Enhancing Tidal Prediction Accuracy in Singapore Regional Model Using Local Model Approach

YABIN SUN[1], PIYAMARN SISOMPHON[2], VLADAN BABOVIC[1,3], ENG SOON CHAN[1]
[1]Department of Civil Engineering
National University of Singapore
SINGAPORE

[2]Strategic and Research Development
Deltares, Delft
NETHERLANDS
http://www.deltares.nl
[3]Singapore-Delft Water Alliance
National University of Singapore
SINGAPORE

Abstract: - With the objective to provide hydrodynamic information of the water around Singapore, the Singapore Regional Model (SRM) has been developed within the Delft3D numerical modelling system. The results of this large-domain numerical model are necessarily a balance between the choices about domain, local resolution, model parameter settings and representation of forcing. Especially in the complex nearshore area around Singapore Island, however, high accuracy in prediction of water levels is required. To further improve the results of the deterministic Singapore Regional Model in the coastal area, an error correction scheme based on the local model (LM) approach is carried out, which is inspired from chaos theory and capable of forecasting the time series based on the underlying mechanism that may not be revealed in the deterministic model simulation. The efficiency of the error correction scheme has been tested on 3 stations in the Singapore Regional Model domain with 4 prediction horizons ranging from 2 hours to 96 hours. It is found that the error correction scheme significantly improves the accuracy of the tidal prediction with more than 70% of the root mean square error removed for 2-hour tidal forecast and around 50% for 96-hour tidal forecast.

Key-Words: - Local model; chaos theory; error correction; Singapore Regional Model

1 Introduction

In ship navigation and offshore operations, tidal prediction is of prime importance. Prediction could be made for time horizons from several hours to several days, which is helpful in the finalization of operational schedules and other coastal activities. The current practice of tidal prediction is undertaken by either using a tidal predictive deterministic model [1,2] or by a time series forecasting model [3,4].

If the deterministic equations underlying the physical phenomena are known, in principle, they can be numerically solved to forecast the future state based on the knowledge of the initial conditions and the time evolution of forcing terms with a high prognostic capability. However, model resolution, parameter uncertainty and also the absence of or uncertainty in the prescribed forcing contribute to model errors. Numerical models tend to produce imperfect results despite our perfect knowledge of the governing laws [5].

The linear models, such as autoregressive moving average (ARMA) models, have dominated the field of time series forecasting for more than half a century. The signals are transformed into a small number of coefficients plus residual white noise in the linear models. However, such appealing simplicity can be entirely misleading even when weak nonlinearities occur. The inherent linearity assumptions may not be applicable to correct the numerical model simulation, where the model errors come from highly nonlinear resources.

Local model originates from chaos theory, and has gained popularity in nonlinear time series forecasting [6-8]. In the present study, an error correction scheme based on local model is applied to correct the outputs of the Singapore Regional Model. The improvements in the prediction accuracy over the numerical model are discussed in detail.

2 Singapore Regional Model

Singapore Strait is one of the busiest shipping routes in the world. Since the 1960s, the coastal area has been heavily utilized as ports or related industrial facilities with rapid economic development. The dedicated Singapore Regional Model has been constructed within the Delft3D modelling system, with the intention to provide hydrodynamic information of the water surrounding Singapore for accurate scheduling of harbor facilities, docking and sailing times [9].

As shown in Figure 1, the model domain coves large parts of seas around Singapore. Its open boundaries are located in the Andaman Sea, in the South China Sea and in the Java Sea. The Singapore Regional Model grid consists of around 38,500 grid cells in the horizontal plane. Grid sizes vary from about 200 m to 300 m around Singapore up to over 15 km at the open boundaries. Singapore is located between two large water bodies: the South China Sea on the east and the Andaman Sea on the west. The water motion in Singapore Strait is driven by tides coming from both sides, by mean sea level differences between seas and by the wind. Therefore, the hydrodynamics of Singapore water is complex. Figure 1 also presents 3 stations studied in this paper, which are located at Jurong (1.31 N, 103.72 E), Tanjong Pagar (1.26 N, 103.85 E) and Bukom (1.23 N, 103.78 E).

The numerical tidal simulation covers a one year period from 1st January 00:00 1999 to 31st December 23:00 1999, producing time series of 8760 hourly data for all grid points. Hourly tide data from the 1999 Singapore Tide Tables and Port Information [10] for the same period are used as 'observations'. After neglecting the first 480 data points to avoid initialization effect (in hindsight, 120 hours would have been sufficient), the Singapore Regional Model outputs are compared with the observations at these 3 stations. The results are shown in Table 1 in terms of root mean square error (RMSE) and correlation coefficient (r). Examples of the model outputs are plotted in Figure 2, accompanied by the observations and the model errors, from which the discrepancies between the model outputs and the observations can be noticed especially at the tidal level extrema.

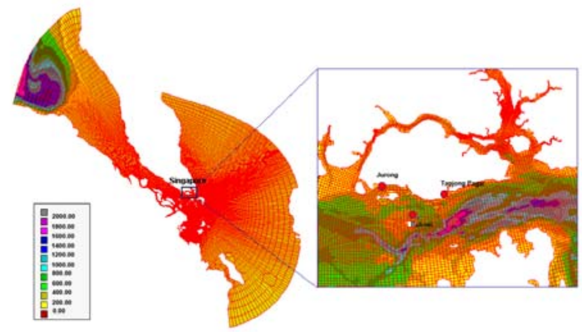

Figure 1. Singapore Regional Model domain and the location of measurement stations.

Table 1. Statistics of the model errors at the measurement stations.

	Jurong	Tanjong Pagar	Bukom
RMSE(cm)	18.30	17.57	17.54
r	0.91	0.91	0.91

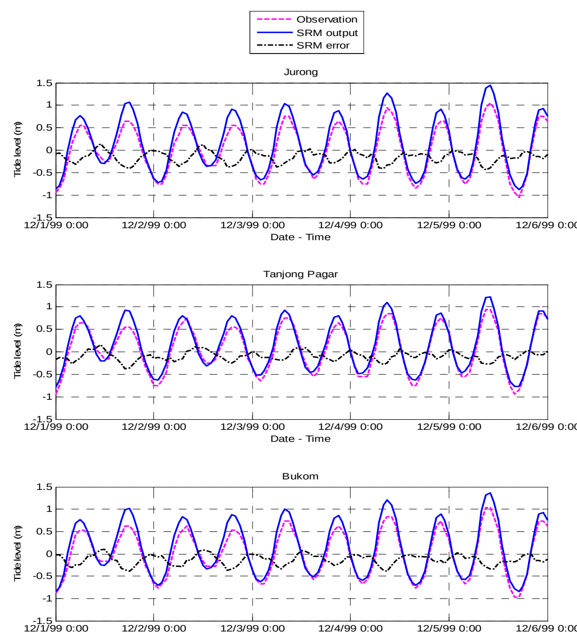

Figure 2. SRM outputs, observations and model errors at the measurement stations.

3 Chaos Theory

Recent developments in nonlinear dynamics have demonstrated that irregular or random behavior in natural systems may arise from purely deterministic dynamics with unstable trajectories. Such types of nonlinear dynamical systems, which are highly sensitive to initial conditions, are popularly known as chaotic systems. According to Williams [11], chaos is a sustained and disorderly-looking evolution that satisfies certain special mathematical criteria and occurs in a deterministic non-linear system.

3.1 Embedding theorem

Takens' time-delay embedding theorem [12] paved the way for the analysis of chaotic time series in chaotic systems. The theorem essentially states that the underlying structure of a complex, multi-dimensional system can be equivalently viewed using a projection from a single variable in the phase space, which is an embedded space with dimensions consisting of various time lags of the variable itself.

Given a scalar time series x_i from a dynamical system, it is possible to reconstruct a phase space in terms of the phase space vector $\mathbf{x}_i$ expressed as

$$\mathbf{x}_i = \left(x_i, x_{i-\tau}, \cdots, x_{i-(m-1)\tau} \right) \tag{1}$$

where m is the embedding dimension, and τ is the time delay. According to the embedding theorem, the underlying structure cannot be seen in the space of the original scalar time series, rather only when unfolded into a phase space, or embedded space.

In the phase space prediction model, the basic idea is to set a functional relationship between the current state $\mathbf{x}_t$ and the future state $\mathbf{x}_{t+T}$ in the form

$$\mathbf{x}_{t+T} = f_T\left(\mathbf{x}_t\right) \tag{2}$$

where T is referred to as lead time or prediction horizon. The problem now is limited to find a good expression for the mapping function f_T.

3.2 Local model

Local model is an effective method of simulating the evolution of a dynamical system by means of local approximation, using only the most similar trajectories from the past to make predictions for the future [6-8]. Steps in the local model approach can be described as follows,

- Step 1. Embedding the time series into a phase space
- Step 2. Finding k nearest neighbors in the phase space

  To predict a future state $\mathbf{x}_{t+T}$, a Euclidean metric is imposed on the phase space to find the k nearest neighbors of the current state $\mathbf{x}_t$, denoted by $\mathbf{x}_n$ $\left(n = 1, 2, \cdots, k\right)$.
- Step 3. Calculating the 'expected' future state

  Having constructed the phase space and pooled the k nearest neighbors of the current state $\mathbf{x}_t$, the 'expected' vector of the future state $\mathbf{x}_{t+T}$, denoted as $\hat{\mathbf{x}}_{t+T}$, can be estimated through averaging as

  $$\hat{\mathbf{x}}_{t+T} = \left(\sum_{n=1}^{k} \mathbf{x}_{n+T} \right) \Big/ k \tag{3}$$
- Step 4. Deriving the forecast scalar value

  In the phase space, the 'expected' future state $\hat{\mathbf{x}}_{t+T}$ can be expressed in the form of Equation (1) as

  $$\hat{\mathbf{x}}_{t+T} = \left(\hat{x}_{t+T}, \hat{x}_{t+T-\tau}, \cdots, \hat{x}_{t+T-(m-1)\tau} \right) \tag{4}$$

  The predicted scalar values $\hat{x}_{t+T}$, $\hat{x}_{t+T-\tau}$, … in the time series x_i can be retrieved according to the structure.

When making a local model forecast, the first step is to unfold the time series into a phase space, which typically involves the selection of an embedding dimension m and a time delay τ. In the traditional standard approach, false nearest neighbors (FNN) and average mutual information (AMI) analyses are recommended to determine these parameters [13]. However, the standard approach has shown to provide suboptimal choices of the embedding parameters. Therefore, in this paper, an alternate inverse approach based on genetic algorithm (GA) is employed, which has demonstrated significant improvements over the standard approach [14,15].

3.3 Inverse approach

In the inverse approach, genetic algorithm is used to simultaneously optimize the embedding dimension m, the time delay τ and the number of nearest neighbors k. Genetic algorithm is a parameter search procedure based upon the mechanics of natural genetics, which combines the Darwinian theory of

ISSN: 1790-2769

ISBN: 978-960-474-012-3

evolution with a random, yet structured information exchange among a population of artificial chromosomes [16].

The evolving process in genetic algorithm is illustrated in Figure 3, while Figure 4 presents the flow diagram for genetic algorithm. In principle, an initial population of chromosomes $P_i = \{m, \tau, k\}$, where m, τ and k are represented by binary bits, is randomly generated within the specified ranges of parameters, and is allowed to evolve through the following process:

- Selection: A scheme is employed to select the chromosomes to reproduce offspring according to their respective fitness. The chromosome with higher fitness has a better chance of being selected. For every chromosome, local model is executed to evaluate the fitness in terms of root mean square error.
- Crossover: Some portion of a pair of chromosomes selected from the population is exchanged according to some constraints in order to generate two new sets of parameters.
- Mutation: One individual chromosome selected from the population is transformed to a new individual by inverting some of its binary values.

The process is continued until an entirely new population is generated with the hope that the fitter parents will create a better generation of children, such that the average fitness of the population will tend to increase with each new generation. The fitness of each child in the new generation is evaluated, and the process of selection, crossover and mutation is repeated. Successive generations are created until the user-defined threshold for the fitness or number of maximum generation is reached.

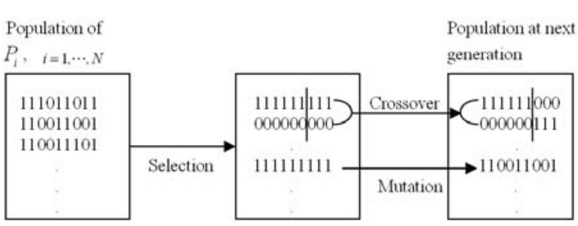

Figure 3. Schematic illustration of evolving process in genetic algorithm.

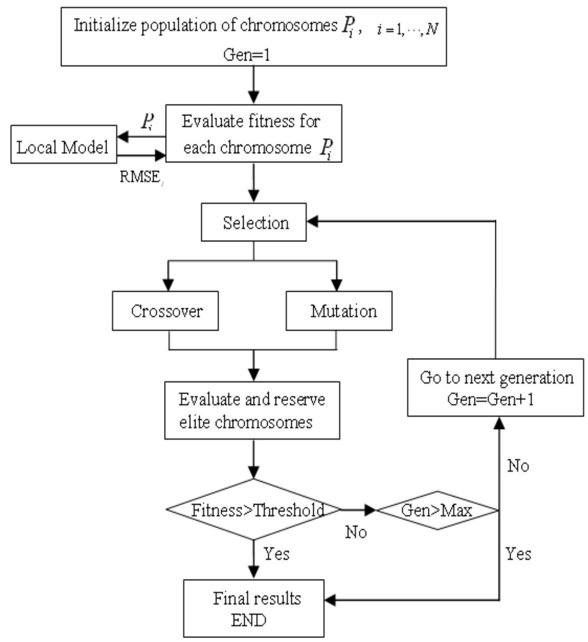

Figure 4. Flow diagram for genetic algorithm.

4 Results and Discussion

In the proposed error correction scheme, the first 480 data points (1st January 0:00 – 20th January 23:00) are discarded from the data sets due to the initialization effect in the Singapore Regional Model. The data points from 481 to 7320 (21st January 0:00 – 1st November 23:00) are used to train the local model in determining the optimal m, τ and k for each prediction horizon required at each station, while the remaining data from 7321 to 8760 (2nd November 0:00 – 31st December 23:00) are used as validation data to testify the efficiency of local model in error prediction.

Table 2 summarizes the error forecasting efficiency of the local model approach through evaluating the respective residual root mean square error as well as the correlation coefficient after error correction. For the prediction horizon T=2 hours, more than 70% of errors have been removed from the Singapore Regional Model outputs at the measurement stations. Figure 5 shows examples of the error forecasting using local model with prediction horizon fixed to 2 hours. The 2-hour forecast successfully resolves the rising and falling tendencies of the model errors, generating trivial residual errors oscillating about zero. The corresponding scatter diagrams are depicted in Figure 6. The Singapore Regional Model is found to over-predict the reality, while the scatter is reduced in the local model corrected tidal levels. When the prediction horizon progresses, it becomes more intractable to capture the trajectories of the state vectors in the unfolded phase space. This makes the model error time series less predictable. Therefore, as anticipated, the accuracy of prediction decreases when T increases. However, even for T=96 hours, the local model forecast successfully removes almost 50% of the errors. Moreover, the correlation coefficient between the corrected model outputs and the observations remains larger than 0.97 for all the prediction horizons compared to the original 0.91 without error correction. The plots of residual root mean square errors against forecast horizons at the measurement stations are shown in Figure 7, together with the root mean square error before correction. A slight increasing trend is observed from the curve.

Table 2. Residual RMSE and correlation coefficient at the measurement stations.

	Jurong		Tanjong Pagar		Bukom	
	RMSE (cm)	r	RMSE (cm)	r	RMSE (cm)	r
SRM	18.30	0.91	17.57	0.91	17.54	0.91
T=2 Hr	5.24	0.99	5.09	0.99	4.87	0.99
T=24 Hr	6.08	0.99	6.31	0.99	5.54	0.99
T=48 Hr	7.64	0.98	8.51	0.98	6.84	0.98
T=96 Hr	9.91	0.97	9.06	0.97	9.07	0.97
Average	7.22	0.98	7.24	0.98	6.58	0.98

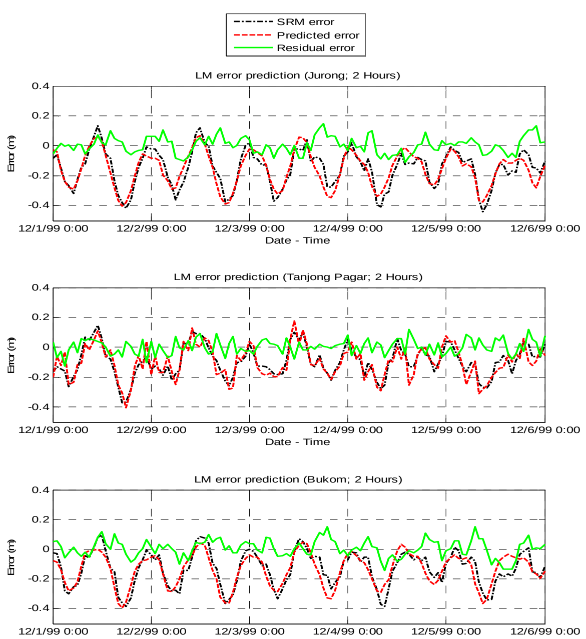

Figure 5. Error forecasting at the measurement stations using local model when T=2 hours.

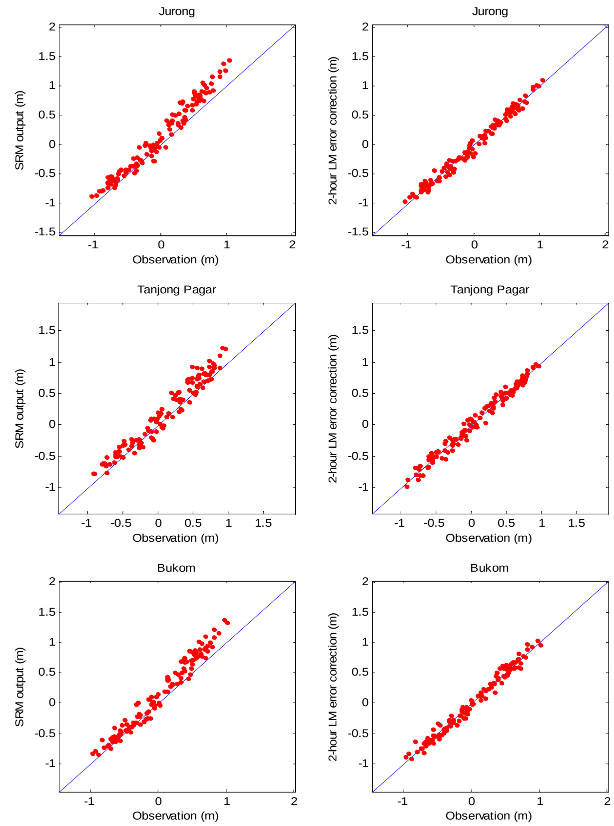

Figure 6. Scatter diagrams of SRM outputs, LM corrected tidal levels from observations at the measurement stations.

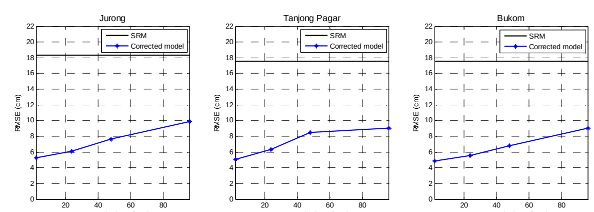

Figure 7. RMSE vs. prediction horizon at the measurement stations.

5 Conclusion

The concept of improving the tidal prediction accuracy of a deterministic numerical model using the local model approach, which is a nonlinear time series predictive algorithm inspired from chaos theory, is discussed in this paper. The embedding parameters required for the phase space reconstruction and the local model prediction are optimized using genetic algorithm. By predicting the model errors at the measurement stations, systematic errors can be modeled by the error correction scheme, while the physical dynamics remain described by the deterministic model. The accuracy of tidal prediction is significantly enhanced. More than 60% of the errors have been removed on average. The performance of local model deteriorates slightly with increasing prediction horizons.

In spite that local model is more efficient in terms of accuracy and computational cost compared to the Singapore Regional Model, nonlinear time series forecasting is not an alternative to the numerical models. Local model forecasting can only be carried out where observations are available with a rather limited prediction horizon, while the numerical model is able to predict over the entire model domain in a finer grid structure with a higher prognostic capability, and hence provides good understanding of the physics of the ocean flows. However, local model is the starting point for an effective data assimilation scheme. Once the forecast at a few measurement stations are carried out with a higher accuracy, the entire domain can be benefited by distributing the errors from the limited number of stations using an optimal interpolation algorithm, such as the Kalman filter (KF) and Artificial Neural Networks (ANN).

Acknowledgement:
The authors gratefully acknowledge the support and contributions of the Singapore-Delft Water Alliance (SDWA). The research presented in this work was carried out as part of the SDWA's "Must-Have Box" research programme (R-264-001-003-272).

References:

[1] Pugh DT. *Tides, surges and mean sea-level*. John Wiley & Sons: Chichester, 1996.

[2] Palacio C, Winter C, Mayerle R. Set-up of a hydrodynamic model for the meldorf bight. In *Proceedings of World Water & Environmental Resources Congress*, Orlando, 2001.

[3] Babovic V. A data mining approach to time series modelling and forecasting. In *Proceedings of the Third International Conference on Hydroinformatics*, Copenhagen, Denmark, Babovic V, Larsen LC (eds). Balkema: Rotterdam, 1998.

[4] Babovic V, Keijzer M. Forecasting of river discharges in the presence of chaos and noise. In *Coping with Floods: Lessons Learned from Recent Experiences*, Marsalek J (ed.), NATO ARW Series. Kluwer: Dordrecht, 1999.

[5] Babovic V, Canizares R, Jensen HR, Klinting A. Neural network as routine for error updating of numerical models. *Journal of Hydraulic Engineering* 2001; **127**:181-193.

[6] Keijzer M, Babovic V. Error correction of a deterministic model in Venice lagoon by local linear models. In *Proceedings of Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione*, Venice, 1999.

[7] Babovic V, Keijzer M, Bundzelm M. From global to local modelling: a case study in error correction of deterministic models. In *Proceedings of the 4th International Conference on Hydroinformatics*, Iowa City, 2000.

[8] Babovic V, Sannasiraj SA, Chan ES. Error correction of a predictive ocean wave model using local model approximation. *Journal of Marine Systems* 2005; **53**:1-17.

[9] Kernkamp H, Zijl F. Further Hydraulic model Studies for Pulau Ubin & Pulau Tekong reclamation scheme. *Interim Report on Hydrodynamic Modelling – Model Set-up and Calibration*, Delft Hydraulics, Delft, 2004.

[10] Hydrographic Department of Singapore. *Singapore tide tables and port information*, ISSN: 0129-0282. Singapore, 1999.

[11] Williams GP. *Chaos theory tamed*. Taylor & Francis: London, 1997.

[12] Takens F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, Rand DA, Yong L-S (eds), Lecture Notes in Mathematics, vol. 898. Springer: Berlin, 1980; 366-381.

[13] Abarbanel HDI. *Analysis of Observed Chaotic Data*. Springer-Verlag: New York, 1998.

[14] Babovic V, Keijzer M, Stefannson M. Optimal embedding using evolutionary algorithms. In *Proceeding of the 4th International Conference on Hydroinformatics*, Iowa City, 2000.

[15] Phoon KK, Islam MN, Liaw CY, Liong SY. A practical inverse approach for forecasting nonlinear hydrological time series. *Journal of Hydrologic Engineering* 2002; **7**:116-128.

[16] Holland JH. *Adaptation in Natural and Artificial Systems*. University of Michigan Press: Ann Arbor, MI, 1975.