

Article

Innovative Approaches for Geometric Uncertainty Quantification in an Operational Oil Spill Modeling System

Dongyu Feng ^{*,†} , Paola Passalacqua  and Ben R. Hodges

Center for Water and the Environment, The University of Texas at Austin, Austin, TX 78758, USA

* Correspondence: fengdongyu@utexas.edu

† Current address: Center for Water and the Environment, The University of Texas at Austin, 10100 Burnet Road, Bldg. 119, Austin, TX 78758, USA.

Received: 10 July 2019; Accepted: 3 August 2019; Published: 8 August 2019



Abstract: Reliable and rapid real-time prediction of likely oil transport paths is critical for decision-making from emergency response managers and timely clean-up after a spill. As high-resolution hydrodynamic models are slow, operational oil spill systems generally rely on relatively coarse-grid models to provide quick estimates of the near-future surface-water velocities and oil transport paths. However, the coarse grid resolution introduces model structural errors, which have been called “geometric uncertainty”. Presently, emergency response managers do not have readily-available methods for estimating how geometric uncertainty might affect predictions. This research develops new methods to quantify geometric uncertainty using fine- and coarse-grid models within a lagoonal estuary along the coast of the northern Gulf of Mexico. Using measures of geometric uncertainty, we propose and test a new data-driven uncertainty model along with a multi-model integration approach to quantify this uncertainty in an operational context. The data-driven uncertainty model is developed from a machine learning algorithm that provides *a priori* assessment of the prediction’s confidence degree. The multi-model integration generates ensemble predictions through comparison with limited fine-grid predictions. The two approaches provide explicit information on the expected scale of modeling errors induced by geometric uncertainty in a manner suitable for operational modeling.

Keywords: oil spill; operational system; geometric uncertainty; data-driven model

1. Introduction

Real-time predictions of oil spill transport rely on operational systems that integrate boundary forcing data, hydrodynamic models, oil transport models, and post-processing software. The predictions inevitably have uncertainties ascribed to empirical parameters of the numerical models (e.g., windage factor, oil diffusion coefficient), as well as forecast weather, offshore currents, oil spill conditions and model structural errors [1]. The latter effects are primarily due to the coarse model grid that is required to get rapid results for emergency operations, although the type of numerical model, governing equations, etc. can also contribute. This issue is not unique to oil spill modeling, but also appears wherever fast models are required e.g., for forward integration of climate simulations over centuries. In climate science, quantifying/reducing uncertainty associated with downscaling from a coarse global-scale grid to a fine-scale local grid is an established research area (e.g., [2]), but this area has been neglected in the development of operational oil spill models. Previous oil spill research focused on quantifying the uncertainties associated with empirical parameters or unknown forcing (i.e., forecast uncertainty), typically by applying stochastic methods (e.g., random walk) to account for subgrid processes [3,4] and/or using ensemble averaging as a means to provide greater confidence in a

prediction [5,6]. Thus, previous research has not focused on the uncertainty arising from the structure of hydrodynamic models. Herein, we will focus solely on the uncertainty introduced into an oil spill transport prediction by the geometric approximations inherent in the coarse model grid, which has been defined as “geometric uncertainty” to differentiate it from broader range of uncertainty in model structure [1]. Geometric uncertainty is arguably the most tractable of our uncertainty sources as it is entirely controlled by our modeling approaches and can be directly diagnosed by model-model comparisons. Our goal is to develop methods for quantifying the effects of geometric uncertainty that are computationally efficient and hence practical for operational oil spill modeling. Note that we are not claiming that geometric uncertainty is necessarily the largest or most important uncertainty in an oil spill modeling system. Our point is that this type of uncertainty can be diagnosed with model-model comparisons and hence is readily quantifiable *without* physical drifter experiments or real-world spill observations. In contrast, diagnosing uncertainty through model comparisons to physical observations will integrate *all* sources of uncertainty across a range of parameters and modeling assumptions. As physical experiments are expensive, it is difficult to obtain sufficiently differentiated data to separate different causes of uncertainty and identify remedies that are process-based rather than stochastic. By proposing an approach that isolates geometric uncertainty from other factors, the present work is a step forward that will hopefully provide foundations for future work that quantitatively addresses other uncertainty sources.

The approach we take is based on the proposition that oil spill response managers would like to have some measure of confidence (or its inverse, uncertainty) in a given oil spill transport prediction—but they want a reasonable prediction *now* rather than a better prediction tomorrow. Unfortunately, we do not see any fix for coarse-grid model errors that does not also require some form of fine-grid model for comparison and analyses. Ideally, an oil spill response manager would have high-resolution hydrodynamic models perpetually running over their entire domain of responsibility; however, coarse-resolution models are more practical for continuous operations. In this paper, we show how “on-demand” high-resolution models—which are only started in an area after a spill occurs—can be systematically combined with coarse-grid operational models to provide oil spill transport forecasts with improved confidence. This approach does not address all the sources of uncertainty, but is logically better than relying on coarse-grid models alone.

The principal goal of this research is to develop new approaches to quantify the geometric uncertainty that can be applied operationally to support oil spill decision-making. As such, we focus on how to apply available numerical resources combining coarse-grid and fine-grid models to improve uncertainty quantification. We use a matched set of coarse-grid and fine-grid models to quantitatively analyze errors introduced by the coarse-grid approximations on particle transport models used in oil spill transport prediction. We propose an approach to combine measures of particle separation and diffusion errors into a single confidence measure that represents the geometric uncertainty of the coarse-grid model. We also propose two new approaches to quantify the geometric uncertainty (increasing confidence) that can be readily applied in an operational oil spill system. The first approach uses historical simulations of the coarse and fine models to train a data-driven uncertainty model that provides an insight into the confidence measure. The second approach, a multi-model integration, provides rapid predictions from the coarse-resolution model and, as the slower fine-grid model results become available over time, creates an ensemble prediction that provides a direct expression of this uncertainty. These new tools quantify model performance based on Lagrangian transport behavior rather than Eulerian variables, and thus might have future applications for model designers who seek to develop a model mesh that is optimized for accurate Lagrangian transport.

The first goal of this paper is to establish a statistical relationship between coarse-grid and fine-grid models based on their historical outputs, and apply the data model to future projections, i.e., using the data-driven model-model comparison as a *priori* estimate to assess the geometric uncertainty for real-time predictions. Such a data-driven model, when incorporated into the operational system, can provide emergency response managers guidance or diagnostics about the coarse-grid

model reliability [7]. Thus an operational system can be considered as a joint effort of the numerical models and the data-driven models. This work presents a data-driven model to predict the geometric uncertainty of an operational system and is the first study that applies modern machine learning techniques for addressing uncertainty issues in operational oil spill modeling.

Our second goal is to move beyond the use of a single-model predictions for operational decision-making. We seek methods for real-time generation of model ensemble members with perturbed parameters and/or forcing data that can be built into an operational forecast system. The ensemble results can be represented as a probability cloud of multiple spill trajectories, which allows uncertainty to be expressed directly within the context of an ensemble mean. The challenge addressed herein is in constructing the ensemble that matches the uncertainty problem for oil spills and explaining the probabilistic predictions from the results [8]. These multi-plot approaches to visualizing uncertainty could provide new tools for operational managers, but there remain open questions as to how managers would perceive such data and whether the proposed formats are effective.

A principal limitation of this paper is associated with the state-of-the-art in validating hydrodynamic models. Such models are readily validated for tidal dynamics and salinity transport using sensors emplaced by operational agencies (e.g., [9]). However, what is simply unknown is with what precision must a model capture a limited set of observed tidal elevations, salinity transport, and currents to *also* get the Lagrangian drifter paths correct. Herein, our new methods are developed and tested with the assumption that a fine grid model that has been demonstrated to represent tidal fluxes and salinity transport is also sufficient for Lagrangian drifter transport. This assumption might be someday proven invalid, but the mechanics of the new methods will remain sound, regardless.

In this paper, we start from the background on geometric uncertainty, the data-driven model, and the multi-model integration in Section 2. The uncertainty quantification and the proposed approaches are introduced in Section 3. The corresponding results and the operational applications are discussed in Sections 4 and 5, respectively. The conclusions are provided in Section 6.

2. Background

Geometric uncertainty in modeling is not a new concept, but methods for quantification as part of oil spill operational applications are relatively recent [1,10]. This section describes some measures of uncertainty and discusses two approaches that have been previously used to quantify oil spill uncertainties: data-driven models and multi-model integration, which are based on *a posteriori* analyses of historical data and real-time ensemble predictions, respectively.

The prediction, post-analysis, and uncertainty quantification of an oil spill build upon the near-surface water velocities provided by operational-scale hydrodynamic models [11]. These predicted velocities are less accurate when the scales of velocity variability in the flow are poorly resolved by a coarse model grid, e.g., where complex boundary geometries redirect the flow and create sharp spatial gradients that are smaller than resolvable the grid scale. By definition, for our purposes a “coarse-grid” model is simply one that is unable to resolve small-scale flow features that affect velocities that, in turn, affect oil spill transport. The coarse-grid effects are quantifiable in discrepancies between models at different grid resolutions (e.g., [9,12,13]). The term “geometric uncertainty” was used by [1] to describe the uncertainty arising from use of a grid that is not sufficiently fine to represent the physical processes of interest for the numerical algorithms of a given model. As an example, the eddies formed by starting jet vortices around a narrow ship channel [14] that are unresolved in a coarse-grid model is a predictable form geometric uncertainty that affects prediction of oil spill transport across a bay-coastal shelf interface [15]. A sensitivity study of oil spill uncertainties was conducted in Galveston Bay, Texas (USA) that showed geometric uncertainty can be a significant contributor to uncertainty in estuarine oil spill modeling [10].

Clearly, geometric uncertainty can be reduced (or possibly eliminated) by employing a sufficiently fine grid resolution to obtain a more accurate prediction of velocity fields [12]. Unfortunately, such high-resolution models are not generally practical for operational applications over large coastal areas. Note that although oil spill transport in an estuary is often confined to near-surface regions, capturing the correct surface velocities will typically require three-dimensional (3D) numerical simulations—particularly in the presence of deep ship channels and horizontal salinity gradients [9]. However, the critical issue for modeling the circulation is resolving the horizontal (barotropic) scales of motion. In general, the computational expense for a desired prediction interval increases by a factor of eight when the length scale resolved in the horizontal grid is reduced by 50%. This occurs because the number of grid cells is increased by a factor of four and the model time step must be cut in half. As an example, the fine-grid model in [9] resolves a length scale of about 1/4 of the coarse-grid model of the same estuary but requires a time step of about 1/15 used at the coarser scale. Thus, the fine-grid model nominally requires about $240\times$ the computations of the fine grid model for a given interval of time (the scaling is approximate due to the peculiarities of numerical stability for unstructured grid models). Thus, a coarse-grid model that can run in tens of minutes takes multiple days for the equivalent fine-grid model despite the relatively modest improvement in grid resolution. Furthermore, the fine grid resolution that is “good enough” for oil spill modeling remains an open question as it necessarily involves the integration of error in Lagrangian transport, which is more difficult than matching tidal elevations and currents. We take the view that geometric uncertainty will be with us for the foreseeable future and needs to be quantified for operational managers to understand its effect on predictions and model designers to evaluate the optimum mesh design.

A key challenge to quantifying geometric uncertainty is that it varies across a complex system such as a bay or estuary. A coarse-grid model often has a relatively uniform grid size throughout the domain, but a fine-grid model typically refines the regions with complex geometric features (e.g., deep ship channels) where the effect of small-scale physics is expected to be significant [16]. Thus, there may be broad regions of a bay or estuary where a coarse-grid model introduces relatively little uncertainty and critical choke-points where the uncertainty is dramatically increased. For example, particles mixing within eddies at the tidal inlet can reduce the diffusion area of the predicted spill but this effect is neglected in coarse-grid models [15]—an uncertainty that is irrelevant where such eddies cannot occur. The underestimation of such processes increases the local geometric uncertainty. Additionally, such uncertainty is likely to increase in time/space regions with strong flow dynamics as oil particles will transport and diffuse in a more rapid manner [17].

Across a wide variety of disciplines, data-driven or statistical models have been used for uncertainty quantifications (typically for all sources of uncertainty rather than our specific focus on geometric uncertainty). While the numerical models are derived from the mathematical description of physics, the data models make predictions based on historical datasets by assuming the present state is similar to those in the future. These typically have lower computational costs than the numerical models used in nowcast/forecast simulations. Conventional statistical models have been employed for hazard mapping and risk assessment of oil spills to understand potential spill impacts based on past data [18,19]. A range of data-driven models have been built upon machine learning algorithms and computational intelligence that automatically learn the statistical relationship among variables. These include models in ecology [20], hydrology, [21] and oil spills [22]. For oil spill analyses, the prior data-driven models were driven by *posteriori* analyses of hypothetical spills and past spills to provide real-time predictions from present-state variables (e.g., flow conditions, atmospheric forcing). In operational applications, it is difficult for data-driven models to provide results as accurate as the numerical simulations for time-series predictions [21]. However, we believe these models can be adapted to complement mechanistic models by quantifying the system uncertainty and diagnosing modeling errors.

Another approach to quantifying uncertainty is using predictions from multiple numerical simulations to create ensembles during real-time prediction [23,24]. Such multi-model ensembles have been constructed by using models with different grids, algorithms, physical parameterizations, data assimilation schemes, and/or forecast forcing conditions. The ensemble approach can be used to quantify the prediction uncertainty and determine the relative model performance or bias by displaying predictions with various uncertainty levels. With such tools operational decisions can consider the probabilistic expression of the uncertainty.

3. Methodology

3.1. Study Site and Numerical Model

Evaluating model uncertainty is inherently site-specific, although the tools developed can be applied much more broadly. In this study, we investigate and quantify the geometric uncertainty in a large lagoonal estuary along the northern coast of the Gulf of Mexico (Galveston Bay, Texas, USA). This bay is the shipping access for the Houston metropolitan area and is protected by barrier islands with a deep-dredged ship channel through a narrow entrance. There is another small inlet at the southwest of the bay that may ingest coastal spills or those exited from the main entrance. The flow speed near the main entrance can reach up to 1.5 ms^{-1} during flood/ebb tides, which is an order of magnitude higher than flows in the inner estuary [9]. The greater Port of Houston sees more than 25,000 commercial ship transits and 200,000 barge transits each year [25], with each transit having a risk of collision/grounding and spilling oil—either from cargoes carried for the petrochemical industry or fuel oil used for ship's engines. Fortunately, such spills are relatively rare. In the past two decades the major oils spills caused by ships were 170 m^3 (44,000 gal) in 2001 and 640 m^3 (168,000 gal) in 2014 [26,27]. Despite their rarity, the pollution from these ship events dominates the roughly 275 spills per year in Galveston Bay that are typically less than 4 litres (~ 1 gal) and average about 0.4 m^3 (100 gal) [26], resulting in an accumulation of 60 m^3 (16,000 gal) per year from all other sources.

An unstructured 3D hydrodynamic model, SUNTANS [28], was previously applied to Galveston Bay by [9] using both fine and coarse grid meshes. Their coarse-grid model (GB-C) is time-efficient and can be used operationally for a multi-model ensemble approach (e.g., [24]). Unfortunately, the fine-grid Galveston Bay model (GB-F) is not practical for multi-model ensembles during an emergency. The GB-C has an average horizontal resolution of 400 m (Figure 1a) and allows a model time step of 30 s, as compared to 95 m and 2 s, respectively, in the GB-F. The grid of the GB-F fully refines the Houston Shipping Channel (HSC) that traverses Galveston Bay and the smaller inlet at the southwestern end of the barrier island (Figure 1b). The GB-F horizontal resolution changes smoothly from ~ 100 m along the HSC and the southwest inlet to ~ 2000 m at the open boundary. In contrast the GB-C has roughly uniform resolution of 400 m inside the Galveston and along the coastline with lesser coarsening (compared to GB-F) in the offshore direction. The slightly coarser offshore resolution in GB-F, as well as some marginally coarser areas across Galveston Bay, was an adaption developed and tested by [9]. The GB-F grid optimizes resolution within the ship channel while limiting the overall increase in the number of grid cells by coarsening the grid in areas that did not affect the simulation results. Consequences of the coarser channel resolution on oil spill transport were investigated in [15].

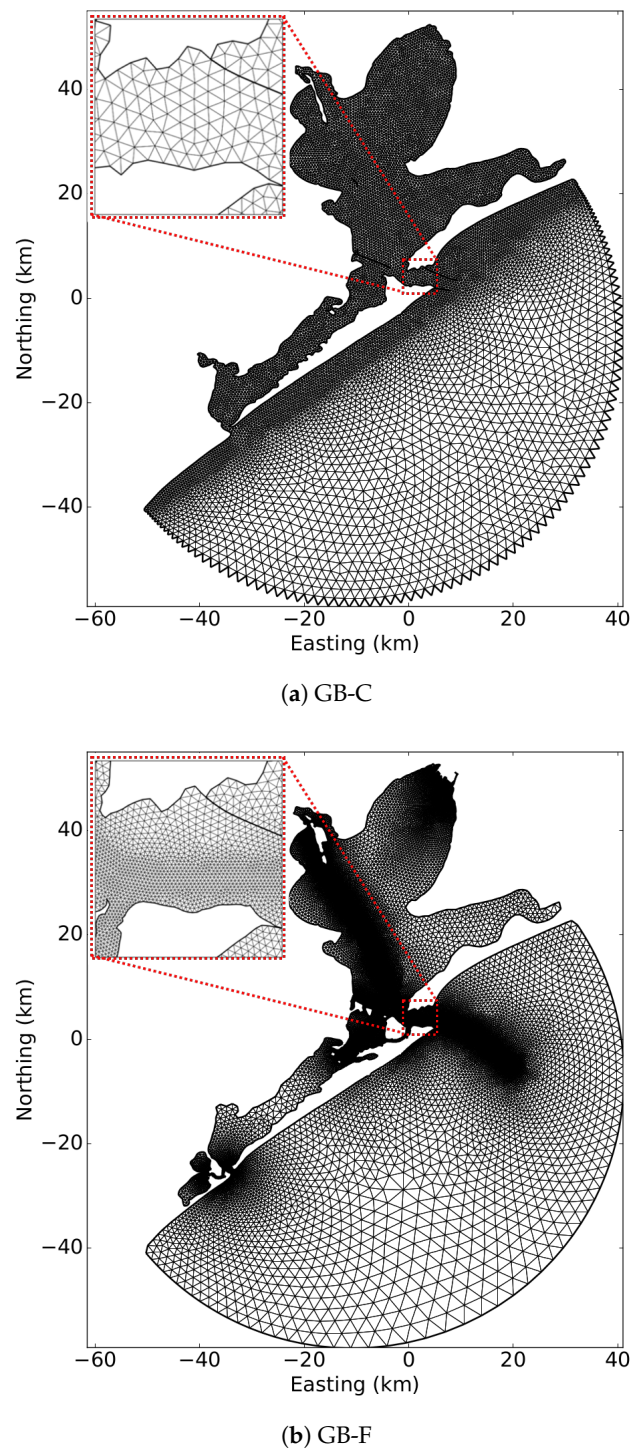


Figure 1. The unstructured model grids of the GB-C and GB-F models and the corresponding magnified views at the inlet.

3.2. Geometric Uncertainty Quantification

The long-term model validation in [9] showed that the GB-F provides a reasonable representation of the physics in Galveston Bay, which allows the model to be used as a “true” baseline to drive Lagrangian particles that represent oil spill motion. We are interested in the errors introduced by the GB-C model relative to the GB-F Lagrangian drifter predictions. In general, particle motions will be subject to two principal forms of error: (i) advective error, where a family of particles moves in the wrong direction or the wrong distance, and (ii) diffusive error, where a family of particles spread out

too quickly (or not quickly enough). Thus, it is useful to define a “separation distance” as a metric of how the advection of a group of particles in GB-C differs from GB-F, and a “diffusion area error” as a metric of how the spread of the group differs. Two statistical metrics are used to quantify the spill advection and diffusion differences between models, respectively: (i) the time-averaged separation distance (\bar{d}) and (ii) the diffusion area error (\bar{e}) [24,29].

To obtain the time-averaged value \bar{d} , it is first convenient to define the instantaneous separation distance at the n th time step (d_n) as:

$$d_n \equiv \frac{1}{N_p} \sum_{p=1}^{N_p} \sqrt{(x_{\text{GB-C},p,n} - x_{\text{GB-F},p,n})^2 + (y_{\text{GB-C},p,n} - y_{\text{GB-F},p,n})^2},$$

$$p \in \{1, 2, \dots, N_p\}; n \in \{1, 2, \dots, N_t\} \quad (1)$$

where x, y are the coordinates of a particle, N_p is the number of particles and N_t is the number of time steps. The instantaneous diffusion area error at the n th time step is:

$$e_n \equiv |A_{\text{GB-C},n} - A_{\text{GB-F},n}|, \quad n \in \{1, 2, \dots, N_t\} \quad (2)$$

where A is the particle diffusion area (i.e., the area of a particle cloud). The time-averaged \bar{d} and \bar{e} values are then computed as $\bar{\phi} = N_t^{-1} \sum \phi_n$ for $\phi \in \{d, e\}$ and $n \in \{1, 2, \dots, N_t\}$. The geometric uncertainty is interpreted by integrating \bar{d} and \bar{e} into a single quantity, i.e., confidence degree (C), which we define as

$$C \equiv \begin{cases} \left(1 - \frac{\bar{d}}{\epsilon_1}\right) \left(1 - \frac{\bar{e}}{\epsilon_2}\right), & (\bar{d} \leq \epsilon_1, \bar{e} \leq \epsilon_2) \\ 0, & (\bar{d} > \epsilon_1) \\ 0, & (\bar{e} > \epsilon_2) \end{cases} \quad (3)$$

where ϵ_1 and ϵ_2 are the tolerance thresholds for \bar{d} and \bar{e} . Herein these are set at $\epsilon_1 = 6$ km and $\epsilon_2 = 16$ km², respectively. These values were determined such that at least 95% of the data were within the thresholds. When \bar{d} or \bar{e} exceed their thresholds the confidence degree is identically zero. Conversely, when there is no difference between the fine and coarse solutions, the confidence is unity. Thus, C as a confidence measure (based solely upon simulations at different model grid resolutions) is inversely related to geometric uncertainty.

Equation (3) provides C of a given spill at a given location for a desired time interval. However, we are interested in quantifying how confidence varies in both space (across the bay) and time (over a tidal period), which can be analyzed through an ensemble of spatially-distributed spill experiments. The tidal scale is likely the shortest time-scale forcing that affects the circulation in the bay, and thus provides the time scale for analyses. The confidence degree (C) in the computational domain can be formulated into a function of space (x_i) and the corresponding tidal elevation (η_i), where the domain $D \subset R^{N_s}$ is discretized into N_s discrete cells $x_i \in D$ ($i = 1, 2, \dots, N_s$). The function of the confidence degree (C) yields:

$$C(x_i, \eta_i) = \left(1 - \frac{\bar{d}(x_i, \eta_i)}{\epsilon_1}\right) \left(1 - \frac{\bar{e}(x_i, \eta_i)}{\epsilon_2}\right). \quad (4)$$

The function C can be obtained using a data-driven model, which is similar to deriving a statistical regression model for an equation, but the data-driven models learn the statistical relationship themselves and do not have a closed-form solution, as discussed in the following section.

3.3. A Data-Driven Uncertainty Model

This section derives a new data-driven uncertainty model based on model-model comparisons, including data collection and a new clustering algorithm with model fitting. The data-driven uncertainty model developed herein corresponds to the function in Equation (4) applied using simulated historical datasets of the confidence degree. That is, our model integrates the information from long-term simulations of likely spills over a range of hydrodynamic conditions into a comprehensive function for future projections. During a real-time simulation, one can obtain the response variable C directly from the data-driven model, as an estimate of the confidence level in a GB-C prediction of a single spill event, given the predictor variables, i.e., the instantaneous spill location x_i and the corresponding tidal elevation η_i predicted by the GB-C.

The simulation data to develop the data-driven model is obtained from matched sets of numerical Lagrangian experiments initiated across the estuary and over multiple tidal cycles with both fine-grid and coarse-grid models. A single Lagrangian experiment consists of two sets of particle trajectories provided by the GB-C and GB-F models, so that we can calculate \vec{d} , \vec{e} and the corresponding C for each experiment. We used the GNOME model as the Lagrangian transport model [3]. For simplicity in this demonstration, direct atmospheric forcing on the particles (windage) and oil weathering were neglected. We characterized Galveston Bay with 1000 uniformly distributed grid cells ($N_S = 1000$), which are illustrated and discussed in Section 4.2. For the Lagrangian experiments we released a set of 120 oil particles (uniformly allocated over a circular area with a 20-m radius) each hour at each cell center. The model training period was from April to June 2009, which was selected as it did not include any extreme events (hurricanes or large-scale hydrologic inflows). Each Lagrangian experiment ran for 10 h. The tidal elevation η_i ($-0.4 \sim 0.4$ m) corresponding to position x_i was the value at the release time of each experiment.

A critical issue associated with this approach is the high spatial dimensionality of the C function of Equation (4). Running the experiments initiated at every cell of the domain is computationally expensive—especially for fine-grid simulations. For example, each hourly release requires 2000 experiments for both coarse-grid and fine-grid models, resulting in over 2 million Lagrangian experiments for the three-month training period, which typically takes ~ 2.5 months on our desktop computer using 8 Intel Xeon E5 CPUs. To address this issue, we developed a new geospatial clustering algorithm to reduce the spatial dimension of C and the data collection time. Based on a small set of randomly selected simulations ($\sim 2 \times 10^5$ simulations), two neighboring cells are merged into a single cell if the C values are highly correlated—i.e., correlation coefficient $r > 0.975$. Using this approach the number of the discretized cells (N_S) is reduced to N'_S , where $N'_S \leq N_S$. Following [30], this geospatial clustering only applies when two components are spatially contiguous and is thus different from dimensional-reduction methods that do not consider spatial relationships between components, e.g., principal component analysis [31].

Our data-driven model uses the gradient boosting regression tree (GBRT) algorithm, a supervised machine-learning approach [32]. The GBRT model can be considered as a combined regression tree model with the boosting operating stagewise for additive simple regression models. The GBRT is effective for handling large datasets and has a few advantages over a single complex decision tree: (i) the model fits in nonlinear relationships of multiple variables; (ii) the gradient boosting is robust to model overfitting; (iii) the GBRT works for both classification problems with discrete predictor variables (e.g., space x) and regression problems with continuous variables (e.g., tidal elevation η). Although the GBRT model is complex, we found it appropriate for uncertainty quantifications of the oil spill modeling as the uncertainty can be related to several parameters. This model was previously used by [33] to identify the probability of shoreline oiling based on multiple oil categories.

We fitted the GBRT model to the simulation dataset and solved for the C function using the *scikit-learn* software package [34]. Conceptually, C is the response variable and the two predictor variables are x_i and η_i as defined in Equation (4). We applied 2000 estimators, a tree complexity of 4, a learning rate of 0.02, and least squares loss functions. These parameters were determined by trial

and error to provide satisfactory model performance. The selection of parameters to fit and evaluate a GBRT model is discussed in [20]. The model is trained on 90% of the dataset and tested on the remainder. We evaluate the GBRT model with the test data using two bulk metrics: the coefficient of determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=0}^{n_s-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_s-1} (y_i - \bar{y})^2}, \quad (5)$$

and the mean squared error (MSE)

$$MSE(y, \hat{y}) = \frac{1}{n_s} \sum_{i=0}^{n_s-1} (y_i - \hat{y}_i)^2, \quad (6)$$

where $\bar{y} = \sum_{i=0}^{n_s-1} y_i / n_s$, y_i is the true value of the i th sample, \hat{y} is the model-predicted value and n_s is the test sample size. These metrics provide a measure of the GBRT model performance on the test samples.

Finally, we test the model by numerical simulation of hypothetical spills driven by the GB-F hydrodynamic model as “observations”. We compare the effective confidence (E) that can be directly estimated from the GBRT model (prediction) compared to that calculated from the simulation (observation). The effective confidence (E) provides the additive confidence of a spill prediction that is reduced during the transport time. For example, E at the second time step of the prediction is always the product of C obtained at the first two steps, and E at the third time step is the product of C at the first three steps, etc. The effective confidence (E) explains how the confidence is reduced over time in a probabilistic nature. The value of E at time $t = n\Delta t$ is defined as:

$$E(t) \equiv \prod_{j=1}^{n(t)} C(x_j, \eta_j) \quad : \quad n = \left\lfloor \frac{t}{\Delta t} \right\rfloor \quad (7)$$

where j is the time step. The hypothetical spills were initiated at representative locations across Galveston Bay every 6 h from 2009 July 17 00:00 to July 19 00:00. This time period was excluded from the training period. Each spill was tracked for 50 h ($t = 50$) with Δt of 10 h.

3.4. A Multi-Model Integration

In this work we demonstrate that geometric uncertainty can be estimated in an operational setting by a multi-model system that uses combinations of coarse-grid and fine-grid models for ensemble predictions. Forecast ensembles (of varying types) are relatively common in other modeling areas to visualize prediction uncertainty, e.g., hurricane tracks [35], but relatively rare in oil spill modeling. We recently proposed an oil spill ensemble approach that generates new ensemble members by continuously replacing prior “poor” forecast data with new “good” hindcast data and analyzing the ensemble of simulations [24]. In this approach, at the initial time when a spill occurs ($t = 0$), there is only forecast data available for the full forecast simulation of period T_s . The hindcast data arrives every few hours t' where $t' < T_s$. A new simulation from $t = 0$ is started with the model initially running with hindcast data in $[0, t')$ and then forecast data over $[t', T_s]$. This pattern continues until the final simulation (at the end of an event) consists of full hindcast data. At any time during an event, the latest results can be represented as ensemble of trajectories that allow an operational manager to understand how the forecasts have been changing. In the present work, we adapt this approach to quantifying geometric uncertainty where the fine-grid model that runs slowly from $t = 0$ provides “good” data at the time interval of t' .

The overall ensemble mechanism of the multi-model integration over the time of an event is illustrated in Figure 2. When a spill occurs, both coarse-grid and fine-grid hydrodynamic models

are started at the initial time ($t = 0$) to provide predictions out to some time horizon (e.g., 2 days). The coarse-grid model runs M times faster than the fine-grid model so the coarse-grid prediction will be available first after its simulation time T . This coarse-grid result is designated the first model, $F(0)$. At $t = (N + M - 1)T/N$, the fine-grid model finishes $1/N$ of the total simulation. We now use the fine-grid modeled near-surface velocity to run particle tracking simulation for the initial $1/N$ fraction of data, and use the coarse-grid modeled velocity to run the remained $(N - 1)/N$ fraction with the initial particles at the last time step of the fine-grid simulation. This hybrid coarse-fine result is designated the second model, $F(1)$. This pattern continues until the fine-grid model completes the entire simulation at $t = MT$, resulting in the last model of $F(N)$, which is in a full fine-grid mode. This multi-model integration allows a new ensemble member $F(i)$ ($i = 1, \dots, N$) to be generated on a time interval of $(M - 1)T/N$ with an increased portion of the fine-grid prediction. The more fine-grid predictions are used, the less geometric uncertainty is in the operational system.

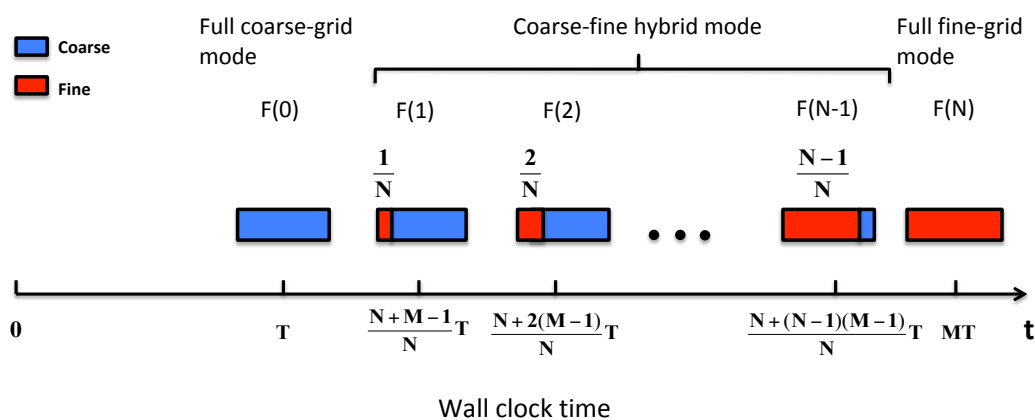


Figure 2. The framework of the multi-model integration. The blue and red squares represent the coarse-grid and fine-grid simulations, respectively.

This multi-model approach addresses an operational concern of emergency response managers, who cannot afford to wait for a fine-grid prediction before deploying assets. For example, our fine-grid GB-F model runs merely 5~10 times faster than real time on 24 Intel Xeon E5 CPUs with 2.6 GHz, so a 3-day forecast takes 7 h. In contrast, the multi-model approach that starts with a GB-C coarse grid can provide an initial 3-day forecast in less than 0.2 h. The important point is that the multi-model integration generates multiple ensemble members moving forward, providing real-time insight into reliability of predictions by comparing the available trajectories with various uncertainties. In general, the ensemble allows a direct expression of the uncertainty and an increased exploitation of the high-performance numerical models in the system. This approach can be generalized to any system with more than one model available.

3.5. Geometric Uncertainty Quantification in the Operational System

The data-driven uncertainty model and the multi-model integration described above have been integrated into the HyosPy system [36] to trial how these models might be applied operationally for geometric uncertainty quantification. The HyosPy oil spill operational system was developed for Galveston Bay and the adjacent Gulf coast [36]. HyosPy is a multi-model Python wrapper, in which forcing data is automatically sourced from online servers, parsed as input files, and controls running of hydrodynamics and oil spill models. The end users are only required to provide the start time, duration, and initial spill conditions to initiate a sequence of hindcast/forecast simulations for the spill incident. HyosPy provides simple integration through an Application Programming Interface (API) for testing new models and approaches.

We created new APIs for the data-driven uncertainty model and the multi-model integration. The application of the uncertainty model conceptually requires two steps, as illustrated in Figure 3: (i) run HyosPy to produce spill tracks, and (ii) assess the GB-C prediction using the uncertainty model. The multi-model integration is employed as a replacement of the coarse-grid hydrodynamic model to provide ensemble predictions of velocity currents for the oil transport model, resulting in ensemble trajectories. Both approaches provide an assessment of the existing modeling errors with respect to the geometric uncertainty, or advocate the use of the fine-grid GB-F model when necessary.

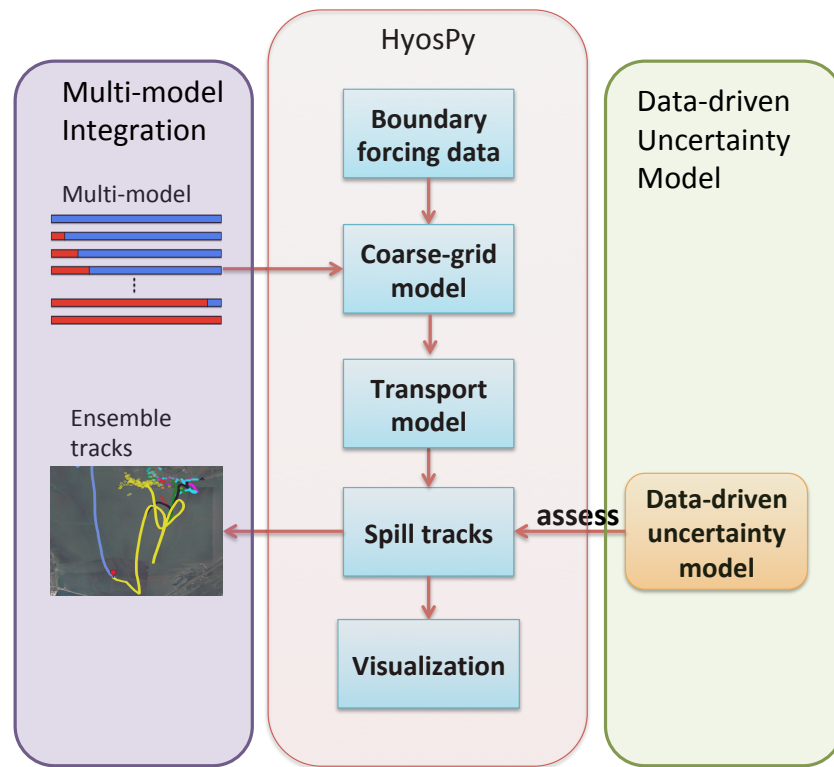


Figure 3. The HyosPy system with the data-driven uncertainty model and the multi-model integration.

4. Results and Discussion

4.1. Evaluation of Geometric Uncertainty

The raw measures of geometric uncertainty are the time-averaged separation distance (\bar{d}), the diffusion error (\bar{e}) and the confidence degree (C) described in Section 3. Although each of these measures are averaged over time, they are defined for hypothetical spills that occur at discrete locations in time and space, and thus these have different spatial distributions associated with the time/space location of an individual hypothetical spill. Figures 4–6 show spatial distributions of these measures during the tested period for Galveston Bay from maximum flood flow to maximum ebb flow, with a moderate flood and an ebb tide near slack water for comparison. In these figures the maximum flood tide is $2.4 \times 10^4 \text{ m}^3\text{s}^{-1}$, the moderate flood tide $1.6 \times 10^4 \text{ m}^3\text{s}^{-1}$, the weak ebb tide (near slack water) $-0.67 \times 10^4 \text{ m}^3\text{s}^{-1}$, and the maximum ebb tide $-2.4 \times 10^4 \text{ m}^3\text{s}^{-1}$, where the negative sign indicates an outflow.

Figure 4 shows the spatial distribution of \bar{d} is greatest at the channel entrances but remains significant (greater than 3 km) through much of the mid-bay. Of particular interest is that the differences are relatively strong throughout the initial 20 km of the Houston Ship Channel. The spatial distribution of \bar{d} also varied depending on the flow conditions such that \bar{d} was higher during high flow conditions, Figure 4a,b,d, than during the lower flows, Figure 4c.

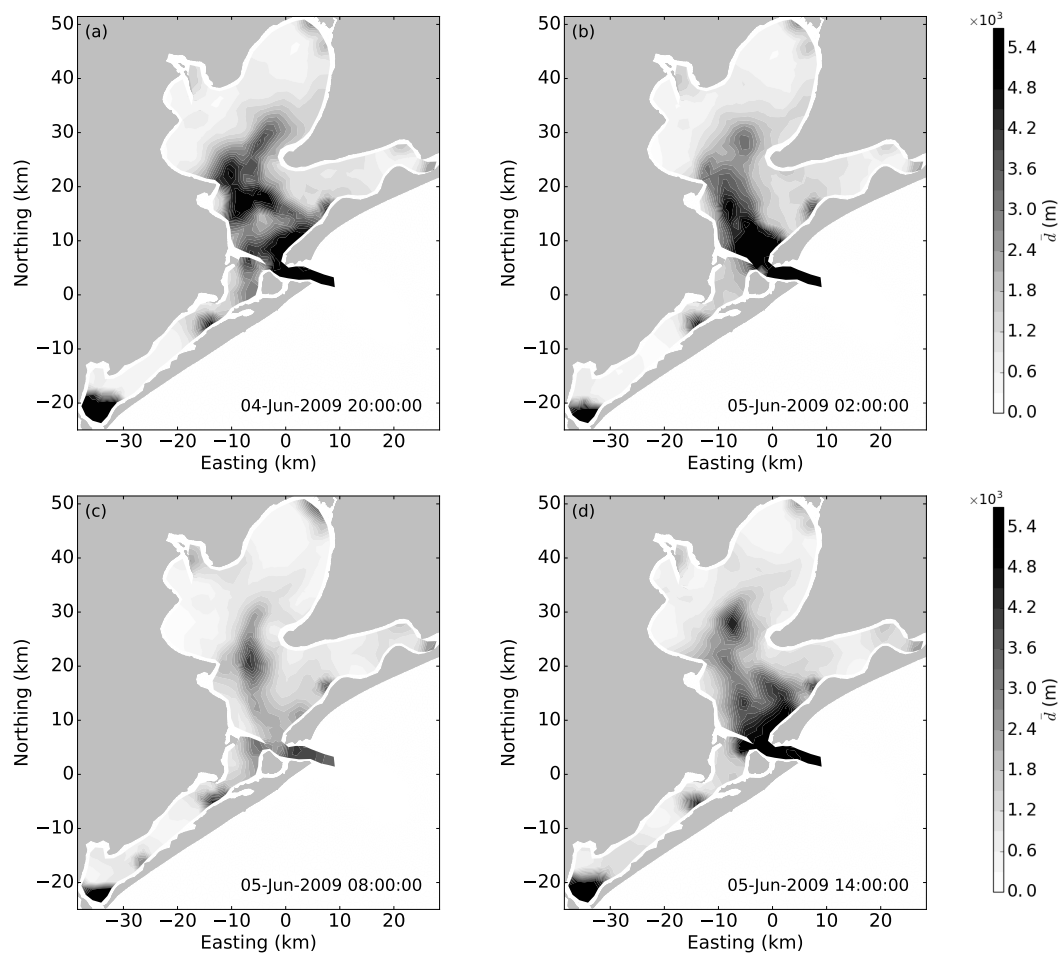


Figure 4. Spatial distribution of separation distance (\bar{d}) at four flow conditions: (a) maximum flood tide, (b) moderate flood tide, (c) weak ebb tide (near slack water), and (d) maximum ebb tide.

In contrast, Figure 5 shows that the geometric uncertainty associated with dispersion of the oil (\bar{e}) is relatively small across much of the bay. Similar to Figure 4, the largest discrepancies are located near the main entrance and the southwest inlet. For comparison, the maximum dispersive error of $6 \times 10^6 \text{ m}^2$ can be thought of as a length scale error of 2.4 km, which is half the maximum error associated with (\bar{d}). These figures illustrate that \bar{e} is strongly affected by the tidal fluxes, with its largest values at peak flows (Figure 5a) and the minimum dispersion during the low flow condition (Figure 5c).

The confidence degree C integrates the information of \bar{d} and \bar{e} in an inverse measure so that a large value indicates high confidence. We see that the C measure shows relatively poor confidence in coarse-grid simulations near the two inlets ($C < 40\%$), unless a spill occurs near low tide. However, inner Galveston Bay generally has much higher confidence, $C > 80\%$. Unfortunately, areas of somewhat reduced confidence ($40\% < C < 80\%$) are coincident with the middle reaches of the Houston Ship Channel. As expected, C is similar to \bar{d} and \bar{e} in that its distribution varies with flow conditions, resulting in greater areas of the poor confidence during high flows in Figure 6a,b,d than those during low flows in Figure 6c.

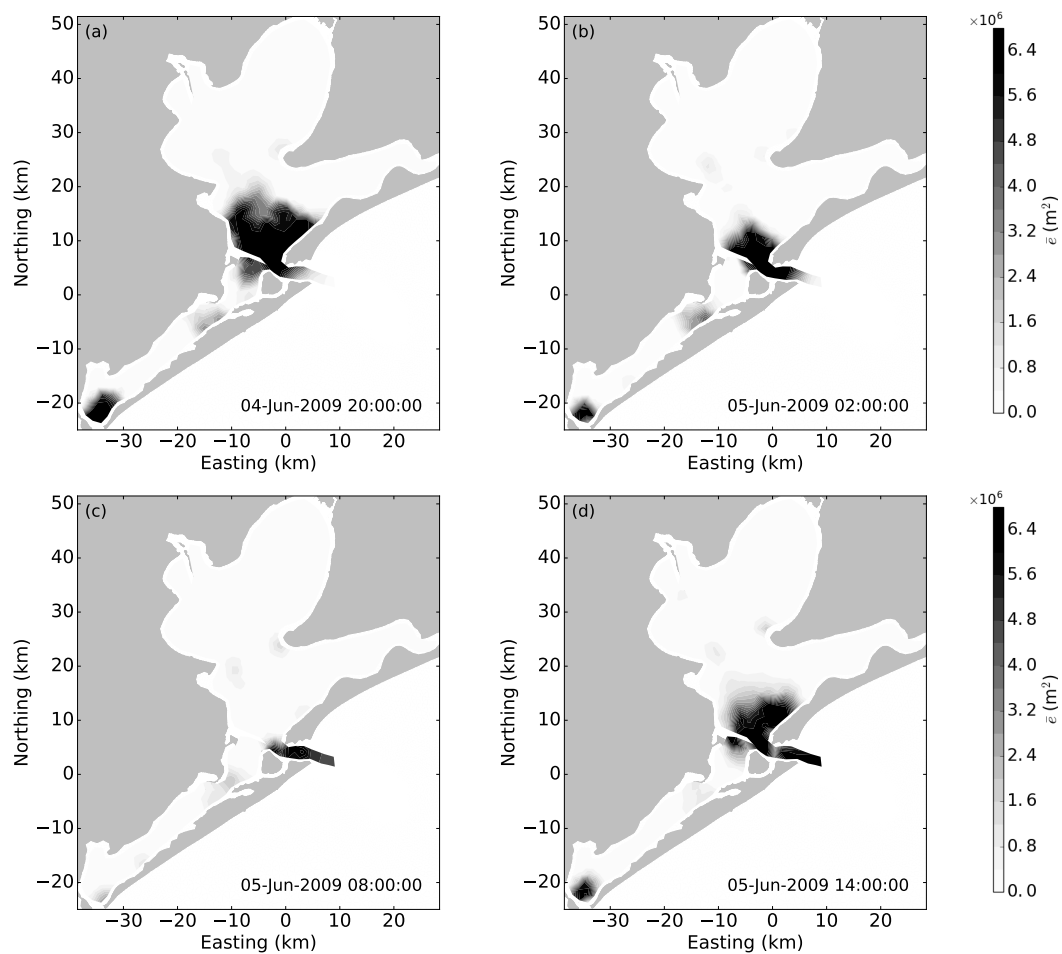


Figure 5. Spatial distribution of the diffusion area error ($\bar{\epsilon}$) at four flow conditions: (a) maximum flood tide, (b) moderate flood tide, (c) weak ebb tide (near slack water), and (d) maximum ebb tide.

The above results indicate the geometric uncertainty varies significantly across Galveston Bay depending on both grid resolution effects and flow dynamics. The highest uncertainty occurs near the bay entrance and along the ship channel during high flow conditions. Although Galveston Bay is a micro-tidal estuary whose interior circulation can be driven by wind [9], the inflow/outflow velocities at the channel are of order 1 ms^{-1} , which is an order of magnitude larger than velocities in the inner estuary. This channel inlet includes the most highly-resolved sections of the GB-F where the length scale of the typical grid cell is one-fourth that of the GB-C. The smaller grid allows the GB-F to resolve smaller-scale flow features, such as vortices at the channel entrance and around headlands that can create eddy effects on particles that are not directly represented in the coarser GB-C model [15].

The quantification of the geometric uncertainty reveals the weakness of the coarse-grid model. The potential modeling error varies system-wide and is subject to tidal oscillations at the timescale of hours. The variability of the geometric uncertainty should be characterized in operational oil spill modeling unless fine-grid models are applied. Additionally, the geometric uncertainty helps improve our physical understanding of the hydrodynamic modeling in the system and provides an alternative criteria that could be used for unstructured grid development. In an estuary system, it is of particular importance to refine grids in sensitive regions with high geometric uncertainties that can cause modeling errors in Lagrangian transport.

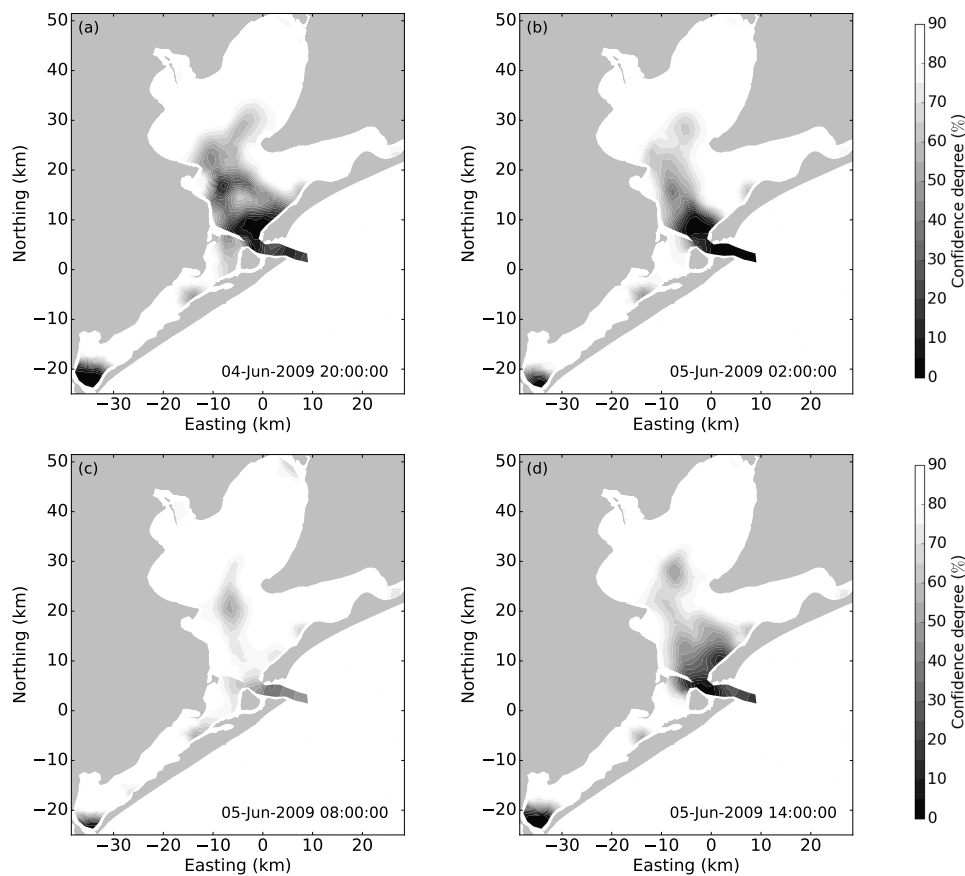


Figure 6. The spatial distribution of confidence (C) at four flow conditions: (a) maximum flood tide, (b) moderate flood tide, (c) weak ebb tide (near slack water), and (d) maximum ebb tide. Note the color scale is inverted so that low confidence is dark grey, which is consistent with large separation and diffusion errors in Figures 4 and 5.

4.2. The Uncertainty Model

The data-driven uncertainty model of Section 3.3 was developed for the geometric uncertainty quantification that provides a *priori* assessment for the operational model. The geospatial data clustering was performed on the grid cells in Galveston Bay (Figure 7a) based on the sampled Lagrangian experiments. At least two grid cells were merged at most regions, while more clustering occurred in the inner Galveston Bay where C was high and the geometric uncertainty had smaller spatial variability. Clustering reduced the overall spatial data from 1000 triangular cells to 382 polygons (Figure 7b), which similarly reduced the computational requirements for analyses to 1/3 of that required for the non-clustered system.

The GBRT model was trained with 90% of the data samples collected from the numerical Lagrangian experiments, which were made on the clustered cell centers through the three-months training period. The model was tested on 10% of the data samples, demonstrating a reasonably good predictive performance with a coefficient of determination (R^2) of 0.82 and a mean squared error (MSE) of 97%. The latter indicates that the predicted C is less than 10% away from the true data. Figure 8 shows the spatial distributions of the predicted the confidence degree (C) for the GBRT model for the same flux ranges described in Section 4.1 and comparable to Figure 6. The model is able to reproduce the variability of C in terms of both spatial distribution and dependency on flow conditions. The predicted C is as high as 90% with little variability in the inner estuary, and varies significantly with η from 10% to 80% towards the bay mouth as a result of the locally dominant tidal forcing.

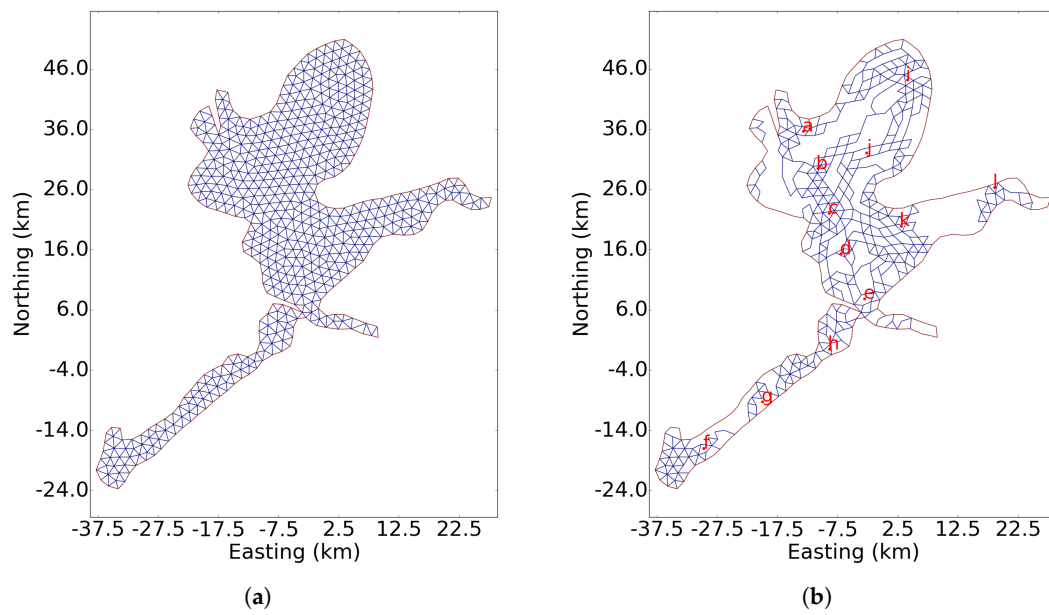


Figure 7. (a) The uniformly distributed cells and (b) the result after clustering.

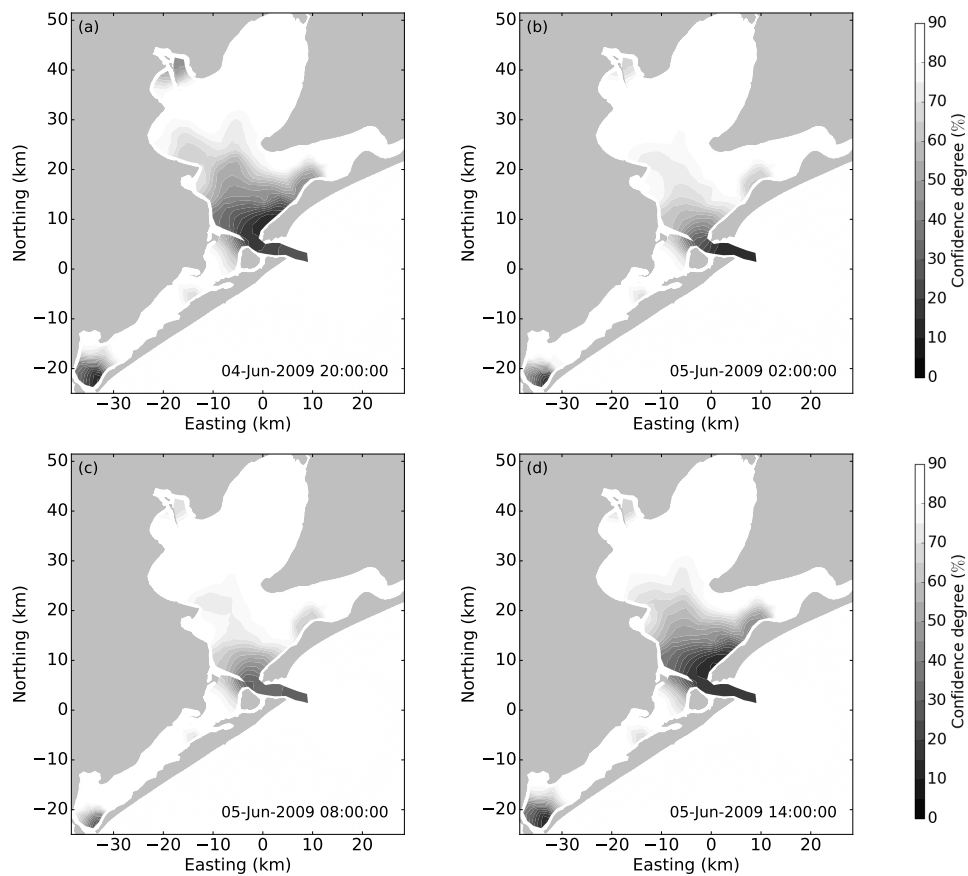


Figure 8. The predicted confidence degree (C) from the GBRT model at four flow conditions: (a) maximum flood tide, (b) moderate flood tide, (c) weak ebb tide (near slack water), and (d) maximum ebb tide.

4.3. Uncertainty Model Evaluation

The GBRT model was also evaluated with the effective confidence degree (E) of Equation (7), which represents the integrated effect of confidence (C) over time for the coarse-grid model transport predictions. Figure 9 shows the relationship between the predicted E based on the GBRT model and the observed value based on numerical transport simulations. Where observed E from the coarse-grid model covers a fairly wide range, i.e., frames (b), (c), (d), (h), and (j) in Figure 9—the GBRT model shows a fairly reasonable trend in the effective confidence behavior, although the prediction spread at any given observed value indicates that the modeled E should be used with caution. Frame (e), corresponding to the region just inside the main ship channel, shows substantial scatter and low observed confidence, which indicates that the GBRT uncertainty model is likely of little use in this dynamic area. The behavior is quite different in frames (a), (f), (g), (i), (k), and (l), around the edges of the domain where the GBRT tends to predict lower E than actually observed. That is, the GBRT in these cases predicts the effective confidence is equal to or less than that observed with the Lagrangian transport simulations, which can be considered a “minimum regret” solution that is generally desirable in uncertainty analysis [3]. A detailed review of the data (not shown) indicates the systemic underestimation of E within the coarse-grid model is primarily due to particles beaching—which causes their stagnation. This behavior is not adequately handled by the uncertainty model. The statistical model prediction is better at the stations in the inner estuaries (e.g., stations a, f and i) with small data variance than those subject to stronger tidal flows (e.g., stations d, e and k). At station e, particles exiting the bay yielded zero E that was predicted as small positive values by the model.

Our results imply that the uncertainty model may be useful in estimating the reliability of the coarse-grid prediction in real time without having to simultaneously compute the fine-grid solution. However, it is clear that further research is needed to reduce the variance between the GBRT model and the Lagrangian transport test simulations. We believe this approach to applying machine learning algorithms to interpret comprehensive *a priori* modeled data sets is a prototype that can be used to improve our understanding of real-time forecasts. This area of application for machine learning is relatively recent, and hence our model is relatively limited. A more comprehensive model would require a larger variety of data (e.g., direct inclusion of observational data such as overflight or satellite oil spills observations [22]), a much longer time interval for data generation, the inclusion of multiple layers or predictor variables (e.g., the ratio of spatial resolutions between different model grids), and a wider variety of detailed physics (e.g., particles beaching, river inflow boundary forcing). In short, the data-driven model provides an approach for extending the analysis of geometric uncertainty to include other forms of uncertainty and more detailed evaluation of effects of fine/coarse grid ratios when sufficient data are available. Note that our uncertainty model is predicated on the assumption that the GB-F model is adequate, which has been demonstrated for salinity distributions and tidal elevations in [9]; however, error in Lagrangian drifter simulations is integrative over time and there does not appear to be a method for directly assessing modeled drifter accuracy without an extensive field study deploying physical drifters.

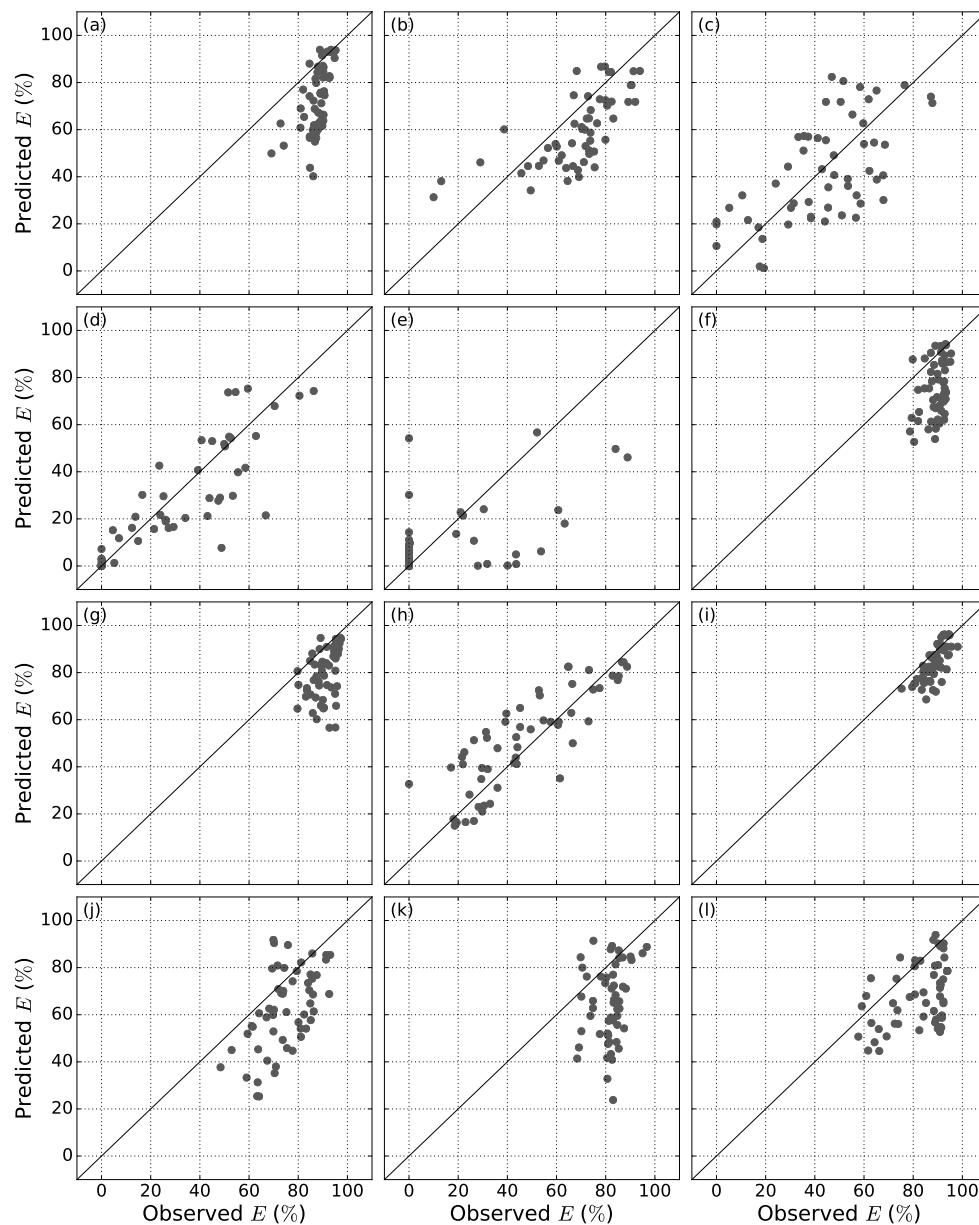


Figure 9. The observations of the effective confidence (E) against its predictions at the 12 selected stations across the bay (noted in Figure 7b). The solid line is the diagonal line.

5. Operational Application to HysoPy

We use two case studies to demonstrate application of the data-driven uncertainty model and the multi-model integration in the HyosPy operational system. Both cases indicate that the proposed approaches were able to quantify the geometric uncertainty, which is critical for evaluating confidence in operational spill predictions. The maps with multiple trajectories or a probability cloud were visualized using new Google Map tools. Please contact the corresponding author for the Google map files.

Hypothetical spills are released at two locations with the spill information provided in Table 1. The spill in Case 1 is initiated along the Houston Ship Channel, analogous to a midnight spill caused by a ship collision, and the spill in Case 2 is released in the upper Galveston Bay to provide a contrast in behaviors. Both spills were driven by the near-surface velocities of GB-C and GB-F models along with winds from the North American Regional Reanalysis (NARR) dataset [37] with a windage of

0.01. The multi-model integration was applied by setting $N = 5$, resulting in 6 different trajectories, which were specified as coarse-grid, 20% fine-grid, 40% fine-grid, 60% fine-grid, 80% fine-grid and fine-grid predictions. The corresponding simulation time for each ensemble member was 0.1, 1.08, 2.06, 3.04, 4.02 and 5 h, respectively.

Table 1. The hypothetical spill information for the two cases.

Case	Initial Location	Release Time	Transport Time (h)
1	29.40° N, 94.83° W	00:00 16 July 2009	54
2	29.63° N, 94.78° W	00:00 17 May 2009	54

It can be seen that the trajectory discrepancies (the separation distance and the diffusion error) are significant between the models, implying a critical geometric uncertainty in the GB-C prediction (Figure 10). In Case 1, the GB-C trajectory extended towards the inner Galveston Bay, which was misleading given the particles beached on the eastern coastline as predicted by the GB-F. In Case 2, the GB-F predicted particles presented a higher probability of beaching on the Texas City Diike, whereas those driven by the GB-C advected towards the bay mouth.

The geometric uncertainty in the GB-C prediction was reflected by the time series of the confidence degree (C) and the effective confidence (E) estimated from the data-driven uncertainty model every 10 h. In Case 1, the confidence degree (C) was low ($\sim 20\%$) at the initial time step and then increased to a high level ($> 80\%$), whereas the integrative error of E never recovered from the initial poor C (Figure 10a). These behaviors imply that the GB-C prediction is not reliable and has non-negligible geometric uncertainties at the beginning, leading to aggregated errors through the simulation period, which explained the overall poor model performance. In Case 2, C fell between 85% and 60%, while E slowly downgraded over time (Figure 10b). The effective confidence suggests that the GB-C is effective for the first 10 h, after which fine-grid simulations should be applied to obtain a more accurate prediction.

The confidence measure emphasizes the need for real-time corrections based on the best available observations—i.e., when E gets low, the GB-C predictions have to be treated as inherently suspect unless re-initialized with new observations of the spill. That is, imagine that in Case 1 we received new data at 10 h as to the general location of this spill, the oil spill simulation could then be re-initialized at the new location with a larger spread. When coupled with the uncertainty model, the operational oil spill system (e.g., HyosPy) is capable of identifying the potential errors and reporting the intermediate model performance to the emergency response managers for decision-making.

The multi-model integration generates multiple trajectories (Figure 11). The prediction generally improved when the GB-F had run a longer time and was employed for a larger portion of the oil spill simulation. The coarse-grid prediction accuracy was significantly increased by the 20% fine-grid prediction, in which the particles advected in a similar direction as the pure GB-F prediction (Figure 11a). When the GB-F was employed for over 40% of the simulation, the particles beached at almost the same location. The predicted trajectories were similar among 40% fine-grid prediction to pure fine-grid prediction, while using the GB-F model for the first 20% did not improve much of the prediction (Figure 11b).

In a real-world application, the prediction ensemble would grow over time and provide important information for making decisions within a limited response time. With the multi-model integration, we can identify the GB-C modeled errors at early stages without having to wait for the pure GB-F prediction to tell where and when the geometric uncertainty was high and the GB-C prediction is no longer adequate. For these experiments, the geometric uncertainty dominated at the beginning in Case 1 and after 20%~40% of the simulation in Case 2.

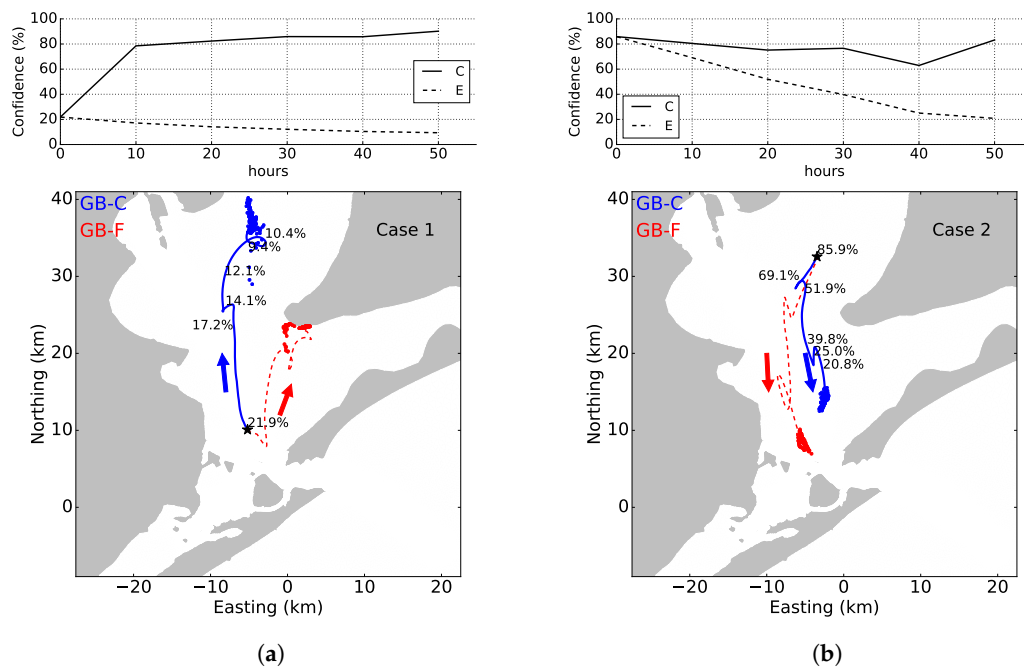


Figure 10. The spill trajectories predicted by GB-C and GB-F models, and the estimated confidence degree (C) and the effective confidence (E) of the uncertainty model for (a) Case 1 and (b) Case 2. The GB-C trajectory is annotated with E. The star is the initial spill location and the arrows represent the general direction of the spill movement.

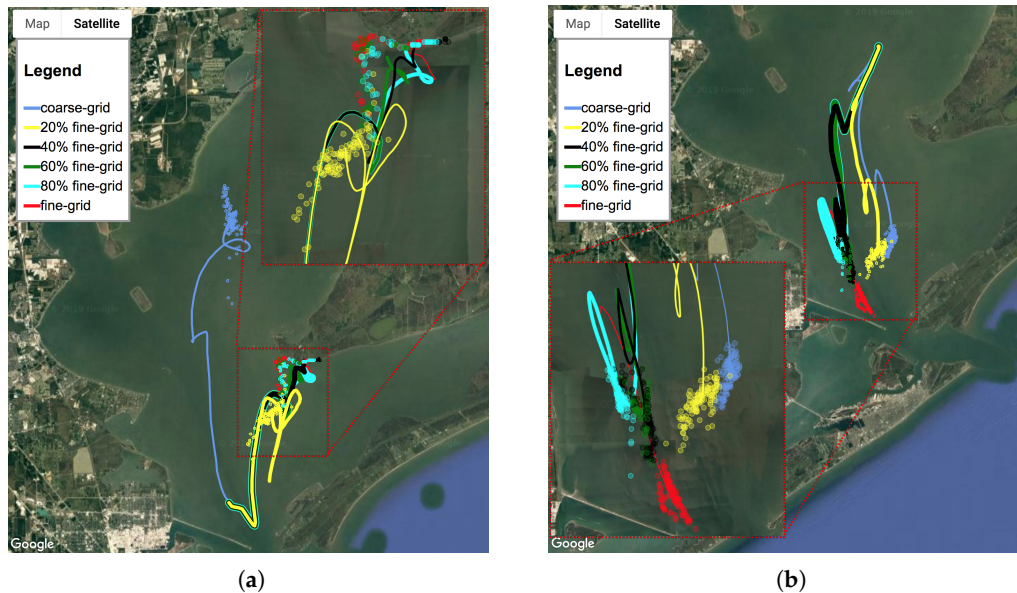


Figure 11. The trajectories of the multi-model integration for (a) Case 1 and (b) Case 2.

We can also use the multiple trajectories to create “probability maps” of the multi-model ensemble as shown in Figure 12. These maps can be used to visualize how the geometric uncertainty is spreading the solution as the ensemble evolves over time. Probability maps describe the normalized (0,1) probability of a point in space occurring in the ensemble. Thus, the red represents clustering of the predicted tracks in some area and *not* necessarily a high probability that the particular area is the best prediction [24]. In Case 1, the high probability (~ 0.2) occurs along the fine-grid prediction and is away from the coarse-grid prediction over the entire simulation period. This result implies that the

geometric uncertainty of the coarse-grid model was high from the beginning. In Case 2, the probability was high for all predictions during the first third of the simulation when we gained confidence that the coarse-grid prediction was adequate. The probability map coincides with the multiple trajectories (Figure 11), but avoids single biased predictions. However, such maps must be used with caution as they are subject to misinterpretation; that is, we can imagine a case where a number of early tracks in the multi-model ensemble fall close together and the most recent track crosses an entirely different space due to a divergence in the latest fine-grid prediction. In such conditions, the early tracks would show up as high probability and the latest track as low probability, despite the fact that the latter is clearly the preferred estimate. Thus, these probability maps should be seen as a way of understanding the geometric uncertainty model behavior but there is a need for further consideration as to how such visualizations could be used in an operational context. In future work, we plan to evaluate a weighting method that ensures that ensemble members with a greater fraction of fine-grid simulation will dominate the probability map. Evaluating how operational managers would respond to such maps and their understanding of weighting functions remains a subject for future research.

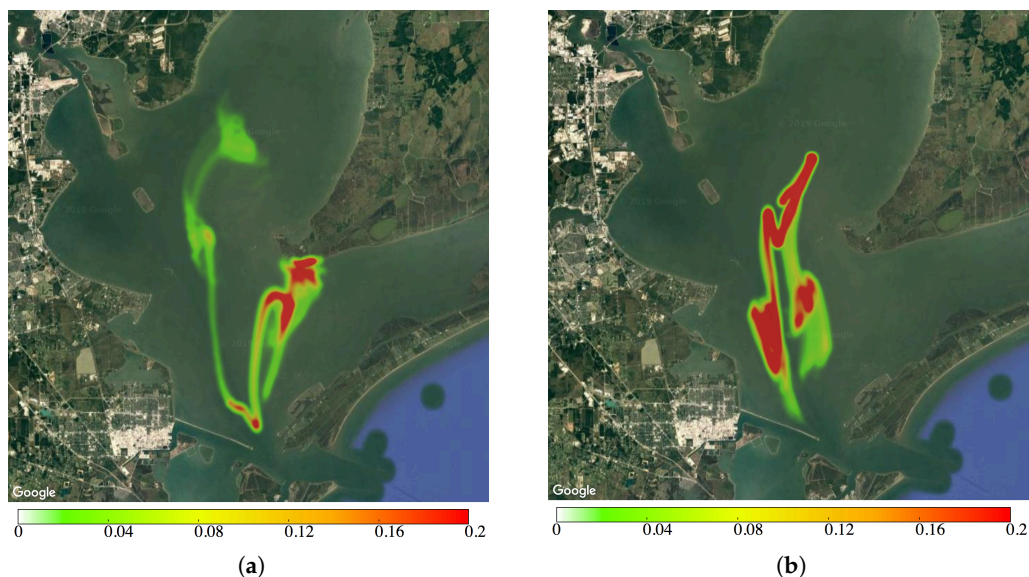


Figure 12. The probability map of the multi-model integration for (a) Case 1 and (b) Case 2.

The two approaches illustrated above can be used to quantify the geometric uncertainty and display the difference between coarse-grid and fine-grid models. The multi-model integration provides an explicit information of where and when the geometric uncertainty of the coarse-grid prediction is high during an event. The ensemble approach of the multi-model system allows generation of a new ensemble member when the fine-grid model has only run a portion (e.g., 20%) of the entire simulation. The ensemble set continues to grow during the fine-grid simulation run, which adds to our knowledge about the prediction uncertainty and provides oil spill response managers with greater insight into the predictions.

6. Conclusions

This work investigated geometric uncertainty of an oil spill modeling system designed for operation with combined coarse-grid and fine-grid models. The system was developed and demonstrated on Galveston Bay using the SUNTANS hydrodynamic model and the GNOME oil spill model within the HyosPy modeling system. The modeled difference in the separation distance and the diffusion error between coarse and fine-grid model results was used as an indicator of the geometric uncertainty for coarse-grid model predictions. The geometric uncertainty was influenced by both the model grid and the flow dynamics, and showed critical variability across the estuary at the

tidal scale of hours. We proposed two approaches to operationally quantify geometric uncertainty: a data-driven uncertainty model and multi-model integration.

The data-driven uncertainty model provides an *a priori* estimate of the coarse-grid model reliability. We applied the GBRT machine learning algorithm to develop the uncertainty model trained on numerical Lagrangian experiments in Galveston Bay over a three-month period. A geospatial data clustering algorithm was derived to reduce the spatial dimension of the model so that data collection for model training was less computationally intensive. The uncertainty model was evaluated with hypothetical spills and showed satisfactory performance for predicting the effective confidence over ~50 h.

The multi-model integration employs partial simulation results from fine-grid model that are available for early times after an event and are then extended to later times with the coarse-grid model. Model runs with different combinations of coarse and fine-grid simulations are used as part of a multi-model ensemble. The resulting ensemble of spill tracks can be visualized as multiple trajectories or as a probability map. Both trajectories and probability maps are integrated with Google Maps so that they can be projected on satellite images. This approach allows emergency response managers to obtain an increased understanding of the geometric uncertainty as an event is tracked and as new model results are produced.

The uncertainty model and the multi-model integration were applied to the HyosPy system in two case studies; we initiated hypothetical spills at different locations to demonstrate the type of information on uncertainty that can be generated. The data-driven uncertainty model enables the HyosPy system to self-identify the uncertainty errors statistically, while the multi-model ensemble allows the emergency managers to quantify the geometric uncertainty without considering its variabilities. Both approaches identify the modeling errors of the coarse-grid prediction, i.e., where and when the coarse-grid model performance starts to decrease significantly.

The uncertainty quantification approaches proposed herein are not necessarily limited to oil spill modeling, and could be applied more broadly when more multiple models are available of a given process. The key challenge is to find the data model appropriate to the physical patterns of the system. Data-driven machine learning models can also be used to quantify other forms of uncertainty associated with forecasting or empirical parameters. In general, the above techniques can be adapted for use whenever historical assessment of model-model comparisons can be used to evaluate uncertainty.

The new methods quantify the error in Lagrangian particle transport associated with the design of a model grid mesh—i.e., the spatial/temporal geometric uncertainty introduced with a coarser grid. Traditional mesh design (e.g., [9]) focuses solely on validating models against Eulerian observations or the grid convergence of Eulerian variables. For our test case in Galveston Bay, the results above show that the existing coarse-grid mesh provides low confidence for Lagrangian transport in the critical area of the ship channel entrance. We believe these tools hold promise for future model developers in designing and testing coarse unstructured meshes that are optimized for Lagrangian particle transport in areas where oil spills are likely to occur.

Author Contributions: Conceptualization, D.F.; methodology, D.F.; validation, D.F.; formal analysis, D.F.; writing—original draft preparation, D.F.; writing—review and editing, D.F. and P.P. and B.R.H.; visualization, D.F.; supervision, P.P. and B.R.H.

Funding: This research was funded partially by the Research and Development program of the Texas General Land Office Oil Spill Prevention and Response Division grant number 18-133-000-A674.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hodges, B.; Orfila, A.; Sayol, J.; Hou, X. Operational oil spill modeling: From science to engineering applications in the presence of uncertainty. In *Mathematical Modelling and Numerical Simulation of Oil Pollution Problems*; Ehrhardt, M., Ed.; The Reacting Atmosphere; Springer: Cham, Switzerland, 2015; Volume 2, doi:10.1007/978-3-319-16459-5_5.
- Benestad, R.; Hanssen-Bauer, I.; Chen, D. *Empirical Statistical Downscaling*; World Scientific Publishing Company: Singapore, 2009; doi:10.1142/6908.
- Zelenke, B.; O'Connor, C.; Barker, C.H.; Beegle-Krause, C.J.; Eclipse, L. (Eds.) *General NOAA Operational Modeling Environment (GNOME) Technical Documentation*; Technical report; The Office of Response and Restoration's (OR&R) Emergency Response Division under NOAA, Emergency Response Division, NOAA: Seattle, WA, USA, 2012.
- Lynch, D.; Greenberg, D.; Bilgili, A.; McGillicuddy, D.; Manning, J.; Aretxabaleta, A. *Particle in the Coastal Ocean: Theory and Applications*; Cambridge University Press: Cambridge, UK, 2015; doi:10.1017/CBO9781107449336.
- Mariano, A.; Kourafalou, V.; Srinivasan, A.; Kang, H.; Halliwell, G.; Ryan, E.; Roffer, M. On the modeling of the 2010 Gulf of Mexico Oil Spill. *Dyn. Atmos. Ocean.* **2011**, *52*, 322–340. [[CrossRef](#)]
- Chang, Y.; Oey, L.; Xu, F.; Lu, H.; Fujisaki, A. 2010 oil spill: Trajectory projections based on ensemble drifter analyses. *Ocean Dyn.* **2011**, *61*, 829–839. [[CrossRef](#)]
- Hart, K.; Steenburgh, W.; Onton, D.; Siffert, A. An evaluation of mesoscale-model-based model output statistics (MOS) during the 2002 Olympic and Paralympic Winter Games. *Weather Forecast.* **2004**, *19*, 200–218. [[CrossRef](#)]
- Roebber, P.; Schultz, D.; Colle, B.; Stensrud, D. Toward Improved Prediction: High-Resolution and Ensemble Modeling Systems in Operations. *Weather Forecast.* **2004**, *19*, 936–949. [[CrossRef](#)]
- Rayson, M.; Gross, E.; Fringer, O. Modeling the tidal and sub-tidal hydrodynamics in a shallow, micro-tidal estuary. *Ocean Model.* **2015**, *89*, 29–44. [[CrossRef](#)]
- Hou, X. Quantitative Risk Analysis in Numerical Oil Spill Modeling System in Offshore Bays and Estuaries. Ph.D. Thesis, The University of Texas at Austin, Austin, TX, USA, 2015.
- Marta-Almeida, M.; Ruiz-Villarreal, M.; Pereira, J.; Otero, P.; Cirano, M.; Zhang, X.; Hetlanda, R. Efficient tools for marine operational forecast and oil spill tracking. *Mar. Pollut. Bull.* **2013**, *71*, 139–151. [[CrossRef](#)] [[PubMed](#)]
- Andrejev, O.; Soomere, T.; Sokolov, A.; Myrberg, K. The role of the spatial resolution of a three-dimensional hydrodynamic model for marine transport risk assessment. *Oceanologia* **2011**, *53*, 309–334. [[CrossRef](#)]
- Neves, A.; Pinardi, N.; Martins, F. IT-OSRA: Applying ensemble simulations to estimate the oil spill risk associated to operational and accidental oil spills. *Ocean Dyn.* **2016**, *66*, 939–954. [[CrossRef](#)]
- Hutschenreuter, K.; Hodges, B.; Socolofsky, S. Simulation of Laboratory Experiments for Vortex Dynamics at Shallow Tidal Inlets using the Fine Resolution Environmental HydroDynamics (Frehd). *Environ. Fluid Mech.* **2019**, *19*, 1–32. [[CrossRef](#)]
- Feng, D.; Hodges, B.; Socolofsky, S.; Thyng, K. Tidal eddies at a narrow channel inlet in operational oil spill models. *Mar. Pollut. Bull.* **2019**, *140*, 374–387. [[CrossRef](#)]
- Greenberg, D.; Dupont, F.; Lyard, F.; Lynch, D.; Werner, F. Resolution issues in numerical models of oceanic and coastal circulation. *Cont. Shelf Res.* **2007**, *27*, 1317–1343. [[CrossRef](#)]
- Azevedo, A.; Fortunato, A.; Epifanio, B.; den Boer, S.; Oliveira, E.; Alves, F.; de Jesus, G.; Gomes, J.; Oliveira, A. An oil risk management system based on high-resolution hazard and vulnerability calculations. *Ocean Coast. Manag.* **2017**, *136*, 1–18. [[CrossRef](#)]
- Liubartseva, S.; Dominicis, M.D.; Oddo, P.; Coppini, G.; Pinardi, N.; Greggio, N. Oil spill hazard from dispersal of oil along shipping lanes in the Southern Adriatic and Northern Ionian Seas. *Mar. Pollut. Bull.* **2015**, *90*, 259–272. [[CrossRef](#)] [[PubMed](#)]
- Guo, W. Development of a statistical oil spill model for risk assessment. *Environ. Pollut.* **2017**, *230*, 945–953. [[CrossRef](#)] [[PubMed](#)]
- Elith, J.; Leathwick, J.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)] [[PubMed](#)]
- Solomatine, D.; Ostfeld, A. Data-driven modelling: Some past experiences and new approaches. *J. Hydroinform.* **2008**, *10*, 3–22. [[CrossRef](#)]

22. Cao, Y.; Xu, L.; Clausi, D. Exploring the Potential of Active Learning for Automatic Identification of Marine Oil Spills Using 10-Year (2004–2013) RADARSAT Data. *Remote Sens.* **2017**, *9*, 1041. [CrossRef]
23. Liu, Y.; Weisberg, R.; Hu, C.; Zheng, L. Trajectory Forecast as a Rapid Response to the Deepwater Horizon Oil Spill. In *Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise*; Geophysical Monograph 195; American Geophysical Union: Washington, DC, USA, 2011; pp. 153–165, doi:10.1029/2011GM001121.
24. Hou, X.; Hodges, B.R.; Feng, D.; Liu, Q. Uncertainty quantification and reliability assessment in operational oil spill forecast modeling system. *Mar. Pollut. Bull.* **2017**, *116*, 420–433. [CrossRef]
25. Seeba, P. Coordinating Chemical Tanker Movements on the Houston Ship Channel. 2016. Available online: <https://txgulf.org/archives/port-bureau-news/coordinating-chemical-tanker-movements-on-the-houston-ship-channel> (accessed on 1 August 2019).
26. Kinney, E.; Gonzalez, L.; Lester, J. Oil Spills—A Continual Risk for Galveston Bay Habitats. 2014. Available online: https://www.harcresearch.org/features/Oils_Spills_a%20Continual_Risk_for_Galveston_Bay_Habitats (accessed on 1 June 2019).
27. Bacosa, H.; Thyng, K.; Plunkett, S.; Erdner, D.; Liu, Z. The tarballs on Texas beaches following the 2014 Texas City “Y” Spill: Modeling, chemical, and microbiological studies. *Mar. Pollut. Bull.* **2016**, *109*, 236–244. [CrossRef]
28. Fringer, O.; Gerritsen, M.; Street, R. An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. *Ocean Model.* **2006**, *14*, 139–173. [CrossRef]
29. Liu, Y.; Weisberg, R.H. Evaluation of trajectory modeling in different dynamic regions using normalized cumulative Lagrangian separation. *J. Geophys. Res.* **2011**, *116*, C09013. [CrossRef]
30. Russ, G.; Kruse, R. Machine Learning Methods for Spatial Clustering on Precision Agriculture Data. In *Eleventh Scandinavian Conference on Artificial Intelligence*; IOS Press: Amsterdam, The Netherlands, 2011; pp. 40–49.
31. Barker, C.; Galt, J. Analysis of Methods Used in Spill Response Planning: Trajectory Analysis Planner TAP II. *Spill Sci. Technol. Bull.* **2000**, *6*, 145–152. [CrossRef]
32. Friedman, J. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
33. Nixon, Z.; Michel, J. Predictive Modeling of Subsurface Shoreline Oil Encounter Probability from the Exxon Valdez Oil Spill in Prince William Sound, Alaska. *Environ. Sci. Technol.* **2015**, *49*, 4354–4361. [CrossRef] [PubMed]
34. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
35. Zhang, Z.; Krishnamurti, T. Ensemble Forecasting of Hurricane Tracks. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 2785–2795. [CrossRef]
36. Hou, X.; Hodges, B.R.; Negusse, S.; Barker, C. A multi-model Python wrapper for operational oil spill transport forecasts. *Comput. Sci. Discov.* **2015**, *8*, 014004. [CrossRef]
37. Mesinger, F.; DiMego, G.; Kalnay, E.; Mitchell, K.; Shafran, P.; Ebisuzaki, W.; Jovis, D.; Woollen, J.; Rogers, E.; Berbery, E.H.; et al. North American Regional Reanalysis. *Bull. Am. Meteorol. Soc.* **2006**, *87*, 343–360. [CrossRef]

