Uncertainty quantification and reliability assessment in operational oil spill forecast modeling system

Xianlong Hou [a], Ben R. Hodges [b], Dongyu Feng [b], Qixiao Liu[a]

[a] Institute of Advanced Computing and Digital Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen, Guangdong 518055, P.R.China

[b] Dept. of Civil, Architectural and Environmental Engineering, Univ. of Texas at Austin, 1 University Station C1786, Austin, TX 78712, USA

Corresponding Author:

Xianlong Hou

Institute of Advanced Computing and Digital Engineering

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

1068 Xueyuan Avenue, Shenzhen University Town

Shenzhen, Guangdong 518055, P.R.China

Tel: +86-17301781023

Email: xl.hou@siat.ac.cn

Uncertainty quantification and Reliability assessment in operational oil spill forecast modeling system

Xianlong Hou [a], Ben R. Hodges [b], Dongyu Feng [b], Qixiao Liu[a]

[a] Institute of Advanced Computing and Digital Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen, Guangdong 518055, P.R.China

[b] Dept. of Civil, Architectural and Environmental Engineering, Univ. of Texas at Austin, 1 University Station C1786, Austin, TX 78712, USA

Abstract

As oil transport increasing in the Texas bays, greater risks of ship collisions will become a challenge, yielding oil spill accidents as a consequence. To minimize the ecological damage and optimize rapid response, emergency managers need to be informed with how fast and where oil will spread as soon as possible after a spill. The state-of-the-art operational oil spill forecast modeling system improves the oil spill response into a new stage. However uncertainty due to predicted data inputs often elicits compromise on the reliability of the forecast result, leading to misdirection in contingency planning. Thus understanding the forecast uncertainty and reliability become significant. In this paper, Monte Carlo simulation is implemented to provide parameters to generate forecast probability maps. The oil spill forecast uncertainty is thus quantified by comparing the forecast probability map and the associated hindcast simulation. A HyosPy-based simple statistic model is developed to assess the reliability of an oil spill forecast in term of belief degree. The technologies developed in this study

create a prototype for uncertainty and reliability analysis in numerical oil spill forecast modeling system, providing emergency managers to improve the capability of real time operational oil spill response and impact assessment.

Keywords: Uncertainty quantification, forecast reliability, oil spill modeling, HyosPy, Monte Carlo simulation, probability map

1. Introduction

When an oil spill occurs at late night in heavily trafficked shipping channels, operational oil spill forecast modeling system provides the spill transport predictions needed for rapidly deploying emergency responses equipment, e.g. booms, dispersant, or skimmer boats. As moving equipment around the margins of an estuary or bay can be time consuming, information on the uncertainty of the forecast spill path could be insightful in deciding whether equipment should be immediately committed or moved to a central location (relative to possible spill paths) to await predictions with greater confidence. Unfortunately, such data is not generally available from existing operational oil spill modeling systems.

1.1 Uncertainty in oil spill modeling

The cause of oil spill forecast uncertainty ranges from the modeling system itself to the forecast model inputs. At the system limit are the interdisciplinary sub-models including chemistry, turbulence, hydrodynamics, meteorology, and hydrology - providing a 2D or 3D oil spill forecast trajectory (You and Leyffer 2011; Zelenke et al. 2012; Mackay et al. 1980; Huang 1983; ASA 1997; Reed 2000). Although some previous studies (e.g. Price et al. 2004; Elliott and Jones 2000; Reinaldo and Henry 1999) proved that numerical formulations would have influence on the performance of oil spill modeling, the forecast uncertainty in this end has been reduced significantly as the evolution of the state-of-the-art models and parallel computing power. At the forecast inputs limit are weather and hydrodynamic forecast time series (e.g. wind and tidal force) required by the oil spill modeling. Presently weather forecasts have qualified predictive capabilities for periods up to 4 days, but it becomes more and more unstable

as time progresses (Sebastiao and Soares 2007; Sebastiao and Soares 2006). The forecast data derived from operational models, such as Texas coastal wind forecasts from the National Centers for Environmental Prediction (NCEP) Eta model (NECP 2015), might have poorer predictive skills for an even shorter forecast period. Unlike the uncertainty from the modeling system limit, the uncertainty from forecast data is inevitable.

NOAA's GNOME (General NOAA Operational Modeling Environment) oil spill transport model developed its oil spill forecast uncertainty assessment package by undertaking self-made assumptions, i.e. modelers should have to guess what the uncertainties of inputs are. The key is to perturb different movers by slightly changing the magnitude or the direction of winds and currents input vectors (Zelenke et al. 2012). The Oil Spill Risk Analysis (OSRA) model estimates the oil spill forecast uncertainty by generating an ensemble of oil spill trajectories over many years of hydrodynamic and meteorological input fields. The forecast uncertainty is assessed by analyzing the difference of the simulated spills under the assumption that the occurrence rates of the spills and the inputs will probably like those that might happen in the future (Price et al. 2003). Drifter modeling along with statistical post-processing is also a practical approach to estimate the forecast uncertainty in most recent studies (Sebastiao and Soares 2006; Price et al. 2006). Many of these methods advocate a minimum regret strategy to deliver the predicted data to the oil spill modeling system (Galt 1997; Galt and Payton 1999). However, all of these methods do not provide explicit information of what can go wrong and how much is it to go wrong, of which are relatively more important issues in operational oil spill forecast modeling. Hence more elaborate

analysis is required to quantify the oil spill forecast uncertainty so that oil spill managers could have a general idea of the forecast quality.

1.2 Reliability of oil spill forecast

Uncertainty in forecast modeling is pervasive; however in most operational engineering, economics, and nature science fields, numerical simulations based on forecast data are the only sources for decision making before hindcast or observed data is available, especially for issues that rapid response is critical. As to oil spill accidents, the observed data is always rare when a spill occurs. The realization of forecast uncertainty in operational oil spill modeling system draws concerns for oil spill managers, who pay high attention to the reliability of the forecast results. Thus, facing to the forecast results of an oil spill modeling, oil spill managers would always ask: how likely is it to go wrong? Or how much can I trust it? This is another uncertainty issue that pertains to the confidence or reliability of a numerical oil spill forecast.

Reliability is the most important forecast quality that measures the degree of the likelihood that a forecast captures the actual event being predicted. Reliability assessment generally involves ensemble forecast, because real physics can only provide a single outcome for a particular forecast, which is impossible to form a probabilistic representation of reliability (Tippett et al. 2014). There are many ways of quantifying forecast reliability. Brown et al. 1997 assessed the reliability of the power distribution system to momentary interruptions and storms by using Monte Carlo simulation. Weisheimer and Palmer 2014 and Ho et al. 2013 analyzed the reliability of seasonal climate forecasts based on reliability diagrams which are tools to visualize and quantify the statistical reliability of a forecast system. Winkler et al. 2010 examined the

reliability of power system during hurricanes via network topology. There are few studies concentrating on the reliability of oil spill modeling (e.g. Abascal et al. 2010; Wang and Zhou 2009), however, these studies aim at specific model or spill event, which simply cannot be applied over a broader sense. A more general approach is required for present operational oil spill forecast modeling where the state-of-the-art numerical models and forecast data sources are changing all the time.

This paper defines several new terms to quantify forecast uncertainty in operational oil spill modeling system. Monte Carlo simulation is applied to evaluate forecast errors so that multiple pseudo-forecast series can be generated to form a time-evolution of forecast probability map for uncertainty quantification. The Hydrodynamic and Oil Spill Python (HyosPy) (Hou and Hodges 2014; Hou et al. 2015; Hodges et al. 2015) is exploited to assess forecast reliability of oil spill predictions in a more general sense.

2. Methods

2.1 Forecast uncertainty probability map (FUPM)

Oil spill forecast uncertinty has two facets - temporal and spatial. Temporal uncertainty originates from the arrival time discrepancy of the surface oil at a specific location (e.g. Abascal et al. 2010); spatial uncertainty emerges from the potential transport track of the spill (e.g. Nelson et al. 2015 for large scale oil spill spatial uncertainty analysis; Sebastiao and Soares 2006 analyzed smaller cases in a coastal zone via a oil spill model).

From an operational response perspective, the critical question within a bay or estuary is *when and where will the spill hit the shoreline?* The *hit time* can be defined as the time that a forecast predicts the spill to hit a particular stretch of beach and the hit

location as the location of beaching. For a given forecast period (T), not all spills will hit the shoreline, so it is also useful to consider a simply binary discriminator of *beaching/no beaching*. The forecast uncertainty can be divided into four categories, that we will quantify as metrics:

1. *Hit time uncertainty* – U_t : the deviation of the transport time between the hindcast and forecast when one or both of them are beaching.
2. *Hit location uncertainty* – U_l : the deviation of possible or simulated-definite landing positions between the hindcast and forecast when one or both of them are beaching.
3. *Transport area uncertainty* – U_a : the deviation of transport directions (represented by area for ease of calculation) between the hindcast and forecast.
4. *Transport speed uncertainty* – U_s : the deviation of transport speed between the hindcast and forecast.

Quantifying the above metrics requires assessing the difference between observations and the ensemble of possible forecasts, i.e. a forecast probability map. These metrics could be developed/used in three different ways: (1) as an a priori exercise with field drifter data as observations to evaluate likely uncertainty in models; (2) as a operational task during a spill, where the latest spill observations are used to rapidly assess evolving uncertainty; or (3) as a model-model comparison where hindcast data driving the model represents the observations, and a range of forecast data driving the model provides the ensemble. The present work demonstrates the technique using the model-model approach, as we do not have access to a data set of drifters or observed oil spill evolution.

The forecast probability map is composed of multiple possible forecasts in the same simulation period. However, a single simulation period could have only one forecast, hence pseudo-forecast need to be created. In this work, the pseudoforecast is developed by identifying input forecast error (ε_k) based on Monte Carlo simulation. Specifically,

$$\varepsilon_k = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_{ik} - h_{ik})^2} \tag{1}$$

where k denotes input class (i.e. wind, tide, river flow); f_{ik} is the input forecast time series; h_{ik} is the input hindcast time series; i = 1, 2, ...N; N is the number of records within the input time series.

The probability density function of ε_k, that is PDF(ε_k), is obtained by applying Monte Carlo simulation on ε_k based on multiple sets of input forecast/hindcast time series in different T. Thus, the pseudo-forecast input time series f'_{ik} for a given T is able to create by adding back a random error generated based on PDF(ε_k) to the f_{ik} for each record:

$$f'_{ik} = f_{ik} + \text{rand}[\text{-1 } or \text{ 1}] \times \text{rand}[\text{PDF}(\varepsilon_k)] \tag{2}$$

where rand[ ] is the random operator that produces random number accordingly. Figure 1 illustrates the mechanism of the pseudo-forecast time series generation. Assuming ε_k follows a normal distribution as the green curves indicated, the purple curve shows the track of the original forecast time series f_{ik}, with which the pseudo-forecast time series f'_{ik} could be generated. The blue dotted lines represent the boundaries of 99.7% confidence intervals ($\pm$ 3 standard deviations away from the f_{ik}) of all the original forecast data points, that is, there is a 93% probability that the new generated pseudo-forecast time series track is within the boundaries if N = 24.

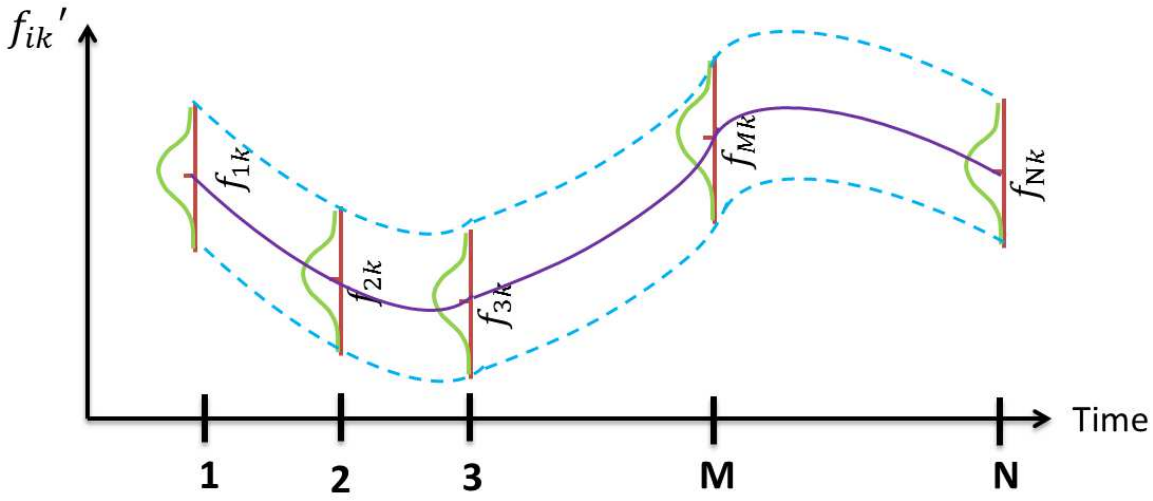


Fig. 1. Pseudo-forecast time series generation mechanism

The forecast probability map can be generated by running multiple models together in the same *T* based on corresponding different sets of pseudo-forecast input time series. Figure 2 shows a simple example of three pseudo-forecasts in the same *T*. The area *oab* is the representation of the ensemble of pseudo-forecasts. For illustrative purpose, diffusion effects are not considered in this example.

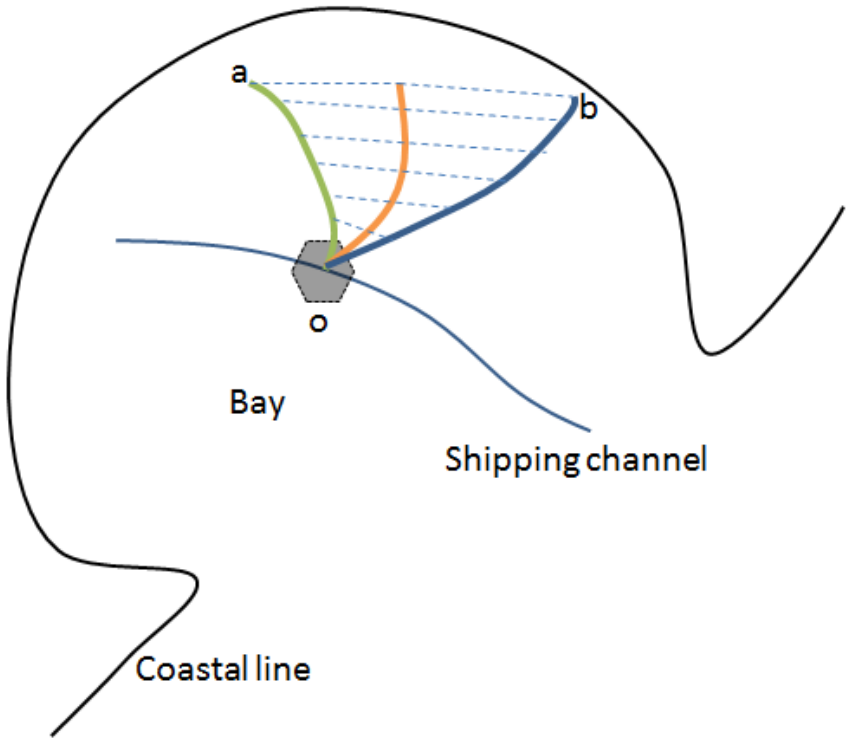


Fig. 2. Forecast probability map example. Position *o* is the initial spill location. The green, orange, and dark blue curves represent three different pseudo-forecasts.

2.2 Different Benching Scenarios

A forecast probability map consists of an approximation of ensemble of pseudo forecasts. Given a forecast probability map and the associated hindcast simulation, the oil spill forecast uncertainty can be quantified by evaluating the forecast uncertainty metrics according to two major possible beaching conditions: (1) forecast of non-beaching and (2) forecast beaching; which are accompanied by two possible observations: (a) observation of non-beaching and (b) observation of beaching.

1. *Forecast of non-beaching and observation of non-beaching*

Under this condition, both the ensemble of pseudo-forecasts (*oab*) and the observed track (*oc*) do not land at the beach within *T*. There are four logical cases where this

might occur: (1) *oc* is inside the range of oab but outside the ab boundary (Figure 3a), (2) *oc* falls entirely within the range of *oab* (Figure 3b), (3) *oc* falls partially within the range of *oab* but crosses over one (or more) of the lateral boundaries (Figure 3c), and (4) oc does not go through any part of *oab* (Figure 3d). For all the four cases, U_t and U_l, are defined to be 0 as these uncertainties require either actual or predicted beaching. For the cases in Figure 3a and Figure 3b, U_a is defined as 0 since *oc* stays inside the bounds of *oab*, i.e. there is no deviation of transport directions between the hindcast and forecasts; for cases in Figure 3c and Figure 3d, U_a is defined as:

$$U_a = \frac{A_{exceed}}{A_{oab}} \tag{3}$$

where A_{exceed} denotes the area that *oc* exceeds *oab*, that is area of *qbc* in Figure 3c, and area of *obc* in Figure 3d; A_{oab} is the area of *oab*.

For case (1), Figure 3a, U_s is defined as:

$$U_s = \frac{1}{V_{median}} \frac{D_{exceed}}{T} \tag{4}$$

where V_{median} is the median speed of the pseudo-forecast spills within T; D_{exceed} is the distance of the outermost observation point to *oab*, that is *cp* in Figure 3a. For cases (2), (3) and (4) U_s is defined as 0 because either the forecast speed is reasonable (Figure 3b), or the uncertainty is dominated by U_a (Figures 3c and d).

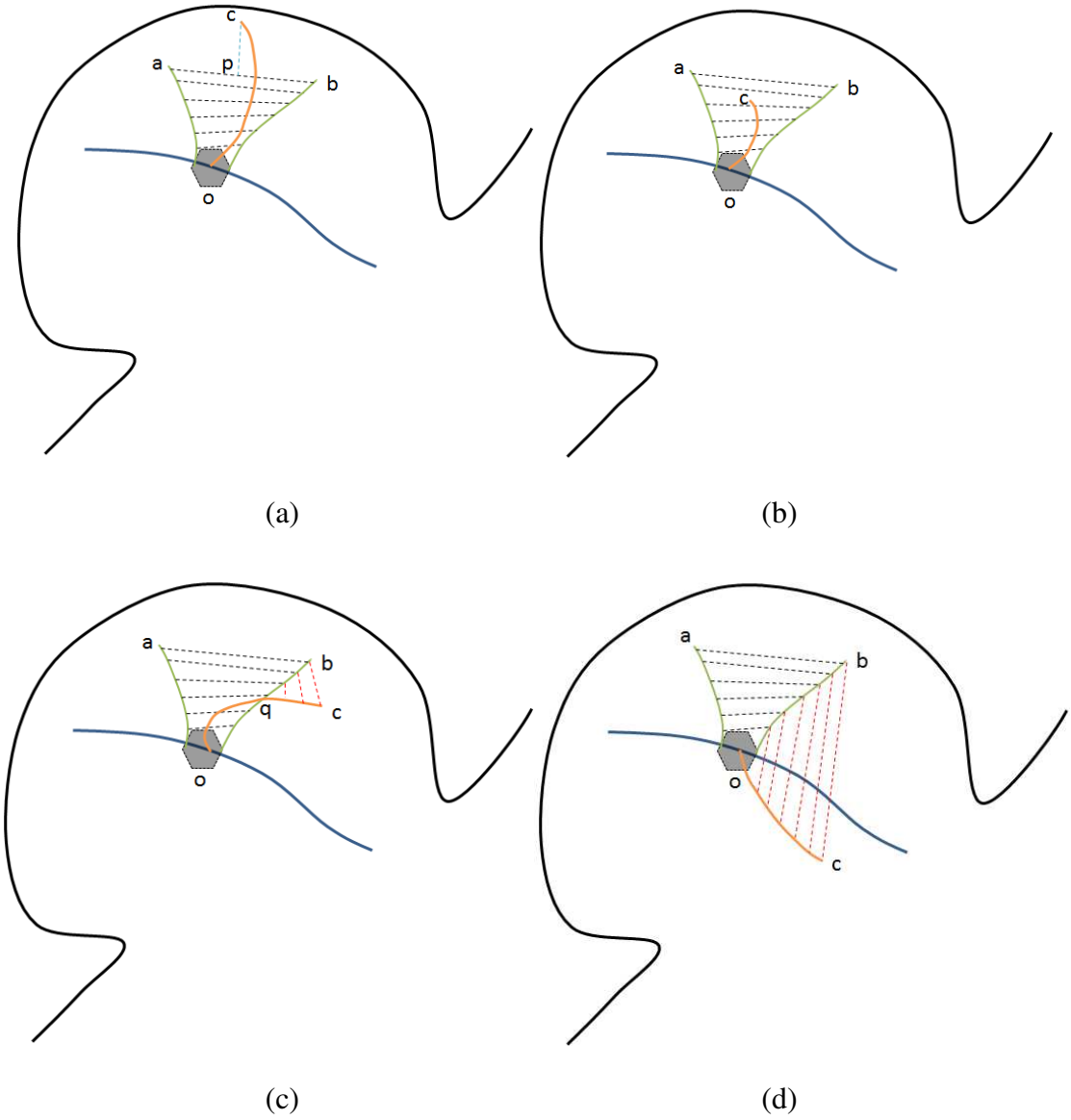


Fig. 3. Forecast of non-beaching and observation of non-beaching scenario

2. *Forecast of non-beaching and observation of beaching*

In this situation, *oab* does not intersect the beach, while oc is observed to intersect the beach within *T*. Three possible cases may happen: (1) *oc* is bounded within *oab* (Figure 4a), (2) *oc* does not go through any part of *oab* (Figure 4b), (3) *oc* falls partially within the range of *oab* but crosses over one (or more) of the lateral boundaries (Figure 4c). In this setting, U_t is declared for all cases as:

$$U_t = \frac{t_0 - T}{T} \tag{5}$$

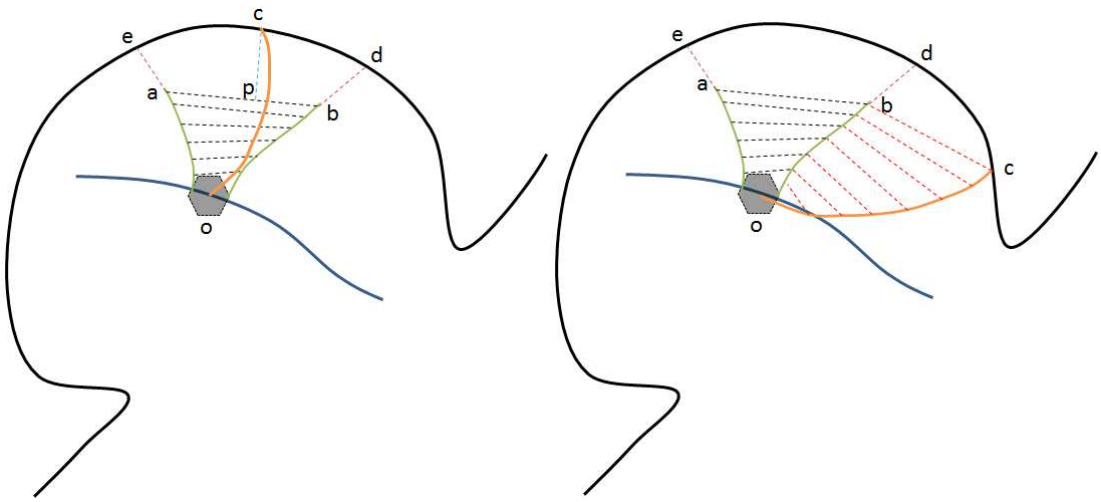
where t_0 is the observation track hit time.

U_l is defined as 0 in case (1) because *oc* does not go beyond the range of the potential landing boundary of the pseudo-forecasts *de*; for case (2) and (3), U_l is defined as:

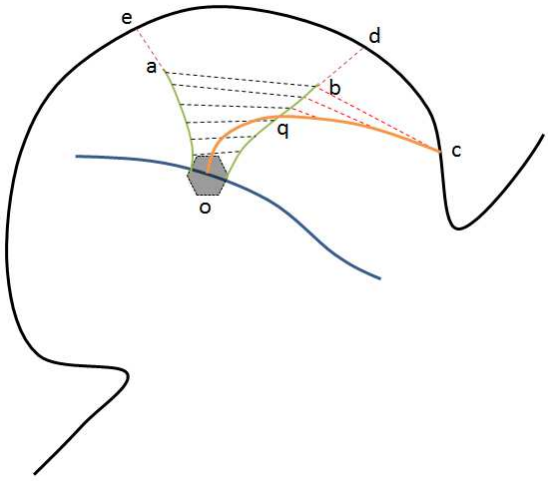
$$U_l = \frac{\mathrm{MIN}(\int_c^d \mathrm{d}s, \int_c^e \mathrm{d}s)}{L_{median}} \tag{6}$$

where L_{median} is the median curve length of the pseudo-forecast tracks.

U_a is defined as 0 in case (1) since *oc* is entirely inside the bounds of *oab*; for case (2) and (3), U_a is defined the same as cases in Figure 3c and Figure 3d based on Equation 3. U_s is 0 in case (2) and (3); for case (1), U_s has the same defination as the case in Figure 3a based on Equation 4.



(a) (b)



(c)

Fig. 4. Forecast of non-beaching and observation beaching scenario. *d* and *e* denote potential landing boundary points for the ensemble of pseudo-forecasts.

3. *Forecast beaching and observation of non-beaching*

In this scenario, the observed *oc* does not intersect the beach but the predicted *oab* does. Three possible cases may happen: (1) *oc* is entirely inside *oab* (Figure 5a), (2) *oc* is entirely outside oab (Figure 5b), (3) *oc* falls partially within the range of *oab* but crosses over one (or more) of the lateral boundaries (Figure 5c). For all cases within this seneario, U_t is defined as:

$$U_t = \frac{T - t_{median}}{T} \quad (7)$$

where t_{median} is the median hit time of the pseudo-forecast spills.

U_a is defined as 0 in case (1); for case (2) and (3), U_a is defined the same as cases in Figure 3c and Figure 3d based on Equation 3. Since *oc* does not land the beach, U_l and U_s are defined to be 0 for all cases.

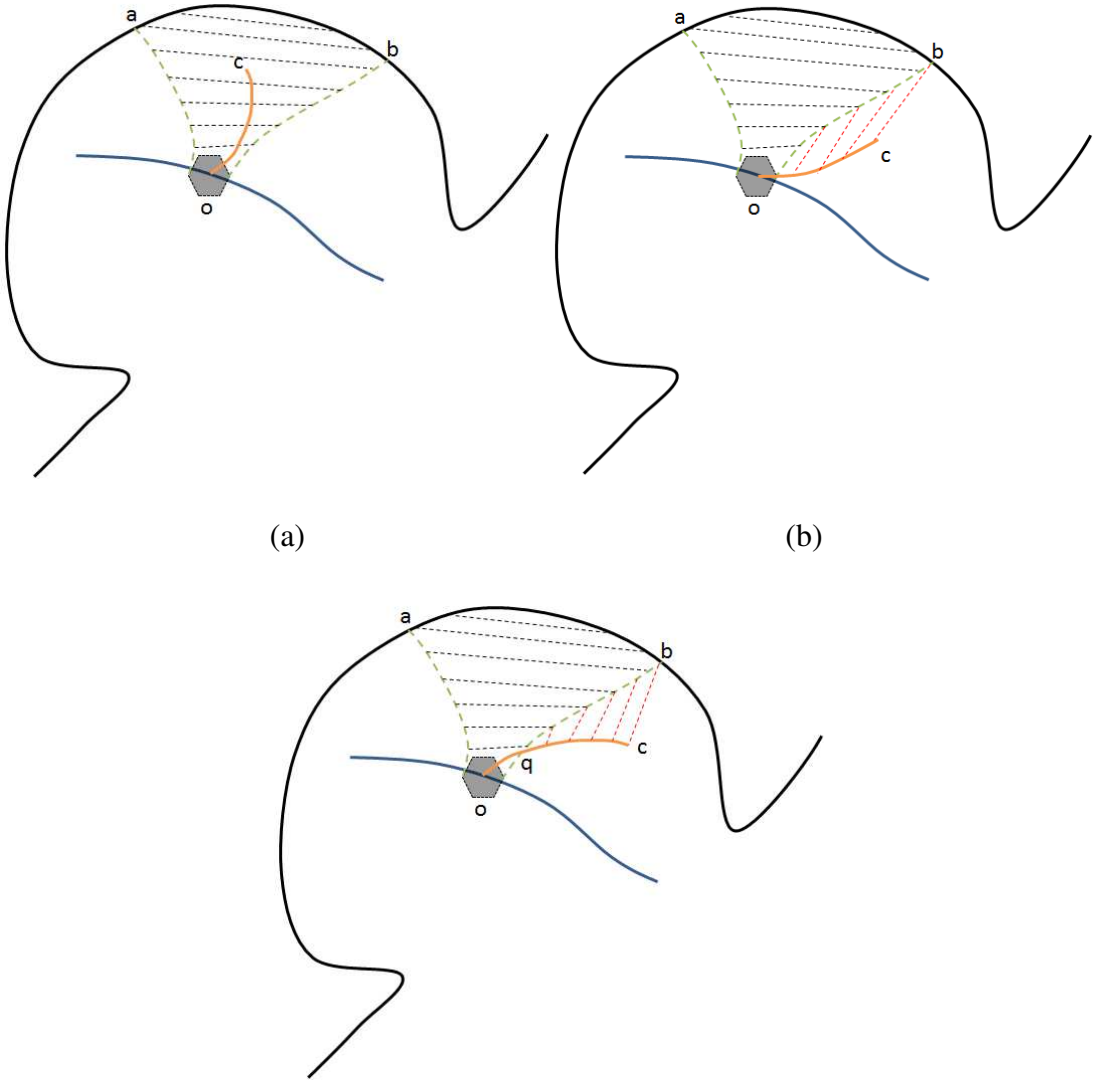


Fig. 5. Forecast beaching and observation of non-beaching scenario

4. *Forecast of beaching and observation of beaching*

It might also happen that both of *oc* and *oab* intersect the beach within *T*. In this circumstance, three possible cases may happen: (1) *oc* is entirely bounded within *oab* (Figure 6a), (2) *oc* is entirely outside *oab* (Figure 6b), (3) *oc* falls partially within the bound of *oab* but crosses over one (or more) of the lateral boundaries (Figure 6c). For all cases, U_s is defined as 0; U_t is defined as:

$$U_t = \frac{t_0 - t_{median}}{T}$$

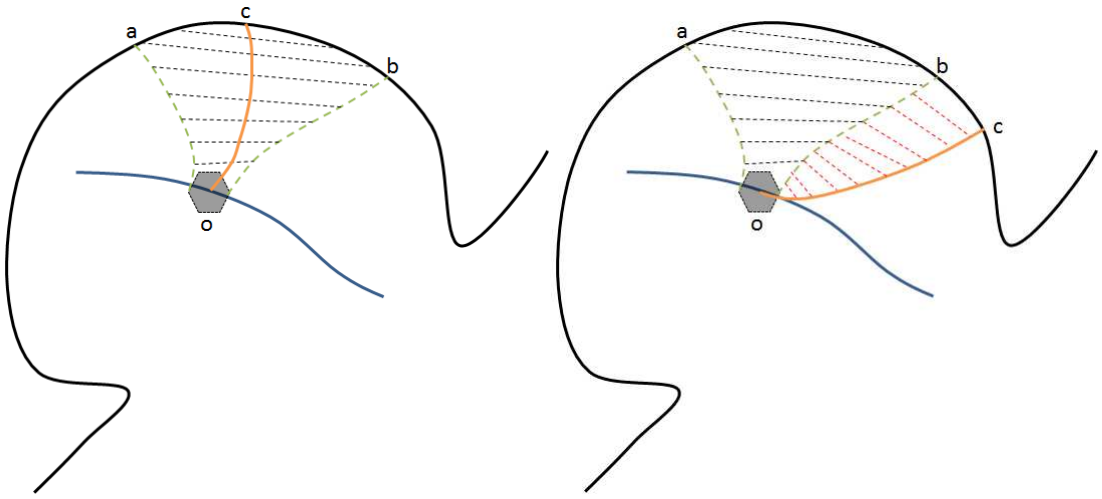
(8)

U_l is set as 0 in case (1); for cases (2) and (3), U_L is defined as:

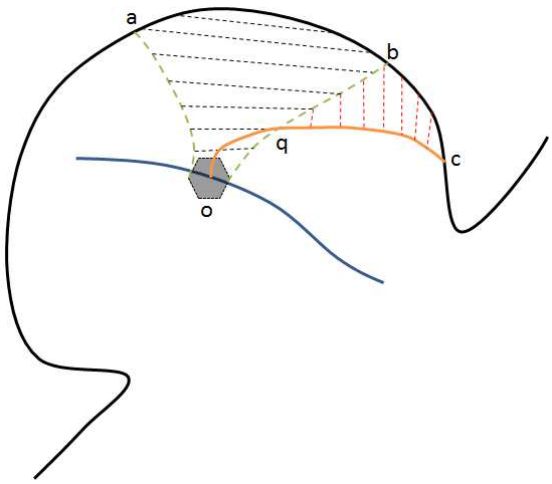
$$U_l = \frac{\mathrm{MIN}(\int_c^b \mathrm{d}s, \int_c^a \mathrm{d}s)}{L_{median}}$$

(9)

U_a is defined as 0 in case (1); for case (2) and (3), U_a is defined the same as cases in Figure 3c and Figure 3d based on Equation 3.



(a) (b)



(c)

Fig. 6. Forecast beaching and observation beaching scenario

2.3 Derivation of the HFRA method

The HyosPy Forecast Reliability Assessment (HFRA) is a posteriori method based on the comparison of the hindcsat and prior forecast simulations. HFRA applies the HyosPy system (Hou et al. 2015; Hodges et al. 2015; Hou and Hodges 2014) to provide data for oil spill forecast reliability assessment in a statistical manner. Figure 7 shows the mechanism of HyosPy simulation within a forecast period *T*. The first model starts at time 0, running in a forecast mode (i.e. using all forecast wind, tide and river inflow data). The second model starts at *t* when *t* hours of hindcast are available; the second model starts from the same start time, so the model is in hindcast mode for the interval [0,*t*] hours and then forecast mode for [*t*,*T*] hours. The third model (started at time 2*t*) does not need to run the time interval from [0,*t*] as it would be identical to the second model; a "hotstart" feature is used to use the results of the second model at the end of its hindcast (*t*) as the starting point for the third model, which is in hindcast for the interval [*t*, 2*t*] and forecast from [2*t*,*T*]. This pattern is continued until the final run is entirely in hindcast mode. In the present work, the complete hindcast simulation provides the benchmark for comparing the reliability of the forecasts. The methods developed herein could be applied with observed spill or drifter data for a better assessment of model reliability.

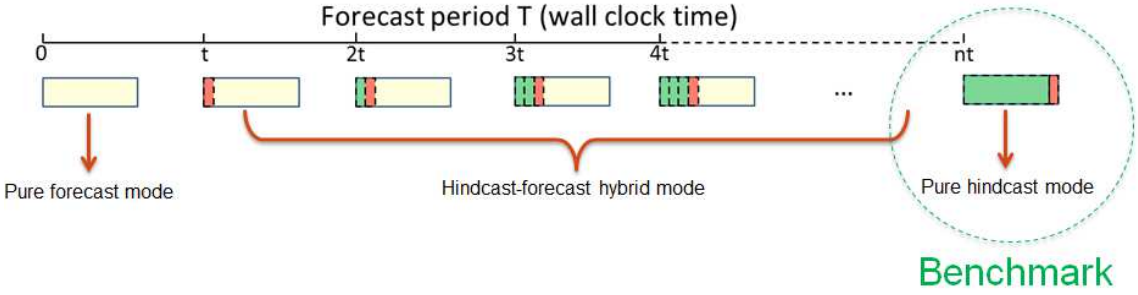


Fig. 7. A single forecast period HyosPy simulation. Light yellow represents forecast simulation. Red represents hindcast simulation. Green denotes the "hotstart" time, i.e. the hindcast simulation which has already been computed by the prior model.

To measure the ensemble forecast uncertainty within a complete single HyosPy simulation, an Error Index (*EI*) is designated based on the Minimum Regret strategy to minimize the worst-case regret (Galt 1997; Galt and Payton 1999). Specifically, given the last model run as benchmark, *EI* adds up all the likely error caused by different forecasts at each time step. The likely error herein is represented by the Mean Distance Error Index (*MDEI*) and the Dispersion Error Index (*DEI*) to measure the average distance and dispersion area difference (respectively) for the modeled spill particles at each time step between a test case and a control case. The *MDEI* and *DEI* for each time step $p = 1, 2, 3...n$ are defined as

$$MDEI_p = \frac{\sum_{q=1}^{m} e_{p,q}}{mS} \tag{10}$$

$$DEI_p = \frac{\sqrt{\Delta A_p}}{S} \tag{11}$$

where m is the number of oil particles for each time step; $e_{p,q}$ denotes the distance of corresponding particle q at time step p between two test cases; ΔA_p is the particle dispersion area difference at time step p between two test cases; S is the curve length of the spill track in the control case. Figure 8 illustrates the quantification process of *MDEI* and *DEI* based on a simple two-case simulation (a test case and a control case).

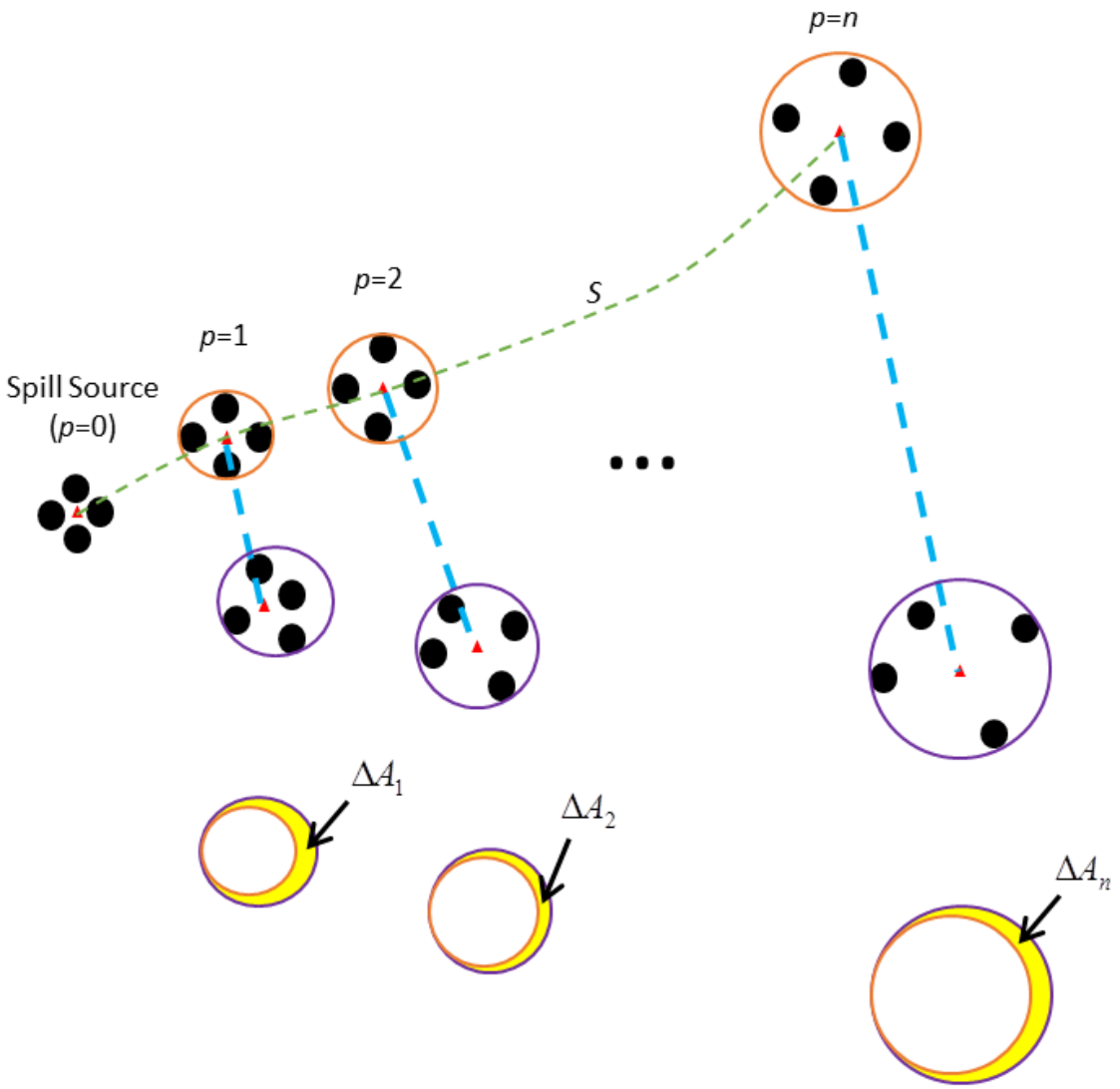


Fig. 8. *MDEI* and *DEI* quantification schematic. Each black dot represents an oil spill particle; red triangles denote central locations of particles at each time step; the orange circles compose the control case; the purple circles compose the test case; the length of the green dash line is *S*; the length of the blue dash line is the mean distance for the modeled spill particles at each time step between the test case and the control case, i.e. $MDEI_p \times S$;; the area of yellow portion is ΔA_p.

Such that

$$EI_p = \sum_{j=1}^{M-1} \sqrt{MDEI_{j,p}^2 + DEI_{j,p}^2}$$

(12)

where j = 1,2,3, ...M − 1; M is the number of models within T.

The oil spill forecast reliability is calculated based on the Critical Error Index (*CEI*) with multiple sets of *EI* as statistic inputs. *CEI* is set as the tolerance, i.e. upper limit of allowable error of a reliable forecast. Because *EI* is the accumulation of n − 1 models' error metrics, *CEI* is defined with the same scale of *EI* based on the n − 1 cumulative of the Spill Scale Benchmark Index (*SSBI*) which is derived from the boom containment lengths of a typical accident. The *SSBI* is defined as

$$SSBI = \frac{B}{L}$$

(13)

where B denotes the theoretical length of boom required to contain free floating oil and L is the length of boom that were actually deployed. More specifically, B = 1.25H, where H is the amount of oil spill in m^3 (PERSGA/UNEP 2003).

Such that

$$CEI = (n-1) \times SSBI \tag{14}$$

With the idea that *EI* below *CEI* is within the tolerance, the forecast reliability η is designed as:

$$\eta = \frac{N_{under}}{N_{total}} \tag{15}$$

where N_{under} denotes the number of *EI* curves that remain under *CEI*; N_{total} is the total number of *EI* curves.

As a simple example, Figure 9 shows the forecast reliability analysis process with 3 different *EI* curves. Before T_1, all of the three curves are under *CEI*, thus the η is 100% during 0 to T_1; during T_1 to T_2, EI_3 has gone beyond *CEI*, hence, the η decreases to 67% (2/3); during T_2 to T_3, only EI_1 remains under *CEI*, thus, the η reduces to 33% (1/3); after T_3, none of the curves are remaining under *CEI*, so that the η is 0%.

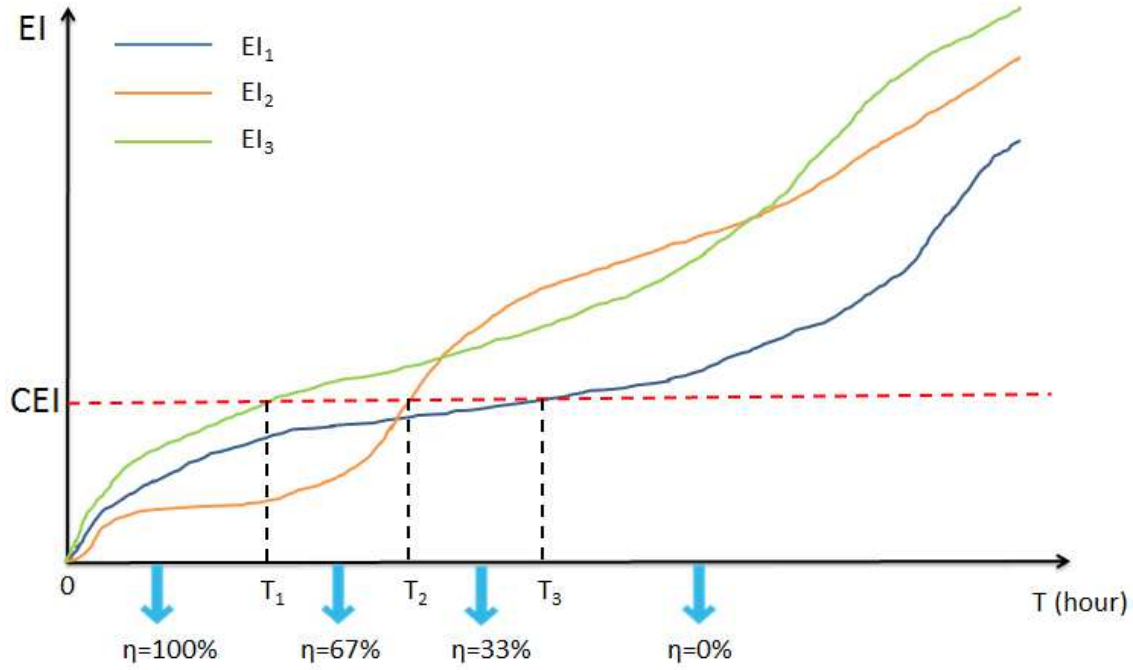


Fig. 9. Forecast reliability example

3. Case study

3.1 HyosPy setup

As discussed in prior section, a single instance of HyosPy running provides multiple forecasts and hindcasts within forecast period *T*. A series of continuous HyosPy simulations would yield statistic sample for *EI* evaluation in HFRA, and Monte Carlo simulations to generate pseudo-forecast series for forecast uncertainty analysis.

In this study, a 48-hr HyosPy simulation with 17 individual model runs at a 3-hr interval (i.e. the first model use complete forecast data; the last one use complete

hindcast data) is used for demonstration (Table 1). This demonstration uses the SELFE - GNOME system, which was previously integrated into HyosPy (Hou et al. 2015; Hou and Hodges 2014).

Table 1. Sequenced model operations

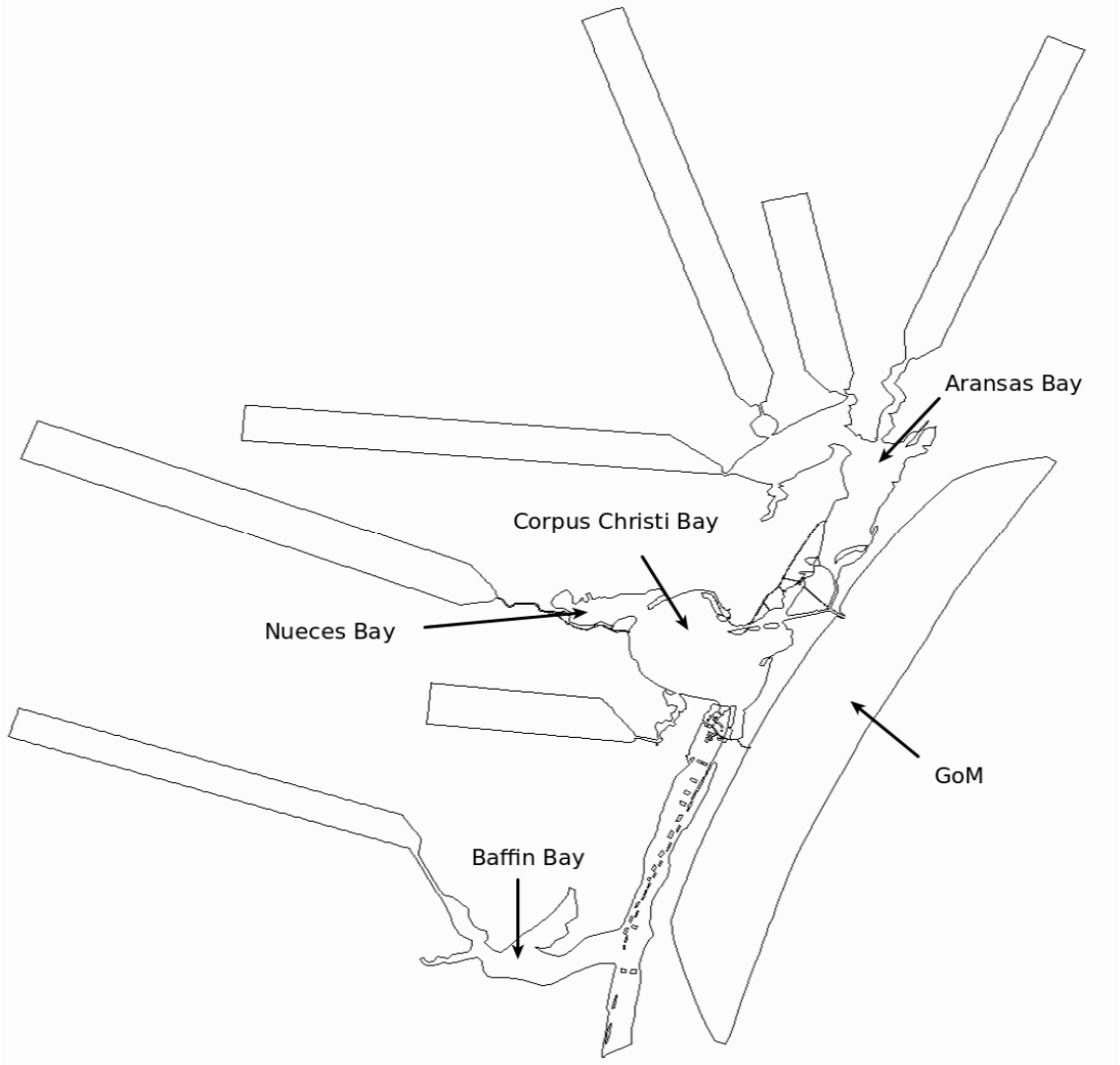
Model #	Start time	Hindcast hour	Forecast hour
1	0:00	0 hour	48 hours
2	3:00	3 hours	45 hours
3	6:00	6 hours	42 hours
…	…	…	…
17	(+2d) 0:00	48 hours	0 hour

The study region is the Corpus Christi Bay portion of the Texas coastal bend region, which is simulated in the Texas General Land Office - Texas Water Development Board (TGLO-TWDB) operational oil spill system. Corpus Christi Bay (Figure 10) is a shallow (3*m* average depth) embayment that is bisected by a shipping channel dredged to 15*m* depth serving the Port of Corpus Christi. The bay is microtidal with typical daily excursions of $O(0.5m)$; however, barotropic tides associated with weather system can cause additional displacements over several days. Wind-driven flows in the bay are dominated by a southerly sea breeze in the afternoon during summer months, but the pattern is periodically disrupted by weather systems with northerly winds (Ward 1997)

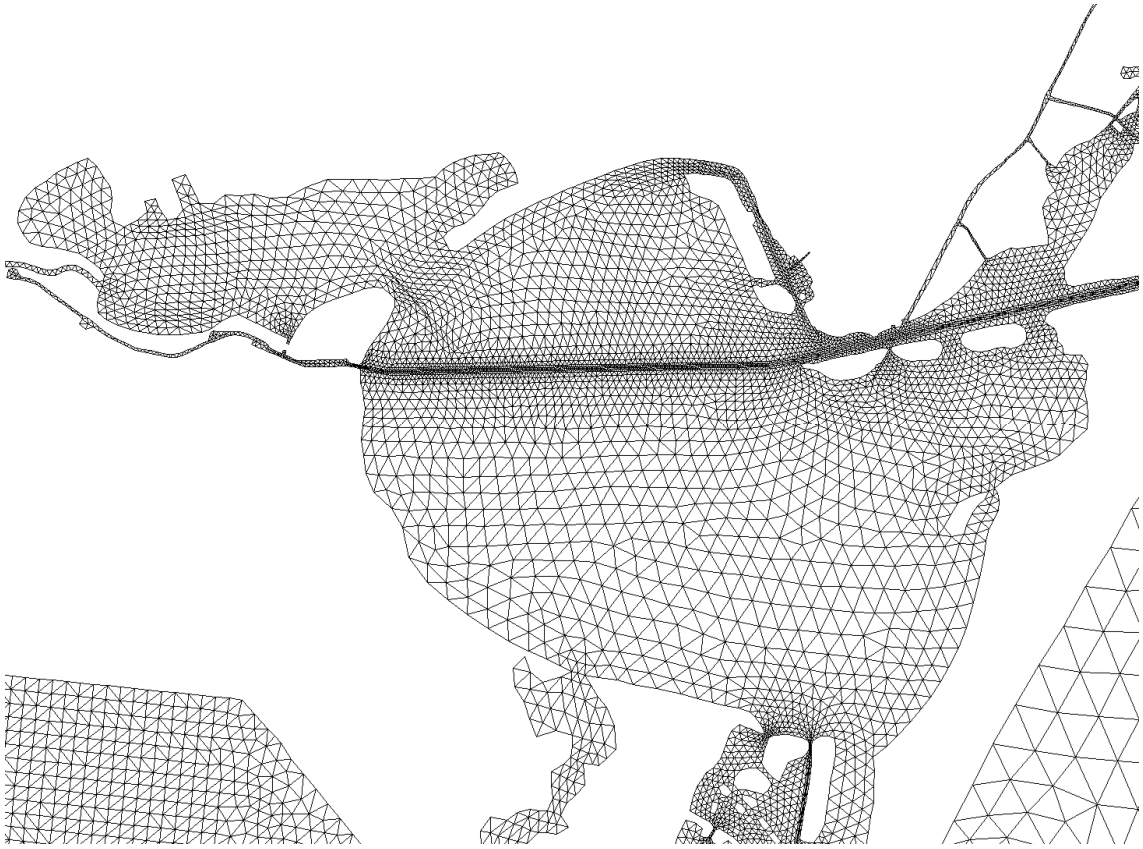
Fig. 10. Corpus Christi Bay

The model grid, Figure 11, was designed by TWDB as part of their modeling program. The computational domain extends from Aransas Bay in the north to just below Baffin Bay in the south, which includes all of Corpus Christi Bay and its neighbors Nueces Bay and Oso Bay. A small offshore portion of the Gulf of Mexico is included for enforcing tidal boundary conditions. To damp wave reflections that would occur if river inflows were enforced directly at the river mouths, large dummy domains (rectangular areas in Figure 11a) are included for each river. Figure 11b shows the portion of the model grid for Corpus Christi Bay discretized by triangular elements in the horizontal direction. The entire model grid contains 23286 nodes (vertex of triangular element) and 41866 elements. The hydrodynamic model time step is set as

180 seconds; the oil spill transport model time step is 900 seconds. Using such setup a 48-hr hydrodynamic simulation takes about 20 minutes of parallel running with 2 logic processors based on MPI protocol on a 24-processor 64 GB-RAM Linux workstation. The corresponding oil spill simulation takes less than 10 seconds.



(a)



(b)

Fig. 11. SELFE horizontal model grid for Corpus Christi bay

Two major sources of forecast model inputs are taken into consideration – wind and tide. Wind forecast are derived from the NAM Model of the NCEP (NECP 2015). The Oceanography Department of Texas A&M University (TAMU) configure the NAM outputs and provide the TWDB with wind hindcast/forecast data from a TAMU server which contains wind data for 241 sites (Figure 12). These data could be used to develop a spatially-varying wind field for the hydrodynamic model; however for the present demonstration, data from a single site (Corpus Christi Bay site 051) are used for a spatially-uniform wind field over the study area. The TAMU server provides wind hindcasts for two months and 4-day wind forecasts on a 3-hour update cycle. Tidal data are obtained from the Texas Coastal Ocean Observation Network (TCOON). This network includes a system of tidal gages along the Texas coast with rapid data availability online. Water level elevations are typically available on six minute

intervals with a real-time lag on the order of 10 minutes. The adjusted harmonic tidal elevation forecast of gage at Bob Hall Pier is used as the offshore water surface elevation to drive the hydrodynamic model (TCOON 2015).

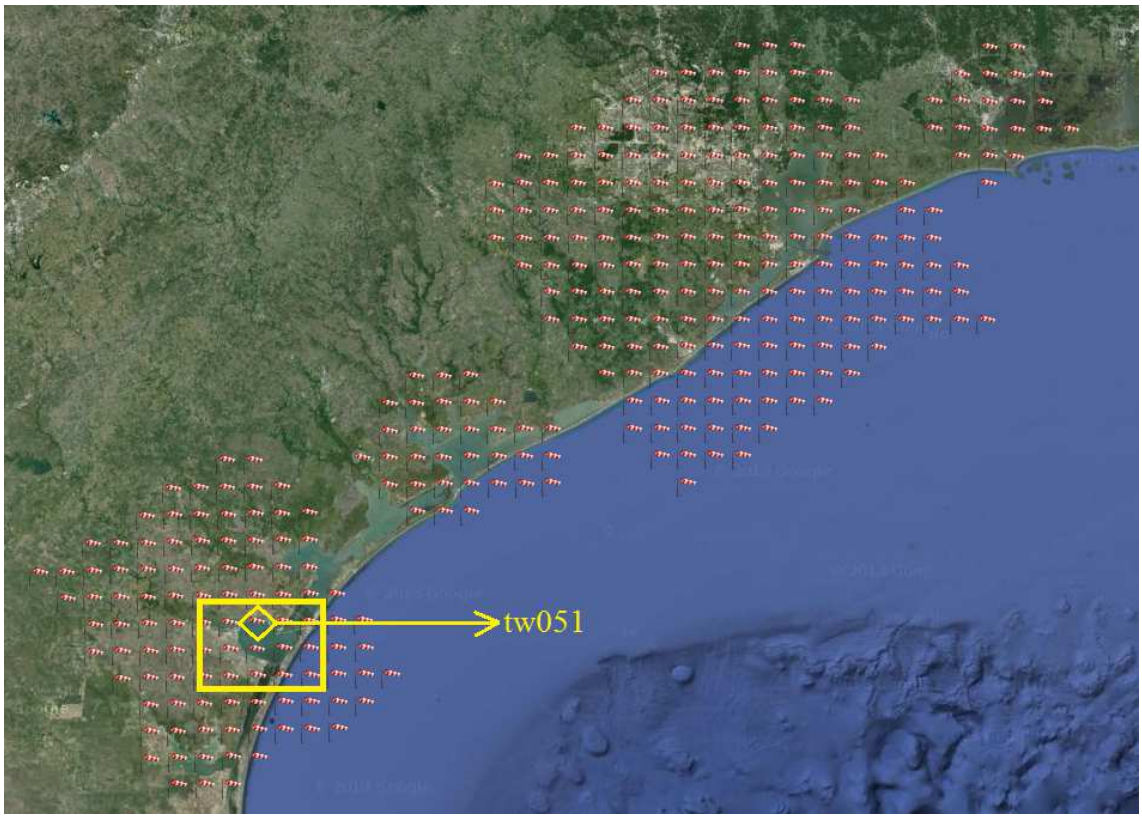
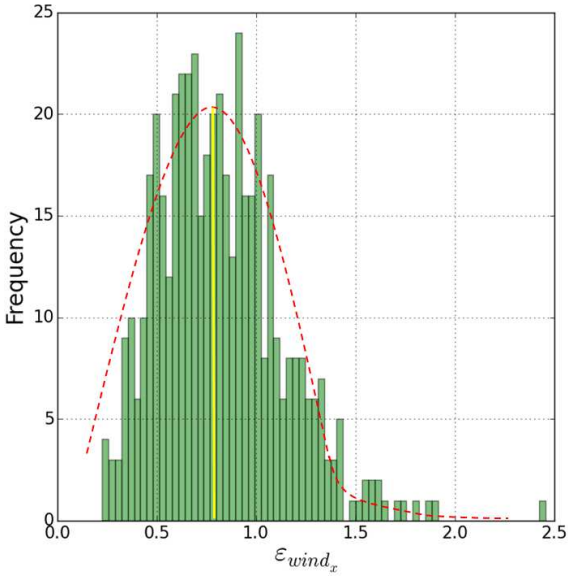


Fig. 12. Distribution of wind data locations

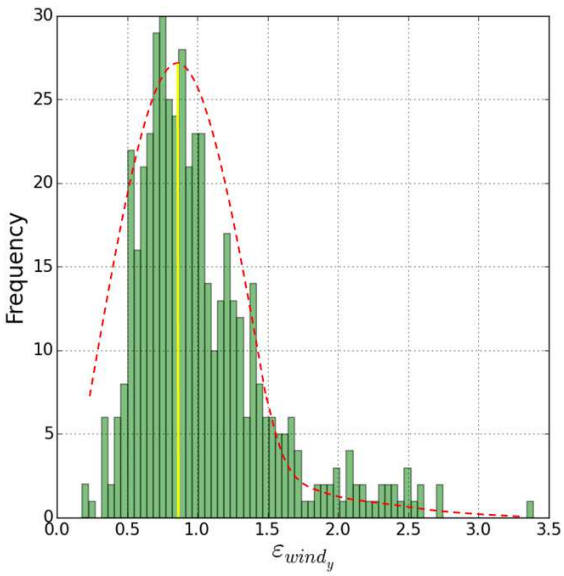
3.2 Forecast uncertainty probability map implementation

The 17-model HyosPy simulation was started on 7/26/2014 and continuously run for a sequence of 2-day forecast simulation run for 2-day forecast periods until 4/11/2015. Because data servers were intermittently down during this time, a total of 66 complete simulation sets out of 128 possible were created. For this investigation, an imaginary oil spill location is selected close to the shipping channel at [27.812 N, -97.309 W] (or UTM 666580*m* E, 3077520*m* N, zone 14).

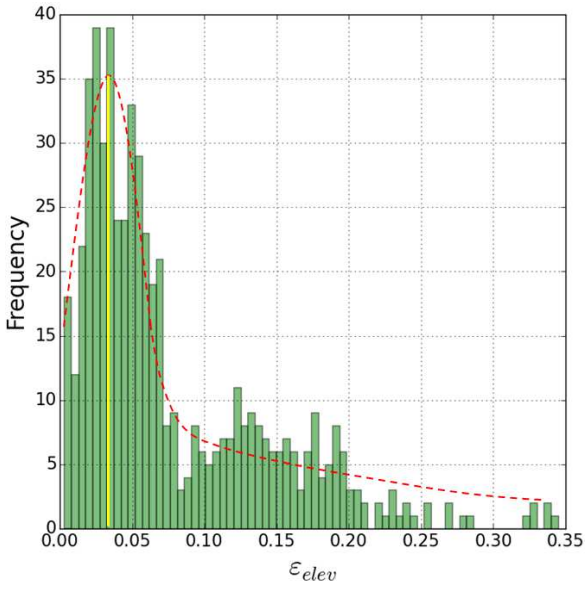
The 66 runs provide forecast/hindcast data series for the input forecast error ε_k (see Equation 1) assessment based on Monte Carlo simulation. Figure 13 shows the modeling results of east-west wind forecast error ε_{windx} (Figure 13a), northsouth wind forecast error ε_{windy} (Figure 13b), and tide elevation forecast error ε_{elev} (Figure 13c). It can be seen that ε_{windx} and ε_{windy} follow a normal distribution approximately; while a student-T distribution is a better fit for ε_{elev} since it has a fatter tail, i.e. more extreme events..



(a) Distribution of ε_{windx}



(b) Distribution of ε_{windy}



(c) Distribution of ε_{elev}

Fig. 13. Monte Carlo simulation for ε_k. The red dash line is the "best fit" curve (PDF curve); the yellow line demonstrates the "best value" (mode of the PDF).

We chose the 8/5/2014 - 8/7/2014 HyosPy simulation results as a typical example to demonstrate the forecast uncertainty quantification process. Given the ε_k, we were able to generate multiple sets of f'_{ik} series based on Equation 2 with which multiple sets of pseudo-forecast spill tracks could be generated, so that an oil spill forecast probability map can be obtained. In this demonstration, we applied 10 pseudo-forecasts to compose the forecast probability map for illustrating purpose (Figure 14).

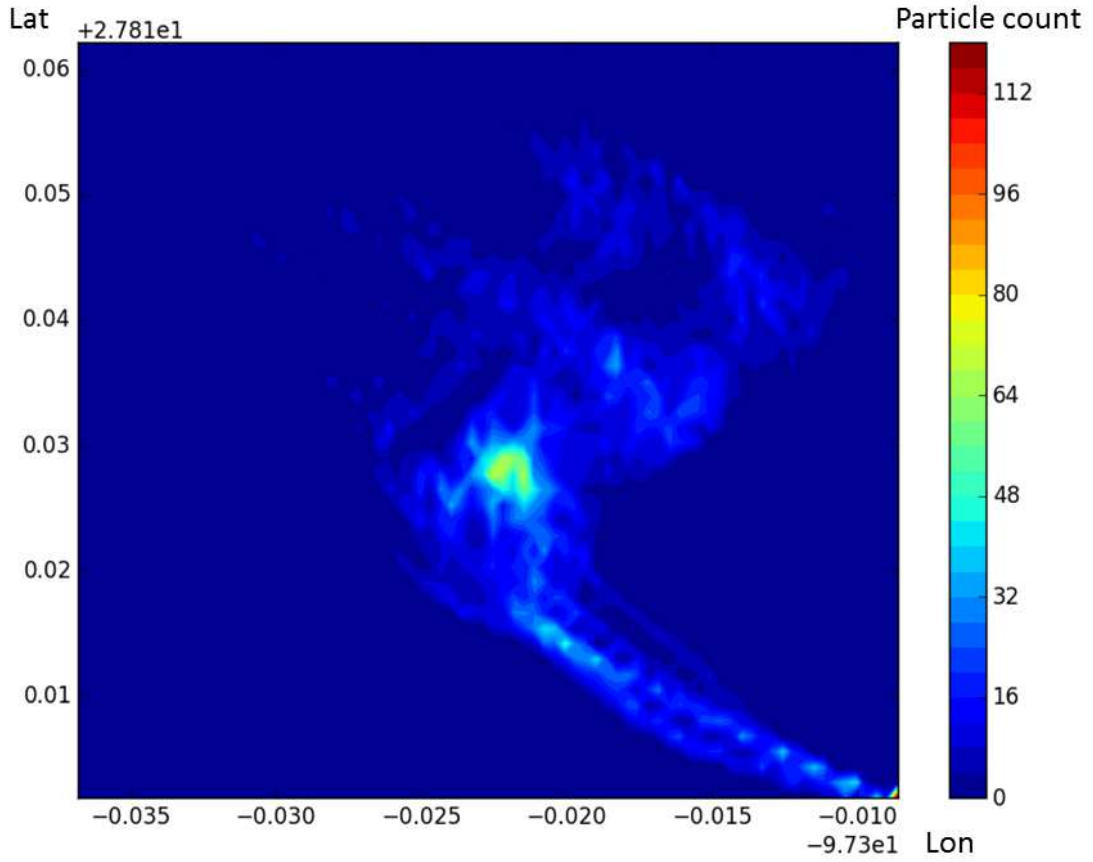


Fig. 14. Probability map based on 10 pseudo-forecasts. Spill source locates at southeast corner.

Figure 15 shows the ensemble of pseudo-forecasts *OAB* (projected on the light yellow polygon) and the associated hindcast simulation *OC* (projected on the green yellow polygon) on 3D Google Earth. In this case, both the hindcast and pseudo-forecast spill tracks land to the beach and the hindcast track is entirely outside the pseudo-forecasts, which is the same as the case in Figure 6b. Thus, U_l is reflected by the ratio of the red line from point *A* to *C* to the median curve length of the pseudo-forecasts; U_a is interpreted by the ration of the orange polygon to the light yellow polygon (*OAB*). Table 2 shows the calculation results.

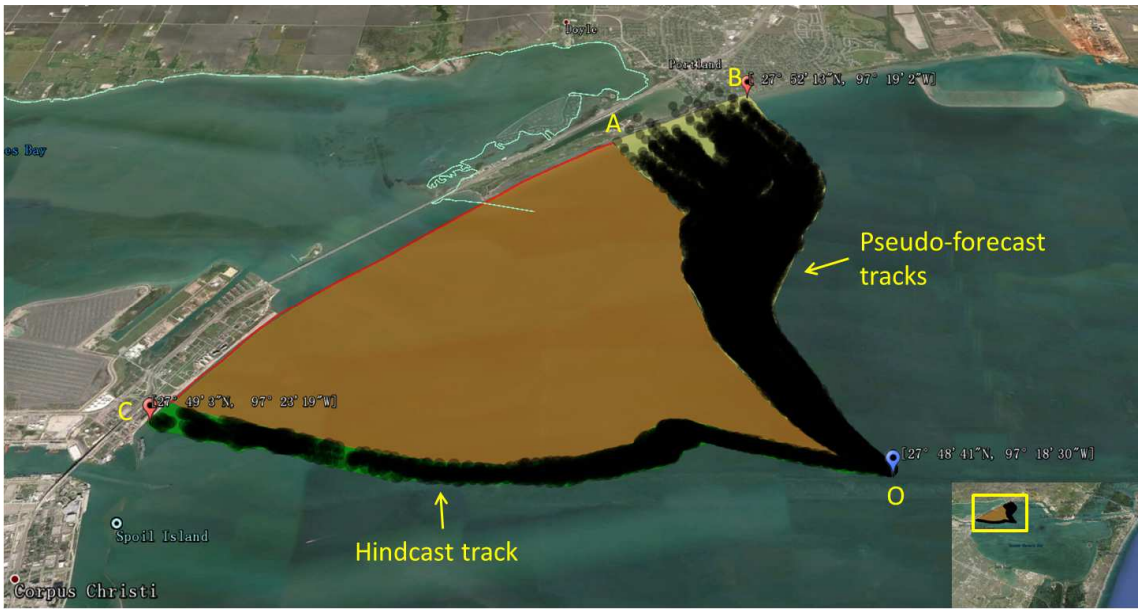


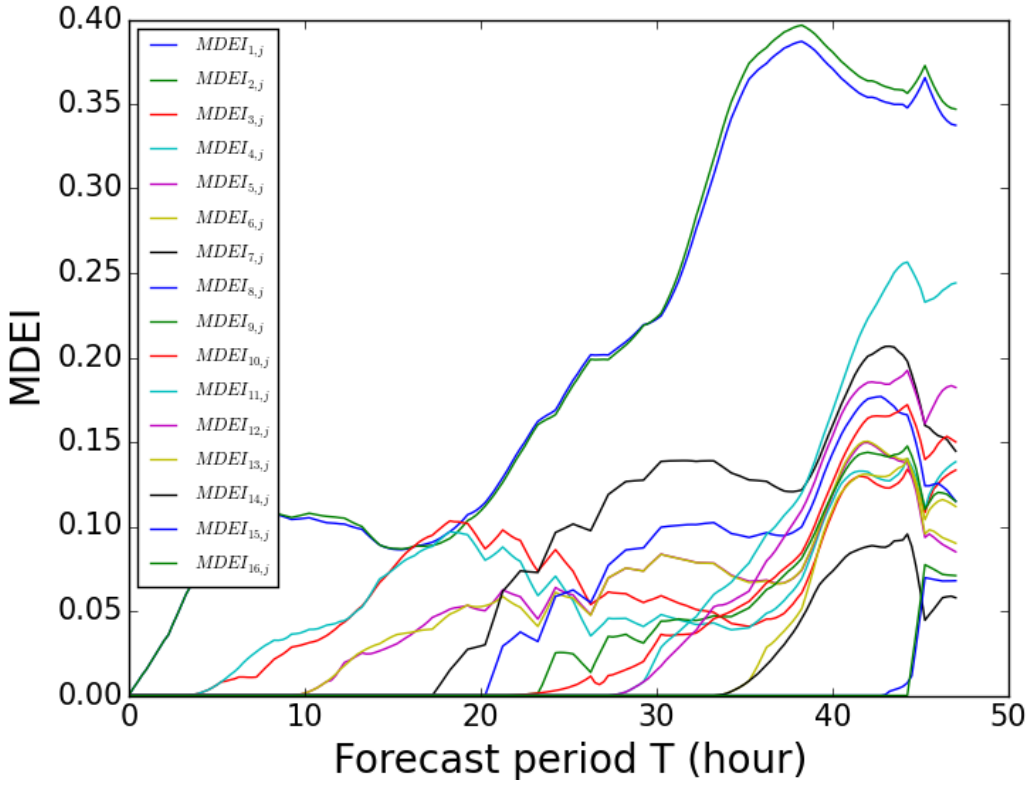
Fig. 15. Forecast uncertainty quantification demonstration on Google Earth. The blue marker marks the oil spill starting position; the red markers marked at boundaries are used as geometric reference points; oil spill tracks are denoted by small black dots.

Table 2. Forecast uncertainty calculation results of 8/5/2014 - 8/7/2014 HyosPy simulation

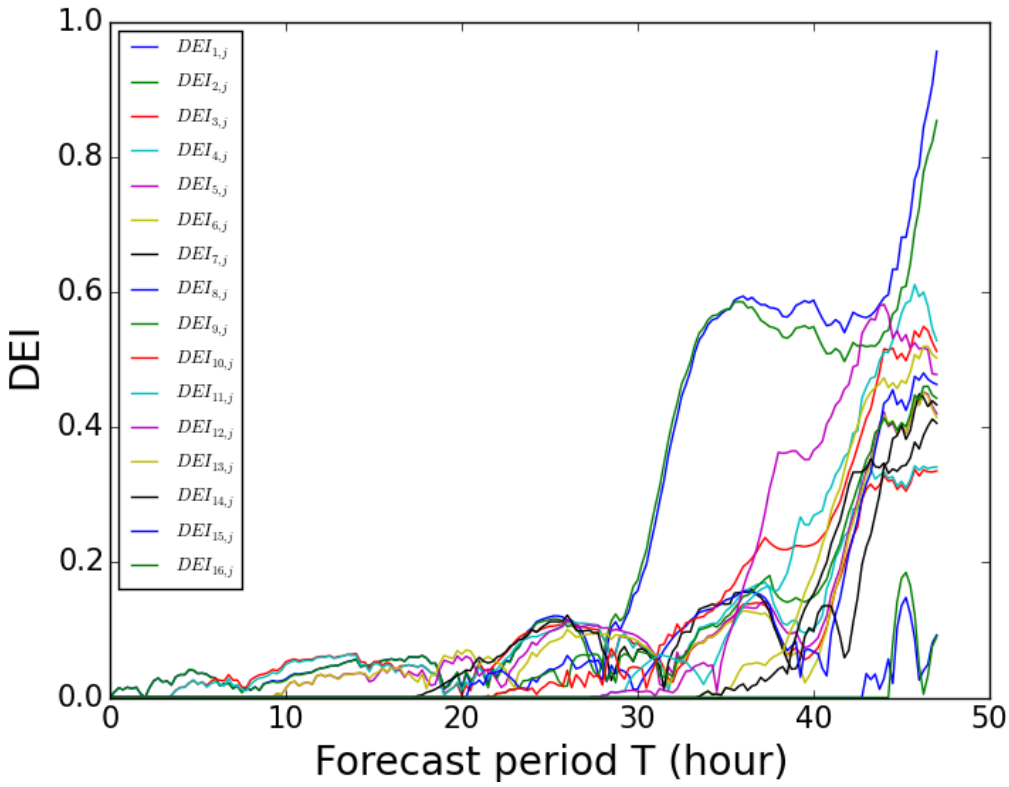
Uncertainty metrics	Value
U_t	0.252
U_l	0.845
U_a	3.157
U_s	0

3.3 HFRA application

The forecast reliability around [27.812 N, -97.309 W] is assessed by statistically analyze the *EI* curves based on the 66 sets of runs. Figure 16 shows examples of *MDEI* and *DEI* curves derived from the 8/5/2014 - 8/7/2014 HyosPy simulation. Generally, the more posterior the model initiates, the "later" the *MDEI* and *DEI* will appear. Because the pure hindcast model (the last model) is set as the benchmark, the more posterior the model initiates, the more hindcast data is able to be used which would result in 0 *MDEI* and *DEI* correspondingly. The *MDEI* and *DEI* tend to increase over the forecast period.


MDEI
Forecast period T (hour)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0
10
20
30
40
50
$MDEI_{1,j}$
$MDEI_{2,j}$
$MDEI_{3,j}$
$MDEI_{4,j}$
$MDEI_{5,j}$
$MDEI_{6,j}$
$MDEI_{7,j}$
$MDEI_{8,j}$
$MDEI_{9,j}$
$MDEI_{10,j}$
$MDEI_{11,j}$
$MDEI_{12,j}$
$MDEI_{13,j}$
$MDEI_{14,j}$
$MDEI_{15,j}$
$MDEI_{16,j}$

(a) *MDEI* plot



(b) *DEI* plot

Fig. 16. *MDEI* and *DEI* curves based on 8/5/2014 - 8/7/2014 HyosPy simulation

Figure 17 shows the *EI* curve of this simulation period. To determine the *CEI*, *SSBI* needs to be quantified. We choose the scale of Texas City Y oil spill accident as the typical case. The volume of oil spilled was $O(1000m^3)$, $B = 1.25 \times 1000 = 1250$ meters of boom; L actually deployed was about 8800 meters from Texas City Dike to the central spill location (Patterson 2014). Thus, *SSBI* is evaluated as about 0.14 based on Equation 13. Thus we conclude that *CEI* is 2.24 based on Equation 14. It is therefore showing that the forecast range from 0 to about 33 hours is below the *CEI*, indicating the forecast quality within 33 hours' simulation is relatively reliable; the rest of the modeling results (33*hr* to 48*hr*) are beyond the tolerance.

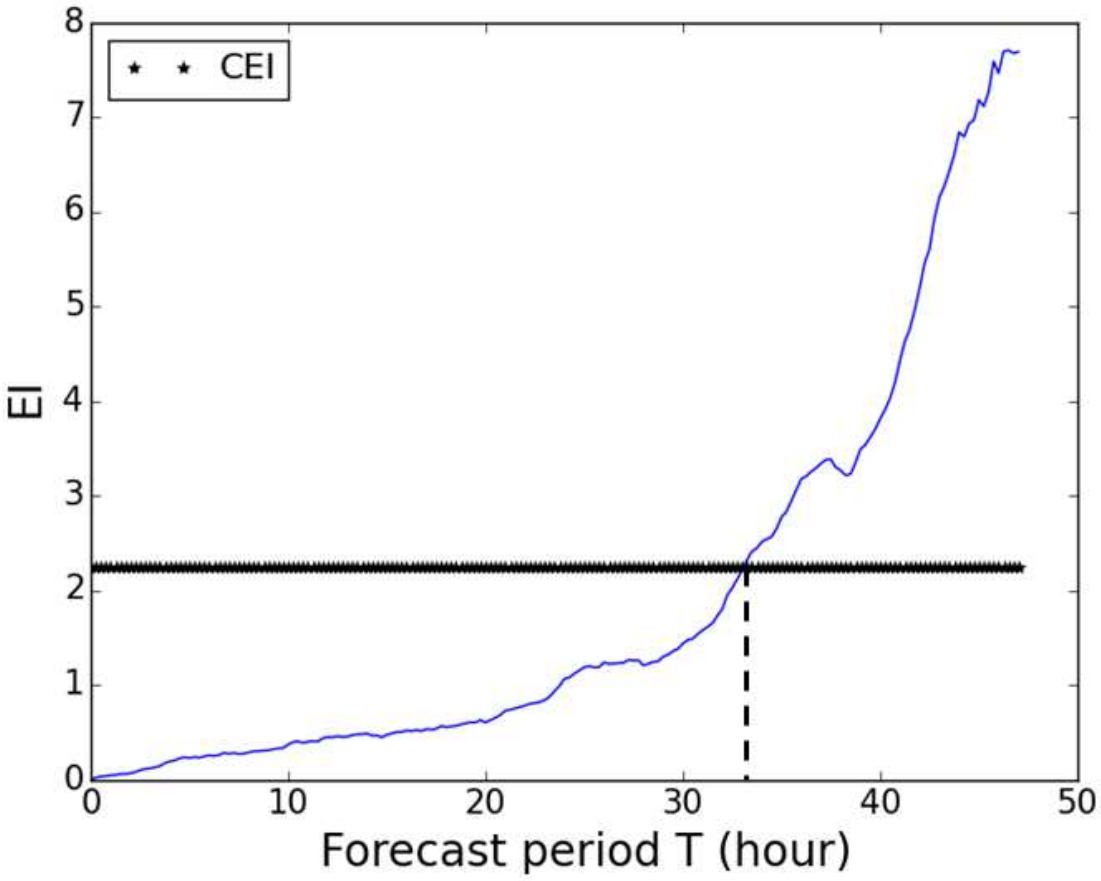


Fig. 17. *EI* curve based on 8/5/2014 - 8/7/2014 HyosPy simulation

Figure 18 shows the forecast reliability assessment results based on the 66 sets of HyosPy runs, with which the reliability of a new forecast can be judged according to any prediction ranges. For example, a 12-hr forecast can be trusted in 100% confidence; while the η of a 30-hr forecast is less than 50%. Naturally, these results are based on a model-model comparison (hindcast vs. forecast) and should not be taken as indicative of actual reliability that would be computed with use of drifter observations or actual oil spill observations.

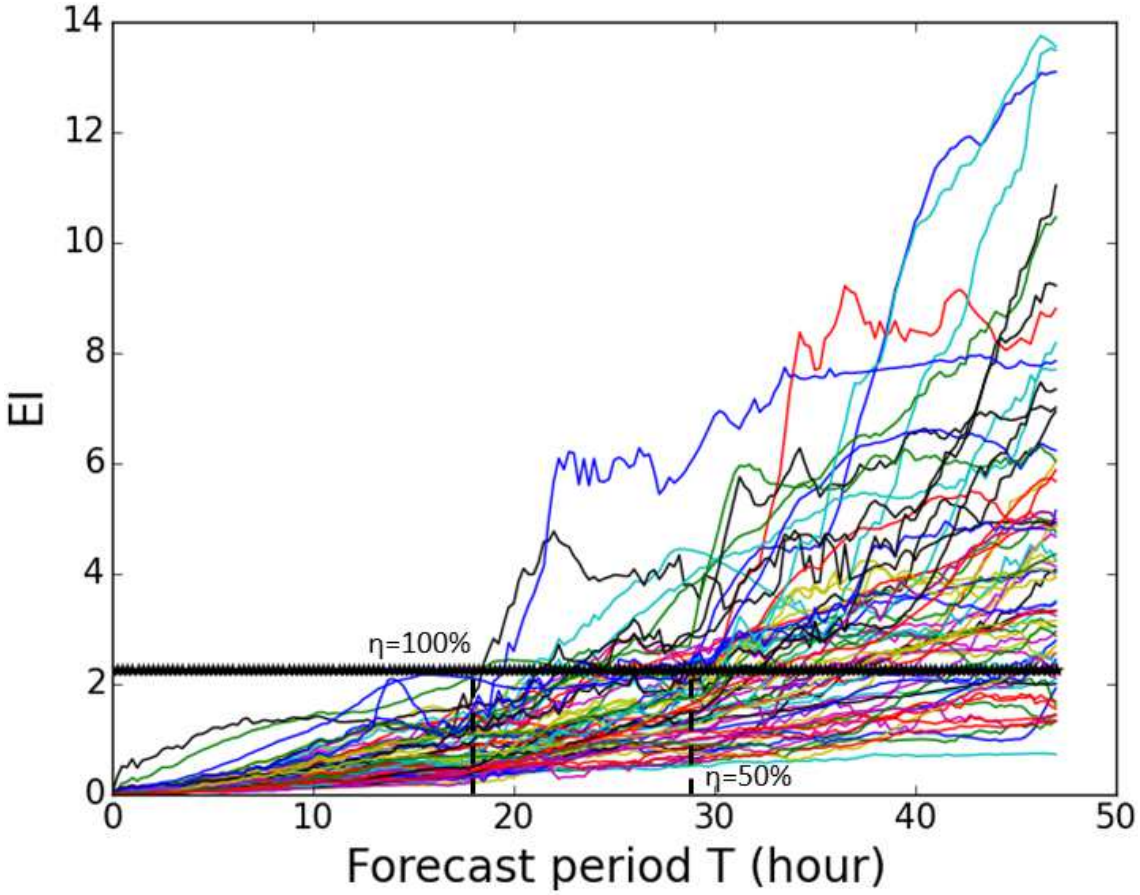


Fig. 18. Overall forecast reliability assessment

4. Conclusion

This study developed new approaches – *FUPM* and *HFRA* to evaluate forecast uncertainty and reliability in operational oil spill modeling system. The *FUPM* combines new developed uncertainty metrics and Monte Carlo simulation to evaluate how uncertainty affects the range of forecasts. A forecast probability map (visually) associated with quantified uncertainty metrics (numerically) can be generated by the *FUPM* method. The *HFRA* provides a new way to integrate reliability assessment directly into of oil spill forecasts. The method can be used to evaluate how far in the future a forecast can be relied upon with confidence, and how rapidly the prediction

quality will likely degrade over time. These new developed approaches can be used to answer the following questions:

(1) *How accurate is an oil spill prediction based on forecast data?*

(2) *At what prediction range (i.e. forecast period) can oil spill forecasts be trusted in certain confidence intervals (e.g. 30%, 60%, and 90%)?*

The demonstration was conducted by a case study in Corpus Christi Bay with an imaginary oil spill at the shipping channel. The Monte Carlo simulation was applied to evaluate hindcast-forecast series errors, with which pseudo-forecast series were generated with a random number generator based on specific error distribution. Multiple set of pseudo-forecast simulations were therefore able to be produced to develop a forecast probability map. Given the associated hindcast simulation, the forecast uncertainties in terms of U_t, U_l, U_a, and U_s were calculated based on four different scenarios. The HyosPy was implemented to provide multiple series of hindcast and forecast for forecast reliability assessment. The HyosPy was continuously running during 7/26/2014 to 4/11/2015 providing data for 66 different forecast periods. The *EI* mechanism was used to evaluate forecast error within one forecast period. The 66 forecast error curves finally provided forecast reliability confidences of any prediction ranges within 48*hr* near [27.812 N, -97.309 W].

The initial motivation and application of this study is developing new methods to quantitatively evaluating modeling uncertainty and assessing forecast reliability in numerical oil spill modeling system, however the newly developed methods are not limited only in oil spill modeling system. In fact, the uncertainty evaluation method provides insights of how explicit forecast error quantification works, which could be benefit for forecast quality assessment of other numerical simulation such as hurricane

forecast and flood forecast. Furthermore, the HyosPy-based forecast reliability method is also extendable to any other numerical simulations (e.g. quantitative finance, weather forecast) that depends on real-time data and is capable of parallel computing.

Acknowledgements

This study is supported by the Research and Development program of the Texas General Land Office Oil Spill Prevention and Response Division under Grant No. 13-439-000-7898, and the BP/The Gulf of Mexico Research Initiative through the Gulf Integrated Spill Research consortium. The authors would like to express grateful thanks to Dharhas Pothina, Solomon Negusse and Dale Crockett at the TWDB; Joseph Zhang at the Virginia Institute of Marine Science; and Chris Barker, Amy MacFadyen, Jasmine Sandhu, and Caitlin O'Connor from NOAA.

References

Abascal, A., Castanedo, S., Medina, R., and Liste, M. (2010). "Analysis of the reliability of a statistical oil spill response model." Marine Pollution Bulletin, 60, 2099–2110.

ASA (1997). "Oilmap for windows technical manual." Report no., Applied Science Associates, Narragansett, RI, USA.

Brown, R. E., Gupta, S., Christie, R. D., Venkata, S. S., and Fletcher, R. (1997). "Distribution system reliability assessment: Momentary interruptions and storms." IEEE Trans. Power Deliv., 12, 1569–1575.

Elliott, A. J. and Jones, B. (2000). "The Need for Operational Forecasting During Oil Spill Response." Marine Pollution Bulletin, 40, 110–121.

Galt, J. A. (1997). “The integration of trajectory models and analysis into spill response information systems.” Spill Science & Technology Bulletin, 4, 123–129.

Galt, J. A. and Payton, D. L. (1999). “Development of Quantitative Methods For Spill Response Planning: A Trajectory Analysis Planner.” Spill Science & Technology Bulletin, 5, 17–28.

Ho, C. K., Hawkins, E., Shaffrey, L., Brocker, J., Hermanson, L., Murphy, J. M., Smith, D. M., and Eade, R. (2013). “Examining reliability of seasonal to decadal sea surface temperature forecasts: The role of ensemble dispersion.” GEOPHYSICAL RESEARCH LETTERS, 40, 57705775.

Hodges, B., Orfila, A., Sayol, J., and Hou, X. (2015). Operational oil spill modeling: from science to engineering applications in the presence of uncertainty. In M. Ehrhardt (Editor), Mathematical Modelling and Numerical Simulation of Oil Pollution Problems. The Reacting Atmosphere, Vol. 2, Springer-Verlag.

Hou, X. and Hodges, B. (2014). “Integrating Google Maps/Earth in an automated oil spill forecast system.” Marine Technology Society Journal, 48(4), 79–85.

Hou, X., Hodges, B. R., Negusse, S., and Barker, C. (2015). “A multi-model Python wrapper for operational oil spill transport forecasts.” Computational Science & Discovery, 8, 014004.

Huang, J. C. (1983). “A review of the state-of-the-art of oil spill fate/behavior models.” Proceeding of the 1983 Oil Spill Conference, Washington, DC, USA, 313–322.

Mackay, D., Paterson, S., and Nadeau, S. (1980). “Calculation of the evaporation rate of volatile liquids.” Proceedings of the National Conference on Control of Hazardous Material Spills, Louisville, KY, USA, 364–368.

NECP (2015). "Wind nowcasts/forecasts derived from eta model, <http://seawater.tamu.edu/tglopu/twdb lc.tar>.

Nelson, J. R., Grubesic, T. H., Sim, L., Rose, K., and Graham, J. (2015). "Approach for assessing coastal vulnerability to oil spills for prevention and readiness using GIS and the Blowout and Spill Occurrence Model." Ocean & Coastal Management, 112, 1–11.

Patterson, J. (2014). "The responder." Report no., Texas General Land Office Oil Spill Prevention and Response Program.

PERSGA/UNEP (2003). "National oil spill contingency plan for sudan." Report No. 6, PERSGA Technical Series.

Price, J. M., Johnson, W. R., Ji, Z., Marshall, C. F., and Rainey, G. B. (2004). "Sensitivity testing for improved efficiency of a statistical oil-spill risk analysis model." Environmental Modelling & Software, 19, 671–679.

Price, J. M., Johnson, W. R., Marshall, C. F., Ji, Z., and Rainey, G. B. (2003). "Overview of the Oil Spill Risk Analysis (OSRA) Model for Environmental Impact Assessment." Spill Science & Technology Bulletin, 8, 529–533.

Price, J. M., Reed, M., Howard, M. K., Johnson, W. R., Ji, Z., Marshall, C. F., Jr., N. L. G., and Rainey, G. B. (2006). "Preliminary assessment of an oil-spill trajectory model using satellite-tracked, oil-spill-simulating drifters." Environmental Modelling & Software, 21, 258–270.

Reed, M. (2000). "A multicomponent 3d oil spill contingency and reponse model." 23rd Arctic and Marine Oilspill Programme (AMOP) Technical Seminar, Ottawa, Canada, 663–680.

Reinaldo, G. and Henry, F. (1999). "Computer Modeling of Oil Spill Trajectories With a High Accuracy Method." Spill Science & Technology Bulletin, 5, 323–330.

Sebastiao, P. and Soares, C. G. (2006). “Uncertainty in predictions of oil spill trajectories in a coastal zone.” Journal of Marine Systems, 63, 257–269.

Sebastiao, P. and Soares, C. G. (2007). “Uncertainty in predictions of oil spill trajectories in open sea.” Ocean Engineering, 34, 576–584.

TCOON (2015). “Texas coastal ocean observation network, <http://lighthouse.tamucc.edu/TCOON/HomePage>.

Tippett, M. K., Delsole, T., and Barnston, A. G. (2014). “Reliabi lity of Regression-C orrected Climate Forecasts.” JOUR NAL OF CLIMATE, 27.

Wang, L. and Zhou, X. (2009). “Organization reliability modeling of ship oil spill emergency management based on bayesian network.” 2009 5th International Conference on Wireless Communications, Beijing, China, Networking and Mobile Computing, 5315–5319.

Ward, G. H. (1997). Processes and Trends of Circulation Within the Corpus Christi Bay National Estuary Program Study Area. Texas Natural Resource Conservation Commission, TX, USA.

Weisheimer, A. and Palmer, T. N. (2014). “On the reliability of seasonal climate forecasts.” J. R. Soc. Interface, 11.

Winkler, J., Duenas-Osorio, L., Stein, R., and Subramaniam, D. (2010). “Performance assessment of topologically diverse power systems subjected to hurricane events.” Reliab. Eng Syst. Saf., 95, 323336.

You, F. and Leyffer, S. (2011). “Mixed-integer dynamic optimization f or oil-spill response planning with integration of a dynamic oil weathering model.” AIChE Journal, 57, 3564.

Zelenke, B., O'Connor, C., Barker, C., Beegle-Krause, C., and (EDs), L. E. (2012). "General NOAA Operational Modeling Environment (GNOME) Technical Documentation." The Office of Response and Restoration's (OR&R) Emergency Response Division under NOAA, Seattle, WA: Emergency Response Division, NOAA.