis and forecasting is the development of an accurate hydrodynamic model to predict the non-tidal barotropic water levels and currents in the region (Ooi et al., 2009, 2011; Kurniawan et al., 2013). However, complex governing mechanisms, multi-scale, multi-dimensional, time varying, and highly non-linear dynamics of the marine systems make the oceanographic modelling efforts much more challenging. Conventional numerical models provide primary solution to this challenging task of characterizing and forecasting ocean weather (mainly water level and flow) by representing the underlying physics in terms of solvable equations. Yet, capturing the ocean dynamics in totality, accounting for the non-tidal anomaly calls for rigorous tuning of the models for further improvement. Such an exercise demands detailed domain knowledge and heavy
computational effort. Hence, there is an increasing need for alternate approaches which can provide vital information leading to better domain knowledge and reduced time and effort required to tune the numerical models.

With the recent advances in measurement and information technology, there is an abundance of data available for analysis and modelling of hydrodynamic systems. Increasing spatial and temporal data coverage, better quality and reliability of data modelling and data driven techniques are becoming more favourable and acceptable by the hydrodynamic community. The data mining tools and techniques are being applied in variety of hydroinformatics applications ranging from simple data mining for pattern discovery to data driven models and numerical model error correction (Babovic et al., 2001, 2009; Sannasiraj et al., 2005; Sun et al., 2010; Rao and Babovic, 2010; Karri et al., 2013, 2014; Wang and Babovic, 2014). The objectives of this paper is to explore the feasibility of applying average mutual information (AMI) theory by evaluating the amount of information contained in observed and prediction errors of non-tidal barotropic numerical modelling (i.e. assuming that the hydrodynamic model by Kurniawan et al. (2013), best represents the physics in the domain of interest given all the available data) by relating them to variables that reflect the state at which the predictions are made such as input data, state variables and model output. In addition, the present study also explores the possibility of employing ‘genetic programming’ (GP) as an offline data driven modelling tool to capture the SLA dynamics and then using them for updating the numerical model prediction in real time applications. To the best knowledge of the authors, no study has been carried out to update the numerical model SLA prediction using combination AMI and GP.

2. Methodology

2.1. Data availability in the study area

The Singapore Regional Waters is defined as the area between 95°E–110°E and 6°S–11°N. It encompasses the two strategic waterways Malacca Strait and Singapore Strait, the central part of the shallow Sunda Shelf which connects the South China Sea (SCS) and the Java Sea, and part of the deep basin of the Andaman Sea. The present work is based on observations data used by Rao and Babovic (2010) and model predictions made in the year 2004 used by Kurniawan et al. (2011) and (2013). Table 1 and Fig. 1 depict the geographical locations for various observation stations used for the present work which are mainly located at Eastern Malaysian Peninsula.

The next sections explore a data model integration approach consisting of two parts to improve the SLA prediction (Fig. 2). The
first part is the data relationship analysis whereas the second part is the actual modelling of SLA prediction error to improve the prediction of SLA. The present works explore the utility of mutual information theory to understand and to generate a data-driven forecast model that links the SLA prediction accuracy at Singapore Strait to the significantly related variables in the South China Sea. The latter modelling of the SLA prediction accuracy is done using appropriate data-driven modelling technique i.e. genetic programming.

2.2. A measure of information

It is common practice to use Pearson’s correlation coefficient as a measure of dependency between variables of interest. The correlation coefficient is a measure of association between variables that are ordinal or continuous. Given a series of observations and model values, the Pearson product-moment correlation coefficient can be used to estimate the correlation between model and observations which is presented below:

\[ r = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sum_{i=1}^{N} (y_i - \bar{y})^2}} \]

where \( x_i \) is the value observed at the \( i \)th time step, \( y_i \) the corresponding model values, \( N \) is the number of time step, \( \bar{x} \) is the mean of observational values and \( \bar{y} \) is the mean values of model values. From its formulation, the correlation coefficient is computed using the mean values of the data (i.e. linear correlation coefficient) which are sensitive to the noise. More importantly, the underlying assumption of linearly structured dependence is contradictory to the development of statistical models of non-linear systems. Hence, alternate measures of non-linear dependencies are necessary to analyse the SLA signals. The present work adopts measures based on mutual information theory. The linear correlation coefficient and mutual information theory are used jointly in the present work.

Mutual information has been found to be a more suitable measure of dependence for analysing the non-linear systems (Gallager, 1968; Abarbanel, 1996; Abebe and Price, 2004), since it is an arbitrary measure and makes no assumption regarding the structure of the dependence between variables. It has also been found to be robust due to its insensitivity to noise and data transformations (Abarbanel, 1996; Abebe and Price, 2004). The average mutual information (AMI) is the measure of information that can be learnt from one set of data having knowledge of another set of data. The AMI does not depend on any particular function and therefore can help to detect both linear and non-linear correlations. It simply connects two sets of data with each other and established a criterion for their mutual dependence based on the notion of information between them.

Detailed fundamental to the notion of information among data (the present work uses time series) which is Shannon’s idea (Shannon, 1948) can be found in Gallager (1968). Given a random output variable \( Y \), there will be some uncertainty surrounding an observation \( y \in Y \), which can be defined according to the Shannon entropy (Shannon, 1948). Following Abarbanel (1996), the present work uses this connection to give precise definition to theoretic fashion to the data \( s(t+T) \) at time \( t+T \). To compute the AMI between a time series \( s(t) \) and its time delayed \( s(t+T) \), the average mutual information for time delay \( T \) is presented below:

\[ AMI(T) = \sum_{s(t),s(t+T)} P(s(t), s(t+T)) \log \left( \frac{P(s(t), s(t+T))}{P(s(t))P(s(t+T))} \right) \]

where \( P(s(t)) \) and \( P(s(t+T)) \) are probability density functions (PDF), whereas \( P(s(t), s(t+T)) \) is joint PDF. From its formulation, it can be seen that when \( T \) becomes large, the chaotic behaviour of the signals makes the data \( s(t) \) and \( s(t+T) \) become independent in a practical sense, therefore \( AMI(T) \) will tend to zero (Abarbanel, 1996). The time lag values are set at \( T = 1, 2, 3, ..., 48 \) h. AMI between the original signal and the time lagged signal is computed for each time lag values. The nature of AMI variations across different \( T \) values are used as basis for establishing the individual SLA temporal patterns.

2.3. Genetic programming and its scope in non-tidal barotropic modelling

Genetic programming (Koza, 1992) is a data-driven evolutionary computing method for automatically generating input–output relations. It differs from other black-box type data-driven modeling techniques, such as artificial neural networks, fuzzy logic, and regression trees, in the sense that it provides mathematically meaningful structures (with optimum parameters) relating input–output variables of the system (Babovic and Keijzer, 2000; Ghorbani et al., 2010; Shiri and Kisi, 2011, Shiri et al., 2011; Kisi et al., 2012; Karimi et al., 2013). Given a set of observed data (training data) on input–output variables, GP generates a population of models with random structure without needing any prior knowledge of the mechanisms governing the process. These models are then evaluated using part of the initial data not used during training (validation data) using a suitable fitness measure. The probability of a given model surviving during the model evolution process is proportional to how well it predicts the output of the validation data. Components of successful models are continuously recombined with those of others to form new models. In each generation, GP optimizes the model structure, with a lower level nonlinear least-squares algorithm harnessed to estimate the associated model parameters. Though, GP has been successfully applied for modelling many hydrological processes (Babovic et al., 2001; Babovic and Rao, 2010; Ghorbani et al., 2010), its application as data assimilation tool in SLA forecasting is indeed unique. A detailed discussion on components of GP tool, its methodology and various applications are detailed in elsewhere (Koza, 1992; Babovic et al., 2001; Babovic, 2009; Babovic and Rao, 2010).

In present study, the focus is given to highlight the utility of this unique modelling tool to address two different classes of data mining problems in non-tidal barotropic modelling which are: (i) using GP as a modelling tool to directly learn the SLA prediction errors and (ii) using GP as a data assimilation tool to update the
Fig. 3 shows a schematic representation of the error-correction strategy for real-time forecast systems. In a real-time setup, say at time \( t \) and at desired location, the primary models (hydrodynamic) are used to forecast the SLA water level for desired forecast horizon \( SLA_{\text{sim}}(t+k) \). The SLA error forecast models (generalized GP models or models designed at location of interest) forecast the SLA error values \( SLA_e(t+k) \) using the past SLA errors values measured (difference between the observed water level and the primary model prediction) at and before time \( 't' \). These SLA error forecasts are then used to update the primary model forecasts to obtain the corrected water level prediction \( SLA_{\text{corrected}}(t+k) \). The following set of equations is used in this analysis.

- \( k = 1, 2, 3, \ldots; \) assuming 1 h sampling time.
- \( SLA_{\text{sim}}(t+k) = H[T_{\text{sim}}]; \) \( H = \) primary model simulation (updated every \( T_{\text{sim}} \) h).
- \( SLA_a(t) = SLA_{\text{obs}}(t) - SLA_{\text{sim}}(t) \).
- \( SLA_a(t+k) = GP[SLA_a(t-1), SLA_a(t-2), \ldots]; \) \( GP = \) SLA error forecast model (every hour).
- \( SLA_{\text{corrected}}(t+k) = SLA_{\text{sim}}(t+k) + SLA_a(t+k) \).

To design a ‘\( k \)’ hours ahead’ SLA prediction error forecast accuracy model, \( \{SLA_a(t-1), SLA_a(t-2), \ldots, SLA_a(t-n)\} \) are treated as input variables and \( SLA_a(t+k) \) as output variable. Here \( 't' \) is the present time and \( (t-n) \) is \( n \) times samples in past (the best value for \( n \) is selected automatically during GP algorithm). For \( t = 1-N_{\text{trg}} \) (the selected number of training samples), the series of input-output data from a single SLA prediction error signal at a given station are extracted.

For example, for time point \( t=20 \) (January 1, 2004 20:00 in annual time series data) the SLA prediction error values between \( t=10 \) and 20 (January 1, 2004 10:00 to 20:00 for \( n=10 \)) will go as input data and (say for \( k = 12 \) h ahead prediction) SLA prediction error value at \( t=32 \) (January 2, 2004 08:00) will be taken as output data. This is repeated \( N_{\text{trg}} \) times for different \( t \) values in order to generate training data matrix. Hence the training dataset consists of input vectors representing the sequence of present and past values at any given \( t \) and the output vector representing \( SLA_a(t+k) \) for the same period. Similar approach is adopted for extracting input–output data for model testing \( N_{\text{test}} \) samples from a specified period of the year. Different time regions are selected for GP model building (training data) and for model testing (validation). In the present study, GP models are built or tested for different forecast horizon \( k \) and validated for data assimilation utility. The previous study on effect of varying dataset sizes (i.e. \( N_{\text{trg}} \) and \( N_{\text{test}} \)) and over-fitting found that there is an improvement of 20–30% in RMSE if the full year sample set is used during training compared to \( N_{\text{trg}} = 1500 \) (Rao and Babovic, 2010). However, the \( N_{\text{trg}} = 2000 \) has been chosen to avoid over-fitting in the prediction results.

GPTIPS (Searson et al., 2010), an open source genetic programming toolbox for multigene symbolic regression is used as GP implementation tool. GPTIPS is an interactive modelling environment with many options to set different GP parameters and data handling. In the present study, the settings used for each GP run are: \( \text{population size} = 500, \text{number of generations} = 150, \text{tournament size} = 12, \text{fitness criteria} = \) multigene symbolic regression (Searson et al., 2010), \( \text{mutation/crossover probability} = 0.25, \) \( \text{terminal set} = \{SLA_a(i) \text{ with } i = t; \text{ present and past time points} \}, \text{output variable} = \{SLA_e(t+k)\}, \text{model type} = \) multivariate–dynamic–algebraic (i.e. one output related to many inputs). Though GP runs can generate models with random structures with varying sizes and element composition, too lengthy and complex structures may lead to data over-fitting. In order to regulate the same, chromosome related parameters are fixed as follows. Maximum number of parameters in a model, \( N_{\text{max}} = 5 \), maximum length and depth of the chromosome restricted to 12 in order to control the model complexity. Restricted functional elements have been set to avoid highly non-linear component interactions (which can enhance the risk of local optimization of model parameters). During each GP run, GPTIPS takes the training data for input–output variables and designs the GP models with the settings explained above. Part of the training data is used for model fitness evaluation using fitness criteria. The selected population evolves over generations retaining the models with best evaluation criteria for GP output and lesser complexity.

The SLA error modelling exercise for SLA error prediction involves following major steps.

- Select the location (i.e. UH699).
- Import the hourly SLA prediction error time series data.
- Set the forecast horizon \( k \), \( N_{\text{trg}} \) and \( N_{\text{test}} \) and choose the time region for datasets.
- For the selected period, extract/store the input–output data matrix for training and testing.
- Use the training data to build GP models.

---

**A. Kurniawan et al. / Computers & Geosciences 72 (2014) 94–104**

---

![Fig. 3. Implementation scheme and data flow for data assimilation strategy using genetic programming based error forecast models](adapted from Babovic and Rao (2010)).
• Import the trained SLA error forecast model code.
• Use the inputs from the test data to predict the SLA output \( \text{SLA}_e(t+k)_{\text{predict}} \).
• Compare \( \text{SLA}_e(t+k)_{\text{predict}} \) and \( \text{SLA}_e(t+k)_{\text{actual}} \) to evaluate the model performance.
• Repeat GP runs to obtain the best local SLA forecast model based.

2.4. Evaluation criteria for GP output

The present study uses two different evaluation criteria for GP model direct forecasting performance in terms of root mean square error (RMSE) and mean absolute error (MAE) expressions for which are presented below:

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_{\text{obs},i} - X_{\text{model},i})^2}
\]

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |X_{\text{obs},i} - X_{\text{model},i}|
\]

where \( X_{\text{obs}} \) is the values observed at \( i \)th time step, \( X_{\text{model}} \) is the modelled values and \( N \) is the number of GP model test.

3. Results and discussion

3.1. Data relationship analysis

The data analysis conducted with the observed data and sea level anomaly (SLA) model simulation for the year 2004 has revealed important points regarding the time dynamics and length scales involved in the interaction between the SLA data observed at various locations around the Singapore Regional Waters, the interaction between meteorological variables and the observed SLA and SLA prediction errors. Fig. 4 shows temporal distribution of autocorrelation and AMI showing analysis of individual observed SLA for temporal memory at observation stations in eastern Malaysian Peninsula within a lag time up to 48 h. As can be seen, Pearson’s correlation coefficient does not take into account nonlinear dynamical correlations, whereas average mutual information (AMI) can help to detect both linear and nonlinear correlations. These confirm previous finding that SLA (wind-driven water level, surge) is strongly non-linear in Singapore Regional Waters (Gerritsen et al., 2009; Rao and Babovic, 2009). Both the AMI and autocorrelation measures show ‘bumps’ at 12 and 25 h, which are reasonably distinctive and it is getting stronger in UH699. Since 12 and 25 h correspond to the periods of...
Fig. 5. Temporal distribution of autocorrelation and AMI showing analysis of individual predicted astronomical tide for temporal memory at observation stations in eastern Malaysian Peninsula within lag times up to 48 h.

Fig. 6. Temporal distribution of observed SLA at UH699 with other observation stations in eastern Malaysian Peninsula with a lag time up to 48 h showing AMI (top) and crosscorrelation (bottom) values.
Fig. 7. Temporal distribution of autocorrelation and AMI showing analysis of individual SLA prediction errors for temporal memory at observation stations in eastern Malaysian Peninsula within lag times up to 48 h.

Fig. 8. same as Fig. 6 but for SLA prediction errors.
the tidal constituents, this raises further interest in the relationship between tidal cycle and the observed SLA in South China Sea (SCS) and Singapore Strait.

Fig. 5 shows temporal distribution of autocorrelation and AMI showing analysis of individual predicted astronomical tide for temporal memory at observation stations in eastern Malaysian Peninsula within lag times up to 48 h. As can be seen, autocorrelation measures shows the diurnal tidal cycle in SCS i.e. UH326 and UH320. The tidal cycle starts to change to mixed tidal cycle from UH322 and it is getting stronger in Singapore Strait (UH699). In addition, AMI values are also affected by these tidal cycle changes. This indicates that there is an obvious influence of the tidal cycle on the observed SLA (i.e. tide-surge interaction). In view of the definitions of the tide and surge such an influence is nonlinear. These findings support the conclusions derived in previous studies by Kurniawan et al. (2013).

The relationship between observed SLA at UH699 and other stations is shown in Fig. 6. As can be seen, the AMI values show immediate response and the correlation coefficients report strong positive correlation between observed SLA at UH699 with observed SLA at other stations in eastern Malaysian Peninsula. These trends also agree with the known physics of the situation and also support previous findings that South China Sea basin triggers sea level anomaly in Singapore Strait (Rao and Babovic, 2009; Ooi et al., 2011; Kurniawan et al., 2013).

The sea level anomaly (SLA) prediction errors at the Singapore Strait, particularly at UH699, are at the centre of all the analyses in the following sections. The previous section based on the observed SLA at UH699 has already provided some insight into the relationship and interaction between various variables and the observed SLA which is solely a function of the physical system. Fig. 7 shows temporal distribution of autocorrelation and AMI showing analysis of individual SLA prediction errors for temporal memory at observation stations in eastern Malaysian Peninsula within lag times up to 48 h. As can be seen, the graphs report the same trends as shown in Fig. 4. Similar conclusions can be drawn here except that in the present case the tidal cycle contributes to SLA prediction errors. The correlation coefficient measure became important since a value of 0.5 corresponding to 25 h lag time implies that the effect of the tide on the error is significant. Moreover, this is also an indication that the information is available if it is intended to forecast the SLA prediction errors, say, 25 h in advance.

Fig. 8 shows temporal distribution of AMI values and cross correlation between SLA prediction errors at UH699 with SLA prediction errors at other stations in eastern Malaysian Peninsula with a lag time up to 48 h. Both the AMI values and the correlation coefficient suggest that the response is distinct and has a delayed peak between 12 and 25 h for other stations in eastern Malaysian Peninsula. In addition, the other stations located near UH699 (i.e. UH324 and UH323) show higher AMI values and stronger positive correlation compared to other stations. These patterns may be due to the fact that these two stations have slightly similar tidal cycle (see Fig. 5). These findings are consistent with the previous analysis and indicate that there is an obvious influence of the tidal cycle on the SLA prediction errors (i.e. tide-surge interaction) and in view of the definitions of the tide and surge such an influence is nonlinear. The data relationship analysis results suggest that past SLA prediction error can be used and the best predictive parameters are 1, 12 and 24 h lag time at UH699, UH324 and UH323.

3.2. Data-driven modelling (genetic programming)

Direct forecast analysis is carried out using SLA prediction error models built using the year 2004 data. The output predicted \( \text{SLA}_{(t+k)} \) is compared with the actual SLA prediction error...
values \( \text{SLA}_{e(t+k)} \). Fig. 9 provides the summary of the results obtained by direct forecast of SLA prediction error models at UH699 using the past SLA prediction errors based on average mutual information (AMI) results. It can be observed that the GP models, built separately for every selected \( k \)-step ahead prediction, perform very efficiently. The minima RMSE and MAE (i.e. 0.049 m and 0.038 m, respectively) are reported by GP error forecast model for 1 h direct forecast which is expected. Whereas, the maxima RMSE and MAE are 0.079 m and 0.062 m for 6 h direct forecast, respectively. It can be seen that the 12 and 24 h direct forecast is better than 4 and 6 h direct forecast in terms of RMSE and MAE values, which supports previous finding that a periodic component of tidal response may be existed in the SLA prediction errors (tide-surge interaction, see Kurniawan et al., 2013). The results suggest that the GP model capture the SLA prediction error satisfactorily.

The larger objective of this investigation is to use the past SLA prediction error capabilities to update the real time forecasting of SLA. Fig. 10 summarizes the performance of SLA prediction error forecast models when used to update the non-tidal barotropic predictions in order to obtain the updated sea level anomaly. The RMSE and MAE have improved for the 1 h forecast up to 50% and 53%, short term forecast by 17% and 26% (up to 12 h forecast) and long term forecast by 22% and 24%, respectively. This is consistent with the previous finding in which the 12 and 24 h direct forecast are slightly more accurate than 2, 4 and 6 h direct forecast as shown in Fig. 10. This is again evidence that tidal response may exist in the SLA prediction errors. In addition, to have a clear displayed of the variation of the values, Fig. 11 illustrates the ability of SLA prediction error forecast to correct the water level prediction during severe SLA events (positive and negative, Kurniawan et al., 2013) which are even more accurate. As shown in Fig. 11, the negative SLA event is significantly better represented by non-tidal barotropic model with GP error model. This clearly establishes the utility of SLA error forecast models as local water level correction tool.

4. Conclusions and recommendations

Data relationship analysis has been carried to evaluate the content, flow and time dynamics of information regarding the
observed sea level anomaly (SLA) and its prediction errors with respect to a number of selected parameters. The analysis has been done both for the observed SLA and the SLA prediction errors. The analysis helps to establish how much information contained in the SLA prediction errors can be traced back to some of the variables.

The significance test is done using average mutual information (AMI) and correlation coefficient measures. Furthermore, the AMI helps to generate a data-driven forecast model that links the SLA prediction accuracy at Singapore Strait to the significantly related variables in the South China Sea. Data driven modelling of the SLA prediction accuracy is carried out using genetic programming (GP).

The results suggest that combination of data relationship analysis (i.e. AMI) and GP help to improve the forecasting ability by providing information of significant predictive parameters. In the final stage it has been found that GP model based SLA prediction error forecast model can provide significant improvement when applied as data assimilation schemes for updating the SLA prediction obtained from primary hydrodynamic models. The results have shown a good performance of non-tidal barotropic numerical modelling and GP error forecast model to forecast the SLA at Singapore Strait.

Acknowledgements

The authors gratefully acknowledge the support and contributions of the Singapore-Delft Water Alliance (SDWA) and Deltares Strategic research funding. The research presented in this work was carried out as part of the SDWA’s “Must-Have Box” research program (R-264-001-003-272).

References

Gerritsen, H., Twigt, D., Mynett, A., Calkoen, C., and Babovic, V. Analysis and prediction of sea level anomalies and associated currents in Singapore and Malaccas straits. In: Proceeding of the 8th International Conference on Hydroinformatics, Chile.

Fig. 11. Comparison of SLA observed, SLA prediction using non-tidal barotropic model without and with GP error forecasting model during positive (top) and negative (bottom) SLA events for 1 h direct forecast windows at UH699.


