Ecological Informatics 91 (2025) 103368

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Improving oil slick trajectory simulations with Bayesian optimization

Gabriele Accarino [a,b], Marco M. De Carlo [a,d], Igor Ruiz Atake [a,*], Donatello Elia [a], Anusha L. Dissanayake [a], Antonio Augusto Sepp Neves [f], Juan Peña Ibañez [f], Italo Epicoco [a,c], Paola Nassisi [a], Sandro Fiore [e], Giovanni Coppini [a]

[a] CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy
[b] Department of Earth and Environmental Engineering, Columbia University, New York, USA
[c] Department Engineering for Innovation, University of Salento, Italy
[d] National Institute of Oceanography and Applied Geophysics - OGS, Italy
[e] Department of Information Engineering and Computer Science, University of Trento, Italy
[f] Orbital EOS - Earth Observation Solutions, Spain

ARTICLE INFO

Keywords:
Oil spill modeling
Bayesian optimization
Stochastic parameterization
Data-driven frameworks
Environmental risks

ABSTRACT

Accurate simulations of oil spill trajectories are essential for supporting practitioners' response and mitigating environmental and socioeconomic impacts. Numerical models, such as MEDSLIK-II, simulate advection, dispersion, and transformation processes of oil particles, but their accuracy depends strongly on the correct tuning of physical parameters, often relying on manual calibration and expert knowledge. This approach is suboptimal, especially in dynamic and uncertain environmental conditions. To overcome these limitations, we couple the MEDSLIK-II oil spill model with a Bayesian optimization framework to iteratively estimate the optimal values of key parameters, such as the horizontal diffusivity, wind angle and wind drag, in order to obtain simulation closer to satellite observations of the slick. We adopt a stochastic parameterization strategy, which probabilistically explores the parameter space to enhance simulation skill. To this end, the Fraction Skill Score (FSS) is maximized to evaluate spatial-temporal overlap between simulated and observed oil distributions. The framework is validated for the Baniyas oil incident that occurred in Syria between August 23 and September 4, 2021, which released over 12,000 m^3 of oil. The approach improves FSS from 7. 97 % to 20. 66 %, on average, compared to control simulations initialized with default parameters. Results demonstrate consistent improvements across time steps, highlighting the method's robustness and suitability for operational oil spill modeling under uncertainty.

1. Introduction

Oil spill pollution has historically been linked to rare but catastrophic events like *Deepwater Horizon oil and gas well blowout*, *Prestige*, *Erika*, and *Jyieh* spills (Boufadel et al., 2023; Balseiro et al., 2003; Daniel et al., 2001; Coppini et al., 2011). Since the 1970s, organizations such as the International Tanker Owners Pollution Federation Limited and the Regional Marine Pollution Emergency Response Centre for the Mediterranean Sea have documented accidental spills, creating a valuable long-term dataset (Polinov et al., 2021). However, limitations in satellite revisit times and coverage hinder real-time detection, especially during initial spill formation, critical for mitigating damage. Thus, numerical models remain essential for simulating oil spill dynamics and assessing broader consequences (Jones and Brischke, 2017; Mills, 2016). While oil spill modeling has advanced (Huang, 1983; Keramea et al., 2021; Reed et al., 1999; Spaulding, 1988, 2017), drift simulation methods, predominantly Eulerian or Lagrangian, have changed little (Barker et al., 2020).

The Lagrangian approach tracks oil as discrete elements, each representing a droplet size class that moves according to physical processes (Barker et al., 2020). The DeepWater Horizon spill highlighted key limitations in existing models, including gaps in observed oil spill in the water column and uncertainties in wind, wave, and current inputs (Spaulding, 2017; Boufadel et al., 2020; Boufadel et al., 2023; Keramea et al., 2023a). However, it also drove the introduction of some advancements, such as Lagrangian coherent structures to capture features

This article is part of a Special issue entitled: 'Ocean Learning' published in Ecological Informatics.

* Correspoding author.
E-mail address: igor.atake@cmcc.it (I. Ruiz Atake).

https://doi.org/10.1016/j.ecoinf.2025.103368
Received 26 February 2025; Received in revised form 9 June 2025; Accepted 1 August 2025
Available online 6 August 2025

like the "tiger tail" (Peacock and Haller, 2013). Despite progress, operational spill modeling still struggles with emergency response demands requiring fast, accurate trajectory predictions (Barker et al., 2020).

Models' effectiveness greatly relies on accurate input and environmental parameters, often derived from theoretical models or expert judgment, introducing uncertainty in representing complex ocean and weather dynamics (Caron et al., 2019; De Dominicis et al., 2012; De Dominicis et al., 2013a; Wunsch, 1998). To account for unresolved processes, parameterization schemes are commonly employed (Christensen and Zanna, 2022). Traditional deterministic parameterizations assume that each grid cell is large enough to average over subgrid phenomena while still capturing local variability (Arakawa and Schubert, 1974). However, the scale invariance of governing equations can give rise to emergent behaviors and power law dynamics (Storer et al., 2022). To better represent the phenomena, stochastic parameterizations have been proposed, showing significant improvements in reducing long-term biases and aligning with theoretical validations for probabilistic modeling (Berner et al., 2017; Gettelman et al., 2021; Gottwald et al., 2017). In this context, refining model calibration becomes crucial for improving forecast reliability and enabling timely emergency response. The increasing availability of high-resolution remote sensing capabilities (Fingas and Brown, 2013; Fingas and Brown, 2018, 2025; Leifer et al., 2012), offers valuable opportunities to analyze more events, thereby supporting the systematic refinement and validation of numerical models.

Conventional calibration methods, such as manual tuning, exhaustive search strategies like Grid Search (Belete and Manjaiah, 2021) and Random Search (Andradóttir, 2006) or genetic algorithms (Bengtsson et al., 2013), are computationally expensive and inefficient in high-dimensional parameter spaces. Bayesian optimization provides a powerful stochastic alternative by efficiently optimizing complex, expensive-to-evaluate functions (Frazier Peter, 2018). It leverages previous evaluations to prioritize promising parameter regions building a surrogate probabilistic model that guides the search by balancing exploration and exploitation. In contrast to other approaches (Shin and Baik, 2022; Wang et al., 2022), the Bayesian framework presented here can capture spatial correlations implicitly through an integration of an online-trained surrogate model, even when working with very sparse observations. This model accounts for uncertainty while preserving physical coherence via carefully constrained parameter bounds. For this reason, the Bayesian optimization is effective for physical parameters calibration where both computational resources and observational data are limited. This approach has proven effective in diverse fields, including materials and drug design (Negoescu et al., 2011; Frazier and Wang, 2016; Packwood, 2017; Zhang et al., 2020), reinforcement learning (Brochu et al., 2009; Lizotte et al., 2010), and hyperparameter tuning in deep learning (Ranjit et al., 2019; Snoek et al., 2012; Snoek et al., 2015).

In this work, we integrate the Bayesian optimization framework with the MEDSLIK-II oil spill model (De Dominicis et al., 2013b) to estimate key physical parameters from initial oil slick observations, offering a scalable solution for environmental applications. By refining physical parameters based on observations, this approach aims to (i) enhance model accuracy and reduce modeling biases in the input forcing, and (ii) improve both rapid-response forecasting and short-term (up to 3 days) environmental assessments, making the approach applicable in an operational context. Oil spill modeling is well-suited for Bayesian Optimization due to the high computational cost of simulations and sparse observational data, allowing efficient parameter tuning via a surrogate model while preserving physical consistency. This has been recently explored in other studies (Liu et al., 2023), where an oil spill model was optimized with a Support Vector Machine (SVM) technique as the model ran, but have never been applied to a community model, where we apply the optimization technique using an "offline" method, coupling a machine learning tool and a physical based oil spill model. Additionally, a recent study by (Breivik et al., 2025) utilized a Bayesian approach to perform and assess backtracking surface Lagrangian trajectory simulations, which can be applied to oil spill simulations considering their inherent uncertainty. While previous studies have explored parameterization spaces and Bayesian techniques, our study focuses on optimizing a model using a Bayesian acquisition function. The article is structured as follows: Section 2 details the data and the methodology used to develop and evaluate the proposed solution, Section 3 presents results, and Section 4 discusses implications, limitations, and future research directions.

2. Materials and methods

2.1. Study area and event

We focus on the Baniyas Thermal Station spill in Syria, caused by prolonged infrastructure neglect after the civil war. Over 12,000 m^3 of oil were released between August 23 and September 4, reaching as far as Cyprus (Abou Samra and Ali, 2024). During this period, satellite imagery, Optical and SAR, captured the evolution of the slick across the Levantine Basin. These data were provided by OrbitalEOS[1] in the context of the iMagine[2] project, published the dataset on Zenodo[3] (Ferrer et al., 2024).

This case was selected due to the unusually dense satellite coverage available, respect to the other cases, along the entire oil slick trajectory. We focus on the first three days of the event (August 24–27, 2021) due to the extension of the oil slick and the complex mesoscale dynamics that created "tiger tail" features, which are elongated oil spill regions caused by sub mesoscale and mesoscale interactions (Özgökmen et al., 2016), which can aggregate oil in thin stripes along the wind direction and transport it further from the oil thick region. Apart from the Syria Oil Spill event, tiger tail features were also seen in one of the most catastrophic oil spills, the Deep Water Horizon (DWH), which happened in the Gulf of Mexico in 2010. (Olascoaga and Haller, 2012). Due to its challenging simulation, for the scope of our work we restricted ourselves to the first 3 days from the initial spill, where the parametrization could make more impact rather than physical behaviors that are not captured by traditional oil spill modeling.

Considering the data and our premises, Fig. 1 presents all the identified and extracted oil slicks over the period of this study.

The study area lies in the Levantine basin, where the flow regime near the spill is shaped by persistent anticyclonic and cyclonic eddies in the Latakia region. These structures, along with the Asia Minor and Libyan-Egyptian currents, drive a net northward transport (Hecht et al., 1988; Keramea et al., 2023b; Robinson et al., 1987; Zodiatis et al., 2023).

Fig. 2. illustrates the sea surface currents at the start and end of the study period, based on Mediterranean Sea reanalysis (Escudier et al., 2020). The main eddies remain present but shift westwards, altering the intensity and direction of surface transport.

2.2. Data

1- Oil spill imagery

The dataset of segmented observations consists of oil slick boundary shapefiles extracted from the raw satellite imagery from 2019 to 2022. The satellite images were preprocessed by selectively enhancing the contrast in those spectral bands where an expert determined the oil spill was most clearly visible. Depending on the satellite sensor and seasonal variations, the specific bands and indices used for contrast enhancement

[1] https://www.orbitaleos.com/

[2] https://www.imagine-ai.eu/case-study/oil-spill-detection-oil-spill-detection-from-satellite-images

[3] https://zenodo.org/records/11354663

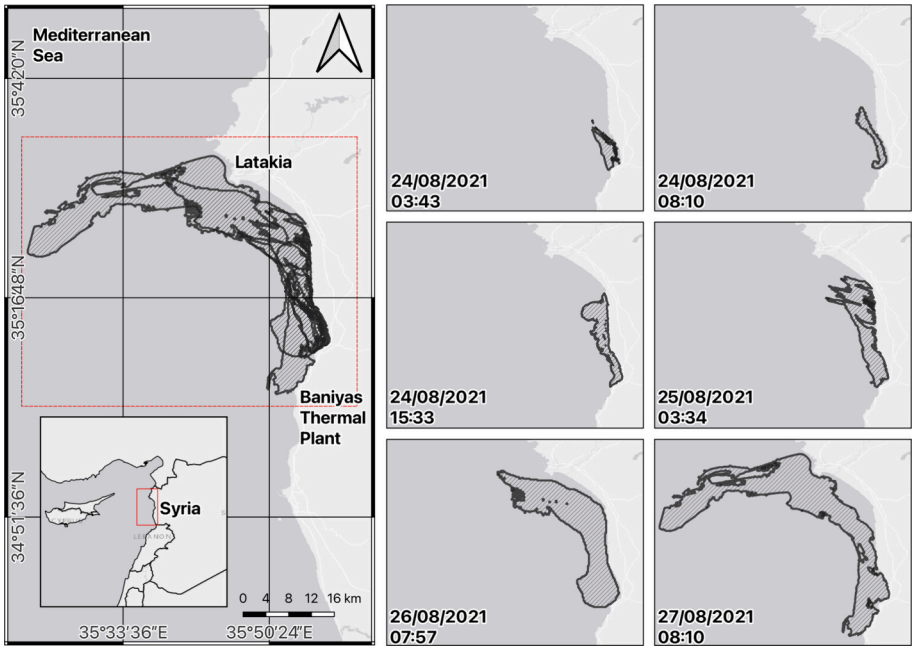

Fig. 1. Oil slicks selected from August 24th to 27th, 2021 targeting the Baniyas oil spill accident in Syria. Each panel reports consecutive stages of the oil slick according to satellite retrievals and revisit times.

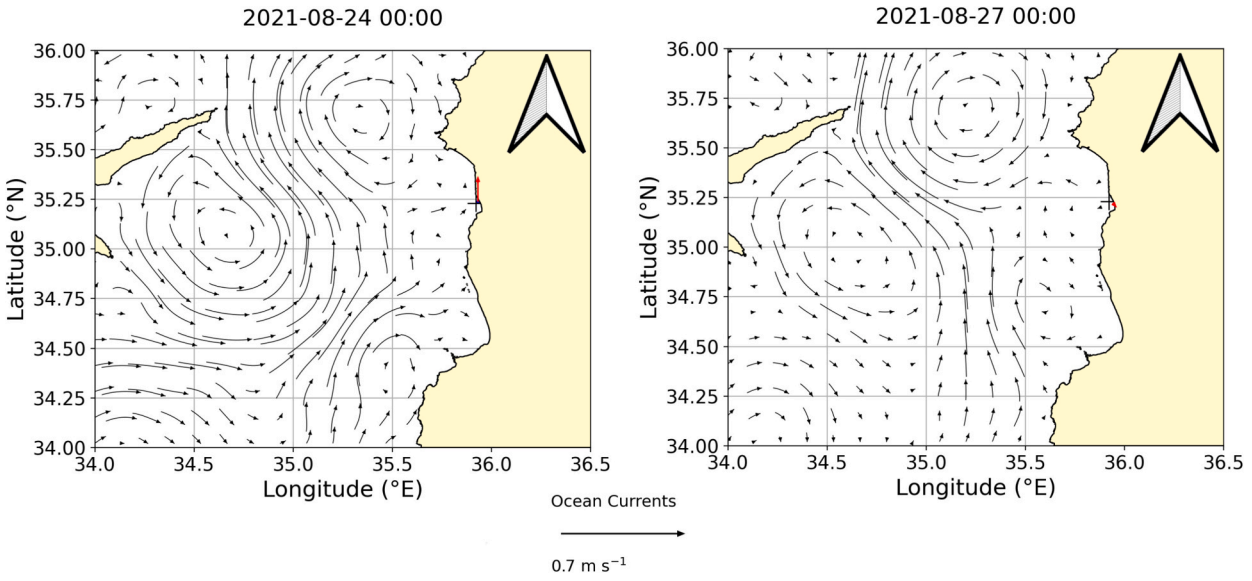

Fig. 2. Surface currents from the Copernicus Marine reanalysis product in the Levantine Basin during the study period. The left panel shows the initial conditions on August 24, 2021, at 00:00, while the right panel displays the final conditions on August 27, 2021, at 12:00.

may be adjusted to account for changing environmental conditions and spectral characteristics. These pre-processing steps were applied to all slicks observed in the above-mentioned period, generating the shapefile dataset (Ferrer et al., 2024) which is available through a data platform hosted at the University of Trento and Zenodo. These were developments made in the context of the iMagine project. Such a platform includes services exposing oil spill data via HTTP (i. e., THREDDS[4]) and a Graphical User Interface (GUI) providing data access, search & discovery as well as advanced data visualization capabilities. Due to page limitations, the GUI is beyond the scope of this paper and will be presented in a future work. Considering the sparse nature of oil spill observations (Shaban et al., 2021) and the requirement of at least two observations for analysis, we focused on the most reliable and consistent data available to evaluate the Bayesian optimization framework in the

Table 1
Remotely sensed oil slick observations provided by Orbital EOS.

Observation ID	Passing time (UTC)	Satellite	Sensor type
O1	24-08-2021 03:43	Sentinel-1	SAR
O2	24-08-2021 08:10	Sentinel-3	SAR
O3	24-08-2021 15:33	Sentinel-1	SAR
O4	25-08-2021 03:34	Sentinel-1	SAR
O5	26-08-2021 07:57	Sentinel-3	Optical
O6	27-08-2021 08:10	Landsat 8	SAR

[4] http://thredds.imagine.disi.unitn.it/thredds/catalog/catalog.html

study area (Fig. 1). Table 1 reports the slicks selected from the dataset.

2- Meteo-oceanographic input

Meteo-oceanographic forcing used to drive the oil spill model was retrieved for the event timelines within a 2–3-degree radius around the first observation's centroid. Surface and subsurface currents and sea surface temperature (SST) were retrieved from the Mediterranean Sea reanalysis dataset (Escudier et al., 2020) at a 1/24° (~4. 5 km) resolution, provided by the European Copernicus Marine Service. Wind data at 10 m were sourced from ERA5 reanalysis single levels reanalysis dataset (Copernicus, 2024) at a 1/4° resolution (~25 × 25 km grid), provided by the Copernicus Climate Change Service.

3- Bathymetry and coastline

Simulations also require bathymetry and coastline data to account for potential shoreline impacts. Sub-ice bathymetry comes from The General Bathymetric Chart of the Oceans 2024 (Mayer et al., 2018) and coastlines from The Global Self-consistent, Hierarchical, High-resolution Geography Database (Wessel and Smith, 1996) at its highest resolution (10–100 m segments).

2.3. MEDSLIK-II oil spill model

The Lagrangian oil spill model MEDSLIK-II (De Dominicis et al., 2013a, 2013b) is an open-source community model that has been used for tracking oil spills in the Mediterranean Sea for nearly 15 years. It discretizes oil slicks into parcels, simulating advection-diffusion and oil weathering processes. Oil transport is driven by water currents, waves, wind, and turbulent fluctuations, modeled via a random walk scheme. MEDSLIK-II includes a proper representation of high-frequency currents and wind fields in the advective components of the Lagrangian trajectory model, the Stokes drift velocity, and the coupling with the remote-sensing data related to observed oil slicks. The governing equation to track oil uses the advection-diffusion equation and is presented in Eq. 2.1:

$$\frac{\partial c}{\partial t} + u \cdot \nabla c = \nabla\left(\vec{k}\nabla c\right) + \sum_{j=1}^{M} r_j(\vec{x}, C, t) \tag{2.1}$$

The term $\frac{\partial c}{\partial t}$ represents the rate of change of concentration over time, with advection captured by the term $u \cdot \nabla C$, where u is the ambient velocity field. The term $\nabla\left(\vec{k}\nabla c\right)$ represents spatially varying diffusivity, and the last term shows the sum of reaction terms accounting for physical-chemical processes such as spreading, evaporation, emulsification, dissolution, photo-oxidation, biodegradation, and sedimentation and adhesion to the coast (De Dominicis et al., 2013b). De Dominicis et al. (2013a) analyzed the sensitivity of the model physical parameters to varying inputs, such as wind drift and wind angle, through five experiments. Their findings highlighted the critical role of high-fidelity and high-resolution currents in improving forecast accuracy over a 1 to 2. 5-day timeframe. The study also suggested the use of ensemble simulations with uncertain physical parameters to enhance predictive performance over deterministic approaches. Building on this work, we extend the analysis to two selected physical parameters, specifically parametrization coefficients within the processes represented in the formulations. In this work we use the MEDSLIK-II model version 2. 01, available as open source at the model website.[5] The two parametrizations we aim to optimize are:

1- Horizontal diffusivity

$$dx_k'(t) = Z_1\sqrt{2K_x dt} = [2r-1]\sqrt{6K_h dt} \tag{2.2}$$

$$dy_k'(t) = Z_2\sqrt{2K_y dt} = [2r-1]\sqrt{6K_h dt} \tag{2.3}$$

where r is a random variable uniformly distributed in $[0, 1]$, and D is the horizontal diffusivity coefficient, varying in the order of magnitude of 1–100 m^2/s (De Dominicis et al., 2012; De Dominicis et al., 2013b).

2- Wind drift and angle

$$U_W = \alpha\left(W_x\cos(\beta) + W_y\sin(\beta)\right) \tag{2.4}$$

$$V_W = \alpha\left(-W_x\sin(\beta) + W_y\cos(\beta)\right) \tag{2.5}$$

where U and V represent the wind zonal and meridional components at 10 m, while α and β correspond to the wind drift and wind angle, respectively. The MEDSLIK-II model typically assumes a wind drift of 3 % and wind angle between 0. 0–15. 0° (Al-Rabeh et al., 2000).

In MEDSLIK-II, other parameters such as evaporation, emulsification and dispersion parameters could be modified to verify the sensibility of the oil slick to its drift and impact how well or worse models represent observed slicks. However, in the context of this work we avoid stressing the oil chemistry parameters to restrict our analysis on a small set of parameters and eventually extend it to other real world slick events, adding more parameters once the relocatability of this approach is validated.

2.4. Bayesian optimization

Bayesian optimization (Jones et al., 1998; Kushner, 1964; Mockus, 1989; Močkus, 1975; Zhilinskas, 1976) is a class of data-driven optimization methods designed for black-box derivative-free global optimization problems. The goal is to maximize an objective function $f(x)$,

$$x^* = \underset{x\in\mathscr{X}}{\operatorname{argmax}} f(x) \tag{2.6}$$

where $x \in \mathscr{X}$ is the physical parameter vector and $\mathscr{X} \subseteq \mathbb{R}^n$ denotes the n-dimensional parameter space (Frazier Peter, 2018). This approach is particularly advantageous when function evaluations are computationally expensive due to the high dimensionality of the parameter space, the computational cost of resource-intensive simulations, or when the analytical form of $f(x)$ is unavailable or unknown. At its core, Bayesian optimization relies on a probabilistic surrogate model, $\widehat{f}(x)$, to approximate the behavior of the objective function. A Gaussian Process (GP) (Naveiro and Tang, 2024; Rasmussen and Williams, 2006) is commonly used as the surrogate model due to its flexibility in modeling a wide class of functions and its ability to quantify uncertainty. The GP defines a prior distribution over functions, which expresses initial beliefs about the function's behavior before any data is observed. Once observations are available, these beliefs are updated to form the posterior distribution. Given a dataset $\mathscr{D} = \{(x_i, f(x_i))\}_{i=1}^m$, the GP posterior $\widehat{f}(x)\,|\,\mathscr{D}$ provides a probabilistic prediction for the function value at a new, unobserved point x. The posterior is fully characterized by its mean $\mu(x)$ and standard deviation $\sigma(x)$, which represent the expected value and uncertainty at each point in the parameter space. Using the posterior, Bayesian optimization selects the next evaluation point based on an acquisition function, which balances (i) Exploration, sampling points with high uncertainty (i. e., large $\sigma(x)$, and (ii) Exploitation sampling points with high predicted performance (i. e., large $\mu(x)$. A common choice for the acquisition function is the Upper Confidence Bound (UCB), defined as:

$$UCB(x) = \mu(x) + k\cdot\sigma(x) \tag{2.7}$$

where $k > 0$ controls the trade-off between exploration and exploitation

[5] https://www.medslik-ii.org/index.html

(Carpentier et al., 2015). Bayesian optimization iteratively refines the surrogate model and acquisition function by incorporating new samples in $\mathscr{D}$. At each iteration, the parameter vector is sampled as:

$$\hat{x} = \underset{x \in \mathscr{X}}{\operatorname{argmax}} UCB(x) \tag{2.8}$$

The objective function is evaluated at $\hat{x}$, and the surrogate model is updated with the new sample point $(\hat{x}, f(\hat{x}))$. This process continues until predefined convergence criteria, such as a maximum number of iterations or a threshold improvement in $f(x)$, are satisfied.

2.5. Coupled Bayesian optimization MEDSLIK-II framework

In this study, we aim to identify the optimal parameter configuration of the MEDSLIK-II model that yields simulations most closely aligned with the spatial-temporal distribution of observed oil slicks, as defined in Section 2. 1. These simulations depend on the specific configuration of a parameter vector. We define the tunable parameter vector as $x = [K_h, \alpha, \beta]^T \in \mathscr{X}$, where $\mathscr{X} \subseteq \mathbb{R}^3$ is the parameter space, K_h represents the horizontal diffusivity, α denotes the wind drift and β the wind angle. The ranges for the tunable parameters are defined as $K_h \in [0.0, 20.0]$ (m^2/s), $\alpha \in [0.0, 0.05]$ (m/s) and $\beta \in [0.0, 15.0]$ (◦) after an initial sensitivity test, but also respecting the boundaries found in the literature (De Dominicis et al., 2012).

To evaluate the agreement between simulated and observed oil slick distributions, we adopt the Fractions Skill Score (FSS) as the optimization metric (Roberts, 2008; Roberts and Lean, 2008), which is further explained in Section 2. 6. It represents the fractional coverage of simulated and observed oil particles within a specified spatial neighborhood, ranging from 0 to 1, with higher values indicating better spatial alignment. Hence, the Bayesian optimization task is formulated to maximize this score:

$$x^* = \arg \max_{x_i \in \mathscr{X}, i=1,\ldots,N} FSS\left(\mathscr{S}_{x_i}, \mathscr{O}_{t+1}\right) \tag{2.9}$$

where $FSS(S_{x_i}, \mathscr{O}_{t+1})$ serves as the objective function to be maximized. Here, S_{x_i} represents the MEDSLIK-II simulation initialized at the observation O_t, based on the corresponding atmospheric and oceanic forcing at that time, along with the physical parameter vector x_i. The term $\mathscr{O}_{t+1}$ denotes the subsequent observed oil slick data used for comparison, while x^* represents the optimal parameter vector obtained upon convergence. In addition to physical parameters, the MEDSLIK-II simulation includes oil-specific information, such as the number of simulated particles and oil type. We used 90. 000 particles and selected the same oil type for all the simulations: Safaniya, an Arabian heavy crude with an API gravity of 27. 9. To maintain consistency, we did not alter values related to oil chemistry, instead focusing on the trajectory of the oil spill while keeping its chemical processes fixed. In the experiments, we set $t \in [1, \ldots, 5]$, aligning with the available observations (Table 1), and limit the maximum number of iterations in the Bayesian optimization framework to $N = 100$. The process requires the evaluation of an objective function, using a "one-step" assumption, which quantifies the accuracy of the MEDSLIK-II model's output relative to target observation(s). To approximate the behavior of $FSS(S_{x_i}, \mathscr{O}_{t+1})$, a Gaussian Process (GP) is employed as a surrogate model. At each iteration i, the MEDSLIK-II model is initialized with the current parameter vector x_i, and the corresponding $FSS(S_{x_i}, \mathscr{O}_{t+1})$ is computed. For the first iteration $(i = 1)$, the parameter vector is initialized to default values, $K_h = 2.0$, $\alpha = 0.0$ and $\beta = 0.0$ (De Dominicis et al., 2013b). The parameter vector for the subsequent iteration $i+1$ is determined by maximizing the acquisition function UCB as in eq. 2.7, with $k = 2.576$ (Jones et al., 1998; Srinivas et al., 2010; Frazier Peter, 2018). After evaluating the objective function at x_{i+1}, the GP is updated with the new sample $(x_{i+1}, FSS(S_{x_i}, \mathscr{O}_{t+1}))$. We repeat this process until we reach N for a given observation. In addition to using the acquisition function for selecting new points, we also leverage the GP posterior to estimate uncertainty around the current best set of parameters x^*. Specifically, we sample 1. 000 functions from the GP posterior conditioned on x^*, yielding a distribution of possible objective values. From this distribution, we extract the 5^{th} and 95^{th} percentiles to form a confidence interval around the predicted FSS. This provides insight into the robustness and uncertainty of the best-performing simulation parameters. The overall workflow is presented in Fig. 3. For the implementation of Bayesian optimization, we used the BayesOpt python library,[6] which was customized to meet the specific needs of this study.

2.6. Experimental settings

This section outlines the experimental settings used to evaluate the skills of the proposed approach across six sequential observations (Table 1). In order to compare simulations with Observations, MEDSLIK-II simulations were run using both default physical parameters (control phase) and the ones estimated through Bayesian optimization (validation phase). In the control phase, default physical parameters were kept fixed across all simulations, mimicking standard model usage. In the validation phase, two strategies were tested:

i. *One-step Optimization (OSO)*: MEDSLIK-II is initialized with observation O1, as the initial condition, and Bayesian optimization is then performed using observation O2 to determine the optimal physical parameters. These parameters are then used unchanged in all subsequent simulations for all future time steps.
ii. *Multi-step Optimization (MSO)*: Bayesian optimization is applied iteratively between each pair of observations (e. g., O1 → O2, O2 → O3), updating parameters step by step.

While OSO prioritizes computational efficiency, for operational use, MSO improves accuracy by continuously recalibrating the model physical parameters, enhancing flexibility and adaptability by continuously updating conditions. The overview of the experimental settings is depicted in Fig. 4.

Oil Slick simulation accuracy was assessed using two metrics:

1. Fractions Skill Score (FSS)

The FSS (Roberts and Lean, 2008), used as the objective function of the Bayesian optimization framework, was originally introduced for comparing observed and modeled rainfall accumulation. Later, it was also applied in oil spill modeling, demonstrating the suitability for operational oil spill tracking purposes (Simecek-Beatty and Lehr, 2021). In this work, the FSS compares spatial distributions (unlike the traditional overlay method) of oil spill simulations by comparing them against a ground truth (i. e., satellite observations) on a 150 m × 150 m grid.

$$FSS = 1 - \frac{\sum_{i=1}^{n} (f_i - o_i)^2}{\sum_{i=1}^{n} \left(f_i^2 + o_i^2\right)} \tag{2.10}$$

where f_i is the forecast fraction, o_i is the observed fraction, and the summation runs over all grid points i. The FSS ranges from 0 to 1, where 1 indicates a perfect match between the forecast and observation, and 0 indicates no skill.

2. Centroid Skill Score (CSS)

The Centroid Skill Score (Liu and Weisberg, 2011) measures the normalized Lagrangian separation distance between the centroids of simulated and observed slicks:

[6] https://github.com/bayesian-optimization/BayesianOptimization

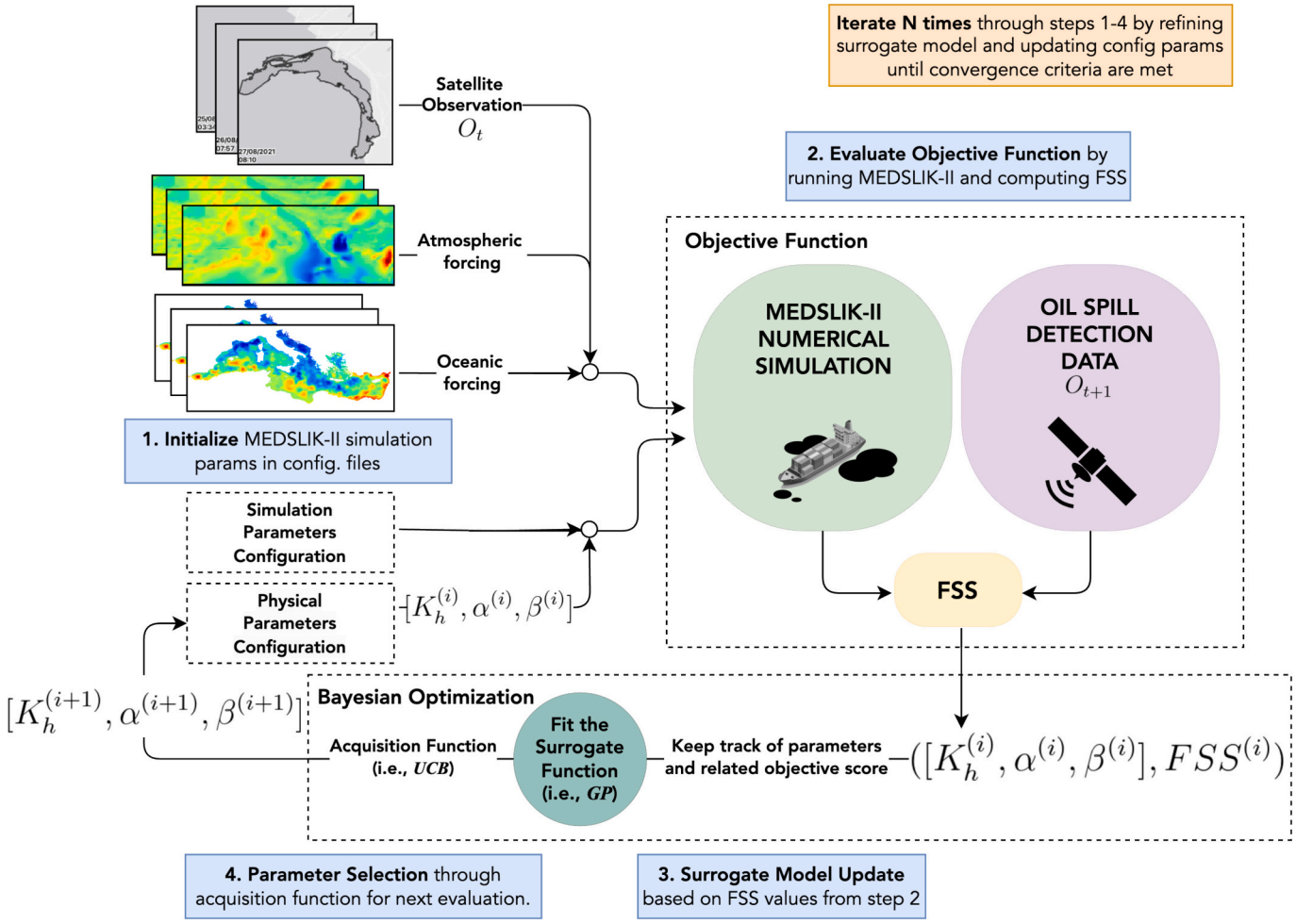

Fig. 3. Overview of the Bayesian optimization workflow designed for improving the accuracy of the MEDSLIK-II oil spill model simulations. The workflow iteratively refines the simulation configuration parameters by evaluating the objective function and optimizing the acquisition function until a convergence criterion is met. Here, we denote the iteration number with the superscript (i), for the sake of clarity.

$$CSS = \begin{cases} 1 - \dfrac{CI}{CI_{thr}} & \text{if CI} < C_{thr}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

where $CI = \frac{\Delta X}{L_{obs}}$ the Lagrangian separation distance (in km) between SAR-derived and modeled oil slick at a given time and L_{obs} is the diagonal length (in km) of the bounding box enclosing the observed slick. A CSS of 1 indicates a perfect match while a CSS of 0 means the centroids are distant relative to the observation size. Following (De Dominicis et al., 2014), the threshold C_{thr} is set to 1.

3. Results

Table 2 presents the quantitative results for all tested experiments, showcasing each simulation metric across different settings: control, OSO, and MSO. For each observation, we report the FSS and CSS scores along with their percentage difference over the control setting when applying the OSO and MSO strategies. The specific physical parameters used in each MEDSLIK-II simulation are detailed in Table 3. The default values come from the standard MEDSLIK-II model, while the OSO parameters were optimized using Bayesian optimization between observations O1 and O2. The MSO parameters, on the other hand, were refined through continuous training over successive observation pairs.

Fig. 5 illustrates the results for the control, OSO, and MSO approaches, using observation O3 (see Table 1) as the initial condition. Each evaluation metric assesses the accuracy of MEDSLIK-II simulations against satellite observations of the oil slick trajectories, following temporal alignment. Since the observations were provided as shapefiles without oil slick thickness data, we focused on the slick shapes for area comparison. Consequently, we employed the geographical centroid in our analysis, as the center of mass could not be determined from the available data. Because the optimization process targeted the FSS metric, we consistently observe an improvement from control to OSO and further to MSO. However, this pattern is not always reflected in CSS scores. For example, in observation O3, both FSS and CSS reached their highest values in MSO. A closer look at observation O4 in Fig. 5 reveals a clear skill enhancement from the control setting to OSO and MSO across all metrics. In the control run, the simulated oil slick appears narrower, remaining closer to the coast and failing to move as far north as in the OSO and MSO experiments. Comparing OSO and MSO, we see that MSO achieves a slightly better match, particularly in the southern region of the observed slick. These differences can largely be attributed to variations in horizontal diffusivity. In the control setting (horizontal diffusivity set to 2.0 m^2/s), the slick is advected with minimal spreading. In contrast, the optimized horizontal diffusivity values for OSO (7.3020 m^2/s) and MSO (15.8515 m^2/s) allow the slick to spread more effectively, covering a larger portion of the observed area and improving accuracy.

Our simulations, initialized with the Bayesian optimization estimated parameters, usually achieved improvements across both settings. The FSS metric in the OSO configuration, outperformed the control

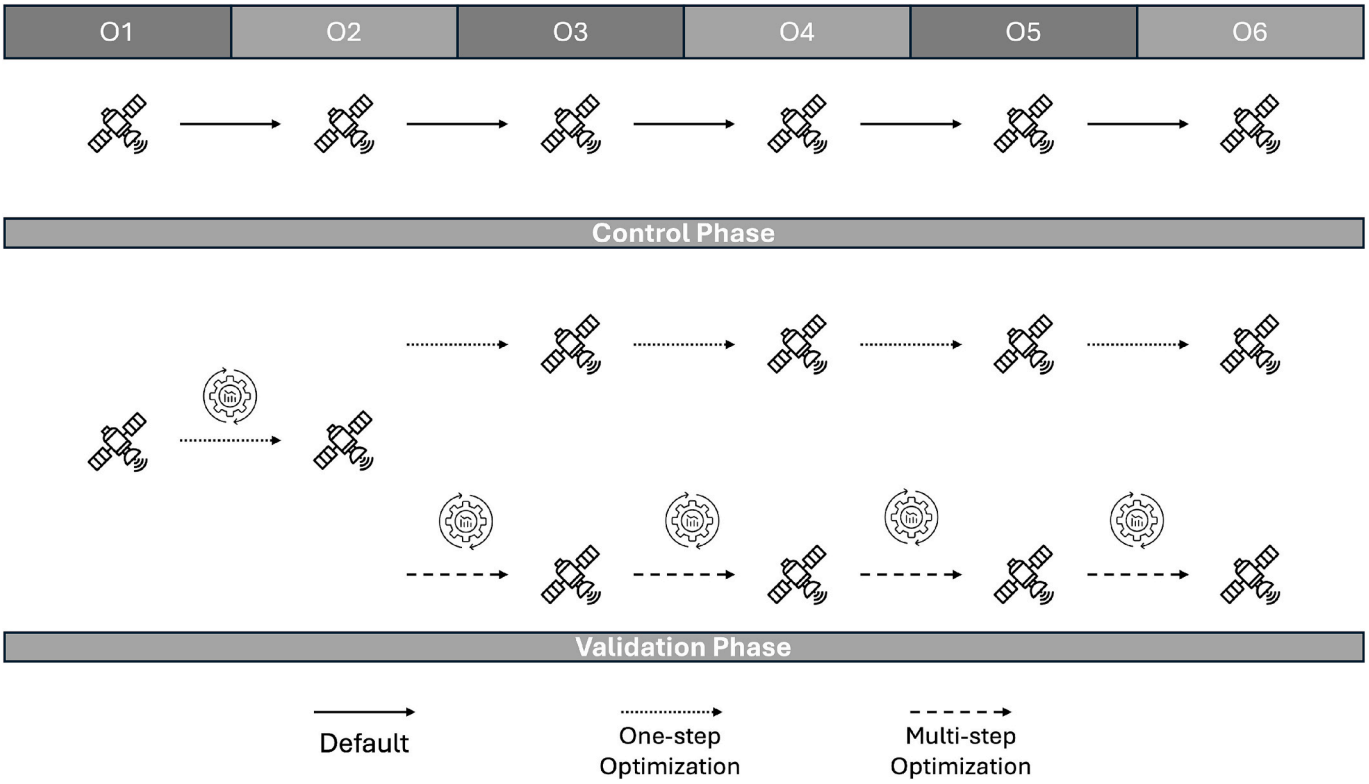

Fig. 4. Experimental settings. The control phase consists of running MEDSLIK-II simulations with standard physical parameters. The validation phase is divided into two approaches: One-step Optimization and Multi-step Optimization. In the former, optimization is applied only from observation O1 to O2, with the resulting physical parameters used for all subsequent simulations. In the latter, optimization is performed at each observation step, to capture potential variations in model physical parameters.

Table 2
FSS and CSS results for control, OSO, and MSO settings across the assessed observations. The FSS values are accompanied by 5^{th} to 95^{th} percentile confidence intervals, derived from 1000 samples drawn from the GP's posterior distribution. Additionally, we report the percentage improvement for both OSO and MSO relative to the control setting (OSO% and MSO%) towards the maximum scores, as well as the average values across the different observations and settings.

Obs	Metric	Control	OSO	OSO%	MSO	MSO %
O2	FSS	0. 5594	0. 6624	23. 37	0. 6624 (0. 6621, 0. 6626)	23. 37
	CSS	0. 8595	0. 8787	13. 66	–	–
O3	FSS	0. 3258	0. 5036	26. 38	0. 5646 (0. 5643, 0. 5649)	35. 42
	CSS	0. 8233	0. 8717	27. 22	0. 9373	64. 51
O4	FSS	0. 4263	0. 5105	14. 67	0. 545 (0. 5447, 0. 5454)	20. 7
	CSS	0. 9617	0. 9665	12. 53	0. 9744	33. 15
O5	FSS	0. 5764	0. 4704	−25. 03	0. 627 (0. 6266, 0. 6274)	11. 95
	CSS	0. 8418	0. 8646	14. 41	0. 8621	12. 83
O6	FSS	0. 5449	0. 5471	0. 46	0. 5989 (0. 5989, 0. 5989)	11. 85
	CSS	0. 8678	0. 8629	−3. 70	0. 8549	−9. 75
Avg.	FSS	0. 4866	0. 5388	7. 97	0. 5996	20. 66
	CSS	0. 8708	0. 8888	12. 82	0. 9014	25. 18

setting by 7. 97 % on average, while the MSO configuration improved even further by 20. 66 %. Regarding the CSS metric, the optimization also enhanced the average accuracy by 12. 82 % for the OSO strategy and 25. 18 % for MSO, further underscoring the advantages of Bayesian optimization. However, despite its superior accuracy, MSO comes with added complexity and computational time, making its implementation in operational settings more challenging, as previously discussed. Nevertheless, when comparing OSO and MSO, it is evident that the latter consistently outperforms both the control and OSO settings for FSS, demonstrating a sustained improvement in accuracy (see Fig. 6). However, when observing the CSS metric, the improvement is not always verified, such as in O6, where both OSO and MSO strategies were outperformed by Default settings.

When comparing the visual characteristics of the oil spill in Fig. 5 with Figs. 7, 8, 9, and 10 in the Appendix A, it becomes clear that the modeled results better match the observed oil slick extent in the first three simulations.

4. Discussion

4.1. Methodological considerations and practical implications

Bayesian Optimization provides a principled approach to calibrating model parameters, making it particularly suitable for data-driven oil spill forecasting. Conventional parameter settings in oil spill models, often derived from expert knowledge, may not generalize well across varying environmental conditions. The core advantage of Bayesian Optimization lies in its ability to iteratively explore the parameter space in a stochastic yet informed manner, balancing exploration of poorly known regions and exploitation of promising configurations. This feature is particularly well-suited to oceanographic simulations, where each model run is computationally expensive and an exhaustive search

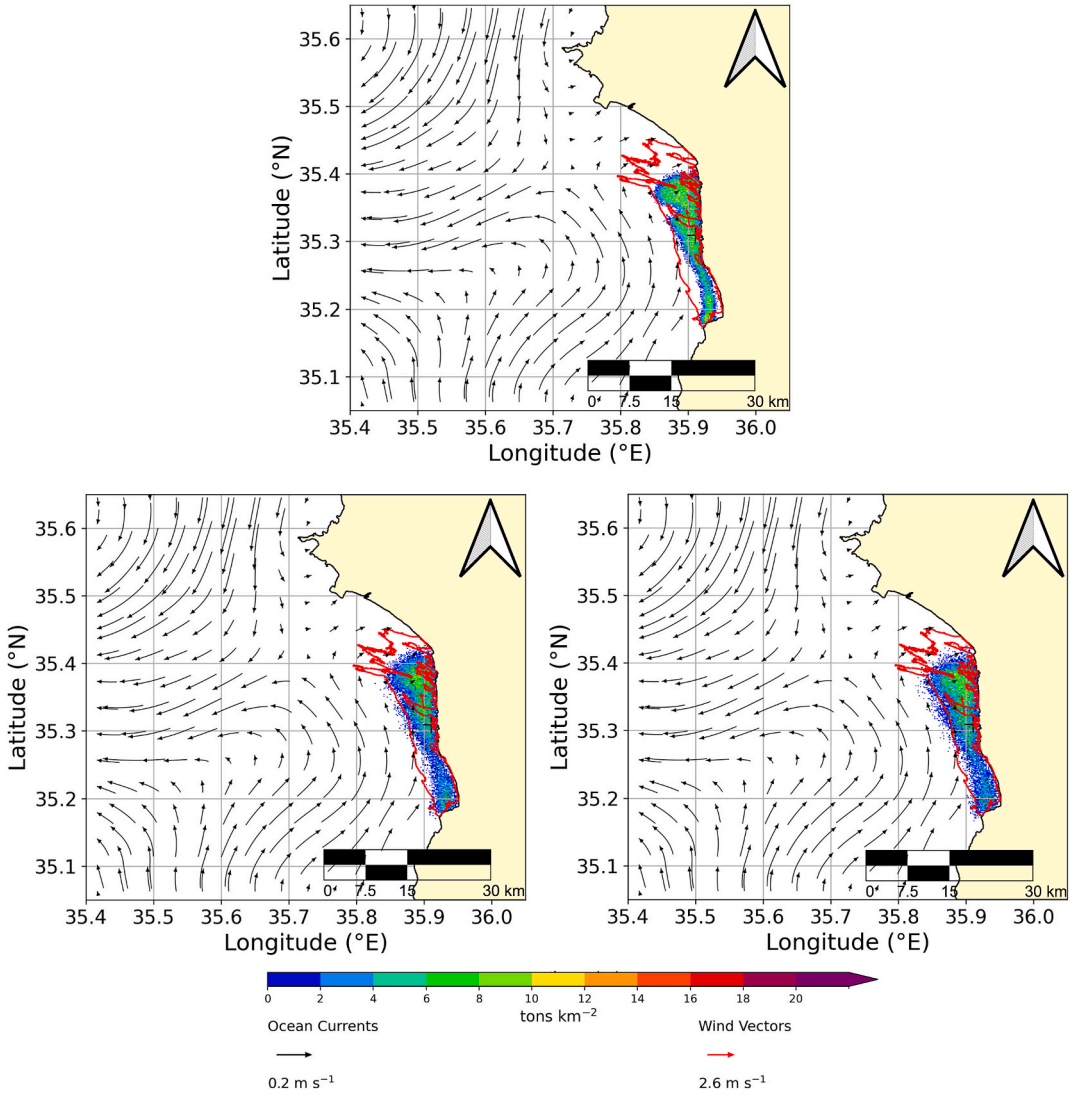

Fig. 5. MEDSLIK-II surface oil concentrations starting from O3 (24/08/2021 15:33) as initial condition and stopping at O4 (25/08/2021 03:34). The upper panel represents control simulation with default physical parameters. Bottom panel represents the optimized simulation using the OSO (left) and MSO (right) settings.

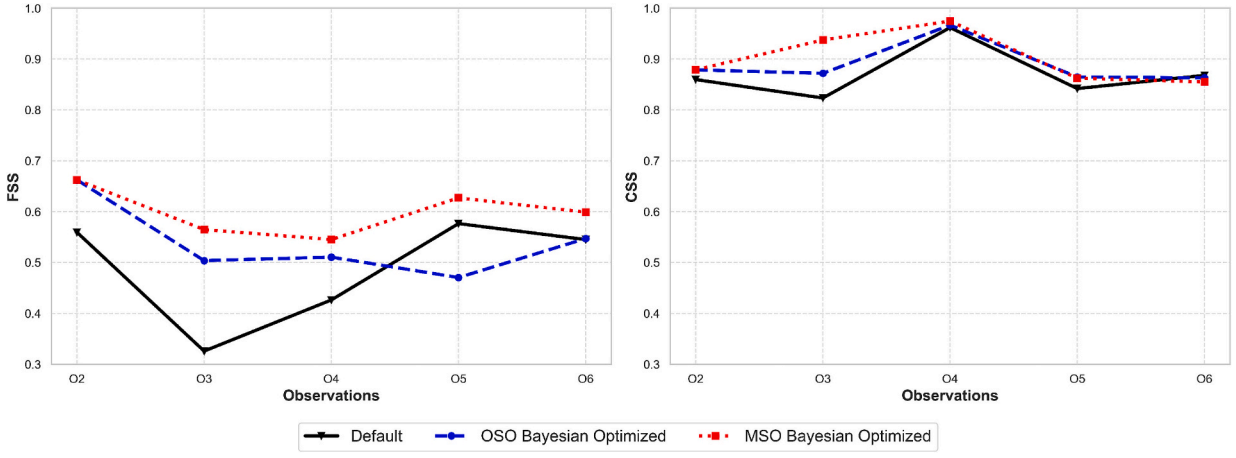

Fig. 6. Comparison of the FSS (left panel) and CSS (right panel) metrics between the simulation using optimal parameters identified through Bayesian optimization and the simulation with default MEDSLIK-II parameters. The x-axis reports each observation (see Table 1), while the y-axis reports the values of each metric, both bounded in $[0, 1]$ (higher values are better). Here, we do not report observation O1 which is instead used for initializing the MEDSLIK-II simulations in all the settings.

is infeasible. Moreover, an online-trained surrogate model can dynamically capture spatial correlations in the simulation outputs, enabling more informed and adaptive sampling. By leveraging a surrogate probabilistic model of the response surface (e. g., a Gaussian Process), Bayesian Optimization can guide the selection of new parameter sets in a way that is both efficient and uncertainty aware. It is worth noting, however, that the performance of Bayesian Optimization depends on the quality of both the surrogate model and the acquisition function, underlining the importance of designing acquisition strategies tailored to the specific landscape of the response surface, which is often non-convex and noisy. Overall, this study demonstrates the potential of Bayesian Optimization as a powerful tool for adaptive parameter tuning. Beyond oil spill modeling, the same methodology can be extended to a wide range of simulation-based models characterized by uncertain parameters, sparse observations, and high computational costs.

4.2. Considerations around metrics and environmental data used

The results presented in Section 3 indicate that simulation accuracy, measured by the optimized metric (FSS), consistently improved when the Bayesian optimization workflow was applied. Both the MSO and OSO approaches led to higher FSS values compared to the control simulation setup, with MSO generally outperforming OSO. However, a few exceptions were observed when considering the CSS metric. This does not necessarily imply poorer results, as area-based metrics like FSS and centroid-based metrics such as CSS complement each other in assessing oil spill modeling accuracy. While CSS primarily indicates the accuracy of the oil spill's location, FSS evaluates how well the model captures spreading and weathering processes (Dearden et al., 2021). Additionally, FSS scores are typically lower than CSS scores, as accurately matching the precise area and shape of an oil slick is inherently more challenging than achieving strong performance on geographical centroids, even after optimization. Although the number of variables was limited in all experiments, other inputs could have significantly influenced the results. For instance, additional MEDSLIK-II parameters—such as the spreading rate of thick and thin slicks or evaporation parameters—different environmental forcings, including regional or local circulation models tailored to the period of interest, or more detailed observations of the oil slick characteristics (e. g., oil thickness) could have been considered. As shown in the surface current data in Fig. 2, the primary oceanographic features are two eddies that drive a northward current pattern. While these features are captured in the forcing data, the accident occurred along the coastline, far from these mesoscale structures. Additionally, the reanalysis data used for coastal conditions lacks the resolution necessary to accurately represent coastal currents at the scale at which this circulation model was generated. MEDSLIK-II can extrapolate currents from areas with available data to cells where no current values are provided. This extrapolation averages currents in ocean cells until land cells are reached (De Dominicis et al., 2012). However, this method does not ensure that the real circulation features are accurately represented in the extrapolated region. The combination of general meteo-oceanographic conditions, data limitations, and extrapolation techniques adds to the inherent uncertainties of oil spill advection and diffusion, which relies on a random tensor to introduce stochasticity in the movement of Lagrangian particles. The Bayesian optimization method, through its acquisition function, seeks to minimize the score without directly accounting for the specific oceanic conditions in which the oil spill simulations were conducted. Given that the currents in the affected area may be weaker than they should be, advection is likely underestimated, preventing the slick from being transported as far north as it would in real conditions. As a result, diffusion, which is modulated by horizontal diffusivity (K_h) becomes the dominant process in achieving a better match between observations and simulations. The assumption for this case may not prove right for other occasions, where wind can have a bigger effect in the slick drift. Additionally, we used ERA5 winds, which have an approximate 25 km resolution, not capturing the precise local behavior of the wind into the oil slick. Wind drift parameter varied from 1. 29 % to 3. 31 % meanwhile wind angle acted varied from 0. 27 to 6. 98 degrees (check 3). These values acted more as calibrators of the slick overall position since the spreading of the slick was more important for matching the modeled and observed areas. Beyond the lack of information on slick thickness, another observational limitation stems from the use of three different satellite sensors, which could influence the results due to variations in oil slick detection across sensors. The only set of experiments using the same sensor was from observation O3 to observation O4, which also coincided with the best results in terms of the CSS metric. Furthermore, when comparing the observed evolution of the oil slick with FSS and CSS metrics, there are cases where the modeled slick performs well in one metric but not in the other. This was particularly evident in observations O5 and O6, where the slick area was larger than in previous observations. While the optimization improved the FSS metric for these cases, the same improvement was not reflected in the CSS metric. Beyond the above-mentioned aspects, the findings of this work are restricted to a single oil spill event, despite abundant in terms of observations, it is not possible yet to imply that the same method and relation between parameters could explain the optimization for other oil spill events. Future works will apply the same approach to other observed oil slick events to assess the relocatability of the Bayesian Optimization considering other locations and environmental data.

4.3. Considerations on experimental settings

All experiments were conducted as sequential runs on a compute node of the JUNO Hybrid Cluster at the High-Performance Computing Center of the Euro-Mediterranean Center on Climate Change (CMCC).[7] One of the key motivations for proposing OSO and MSO was the high computational cost of a single optimization process. The process is performed online, iteratively, updating the surrogate model with new observations, making its execution time dependent on the simulation model's computational cost. In the worst-case scenario for the analyzed event, a single model run required approximately 295 s (∼5 min) of CPU time, with a maximum memory usage of 292 MB and an average of 228. 83 MB. During the training process, CPU time increased to 26,970 s (∼7. 49 h), with a peak memory usage of 568 MB and an average of 484. 36 MB. The OSO procedure involved one training step and four single model runs, whereas the more computationally demanding MSO required five training steps, further increasing resource consumption. For operational responses to oil spill incidents, computational time and resource usage could be limiting factors, as optimizing model performance may come at the expense of a rapid response. However, OSO demonstrated that even a single optimization step was sufficient to achieve improvements across all evaluated metrics. An added advantage of the proposed coupled workflow is its ability to provide an integrated solution for parameter selection, model execution, and result generation. This process can be easily automated, for instance, when a new observation becomes available, thereby reducing overall response time. Lastly, optimizing for FSS did not automatically lead to improvements in the CSS metric. However, the flexibility of the system allows the optimization metric to be modified, potentially yielding a different combination of MEDSLIK-II parameters better suited for specific evaluation criteria.

4.4. Usage on an operational oil spill modeling context

The usage of the Bayesian Optimization on an operational context is challenging, but not impossible. The method works well when hindcast data is available for the training, which is usually not the case when a

[7] https://www.cmcc.it/what-we-do/high-performance-computing-center-hpcc

real oil spill emergency is happening. The usage of the application is only possible considering the following aspects:

- Forecast availability - Sea currents, sea surface temperature and wind at 10 m height. Which is the basis for an oil spill simulation even without optimization.
- One segmented oil spill observation - In this case if the source of the oil spill is known, the Bayesian Optimization can be applied from the point source to the identified oil slick, allowing future slicks to use the optimized parameters in the event area.
- Two segmented oil spill observations - The most common case where source is unknown. Due to this, it is important to check the state of the slick in two distinct moments, so the optimization can be applied

Therefore, on a scenario where the accident has just happened and no observations are still available, the optimization cannot be applied, and the simulation skill will be based in the available data and the skill of the oil spill modeler to represent that event.

4.5. Open questions, limitations, and future directions

While the proposed Bayesian optimization framework demonstrates clear advantages in parameter tuning for oil spill models, some limitations should be acknowledged. The surrogate model's performance strongly depends on the initial sampling strategy and the choice of prior assumptions, which may influence convergence and robustness. The availability of very sparse observations also constrains the complexity of the surrogate model, potentially limiting its expressiveness. Another question that remains unanswered in this study is which metric is most useful for characterizing oil slicks under different conditions. For example, would optimizing CSS better represent oil slicks in a coastal regime, as considered in this study? Conversely, how would the method perform for offshore slicks exhibiting tiger-tail features? While this remains unclear, the proposed approach allows for further exploration of these questions, even within the same set of experiments presented in this study.

Beyond the slick conditions, it is important to highlight that we have chosen a single data source for currents, sea surface temperature and winds. More experiments could have been done in the Syria Oil Spill case changing these data and checking how the parameterizations would answer to this stress. Due to its architecture, the system would search for the values, but it is unknown, for instance, if horizontal diffusion would be smaller than in the present optimization, or even if wind drag would have a bigger impact in the oil drift under this scenario.

One of the limitations of the optimization process is that it does not account for the amount of oil being beached, nor does it consider oil subsurface naturally entrained oil or the fraction sedimented. We also did not change the oil type in this study, which can alter significantly the characteristics and dispersion of the modeled oil slick, while also being a very plausible scenario, since it is common to not know what type of oil is being observed from satellite imagery if the source is not known. Further, the smallest thickness of oil slick that can be detected from an observation satellite that defines the shape, and the area of the slick may differ from one satellite to another. This plays a role in our study here, because the observation sources used are not from the same source throughout the optimization process. It remains uncertain how the optimization process could be effectively applied to all these aspects. Consequently, our approach focuses on optimizing surface oil concentrations, specifically in terms of centroid and shape, while disregarding mass balance considerations.

Despite these limitations, the results of this study demonstrate that Bayesian optimization successfully enhanced simulation accuracy. On average, the FSS metric improved by 7. 97 % with OSO and 20. 66 % with MSO, while for CSS, the improvements were 12. 82 % and 25. 18 %, respectively. These results highlight that even a single optimization step can improve the modeled oil slick patterns in future simulations, as biases in the forcing data used for MEDSLIK-II become better incorporated into the optimization strategy.

Future research should explore the integration of additional prior knowledge to further improve sampling efficiency and model fidelity. The findings also indicate that further experiments could introduce additional complexity to the system by modifying parameters and refining the methodology. Testing the workflow on additional real-world oil spill events will be essential to assess whether similar or even greater improvements can be achieved under different environmental conditions, ultimately demonstrating the generalizability of the approach. Finally, hybridizing the Bayesian framework with data-driven emulators, when sufficient observations are available, may enhance both accuracy and computational performance.

Despite these constraints, this study successfully demonstrates an automated method for improving oil spill simulations with MEDSLIK-II using Bayesian optimization, offering a systematic and efficient way to enhance model accuracy.

5. Conclusions

In this study we propose a novel approach by using BayesOpt with the MEDSLIK-II model to simulate the oil slick behavior in Baniyas Thermal Station accident in 2021. We demonstrated that such an optimization framework can be effectively coupled with numerical simulation to increase the accuracy of the results when considering the FSS metric. It is important to remark that, while the experimental evaluation presents a single study area using only a part of the satellite imagery available for the same accident, the framework is flexible enough for testing a variety of optimization scenarios. For example, it could be applied to different sets of environmental data (currents, winds and sea surface temperature) and considering additional MEDSLIK-II model parameters that were not tested in this first experiment. The capacity to change and test different components of the framework opens further enhancement possibilities for this same experiment and for other real-world scenarios. Moreover, the workflow presented in this study is suitable for being used in an operational setting, when considering the OSO approach. A promising direction for future work is to develop a unified objective function that integrates multiple metrics, such as FSS and CSS, allowing the optimization process to extract more comprehensive information from multiple evaluations simultaneously. This approach would enhance the robustness of the optimization by balancing different performance aspects. Future developments could also include relocating the framework to other oil spill events to assess its generalizability, which will ensure that the framework is reliable in an operational environment. Moreover, further evaluation of the approach by testing different acquisition functions to potentially improve the optimization process, expanding the temporal horizon to capture longer-term spill dynamics, and extending the number of parameters subject to optimization (we currently focus on only two) could also be explored. Additionally, investigating alternative parametrizations might further refine simulation accuracy.

CRediT authorship contribution statement

Gabriele Accarino: Writing – review & editing, Visualization, Methodology, Conceptualization, Writing – original draft, Software, Formal analysis. **Marco M. De Carlo:** Writing – original draft, Software, Investigation, Visualization, Methodology, Formal analysis. **Igor Ruiz Atake:** Writing - original draft preparation, Software, Investigation, Visualization, Methodology, Formal analysis, Conceptualization. **Donatello Elia:** Writing – original draft, Methodology, Visualization, Conceptualization. **Anusha L. Dissanayake:** Visualization, Writing – original draft, Validation. **Antonio Augusto Sepp Neves:** Writing – review & editing, Methodology, Conceptualization, Software, Data curation. **Juan Peña Ibañez:** Writing – review & editing, Data curation. **Italo Epicoco:** Validation, Methodology, Writing – review & editing,

Supervision. **Paola Nassisi:** Validation, Writing – review & editing, Funding acquisition. **Sandro Fiore:** Data curation, Writing – review & editing. **Giovanni Coppini:** Validation, Conceptualization, Writing – review & editing, Funding acquisition.

Funding

This work has been in part supported by the iMagine project. The project iMagine receives funding from the European Union's Horizon Europe Research and Innovation Program under grant agreement number 101058625.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

G. Accarino's research has been supported by the NSF through the Learning the Earth with Artificial intelligence and Physics (LEAP) Science and Technology Center (STC) (Award #2019625). M. De Carlo has been also supported in this work by OGS and CINECA under HPC-TRES award number 2023-05.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103368.

Data availability

The data used in this study is available on Zenodo at: https://zenodo.org/records/11354663.

References

Abou Samra, R.M., Ali, R.R., 2024. Tracking the behavior of an accidental oil spill and its impacts on the marine environment in the eastern Mediterranean. Mar. Pollut. Bull. 198, 115887.

Al-Rabeh, A.H., Lardner, R.W., Gunay, N., 2000. Gulfspill version 2. 0: a software package for oil spills in the Arabian gulf. Environ. Model. Software 15, 425–442.

Andradóttir, Sigrún, 2006. Chapter 20 an overview of simulation optimization via random search. In: Handbooks in Operations Research and Management Science, 13. https://doi.org/10.1016/S0927–0507(06)13020–0. December.

Arakawa, Akio, Schubert, Wayne H., 1974. Interaction of a cumulus cloud ensemble with the large-scale environment, part I. J. Atmos. Sci. 31 (3), 674–701. https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Balseiro, C.F., Carracedo, P., Gómez, B., Leitão, P.C., Montero, P., Naranjo, L., et al., 2003. Tracking the prestige oil spill: an operational experience in simulation at MeteoGalicia. Weather 58, 452–458. https://doi.org/10.1002/wea.6080581204.

Barker, C.H., Kourafalou, V.H., Beegle-Krause, C.J., Boufadel, M., Bourassa, M.A., Buschang, S.G., Androulidakis, Y., Chassignet, E.P., Dagestad, K.-F., Danmeier, D.G., 2020. Progress in operational Modeling in support of oil spill response. J Mar Sci Eng 8, 668.

Belete, Daniel, Manjaiah, D.H., 2021. Grid search in Hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. Int. J. Comput. Appl. 44 (September), 1–12. https://doi.org/10.1080/1206212X.2021.1974663.

Bengtsson, L., Steinheimer, M., Bechtold, P., Geleyn, J.-F., 2013. A stochastic parametrization for deep convection using cellular automata. Q. J. Roy. Meteorol. Soc. 139 (675), 1533–1543. https://doi.org/10.1002/qj.2108.

Berner, J., Achatz, U., Batté, L., Bengtsson, L., Cámara, A.d.l., Christensen, H.M., Colangeli, M., et al., 2017. Stochastic parameterization: toward a new view of weather and climate models. Bull. Am. Meteorol. Soc. 98 (3), 565–588. https://doi.org/10.1175/BAMS-D-15-00268.1.

Boufadel, M., Liu, R., Zhao, L., Lu, Y., Özgökmen, T., Nedwed, T., Lee, K., 2020. Transport of oil droplets in the upper ocean: impact of the Eddy diffusivity. J. Geophys. Res. 125 (2). https://doi.org/10.1029/2019JC015727 e2019JC015727.

Boufadel, M.C., Özgökmen, T., Socolofsky, S.A., Kourafalou, V.H., Liu, R., Lee, K., 2023. Oil transport following the deepwater horizon blowout. Ann. Rev. Mar. Sci. 15 (1), 67–93. https://doi.org/10.1146/annurev-marine-032122-112316.

Breivik, Øyvind, Moerman, Bente, Dagestad, Knut-Frode, Nordam, Tor, Hope, Gaute, Hole, Lars Robert, Allen, Arthur A., Stone, Lawrence D., 2025. The Bayesian backtracking problem in oceanic drift modelling. Ocean Model. 194, 102505.

Brochu, E., Cora, M., de Freitas, N., 2009. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. TR-2009-023.. Department of Computer Science, University of British Columbia.

Caron, S., Heskes, T., Otten, S., et al., 2019. Constraining the parameters of high-dimensional models with active learning. Eur. Phys. J. 79 (944). https://doi.org/10.1140/epjc/s10052-019-7437-5.

Carpentier, Alexandra, Lazaric, Alessandro, Ghavamzadeh, Mohammad, Munos, Rémi, Auer, Peter, Antos, András, 2015. Upper-confidence-bound Algorithms for Active Learning in Multi-armed bandits (https://arxiv.org/abs/1507.04523).

Christensen, Henrik, Zanna, Laure, 2022. Parametrization in Weather and Climate Models. Oxford University Press, Oxford.

Copernicus, 2024. ERA5 Post-Processed Daily-Statistics on Single Levels from 1940 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.4991cf48.

Coppini, G., De Dominicis, M., Zodiatis, G., Lardner, R., Pinardi, N., Santoleri, R., Kallos, G., 2011. Hindcast of oil-spill pollution during the Lebanon crisis in the eastern Mediterranean, July–august 2006. Mar. Pollut. Bull. 62 (1), 140–153. https://doi.org/10.1016/j.marpolbul.2010.09.015.

Daniel, P., Josse, P., Dandin, P., Gouriou, V., Marchand, M., Tiercelin, C., 2001. Forecasting the Erika oil spills. In: International Oil Spill Conference, 2001. American Petroleum Institute, pp. 649–655.

De Dominicis, M., Leuzzi, G., Monti, P., Pinardi, N., Poulain, P., 2012. Eddy diffusivity derived from drifter data for dispersion model applications. Ocean Dynam. 62, 1381–1398. https://doi.org/10.1007/s10236-012-0564-2.

De Dominicis, M., Pinardi, N., Zodiatis, G., Archetti, R., 2013a. MEDSLIK-II, a Lagrangian marine surface oil spill model for short-term forecasting – part 2: numerical simulations and validations. Geosci. Model Dev. 6, 1871–1888.

De Dominicis, M., Pinardi, N., Zodiatis, G., Lardner, R., 2013b. MEDSLIK-II, a Lagrangian marine surface oil spill model for short term forecasting – part 1: theory. Geosci. Model Dev. 6, 1851–1869.

De Dominicis, M., Falchetti, S., Trotta, F., Pinardi, N., Giacomelli, L., Napolitano, E., Fazioli, L., et al., 2014. A relocatable ocean model in support of environmental emergencies – the Costa Concordia emergency case. Ocean Dyn. 64, 667–688.

Dearden, Chris, Culmer, Tim, Brooke, Richard, 2021. Performance measures for validation of oil spill dispersion models based on satellite and coastal data. IEEE J. Ocean. Eng. 47 (1), 126–140.

Escudier, R., Clementi, E., Omar, M., Cipollone, A., Pistoia, J., Aydogdu, A., Drudi, M., et al., 2020. Mediterranean Sea Physical Reanalysis (CMEMS MED-currents) (Version 1). Available at: doi:10. 25423/CMCC/MEDSEA_MULTIYEAR_PHY_006_004_E3R1; Copernicus Monitoring Environment Marine Service (CMEMS).

Ferrer, S., Mengual, A., Sepp Neves, A.A.. iMAGINE UC4 - Segmented Oil Spills (0. 1. 1). https://zenodo.org/record/11354663.

Fingas, M., Brown, C., 2013. Oil spill remote sensing. In: Earth System Monitoring. Springer, New York, NY, pp. 337–388.

Fingas, M., Brown, C.E., 2018. A review of oil spill remote sensing. Sensors 18 (91).

Fingas, M., Brown, C., 2025. Review of oil spill remote sensing: the current state of the art. Oil Spill Sci. Technol. 309–358.

Frazier Peter, I., 2018. A Tutorial on Bayesian Optimization. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1807.02811. July.

Frazier, P.I., Wang, J., 2016. Bayesian optimization for materials design. In: Lookman, T., Alexander, F.J., Rajan, K. (Eds.), Information Science for Materials Discovery and Design. Springer, pp. 45–75.

Gettelman, Andrew, Gagne, David J., Chen, Chia-Cheng, Christensen, Michael W., Lebo, Zachary J., Morrison, Hugh, Gantos, George, 2021. Machine learning the warm rain process. J. Adv. Model. Earth Syst. 13 (2). https://doi.org/10.1029/2020MS002268 e2020MS002268.

Gottwald, Georg A., Crommelin, Daan T., Franzke, Christian L.E., 2017. In: Franzke, Christian L.E., O'Kane, Terence J. (Eds.), Stochastic Climate Theory. In Nonlinear and Stochastic Climate Dynamics. Cambridge University Press, pp. 209–240. https://doi.org/10.1017/9781316339251.009.

Hecht, A., Pinardi, N., Robinson, A.R., 1988. Currents, water masses, eddies, and jets in the Mediterranean Levantine Basin. J. Phys. Oceanogr. 18, 1320–1353. https://doi.org/10.1175/1520-0485(1988)018<1320:CWMEAJ>2.0.CO;2.

Huang, J.C., 1983. A Review of the State-of-the-art of Oil Spill Fate/Behavior Models. American Petroleum Institute, pp. 313–322.

Jones, D., Brischke, C., 2017. 8 - modelling. In: Jones, D., Brischke, C. (Eds.), Performance of Bio-based Building Materials. Woodhead Publishing, pp. 483–546. https://doi.org/10.1016/B978-0-08-100982-6.00008-2.

Jones, Donald R., Schonlau, Matthias, Welch, William J., 1998. Efficient global optimization of expensive black-box functions. J. Glob. Optim. 13 (4), 455–492. https://doi.org/10.1023/A:1008306431147.

Keramea, P., Spanoudaki, K., Zodiatis, G., Gikas, G., Sylaios, G., 2021. Oil spill Modeling: a critical review on current trends, perspectives, and challenges. J Mar Sci Eng 9 (2), 181.

Keramea, P., Kokkos, N., Zodiatis, G., Sylaios, G., 2023a. Modes of operation and forcing in oil spill modeling: state-of-art, deficiencies and challenges. J Mar Sci Eng 11 (6), 1165.

Keramea, P., Kokkos, N., Zodiatis, G., Sylaios, G., Coppini, G., Peña, J., Benjumeda, P., et al., 2023b. Satellite imagery in evaluating oil spill modelling scenarios for the Syrian oil spill crisis, summer 2021. Front. Mar. Sci. 10, 1264261.

Kushner, H.J., 1964. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. J. Basic Eng. 86 (1), 97–106. https://doi.org/10.1115/1.3653121.

Leifer, I., Lehr, W.J., Simecek-Beatty, D., Bradley, E., Clark, R., Dennison, P., Wozencraft, J., 2012. State of the art satellite and airborne marine oil spill remote sensing: application to the BP Deepwater horizon oil spill. Remote Sens. Environ. 124, 185–209.

Liu, Y., Weisberg, R.H., 2011. Evaluation of trajectory Modeling in different dynamic regions using normalized cumulative Lagrangian separation. J. Geophys. Res. 116, C09013. https://doi.org/10.1029/2010JC006837.
Liu, Darong, Li, Yan, Mu, Lin, 2023. Parameterization Modeling for wind drift factor in oil spill drift trajectory simulation based on machine learning. Front. Mar. Sci. 10, 1222347.
Lizotte, Daniel, et al., 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. arXiv. https://arxiv.org/pdf/1807.02811.
Mayer, L., Jakobsson, M., Allen, G., Dorschel, B., Falconer, R., Ferrini, V., et al., 2018. The Nippon foundation—GEBCO seabed 2030 project: the quest to see the world's oceans completely mapped by 2030. Geosciences 8 (2), 63. https://doi.org/10.3390/geosciences8020063.
Mills, D., 2016. Unknown Title. Butterworth-Heinemann. https://doi.org/10.1016/C2014-0-02678-0.
Močkus, J., 1975. On bayesian methods for seeking the extremum. In: Marchuk, G.I. (Ed.), Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 400–404.
Mockus, Jonas, 1989. Bayesian approach to global optimization. In: Mathematics and Its Applications. Springer Nature (Netherlands). https://doi.org/10.1007/978-94-009-0909-0.
Naveiro, Roi, Tang, Becky, 2024. Simulation Based Bayesian Optimization. arXiv preprint arXiv:2401. 10811v1. https://arxiv.org/abs/2401.10811v1.
Negoescu, Diana M., Frazier, Peter I., Powell, Warren B., 2011. The knowledge-gradient algorithm for sequencing experiments in drug discovery. INFORMS J. Comput. 23 (3), 346–363. https://doi.org/10.1287/ijoc.1100.0417.
Olascoaga, M.J., Haller, G., 2012. Forecasting sudden changes in environmental pollution patterns. Proc. Natl. Acad. Sci. 109 (13), 4738–4743. https://doi.org/10.1073/pnas.1200398109.
Özgökmen, Tamay M., Chassignet, Eric P., Dawson, Clint N., Dukhovskoy, Dmitry, Jacobs, Gregg, Ledwell, James, Garcia-Pineda, Oscar, et al., 2016. Over what area did the oil and gas spread during the 2010 Deepwater horizon oil spill? Oceanography 29 (3), 96–107.
Packwood, Daniel, 2017. Bayesian Optimization for Materials Science, vol. 3. Springer.
Peacock, T., Haller, G., 2013. Lagrangian coherent structures: the hidden skeleton of fluid flows. Phys. Today 66 (2), 41–47.
Polinov, S., Bookman, R., Levin, N., 2021. Spatial and temporal assessment of oil spills in the Mediterranean Sea. Mar. Pollut. Bull. 167, 112338. https://doi.org/10.1016/j.marpolbul.2021.112338.
Ranjit, Prasanna, Mercy, Ganapathy, Gopinath, Sridhar, Kalaivani, Arumugham, Vikram, 2019. Efficient deep learning hyperparameter tuning using cloud infrastructure: intelligent distributed hyperparameter tuning with bayesian optimization in the cloud. In: 2019 IEEE 12th International Conference on Cloud Computing (CLOUD), pp. 520–522. https://doi.org/10.1109/CLOUD.2019.00097.
Rasmussen, Carl E., Williams, Christopher K.I., 2006. Gaussian Processes for Machine Learning, vol. 1. Springer. In: http://www.gaussianprocess.org/gpml/.
Reed, M., Johansen, Ø., Brandvik, P.J., et al., 1999. Oil spill Modeling towards the close of the 20th century: overview of the state of the art. Spill Sci Technol Bull 5 (1), 3–16. https://doi.org/10.1016/S1353-2561(99)00011-9.
Roberts, Nigel, 2008. Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. Meteorol. Appl. 15 (1), 163–169. https://doi.org/10.1002/met.57.
Roberts, Nigel M., Lean, Humphrey W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Weather Rev. 136 (1), 78–97. https://doi.org/10.1175/2007MWR2123.1.
Robinson, A., Hecht, A., Pinardi, N., et al., 1987. Small synoptic/mesoscale eddies and energetic variability of the eastern Levantine Basin. Nature 327 (6122), 131–134. https://doi.org/10.1038/327131a0.
Shaban, M., Salim, R., Abu Khalifeh, H., Khelifi, A., Shalaby, A., El-Mashad, S., El-Baz, A., 2021. A deep-learning framework for the detection of oil Spills from SAR data. Sensors 21 (7), 2351. https://doi.org/10.3390/s21072351.
Shin, Jaehyun, Baik, Jong-Jin, 2022. Parameterization of stochastically entraining convection using machine learning technique. J. Adv. Model. Earth Syst. 14 (5). https://doi.org/10.1029/2021MS002817 e2021MS002817.
Simecek-Beatty, D., Lehr, W.J., 2021. Oil spill forecast assessment using fractions skill score. Mar. Pollut. Bull. 164, 112041. https://doi.org/10.1016/j.marpolbul.2021.112041.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms. In: Advances in Neural Information Processing Systems, pp. 2951–2959.
Snoek, Jasper, Larochelle, Hugo, Adams, Ryan P., 2015. Scalable Bayesian optimization using deep neural networks. In: Proceedings of the 32nd International Conference on Machine Learning, 37. PMLR, pp. 2167–2175. In: https://proceedings.mlr.press/v37/snoek15.pdf.
Spaulding, M.L., 1988. A state-of-the-art review of oil spill trajectory and fate modeling. Oil Chem. Pollut. 4, 39–55. https://doi.org/10.1016/0269-8579(88)90017-7.
Spaulding, M.L., 2017. State of the art review and future directions in oil spill modeling. Mar. Pollut. Bull. 115 (1–2), 7–19. https://doi.org/10.1016/j.marpolbul.2016.10.054.
Srinivas, Niranjan, Krause, Andreas, Kakade, Sham, Seeger, Matthias, 2010. Gaussian process optimization in the bandit setting: no regret and experimental design. In: ICML 2010 - Proceedings, 27th International Conference on Machine Learning, pp. 1015–1022.
Storer, Benjamin A., Buzzicotti, Michele, Khatri, Hemant, Griffies, Steven M., Aluie, Hussein, 2022. Global energy spectrum of the general oceanic circulation. Nat. Commun. 13 (1), 5314. https://doi.org/10.1038/s41467-022-33031-3.
Wang, Peidong, Yuval, Johannes, O'Gorman, Paul A., 2022. Non-local parameterization of atmospheric subgrid processes with neural networks. J. Adv. Model. Earth Syst. 14 (10). https://doi.org/10.1029/2022MS002984 e2022MS002984.
Wessel, P., Smith, W.H., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res. Solid Earth 101 (B4), 8741–8743. https://doi.org/10.1029/96JB00104.
Wunsch, Carl, 1998. The work done by the wind on the oceanic general circulation. J. Phys. Oceanogr. 28 (12), 2332–2340. https://doi.org/10.1175/1520-0485(1998)028<2332:TWDBTW>2.0.CO;2.
Zhang, Yichi, Apley, Daniel, Chen, Wei, 2020. Bayesian optimization for materials design with mixed quantitative and qualitative variables. Sci. Rep. 10 (march). https://doi.org/10.1038/s41598-020-60652-9.
Zhilinskas, A.I., 1976. Single-step Bayesian search method for an extremum of functions of a single variable. Cybernetics 11 (1), 160–166. https://doi.org/10.1007/bf01069961.
Zodiatis, G., Brenner, S., Gertman, I., Ozer, T., Simoncelli, S., Ioannou, M., Savva, S., 2023. Twenty years of in-situ monitoring in the south-eastern Mediterranean Levantine Basin: basic elements of the thermohaline structure and of the mesoscale circulation during 1995-2015. Front. Mar. Sci. 9, 1074504. https://doi.org/10.3389/fmars.2022.1074504.