SCIENTIFIC JOURNAL OF MARITIME RESEARCH [POMORSTVO]
University of Rijeka FACULTY OF MARITIME STUDIES

https://doi.org/10.31217/p.39.2.12

Improving the accuracy of the oil spill model: A case study of the Hai Phong Sea area

Do Van Cuong[1], Nguyen Xuan Long[1*], Nguyen Manh Cuong[1], Do Trung Kien[2]

[1] Navigation Faculty, Vietnam Maritime University; 484, Lach Tray Str., Haiphong City, Vietnam, e-mail: dovancuong@vimaru.edu.vn; nguyenxuanlong@vimaru.edu.vn

[2] Department of International Relations, Vietnam Maritime University; 484, Lach Tray Str., Haiphong City, Vietnam, e-mail: dotrungkien@vimaru.edu.vn

[*] Corresponding author

ARTICLE INFO

Original scientific paper
Received 9 May 2025
Accepted 20 June 2025

Key words:
Oil spill modeling
Hydrodynamic calibration
Coastal pollution simulation

ABSTRACT

Oil spill modeling is a crucial tool in marine environmental management, enabling early prediction and mitigation of pollution impacts on coastal ecosystems. This study presents a high-resolution oil spill simulation framework developed for the Hai Phong coastal region in northern Vietnam, where risks from maritime activities are significant. The proposed approach utilizes a Lagrangian particle tracking method to simulate the transport and dispersion of oil particles driven by wind, wave, and tidal forces. A hydrodynamic model underpins the simulation, with its input parameters systematically calibrated to improve accuracy. Model calibration was guided by widely accepted error assessment metrics, including Root Mean Square Error, Nash–Sutcliffe Efficiency, and Percent Bias. These indicators were used to iteratively refine the simulation until a satisfactory match with empirical observations was achieved. Unlike traditional models that often rely on static inputs, this study incorporates dynamic environmental forcing and applies adaptive calibration strategies, allowing for reliable forecasts even when the spill source is uncertain or data availability is limited. The model demonstrates strong predictive capacity during the critical early hours following an oil spill, highlighting its potential for integration into rapid-response systems. Overall, the framework contributes to advancing oil spill preparedness and response strategies in data-scarce coastal regions, supporting both ecological protection and maritime risk management.

1 Introduction

Oil spills in marine environments pose severe environmental threats, disrupting ecological balance, damaging fisheries, and undermining the socio-economic stability of coastal communities. The Hai Phong coastal region in northern Vietnam—home to one of the country's most active port systems and dense industrial zones—is particularly at risk from oil pollution due to increasing maritime traffic and coastal development.

Accurate simulation of oil spill behavior is critical for timely and effective response, enabling mitigation strategies to be deployed before irreversible environmental damage occurs. While numerous oil spill models have been developed globally, their application in complex nearshore environments is often hindered by challenges such as limited spatial resolution, uncertainties in input data, and the inability to simulate dynamic interactions between wind, tides, and waves with sufficient fidelity.

To overcome these limitations, the integration of observational data with advanced numerical modeling techniques has become increasingly important. Recent studies have demonstrated the value of assimilating real-time environmental data into simulation frameworks. For example, Wilson et al. [1] emphasized data assimilation to enhance model reliability, while Nordam [2] developed a scenario-specific model for Arctic conditions using the OSCAR platform. Dinu et al. [3] integrated real-time environmental data—such as wind, wave, and current information—which significantly improved the

accuracy of oil slick trajectory predictions. However, their model was developed based on the proprietary ADIOS2 platform. Andrés et al. [4] enhanced model accuracy by optimizing grid resolution and the number of Lagrangian particles, effectively balancing simulation precision and computational performance. Konstantinos et al. [5] employed ECMWF wind datasets to assess uncertainty arising from atmospheric forcing, demonstrating the effectiveness of the ensemble approach in environmental risk forecasting. Nevertheless, their work was also built upon the commercial MEDSLIK-II oil spill modeling framework applied in the Aegean Sea. Neda et al. [6] calibrated the Delft3D FM hydrodynamic model using both Eulerian and Lagrangian data (including ADCP drifter measurements), which improved simulation accuracy. However, the validation relied solely on the R^2 metric, which may not provide a sufficiently objective or comprehensive assessment of model performance.

In Vietnam, Doan Quang Tri et al. [7] applied the MIKE 21 SA model to simulate oil dispersion in Lach Huyen Port, and Nguyen [8] introduced a reverse-time tracking method to identify spill origins in the East Sea.

The program, independently developed by the author, is capable of accurately predicting the drift of oil slicks, as validated through comparison with actual oil spill scenarios simulated using the MIKE software. This work contributes significantly to the scientific advancement of oil spill modeling research in both Vietnam and the global context.

The paper presents the development of an oil spill simulation model based on the Lagrangian approach, tailored to the natural conditions of the Vietnamese marine environment, with a pilot application in the Hai Phong coastal region. The model allows for flexible adjustment of input parameters in the hydrodynamic module and enables control of output reliability through error evaluation indicators: Root Mean Square Error (RMSE), which measures the average deviation between observed and modeled values; Nash–Sutcliffe Efficiency (NSE), which evaluates predictive skill; and Percent Bias (PBIAS), thereby ensuring the credibility of the simulation results.

2 Methods

2.1 Oil spill dispersion model

2.1.1 Advection and diffusion processes

This equation is formulated to simulate the transport processes of substances in both surface and subsurface water layers. The substances are discretized into individual particles, which are influenced by transport mechanisms arising from the surrounding mass movement.

a) Transport equation

The governing principle is defined by the following equation [9-11]:

$$\frac{\partial(ch)}{\partial t}+\frac{\partial(uch)}{\partial x}+\frac{\partial(vch)}{\partial y}=\frac{\partial}{\partial x}\left[hD_x\frac{\partial c}{\partial x}+hD_{xy}\frac{\partial c}{\partial y}\right]+\frac{\partial}{\partial y}\left[hD_y\frac{\partial c}{\partial y}+hD_{yx}\frac{\partial c}{\partial x}\right] \quad (2.1)$$

In this equation, $\boldsymbol{c}(x,y,t)$ represents the mean concentration; $\boldsymbol{u}(x,y,t)$ and $\boldsymbol{v}(x,y,t)$ denote the velocity components, $\boldsymbol{h}(x,y,t)$ is the water depth (relative to MSL); $\boldsymbol{D}_x(x,y,t)$, $\boldsymbol{D}_y(x,y,t)$, $\boldsymbol{D}_{xy}(x,y,t)$ and $\boldsymbol{D}_{yx}(x,y,t)$ are the diffusion coefficients. The angle between the intrinsic current direction and the x-axis is represented by:

$$\theta=\operatorname{atan}\left(\frac{\mathrm{v}(x(t),t)}{u(x(t),t)}\right) \quad (2.2)$$

After converting coordinate system:

$$D_x=D_L cos^2(\theta)+D_T\, sin^2(\theta) \quad (2.3)$$

$$D_y=D_L sin^2(\theta)+D_T cos^2(\theta) \quad (2.4)$$

$$D_{xy}=(D_L-D_T)\, sin(\theta) cos(\theta) \quad (2.5)$$

with, $\boldsymbol{D}_L$ and $\boldsymbol{D}_T$ are diffusion coefficients respectively.

b) Langevin Equation

The equation describes the motion of an oil particle at position $x(t)$ [12-14]:

$$\frac{dx}{dt}=A(x,t)+B(x,t)\xi(t) \quad (2.6)$$

Here $\boldsymbol{A}(x,t)$ represents the current speed field; $\boldsymbol{B}(x,t)$ denotes the stochastic component simulating the diffusion process of the oil; $\xi(t)$ is a Gaussian white noise term (mean 0, variance 1), generated by a random number generator. Equation (2.6) can be rewritten as follows:

$$dx=(A(x(t),t)dt+B(x(t),t)dW(t) \quad (2.7)$$

With, $dW(t)$ is a Wiener random process.

The discrete method for computing (2.7) uses the Euler implicit method:

$$\Delta x_n=x_n-x_{n-1}=A(x_{n-1},t_{n-1})\Delta t+B(x_{n-1},t_{n-1})\sqrt{\Delta t}Z_n \quad (2.8)$$

Where, x_n is the particle position at the current time step t_n; x_{n-1} is the particle position at the previous time step t_{n-1}; Δx_n is the incremental displacement of the particle over the time step Δt; $A(x_{n-1},t_{n-1})$ is the drift term, representing the deterministic component of motion,

typically associated with the advective velocity field; $B(x_{n-1}, t_{n-1})$ is the diffusion coefficient, accounting for the magnitude of stochastic fluctuations due to turbulence or unresolved subgrid processes; Zn is a standard Gaussian random variable drawn from N (0,1), used to represent random perturbations; $\sqrt{\Delta t} Zn$ is a discrete approximation of the Brownian motion increment $\Delta Wn \sim N$ (0,Δt), used to model the stochastic diffusion process.

The equation describes the motion of an oil particle at position $y(t)$ is the same method:

$$\frac{dy}{dt} = A(y,t) + B(y,t)\xi(t) \tag{2.9}$$

$\boldsymbol{A}(y,t)$ represents the current speed field in the y direction; $\boldsymbol{B}(y,t)$ denotes the stochastic component simulating the diffusion process of the oil in the y direction; $\xi(t)$ is a Gaussian white noise term (mean 0, variance 1), generated by a random number generator.

c) Advection process

This process is primarily driven by currents and wind, causing particles to move horizontally. The drift velocity is typically calculated as follows [15]:

$$U_{tot} = c_w(z)U_w + c_a(z)U_a \tag{2.10}$$

Where: $\boldsymbol{U}_w$ is the wind velocity vector at 10 meters above the sea surface; $\boldsymbol{U}_a$ is the depth-integrated current velocity vector; $c_w(z)$ is the wind drift coefficient at depth z; and $c_a(z)$ is the advective current drift coefficient at depth z.

2.1.2 Mechanical Spreading

This process occurs immediately following an oil spill and is primarily driven by two factors:

– The physicochemical properties of the oil (such as differences in density and surface tension);
– The diffusion process induced by waves, wind, and currents.

The oil slick is typically treated as a homogeneous mass, and the spreading process leads to thinner layers over time. The surface area of the spill evolves as a function of time t [16]:

$$\frac{dA_{oil}}{dt} = K_a A_{oil}^{1/3} \left(\frac{V_{oil}}{A_{oil}}\right)^{4/3} \tag{2.11}$$

In which, K_a is a constant; A_{oil} is the area of oil spill; V_{oil} is the volume of oil; R_{oil} is the radius of oil spill.

$$A_{oil} = \pi R_{oil}^2 \tag{2.12}$$

The volume of the oil spill is:

$$V_{oil} = \pi R_{oil}^2 . h_s \tag{2.13}$$

Where, h_s is the initial thickness of the oil slick.

2.2 Material transport model

2.2.1 Theoretical basis of Lagrange method

Building upon the work of Garcia [17], the author simulates particle trajectories using three different schemes to evaluate the advantages and limitations of each approach. A two-dimensional flow field is considered:

$$dX_t = a(t, X_t)dt + b(t, X_t)\xi_t dt \tag{2.14}$$

$$X_{n+1} = X_n + a(t)dt + b(t)\xi dt \tag{2.15}$$

The material transport process is approximated using the Euler scheme for Y over time, and is thus expressed as follows:

$$Y_{n+1} = Y_n + a(t)dt + b(t)\xi dt \tag{2.16}$$

For the diffusion component:

– Time from start to 12 hours:

$$b(t) = b*(t/24)^{0.1} \tag{2.17}$$

– Time greater than 12 hours:

$$b(t) = b*(e^{(48-t)/48} - 1.7)^{2.75} \tag{2.18}$$

Where b is the dispersion coefficient and t is time (in hours).

The random process ξ is defined to vary within the range of -1 to 1.

In the adopted material transport framework, the transported matter is discretized into multiple particle classes, each characterized by distinct physical and chemical properties such as decay rate, vertical settling or rising velocity, corrosion behavior, and dispersion coefficients. These attributes are individually defined for each particle class, allowing for a refined representation of heterogeneous material behavior. Additionally, class-specific parameters such as minimum mass thresholds and maximum particle lifetime are specified to facilitate dynamic classification and removal of particles that have degraded below significant mass or exceeded the temporal bounds of interest. This selective deactivation enhances computational efficiency by reducing the total number of active particles within the simulation domain.

The transport dynamics are simulated using a Lagrangian discrete approach, wherein the continuum of material is represented as an ensemble of computational particles. Each particle is assigned a discrete mass and spatial coordinates, and its trajectory and mass evolution are computed as functions of time. Unlike Eulerian methods, particle movement in the Lagrangian framework is not constrained to the fixed nodes of a reference grid. Instead, the grid system is employed solely for the interpolation of environmental parameters—

such as current velocity fields—at the particle's instantaneous position. These interpolated values are then used to determine the drift velocity and subsequent advection of each particle. This approach allows for a more accurate and physically consistent simulation of transport phenomena in spatially and temporally variable environments.

2.2.2 Algorithm structure diagram and computational programs

Based on the computational approach for forecasting marine oil spills using the Lagrangian method, along with the mathematical foundations supporting the discretization process and the necessary conditions for developing a computational model, the author has designed the algorithmic structure as illustrated below. Figure 1 presents the general algorithm of the oil spill simulation model. From this structure, different computational procedures are developed according to various information blocks using the Fortran programming language. The program architecture consists of a main module and several auxiliary subroutines, which reduce complexity and enhance usability.

Figure 2 illustrates the block diagram for input data and initial condition setup. In this stage, each computational component is organized into modular subroutines. The program reads temporal and environmental parameters, including spill timing, location, and release type. Once validated, these inputs initiate synchronized environmental data processing across spatial-temporal nodes for further simulation.

Figure 3 presents the block diagram for updating and processing environmental and oceanographic conditions. Based on specified indices, the program initiates the simulation using various input datasets. As these inputs often differ in spatial and temporal resolution, preprocessing steps are required to harmonize them onto a unified coordinate system and synchronized computational grid. Once standardized, the processed data are passed to subsequent modeling stages.

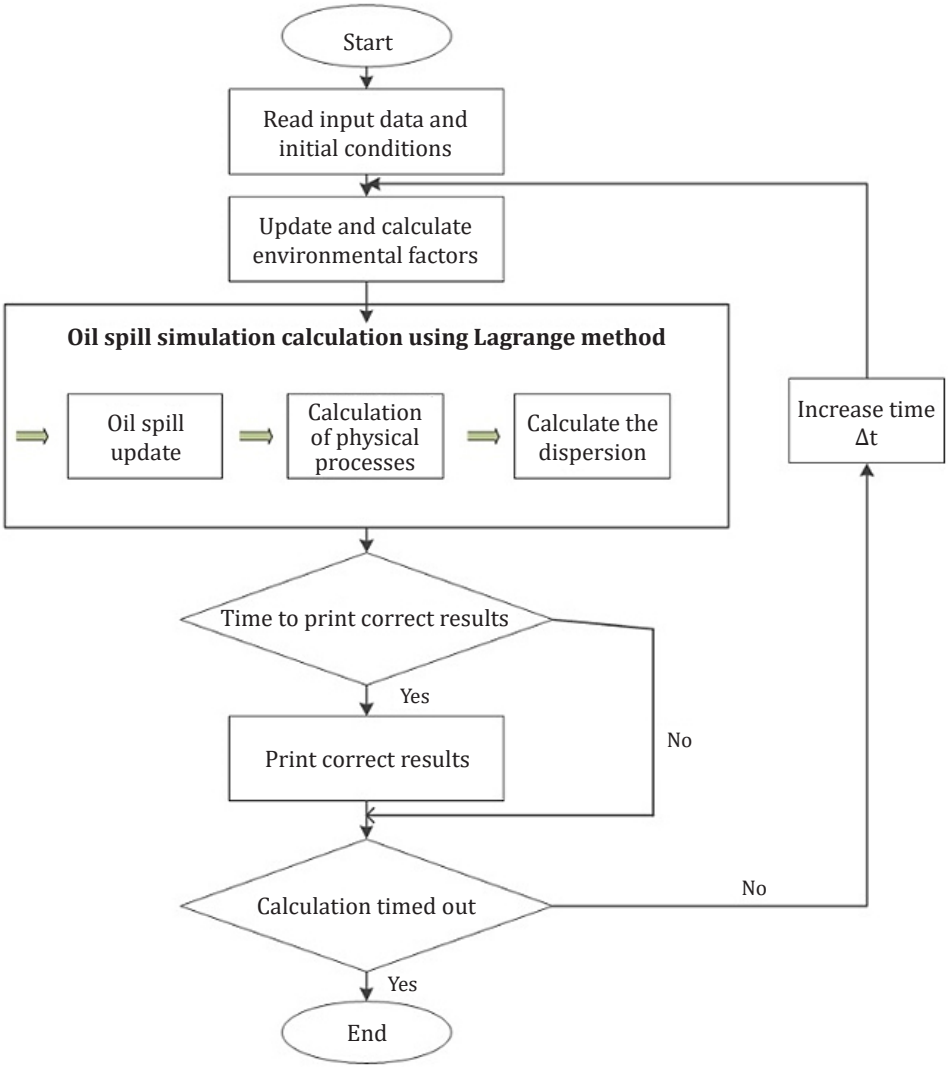

Figure 1 Oil Spill Calculation Program

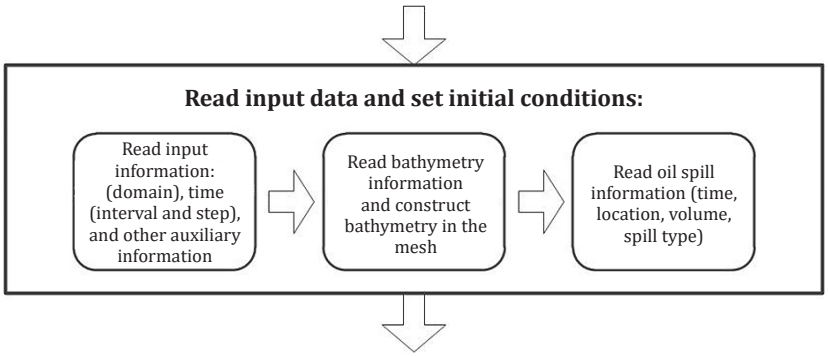

Figure 2 Information processing

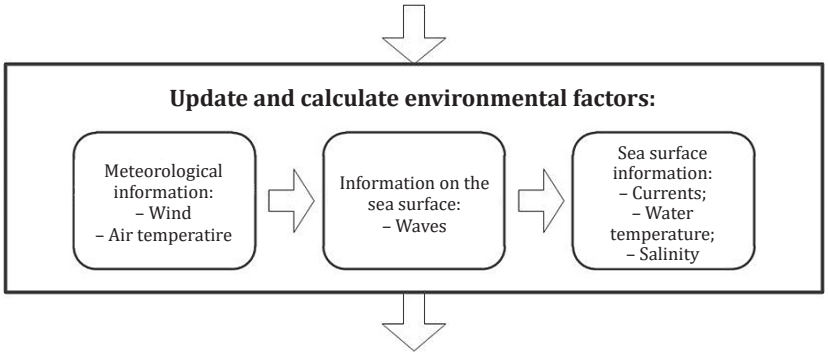

Figure 3 Update and process information on environmental factors

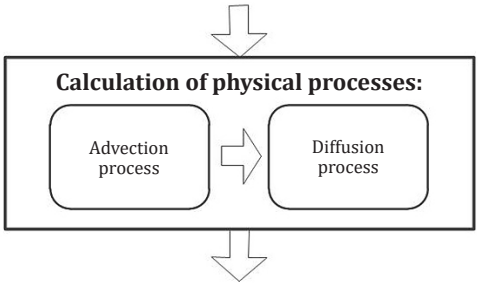

Figure 4 Advection and diffusion processes

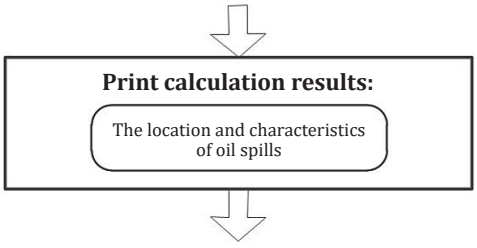

Figure 5 Processing model output

Figure 4 illustrates the block diagram for computing the advection and diffusion processes of surface oil spills. The associated subroutines govern the two-dimensional drift of the oil slick. Advection is modeled using surface dynamic forcing (wind, waves, and currents) represented in two horizontal dimensions. Diffusion is formulated based on oil properties and release volume, incorporating turbulent motion induced by environmental conditions, as described in Equation (2.15) and (2.16). The outputs from this stage serve as inputs for subsequent oil spill dispersion calculations.

Figure 5 presents the output block diagram following the oil spill dispersion computations. This stage enables the extraction of all relevant information regarding parameters and environmental variables involved in both the main and subroutine components of the model.

2.3 Criteria for assessing model accuracy and error

2.3.1 Correlation coefficient (R)

This coefficient is a statistical indicator used to evaluate the degree of association between the reference variable (X_{net}) and the simulated variable (X_{sim}), based on the perspective of proportional deviation magnitude [18]. The formula for calculating this coefficient is expressed as follows:

$$r = \frac{\sum_{i=1}^{n}\left[\left(X_{net} - X_{mean\_net}\right)*\left(X_{sim} - X_{mean\_sim}\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(X_{net} - X_{mean\_net}\right)^2 * \sum_{i=1}^{n}\left(X_{sim} - X_{mean\_sim}\right)^2}}$$

$$X_{mean\_net} = \frac{1}{n}\sum_{i=1}^{n} X_{net}\,; X_{mean\_sim} = \frac{1}{n}\sum_{i=1}^{n} X_{sim} \qquad (2.19)$$

This is a statistical indicator used to assess the discrepancy in the relationship between two variables. Specifically:

- The value of the correlation coefficient lies within the range (−1.0; 1.0). If the simulation results fall outside this range (i.e., greater than 1.0 or less than −1.0), it indicates errors in the correlation measurement process;
- A negative value of the correlation coefficient indicates an inverse relationship between the two variables, commonly referred to as negative correlation. A value of −1 implies a perfect negative correlation;
- A positive value of the coefficient suggests a direct relationship, or positive correlation, with a value of 1 indicating a perfect positive correlation;
- A value of 0 represents statistical independence between the two variables.

2.3.2 Root Mean Square Error (RMSE)

This is an indicator used to measure and evaluate the relationship between the reference variable (X_{net}) and the simulated variable (X_{sim}) based on the mean squared error (MSE) approach [19]. The smaller the value of this indicator, the closer the model output is to the observed data. The formula for calculating this indicator is expressed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_{net} - X_{sim})^2} \qquad (2.20)$$

RMSE is an indicator that measures the root mean square deviation; Here, *Xi* and *Yi* represent the i^{th} values of the observed variable *x* and the simulated variable *y*, respectively; *i* denotes the index within the data series ranging from 1 to *n*; and *n* is the total number of observed values for both variables *x* and *y.*

2.3.3 NASH Index

This index can be calculated by the following equation [20]:

$$NASH = 1 - \left[\frac{\sum_{i=1}^{n}(X_{net} - X_{sim})^2}{\sum_{i=1}^{n}(X_{net} - X_{mean})^2}\right];$$
$$X_{mean} = \frac{1}{n}\sum_{i=1}^{n} X_{net} \qquad (2.21)$$

In this context, X_i and Y_i represent the i^{th} values of the observed variable *x* and the simulated variable *y*, respectively; X_{mean} and Y_{mean} denote the mean values of the two variables; *i* is the index in the data series ranging from 1 to *n*; and *n* is the total number of observations for both *x* and *y*.

Table 1 below presents the classification of model accuracy based on the Nash–Sutcliffe Efficiency (NASH) coefficient.

The Nash–Sutcliffe Efficiency (NSE) coefficient reflects the degree of correlation between observed and simulated data. In principle, the closer the NSE value is to 1, the more reliable the model performance is. The variation of the NSE coefficient can be interpreted as follows:

NSE = 1: The simulation perfectly matches the observed data, indicating absolute accuracy.

NSE = 0: The model has no correlation with the observed data, implying no predictive capability. This case is rare.

0 < NSE < 1: The model shows a certain degree of correlation with the observed data. Depending on the specific value, the performance level can be categorized as:

0.0 < NSE < 0.2: Very low correlation (very poor performance).

0.2 ≤ NSE < 0.4: Low correlation (poor performance).

0.4 ≤ NSE < 0.6: Moderate correlation (moderate performance).

0.6 ≤ NSE < 0.8: High correlation (good performance).

0.8 ≤ NSE < 1.0: Very high correlation (very good performance).

Table 1 Classification of NASH index accuracy level

NASH Index	0.0 – 0.2	0.2 – 0.4	0.4 – 0.6	0.6 – 0.8	0.8 – 1.0
Level	Very low	Low	Moderate	High	Very high

To ensure the reliability of the NSE value, it is essential to use a sufficiently large dataset that captures the full characteristics of the variable under evaluation. The longer and more accurate the dataset, the more meaningful the NSE becomes in assessing model performance.

2.3.4 PBIAS Standard Deviation Index

This is a composite indicator used to assess the degree of agreement between the observed (measured) values (X_{net}) and the simulated values (X_{sim}). The formula for calculating this indicator is expressed as follows [21]:

$$PBIAS = \left[\frac{\sum_{i=1}^{n}(X_{net} - X_{sim})*100}{\sum_{i=1}^{n} X_{net}}\right] \quad (2.22)$$

Here, i denotes the index in the data series ranging from 1 to n, and n is the total number of observations for variables x and y.

PBIAS measures the tendency of the average simulated streamflow values to be greater or smaller than the corresponding observed streamflow values. The optimal value is 0.0. Positive values indicate a model bias toward underestimation, while negative values indicate a bias toward overestimation [22].

Figure 6 illustrates the flowchart of the calibration steps used to verify the reliability of the hydrodynamic module through error evaluation indicators:

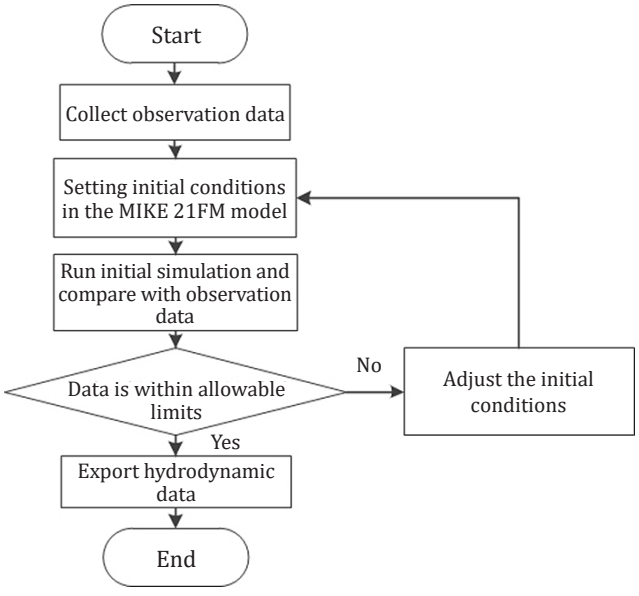

Figure 6 Calibration process block diagram

- Step 1 – Collection of observed data: Gather hydrological data such as water level, discharge, current speed, and other relevant parameters at the study area from monitoring stations or reliable data sources.
- Step 2 – Initialization of model conditions: Input initial values into the model, including water levels, topographic parameters, bed characteristics, friction coefficients, etc., ensuring that these values accurately represent the real-world conditions at the start of the simulation.
- Step 3 – Initial simulation and comparison with observed data: Run the model using the initialized conditions. Compare the simulation results (e.g., water level, current speed) with observed data to identify errors and discrepancies.
- Step 4 – Parameter adjustment: Based on the differences between simulation outputs and observed data, adjust the initial parameters (e.g., modify water levels, friction coefficients) to minimize errors. Rerun the model after each adjustment and continue comparing until the simulation error falls within an acceptable threshold.

2.4 Data base for oil spill modelling

2.4.1 Bathymetry

The bathymetry data for the Hai Phong marine area and its vicinity comprise the following sources:

- Terrestrial and underwater data (including bathymetry and nautical charts) provided by the Department of Survey and Mapping and the Navy at scales of 1:25,000, 1:100,000, and 1:500,000, published in the years 2000, 2005, 2010, and 2015;
- Bathymetric data and nautical charts from the Ministry of Natural Resources and Environment at a scale of 1:50,000, published in 2000 and 2005;
- Online data from GEBCO (General Bathymetric Chart of the Oceans), which aims to provide standardized and publicly available global ocean depth datasets.

These disparate datasets, initially referenced to different coordinate systems, were subsequently transformed and unified into a common geodetic reference framework. The author then utilized the MIKE modeling system to generate the computational mesh and bathymetric field required for subsequent simulations.

2.4.2 Meteorology

The meteorological data utilized in this study include records from a fixed meteorological station managed by the Ministry of Natural Resources and Environment (MONRE) and global reanalysis datasets obtained from international online sources.

- The fixed meteorological station used is the Hon Dau station (as presented in Table 2 and Figure 7).
- The data source employed is the global ERA5 reanalysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), extracted for the location at 107°00'E – 20°30'N (as shown in Figure 8).

Table 2 Information about Hon Dau meteorological and hydrological station

Station Name	Position		
	Place	Longitude (°E)	Latitude (°N)
Hon Dau	Hon Dau island, Do Son district, Hai Phong city	106° 48' 0.0"	20° 40' 0.0"

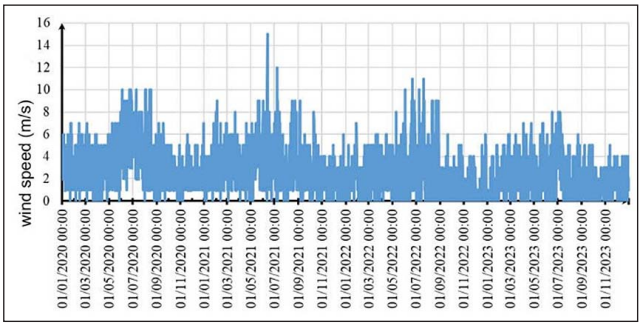

Figure 7 Wind speed variation at Hon Dau Station (Jan 2020 – Dec 2023)

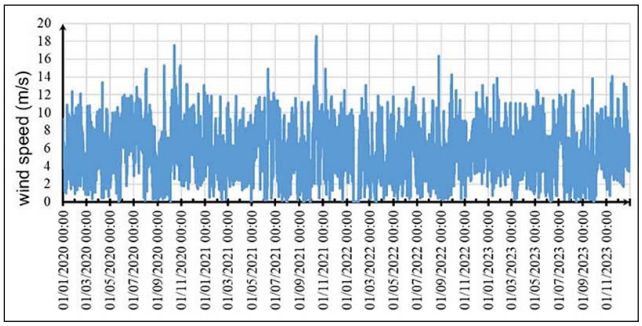

Figure 8 ERA5 global wind speed variation (ECMWF) at 107°00'E – 20°30'N (Jan 2020 – Dec 2023)

As the global wind data from the ERA5 reanalysis by the European Centre for Medium-Range Weather Forecasts (ECMWF) at the location 107°00'E – 20°30'N are available at an hourly temporal resolution and situated near the outer boundary, they were selected for incorporation into the computational model.

2.4.3 Hydrology

The hydrological data consist of records from fixed hydrological stations managed by the Ministry of Natural Resources and Environment (MONRE). The data from these conventional stations were collected on an hourly basis. The hydrological stations utilized within the study area include the Cua Cam station and the Do Nghi station (see Table 3, Figure 9, and Figure 10).

Table 3 Information on Hydrological Stations

Station Name	Position		
	Place	Longitude (°E)	Latitude (°N)
Hon Dau	Hon Dau island, Do Son district, Hai Phong city	106° 48' 0.0"	20° 40' 0.0"
Do Nghi	Tam Hung Ward, Thuy Nguyen City, Hai Phong City	106° 33'	20° 56'

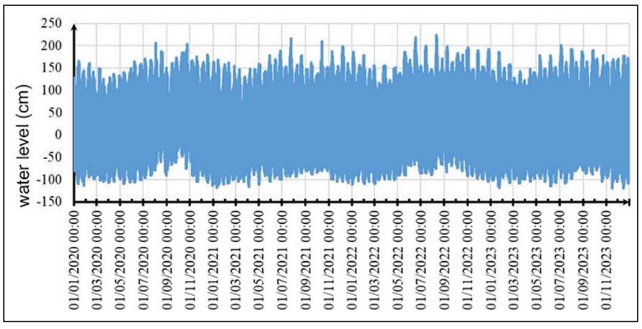

Figure 9 Water level variation at Cua Cam Station (Jan 2020 – Dec 2023)

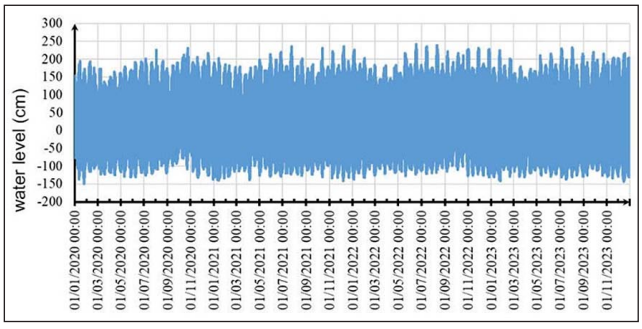

Figure 10 Water level variation at Do Nghi Station (Jan 2020 – Dec 2023)

2.4.4 Oceanography

The hydrographic data includes the fixed hydrographic station used at Hon Dau (Figure 11):

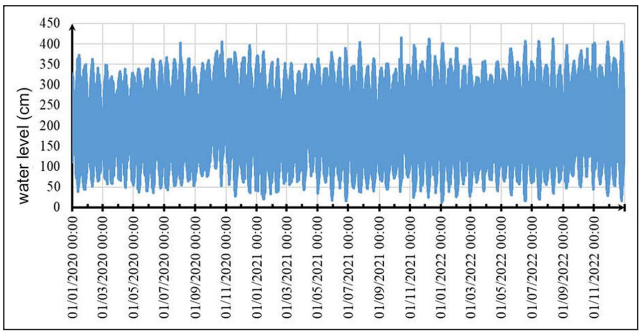

Figure 11 Water level variation at Hon Dau station (Jan 2020 – Dec 2022)

3 Results

3.1 Hydrodynamic simulation results

3.1.1 Domain, mesh and bathymetry

The meteorological, hydrological, and oceanographic data utilized in this dissertation include observations from the Hon Dau Meteorological–Oceanographic Station, the Do Nghi Hydrological Station, and the Cua Cam Hydrological Station, all of which are fixed stations managed by the Ministry of Natural Resources and Environment (MONRE), along with global reanalysis datasets obtained from international online sources.

The generation of the computational mesh is a critical step in the modeling and forecasting of tidal dynamics. The computational domain defined in the model encompasses the Hai Phong coastal area, bounded by longitudes 106°37.2' E to 107°06.6' E and latitudes 20°34.1' N to 20°56.4' N within the UTM Zone 48 coordinate system. The computational grid is an unstructured triangular mesh, with higher resolution applied to areas of elevated oil spill risk such as the Cua Cam estuary, the Do Nghi estuary, and the Hai Phong port region. In these zones, the maximum area of each mesh element does not exceed 1.5 km^2. The study domain consists of 14,522 nodes and 24,851 elements [23], with a minimum allowable internal angle of 32°, as shown in Figure 12.

3.1.2 Simulation calculation conditions and initial conditions

The modeling program adopts the integrated MIKE 21/3 FM system, utilizing two modules: the Hydrodynamic module and the Oil Spill module within ECO Lab. The simulation is configured with two open water boundaries at Cua Cam and Do Nghi, along with the coastal boundaries. In addition, for the oil spill module, it is necessary to specify parameters such as the type of oil, the spill source, and other essential input parameters (Figure 13 and 14).

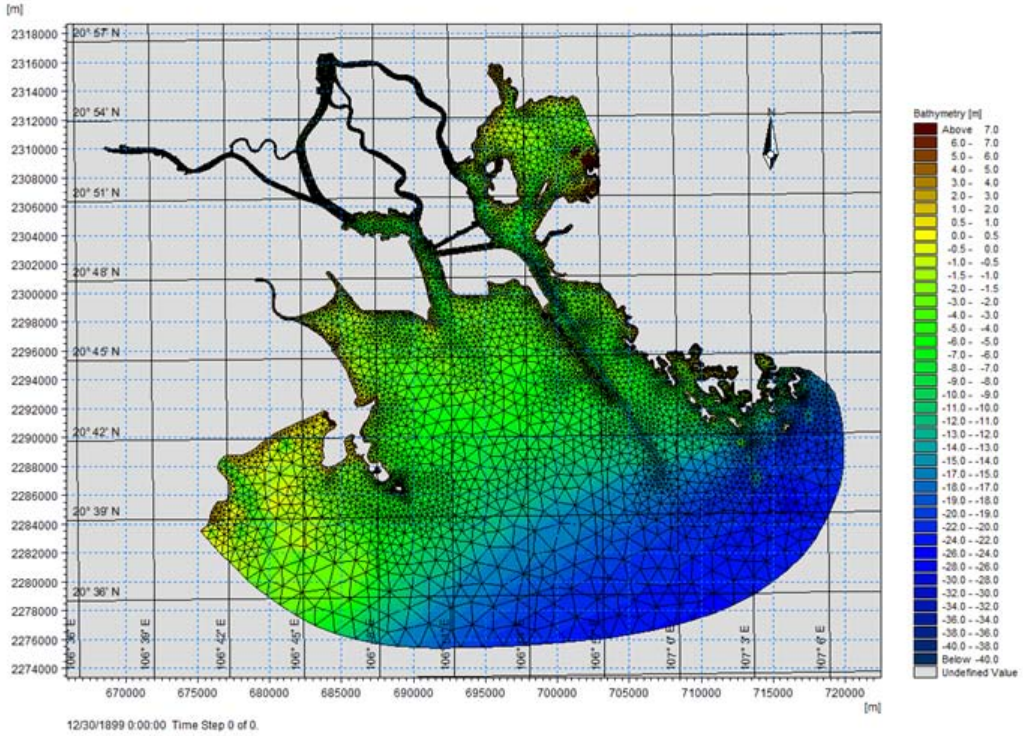

Figure 12 Domain and mesh

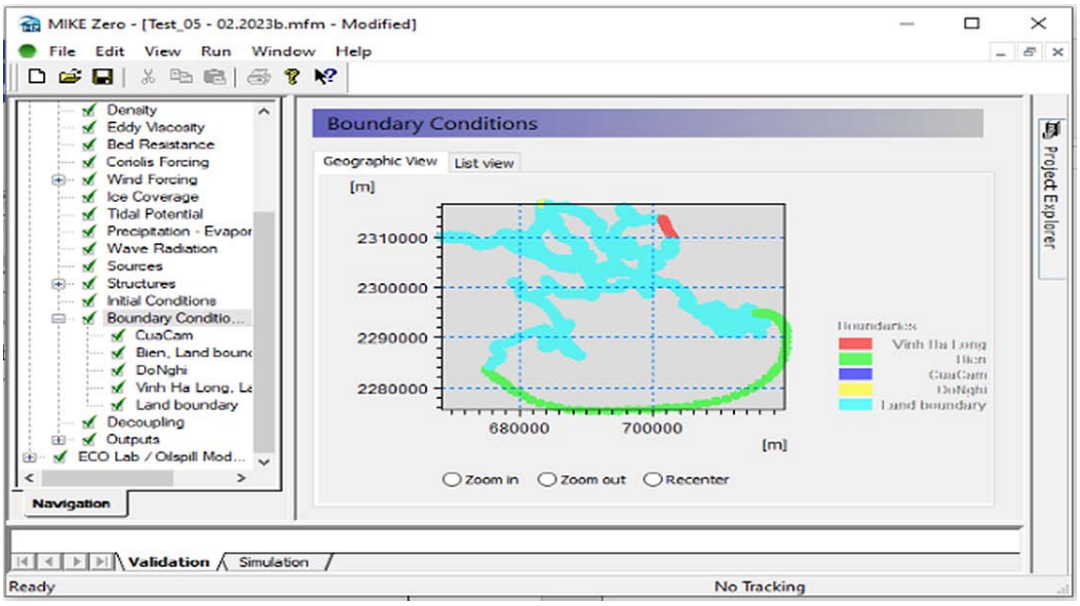

Figure 13 Selection of calculation conditions for the hydrodynamic module

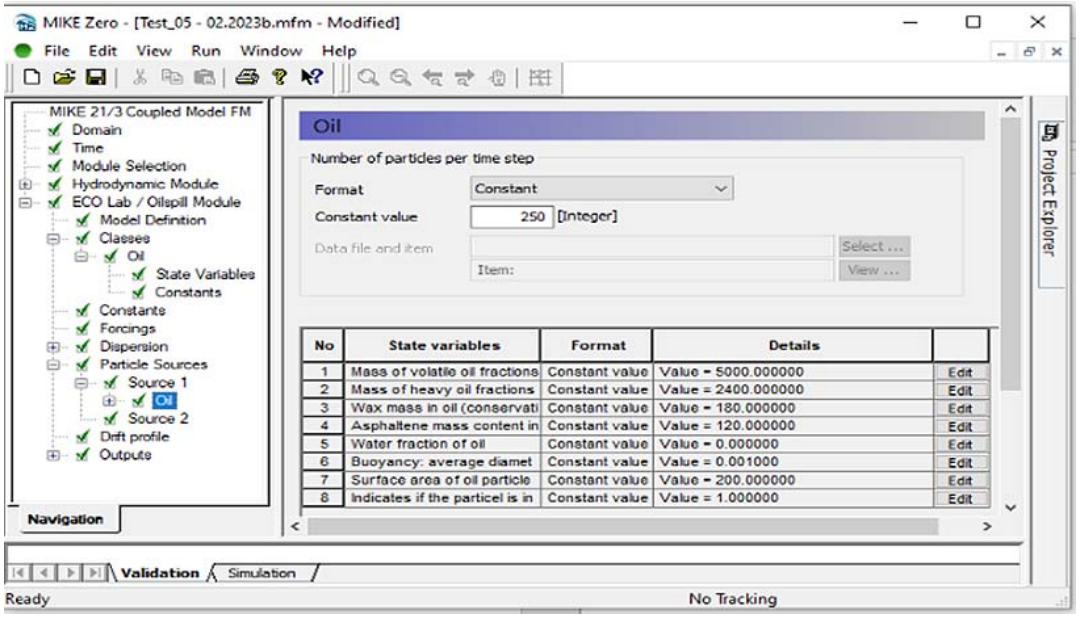

Figure 14 Select calculation conditions for oil spill module

3.1.3 Results of Calibration and Validation of the Hydrodynamic Model

The calibration results were assessed through comparisons between observed water levels from nautical charts (Hobs), national benchmark observations (Hobs_01), and simulated outputs (Hsim). The findings indicate a high level of agreement in both phase and amplitude. In instances where the hydrodynamic simulation did not meet accuracy requirements, further refinements were made to initial input parameters such as time step, wind fields, wave data, viscosity, and bottom friction. The overall outcomes demonstrate the stability of the computational model, reinforcing its reliability and validity.

Figure 15 illustrates the correlation between simulated water levels (Hsim) and observed values from the national monitoring station (Hobs_01) at the Hon Dau station, based on x and y coordinates. While there is a maximum amplitude difference of approximately 2 meters, the trend and magnitude of variation remain highly consistent. According to Table 4, correlation coefficients at all stations exceed 0.95, confirming the high accuracy of the model. Table 4 also presents the average comparison between measured and simulated data at Hon Dau

Table 4 Statistical accuracy assessment of Hon Dau water level (Jan 1–16, 2020)

Station Name	Elevation benchmark	Actual average	Average calculation	Evaluation index			
				R	RMSE	NASH	PBIAS (%)
Hon Dau	Nautical Chart	2.020	0.437	0.985	1.628	-2.718	79.390
	National	0.160		0.985	0.383	0.795	-160.786

Table 5 Statistical accuracy assessment of Hon Dau water level (Feb 26 – Mar 11, 2023)

No.	Characteristic		Hon Dau Station	
			Not calibrated	Calibrated
1	Actual average		1,956	0,096
2	Average calculation		0,153	
3	Evaluation index	R	0,975	0,975
4		RMSE	1,695	0,169
5		NASH	-4,597	0,944
6		PBIAS (%)	92,185	-58,606

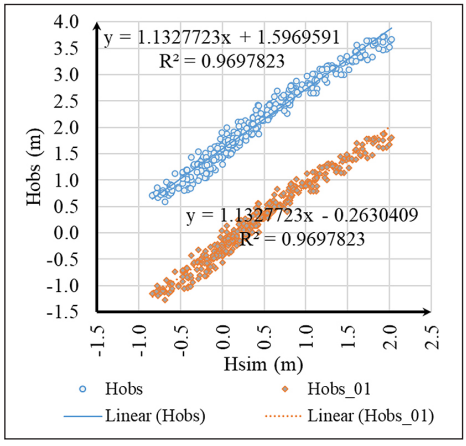

Figure 15 Water level comparison at Hon Dau station (Feb 26 – Mar 11, 2023)

station, with all statistical error indicators meeting acceptable thresholds. Specifically, the Nash-Sutcliffe efficiency coefficient (NASH = 0.795) falls within the range of high model performance (0.6–0.8), the PBIAS reflects minimal bias, and the RMSE value of 0.383 remains well below the 0.5 benchmark, further supporting model precision.

The results obtained from the model calibration and subsequent validation confirm that the hydrodynamic model demonstrates considerable stability and accuracy. The author proceeded to conduct further simulations incorporating the combined effects of wind, tidal forces, and river water level fluctuations at monitoring locations such as Cua Cam and Do Nghi hydrological stations. The alignment between simulated results and observed data at the Hon Dau marine station during the same period was examined through time series comparisons of water level fluctuations, supported by key statistical performance indicators including the correlation coefficient (R), root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE), and percent bias (PBIAS).

On that basis, the hydrodynamic simulation was applied for February 2023, coinciding with the timeframe during which an unidentified oil spill incident was reported near Ben Got, Lach Huyen, Hai Phong. Table 5 presents the simulation outcomes evaluated at Hon Dau station, reflecting both the observed and simulated average values, as well as associated error metrics. Overall, the indicators suggest a high level of model accuracy, with correlation coefficients exceeding 0.95. The NSE value falls within the range of very high predictive performance (0.8–1.0), while PBIAS values also confirm high reliability, showing a tendency toward overestimation (values<0). Additionally, the RMSE of 0.169 remains significantly below the 0.5 threshold, further validating the model's precision.

Figure 16 and 17 illustrates the hydrodynamic simulation results corresponding to high tide and low tide phases. The analysis shows that high-velocity flow zones are predominantly located in regions such as the outer Nam Trieu channel, Lach Tray estuary, Cai Trap canal, and the Lach Huyen fairway. During high tide, current speeds are generally moderate to low. The current direction changes in response to tidal fluctuations, following two distinct patterns:

At high tide, when the tide is rising, seawater intrudes inland, and the current typically flows from the sea toward the river systems.

At low tide, as the water level falls, the current direction reverses, and the current moves from the rivers back out to the sea.

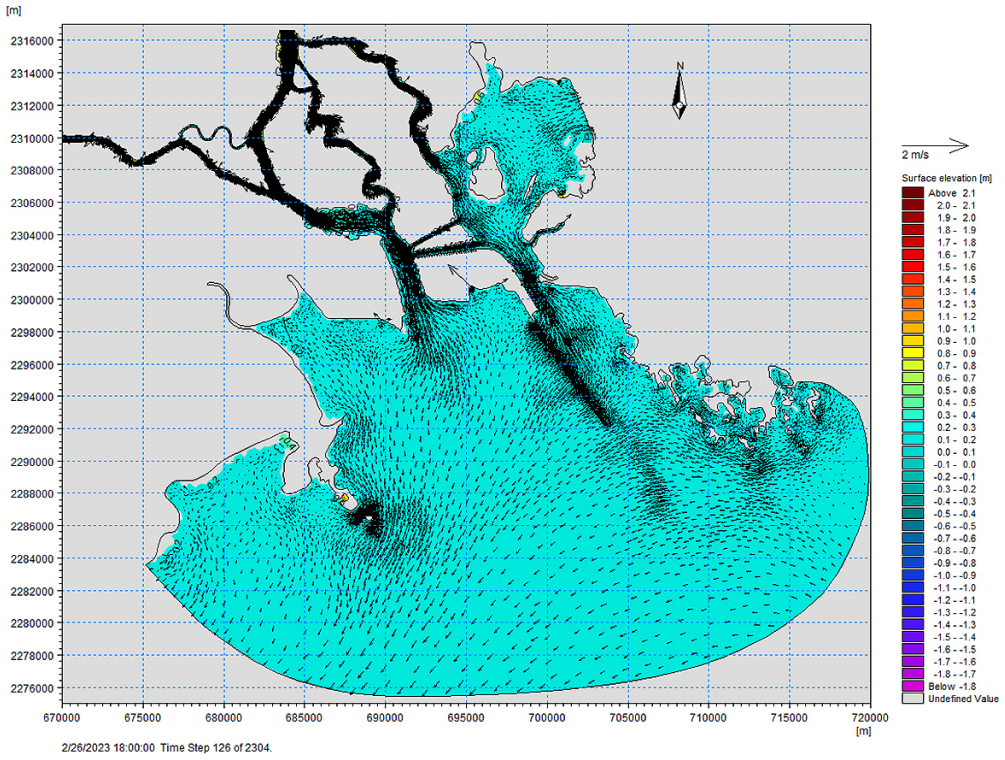

Figure 16 Low tide level at 01:00 on Feb 27, 2023

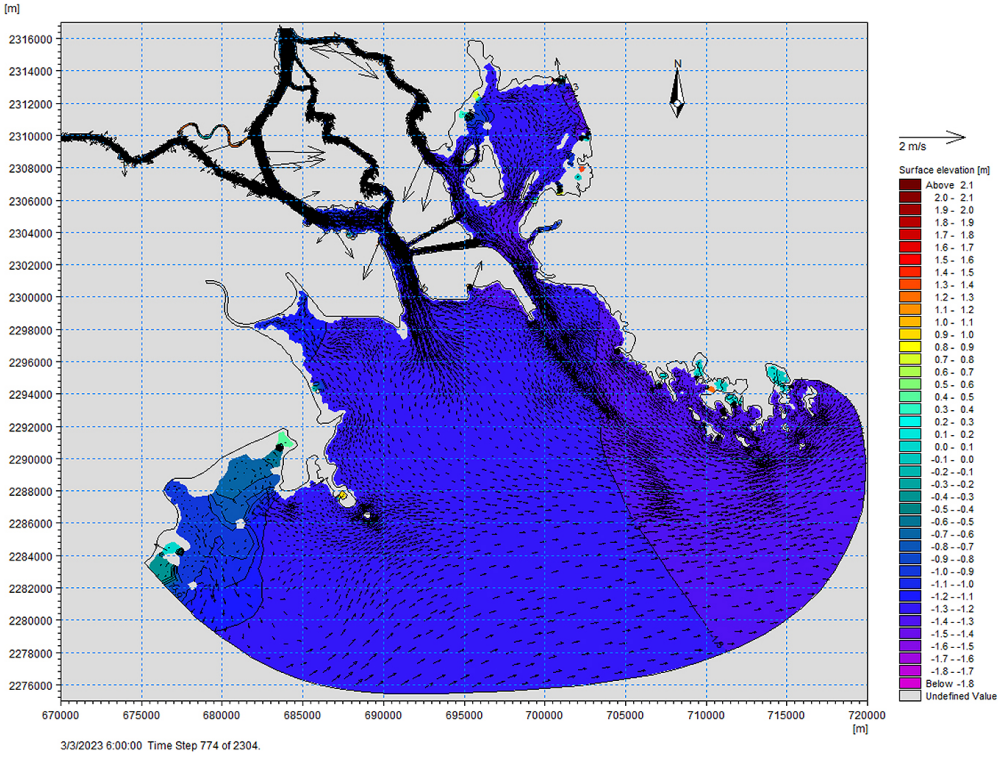

Figure 17 High tide level at 13:00 on Mar 3, 2023

Additionally, current speed varies spatially. Under high tide conditions, current speed tends to increase progressively from the open sea toward the river mouths. Conversely, during low tide, current speed diminishes as it moves seaward from the inner riverine areas.

3.2 Oil spill simulation results

3.2.1 Oil Spill Simulation Results Using the MIKE Model

To simulate the oil spill incident that occurred in Hai Phong in February 2023, the MIKE 21/3 Coupled FM module was utilized. The initial modeling conditions were systematically defined to ensure the accuracy and reliability of the simulation framework (Table 6).

Figure 18 presents a composite trajectory of oil particle movement from February 26 to March 11, 2023. Influenced by external environmental forces, the oil slick was predominantly concentrated in the Nam Trieu channel, Lach Huyen channel, and the Hai Phong Bay area.

In February 2023, an unidentified oil slick was detected near the Lach Huyen port area (Figure 19). Initially influenced by ebb tide conditions, the slick drifted seaward, moving beyond Berths No. 3 and 4. On February 27, the slick showed a tendency to shift in a north-northwest direction, approaching the vicinity of the VinFast factory port. In the following days, the prevailing movement of the oil slick alternated between southeast and northwest directions.

3.2.2 Oil spill calculation program using Lagrange method

Based on the theoretical framework and algorithmic flowchart for forecasting oil spill dispersion in the marine environment using the Lagrangian approach (Figure 1), the author developed a computational application using

Table 6 Selected Parameters of the Oil Spill Transport Model

No.	Content	Parameters
1	Input layers	Standardized according to the model (DHI Oil Weathering Model 2013)
2	Spill Sources	Lach Huyen estuary area near Got wharf
3	Type of oil spill	Oil spill source in first step calculation
4	Advection and Diffusion	The horizontal diffusion coefficient is set at 0.025 m^2/s; The vertical dispersion coefficient is 1.0×10^{-5} m^2/s.
5	Advection	Using data sources from hydrodynamic models
6	Bottom Roughness	Using data sources from hydrodynamic models
7	Output	Based on hydrodynamic data and time step

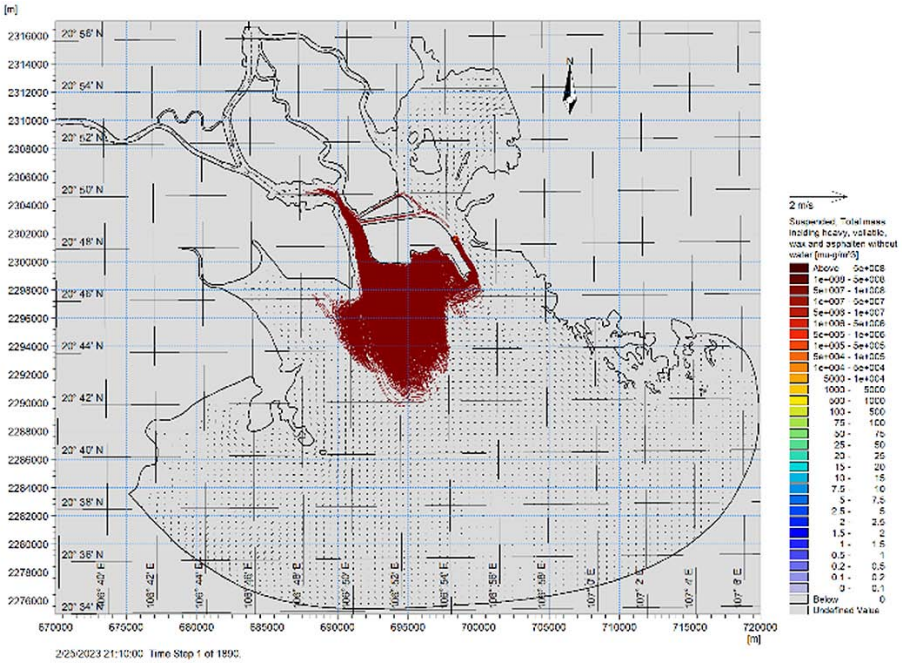

Figure 18 Fate of oil spill from Feb 26 to Mar 11, 2023

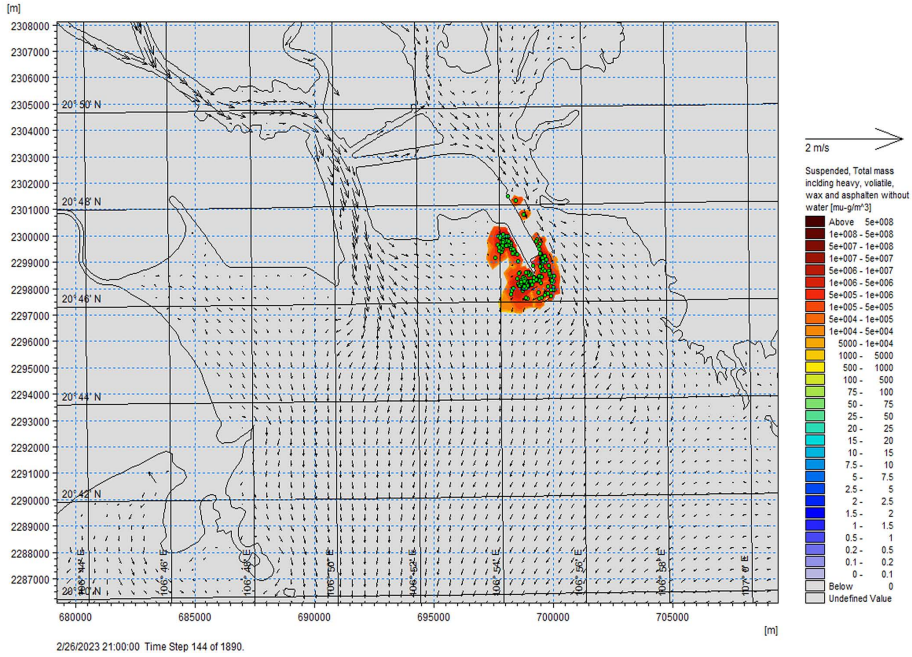

Figure 19 Oil spill and droplet locations at 04:00 on Feb 27, 2023

Fortran language. The model structure is organized into the following components (Figure 20).

According to the block diagram presented in Figure 2, the program begins by reading input data, including computational domain, simulation period and time step, and other auxiliary information required for the simulation (see Figure 21).

As illustrated in the block diagram in Figure 3, the program processes environmental factors (e.g., waves, wind, air temperature) by harmonizing them onto a unified coordinate system and computational mesh, ensuring temporal synchronization across all datasets (see Figure 22).

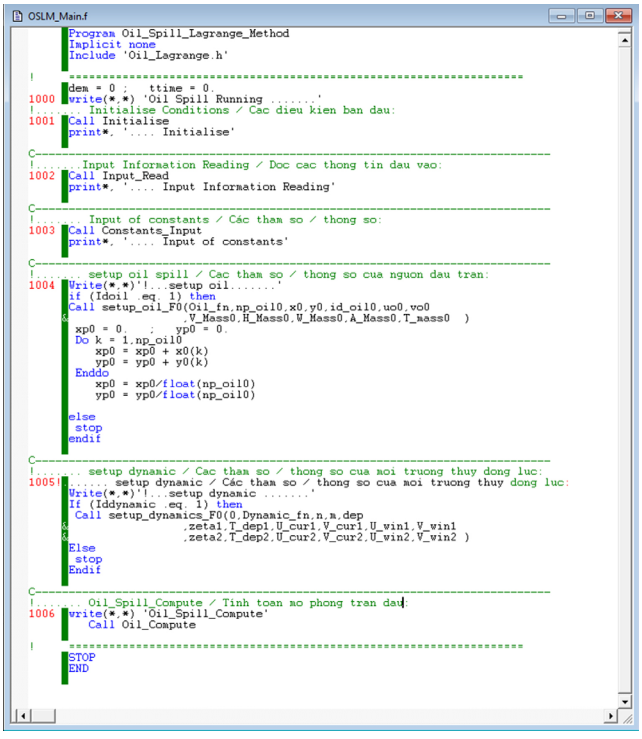

Figure 20 Main oil spill simulation calculation program

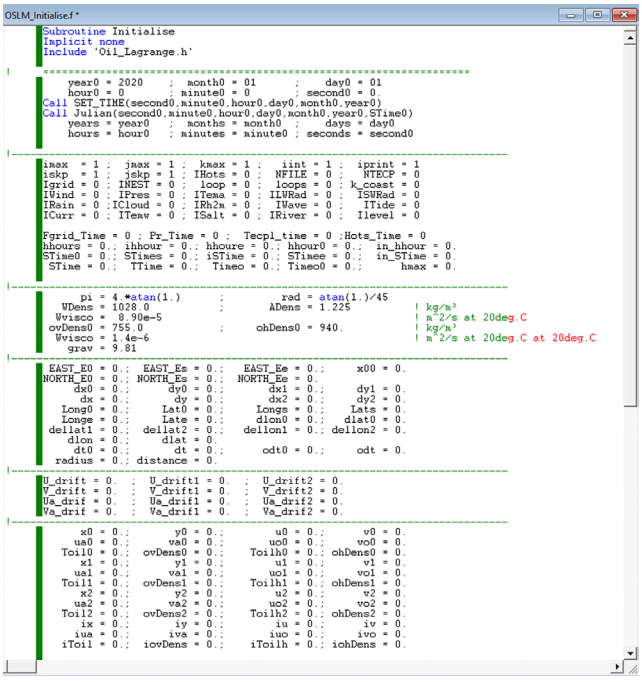

Figure 21 Program to read input information

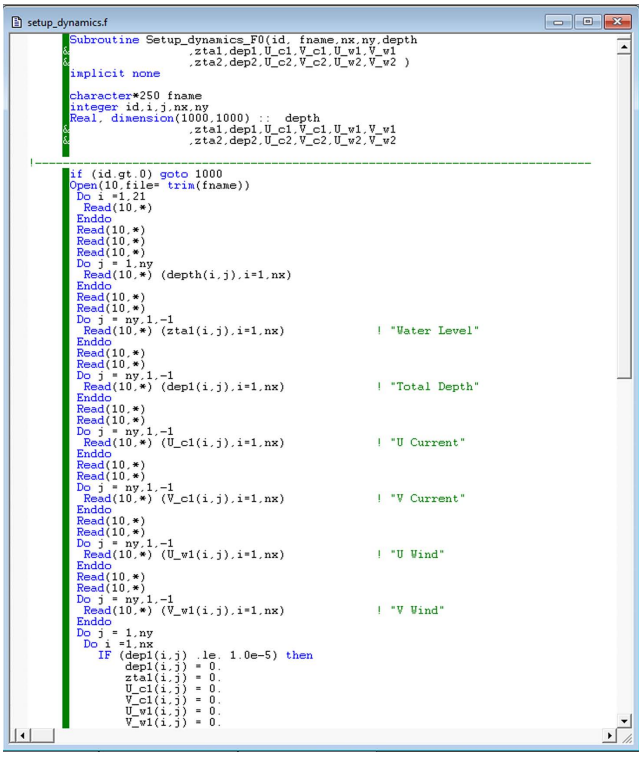

Figure 22 Subroutine for updating and processing environmental factors

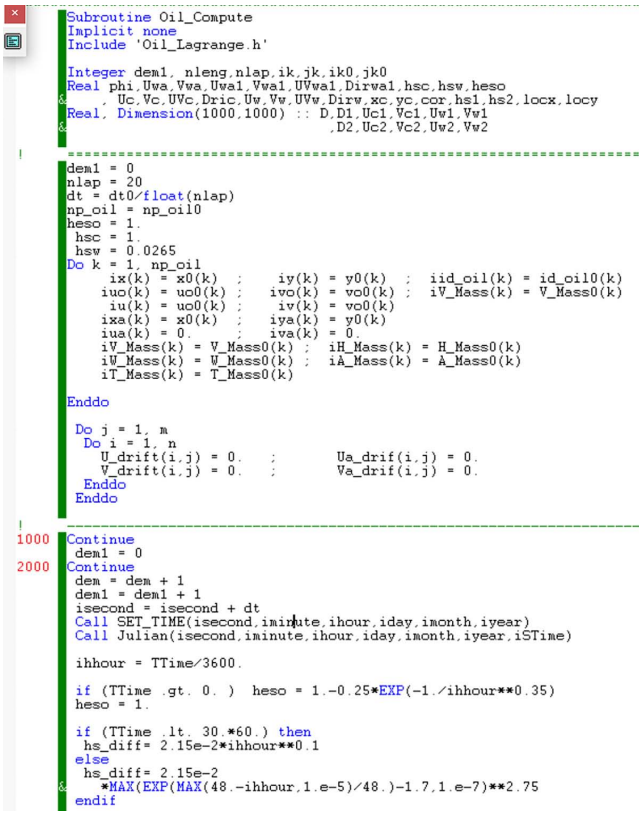

Figure 23 Subroutine for calculating advection and diffusion of oil particles

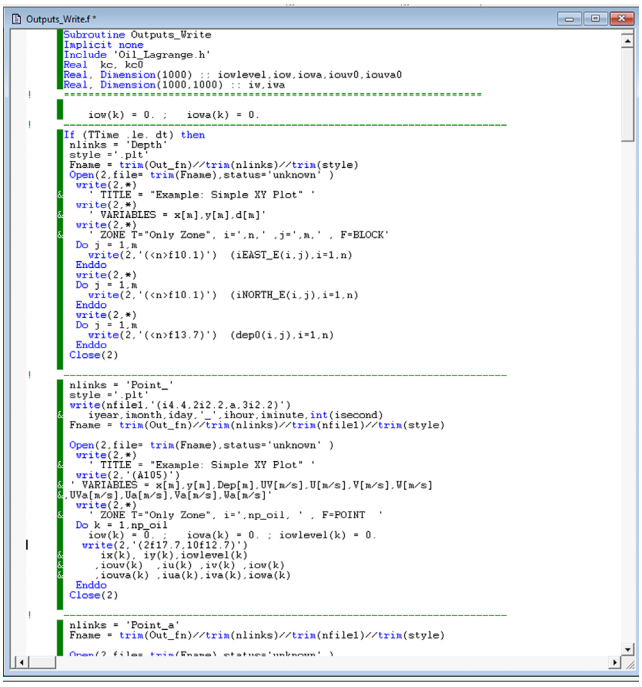

Figure 24 Subroutine to print calculation results

The advection and diffusion processes of surface oil spills are computed using the algorithm illustrated in the block diagram in Figure 4. This module plays a critical role in determining the movement of oil particles under the influence of hydrodynamic environmental factors (see Figure 23).

After computing the movement of oil particles, this subroutine extracts all relevant information related to the oil spill process, as depicted in the block diagram in Figure 5 (see Figure 24).

3.2.3 Analysis and comparison of results of two oil spill models

Figure 25 illustrates a comparative analysis of oil slick trajectories represented by particle positions simulated using the MIKE model (depicted as purple squares) and the custom-built model (represented as green circles) on Feb 27, 2023 (after one day), under the oil spill scenario at Ben Got, Lach Huyen estuary [24]. In the initial phase, the trajectories generated by both models show a high degree of consistency. However, in the subsequent days, particles in the custom model tend to drift southward, spreading more extensively across Hai Phong Bay, whereas particles in the MIKE model exhibit a southward movement predominantly directed into the Nam Trieu channel.

Table 7 presents a detailed comparison of three parameters: the average distance from the spill origin to

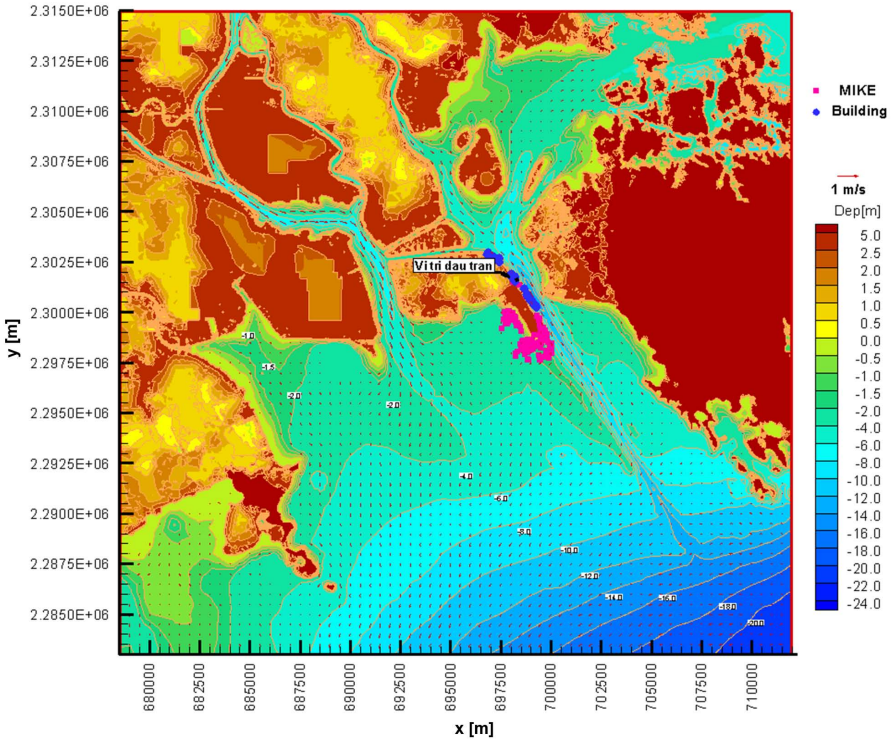

Figure 25 Oil Slick at 04:00 on Feb 27, 2023 (after one day)

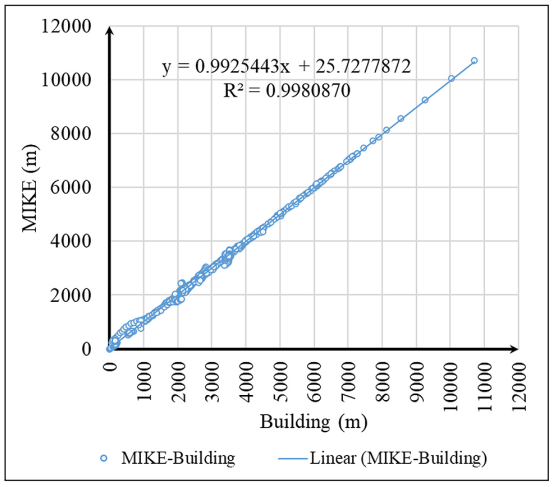

Figure 26 Correlation of oil particle distances: MIKE vs. custom model (Feb 26 – Mar 4, 2023)

Table 7 Comparison of Statistical Accuracy in Oil Particle Distances between the MIKE and Custom Model (Feb 26 – Mar 4, 2023)

Average Distance from the Spill Location to the Nearest Particle Position [m]	MIKE	2966.632
	Building	2962.996
Minimum Distance Between MIKE and Custom Model Particle Positions [m]	Maximum	365.298
	Average	63.647
Evaluation Metrics for the Distance from the Spill Origin to the Nearest Particle Positions between the MIKE and Custom Models	R	0.999
	RMSE	66.158
	NASH	0.998
	PBIAS (%)	0.123

the nearest particle position is 2,966.632 meters for the MIKE model and 2,962.996 meters for the custom model, resulting in a minimal difference of 3.6 meters. The maximum deviation between the nearest particle positions of the two models is 365.298 meters, with an average discrepancy of 63.647 meters. All statistical performance indicators meet the required standards: the correlation coefficient (R) exceeds 0.95; the Nash–Sutcliffe efficiency (NASH) is 0.998, indicating excellent agreement (within the range 0.8–1.0); and the percent bias (PBIAS) is 0.123%, approaching the optimal value of zero. Figure 26 displays a first-order regression plot comparing particle distances from both models. The linear pattern of the plot confirms a high level of consistency between the two simulation approaches.

4 Discussion

This study introduces an effective approach to enhance oil spill simulation accuracy using refined hydrodynamic modeling combined with a Lagrangian discrete particle method. The model leverages multiple statisti-

cal metrics, including the correlation coefficient (R), Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and Percent Bias (PBIAS), ensuring a systematic calibration process. High statistical alignment with observational data was achieved (R > 0.95, RMSE < 0.5, NSE between 0.795 – 0.944), indicating strong model reliability.

The comparative evaluation with the established MIKE 21 FM model confirmed the proposed model's validity, as both models yielded closely aligned initial particle trajectories. Over extended simulations, minor deviations emerged due to intrinsic computational differences. However, the positional discrepancy between models remained minimal, averaging approximately 63.6 meters. This minor divergence reinforces the robustness and practicality of the developed methodology for operational forecasting and environmental risk assessment.

Model accuracy was notably influenced by the quality and resolution of input environmental datasets, including bathymetric, meteorological, and oceanographic data. Hence, continuous updates and integration of real-time observations are recommended to further enhance predictive performance. Future research should extend model validation across diverse environmental conditions, thereby broadening applicability and enhancing preparedness for various oil spill scenarios.

5 Conclusions

This study demonstrates a successful application of advanced hydrodynamic modeling combined with error evaluation techniques to improve oil spill simulations in the Hai Phong coastal region. By integrating statistical metrics (R, RMSE, NSE, PBIAS), the model achieved high accuracy and alignment with observational data, confirming its reliability.

The developed Lagrangian particle-based simulation model produced results closely matching those from the MIKE 21 FM system. Despite minor variations in particle trajectories, the average positional difference remained small (~63.6 m), underscoring the model's robustness.

Accurate environmental datasets played a critical role in enhancing simulation fidelity, highlighting the need for high-resolution and continuously updated input data. This framework supports more reliable forecasting and efficient emergency response in marine environments.

5.1 Limitations of the study

The model has not yet been validated or calibrated using current velocity data; calibration has so far been limited to water level data from observation stations. This limitation is primarily due to the high cost and technical complexity of flow measurement equipment, which makes it difficult to deploy widely across multiple locations. Furthermore, current velocity data are typically collected during short-term measurement campaigns at a limited number of sites and are rarely available as continuous long-term records, posing challenges for comprehensive model validation. The approach of using only water level data for validation is, however, common practice in both national and international hydrodynamic modeling studies.

In addition, the current model focuses solely on the positional accuracy of the oil slick and does not yet incorporate or validate ecological, environmental, or economic impacts resulting from oil spills.

5.2 Recommendations for future research

Future research should focus on enhancing the model's operational capabilities by incorporating real-time environmental data—such as wind, wave, and current forecasts—from observation networks or remote sensing systems. This would enable the model to support near-real-time spill prediction and emergency response applications.

To improve the model's utility in environmental risk assessment and decision-making, future studies should also integrate ecological and environmental impact evaluation modules, allowing the simulation results not only to forecast oil slick trajectories but also to estimate potential damage to marine ecosystems, protected areas, and socio-economic assets (e.g., aquaculture zones, port operations, and tourism infrastructure).

Such advancements will strengthen the model's relevance for comprehensive marine pollution management, strategic planning, and sustainable coastal zone governance.

Funding: The research presented in the manuscript did not receive any external funding.

Acknowledgments: The authors sincerely thank all co-authors for their valuable contributions, insightful discussions, and collaboration throughout this research.

Author Contributions: Research and writing, Do Van Cuong; Data collection, Do Trung Kien; Review and editing, Nguyen Manh Cuong; Verification, Nguyen Xuan Long. All authors have read and agreed to the published version of the manuscript.

References

[1] Wilson, M., et al. Oil spill modelling: From the ocean to humans: Integrated modelling of oil spill impacts, 2021.

[2] Nordam Tor, et al., Improving oil spill trajectory modelling in the Arctic, Marine Pollution Bulletin, 2019, Volume 140, pp. 65-74. https://doi.org/10.1016/j.marpolbul.2019.01.019.

[3] D. Atodiresei, C. Popa, and V. Dobref. Simulating Oil Spill Evolution and Environmental Impact with Specialized Software: A Case Study for the Black Sea. Sustainability, vol. 17, no. 9, Art. no. 3770, 2025. https://doi.org/10.3390/su17093770.

[4] A. Martínez, A. J. Abascal, A. García, G. Aragón, and R. Medina. Lagrangian modelling of oil concentrations at sea: A sensitivity analysis to the grid resolution and number of Lagrangian elements. Marine Pollution Bulletin, vol. 198, Art. no. 115787, Jan. 2024. https://doi.org/10.1016/j.marpolbul.2023.115787.

[5] K. Kampouris, V. Vervatis, J. Karagiorgos, and S. Sofianos. Oil spill model uncertainty quantification using an atmospheric ensemble. Ocean Science, vol. 17, no. 4, pp. 919–934, 2021. https://doi.org/10.5194/os-17-919-2021.

[6] N. Mardani, K. Suara, H. Fairweather, R. Brown, A. McCallum, and R. C. Sidle. Improving the Accuracy of Hydrodynamic Model Predictions Using Lagrangian Calibration. Water, vol. 12, no. 2, Art. no. 575, 2020. https://doi.org/10.3390/w12020575.

[7] Tri, D.Q., Don, N.C., Ching, C.Y. et al. Application of environmental sensitivity index (ESI) maps of shorelines to coastal oil spills: a case study of Cat Ba Island, Vietnam. Environ Earth Sci 74, 2015, pp. 3433–3451. https://doi.org/10.1007/s12665-015-4380-0.

[8] Nguyen Quoc Trinh and Nguyen Quang Vinh. Research and development of backward-in-time oil spill simulation at East Vietnam Sea. PetroVietnam Journal, Vol. 02, pp. 60-68, 2018.

[9] MIKE 21 and MIKE 3 Flow Model FM Hydrodynamic and Transport Module, Scientific Documentation, DHI, 2014.

[10] E. Mignot, N. Riviere, and B. Dewals. Formulations and diffusivity coefficients of the 2D depth-averaged advection-diffusion models: A literature review. Water Resources Research, vol. 59, no. 12, p. e2023WR035053, Dec. 2023. https://doi.org/10.1029/2023WR035053.

[11] J. Kafle, K. P. Adhikari, and E. P. Poudel. Air pollutant dispersion using advection-diffusion equation. Nepal Journal of Environmental Science, vol. 12, no. 1, pp. 1–6, Jul. 2024. https://doi: 10.3126/njes.v12i1.47531.

[12] T. Schlick. Molecular Modeling and Simulation: An Interdisciplinary Guide. Interdisciplinary Applied Mathematics, vol. 21, Springer, 2002, pp. 480–494.

[13] D. L. Ermak and J. A. McCammon. Brownian dynamics with hydrodynamic interactions. The Journal of Chemical Physics, vol. 69, no. 4, pp. 1352–1360, 1978. https://doi: 10.1063/1.436761.

[14] R. J. Loncharich, B. R. Brooks, and R. W. Pastor. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers, vol. 32, no. 5, pp. 523–535, May 1992. https://doi: 10.1002/bip.360320508.

[15] B. Hackett, Ø. Breivik, and C. Wettre. Forecasting the drift of objects and substances in the ocean. Ocean Weather Forecasting: An Integrated View of Oceanography, E. P. Chassignet and J. Verron, Eds. Dordrecht: Springer, 2006, pp. 507–523.

[16] D. Mackay, I. Buist, R. Mascarenhas, and S. Paterson. Oil Spill Processes and Models. Manuscript Report No. EE-8, Environmental Impact Control Directorate, Environment Canada, Ottawa, ON, 1980.

[17] R. García-Martínez and H. Flores-Tovar. Computer modeling of oil spill trajectories with a high accuracy method. Spill Science & Technology Bulletin, vol. 5, no. 5–6, pp. 323–330, 1999.

[18] J. F. Kenney and E. S. Keeping. Linear Regression and Correlation. Mathematics of Statistics, Part 1, 3rd ed., Princeton, NJ: Van Nostrand, 1962, pp. 252–285.

[19] J. F. Kenney and E. S. Keeping. Root Mean Square. Mathematics of Statistics, Part 1, 3rd ed., Princeton, NJ: Van Nostrand, 1962, pp. 59–60.

[20] J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, vol. 10, no. 3, pp. 282–290, Apr. 1970. https://doi: 10.1016/0022-1694(70)90255-6.

[21] E. R. Gundlach. Oil-holding capacities and removal coefficients for different shoreline types to computer simulate spills in coastal waters. Proceedings of the 1987 Oil Spill Conference, Washington, DC, USA, 1987, pp. 451–457.

[22] H. V. Gupta, S. Sorooshian, and P. O. Yapo. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. Journal of Hydrologic Engineering, vol. 4, no. 2, pp. 135–143, Apr. 1999. https://doi: 10.1061/(ASCE)1084-0699(1999)4:2(135).

[23] Mike Zero: Creating 2D Bathymetries, Bathymetry Editor & Mesh Generator, Scientific Documentation, DHI 2017.

[24] https://thanhphohaiphong.gov.vn/hai-phong-co-ban-khac-phuc-xong-su-co-tran-dau-tai-ben-pha-got.html (Last visit on April 2025).