
🌊 Oil Spill Forecasting Model — Full Technical Summary (English Version)

1. Overview

The goal of this project is to develop a **spatio-temporal deep learning model capable of forecasting the evolution of marine oil spills**.
Due to the scarcity of real-world spill datasets, we designed a **physics-informed synthetic data generator** based on peer-reviewed literature, and trained a **ConvLSTM-based predictor** to learn oil drift, diffusion, and degradation patterns.

---

2. System Architecture

The full pipeline consists of three main layers:

Layer 1 — Physics·Chemistry·Biology Mini-Simulator

Files: `utils/physics_ops.py`, `utils/chemistry_ops.py`, `utils/biology_ops.py`
Data generator: `data/make_synthetic_data.py`

Purpose

Real-world oil spill datasets are extremely limited. Thus, we extract **quantitative parameters** from scientific literature and encode them into lightweight mini-models:

✔ Physics

- Oil advection driven by ocean currents

- Wind drag (~2–4%)
- Diffusion coefficients from empirical studies
- Temperature-dependent viscosity

✔ Chemistry

- Evaporation rate constants
- Photooxidation kinetics
- Dissolution behavior
- UV fluorescence ↔ concentration relationships

✔ Biology

- Biodegradation rates
- Oxygen depletion & recovery (DO dynamics)
- Toxicity / plankton decline ratios

Each component provides a step function:

oil_next = step_physics(oil)

oil_next = step_chemistry(oil_next)

oil_next = step_biology(oil_next)

Output

Synthetic training sequences:

(Sequence length = 30 steps, Grid size = 40×40, Channels = 8 features)

Saved as a compressed `.npz` file.

Layer 2 — Deep Learning Forecasting Model (ConvLSTM)

File: `ai_predictor/model_conv_lstm.py`

Why ConvLSTM?

Oil spills evolve in both **space and time**, influenced by currents, diffusion, and non-linear chemical changes.
ConvLSTM is ideal because it:

- Handles **grid-based spatial data** directly
- Learns **temporal evolution** (LSTM)
- Captures **local spatial correlations** (convolution)
- Is widely used for smoke dispersion, weather, and pollutant forecasting

Model structure

- Input: `T_in` past frames (e.g., 10)
- Output: `T_out` future frames (e.g., 5)
- Layers:
    - Multiple ConvLSTM layers
    - 1×1 convolution for channel reduction
- Loss: MSE
- Optimizer: Adam/AdamW

Layer 3 — Training Pipeline

Files:
`ai_predictor/train_predictor.py`,
`ai_predictor/dataset.py`

Function

- Loads synthetic `.npz` sequences
- Splits into train/valid
- Batches into PyTorch DataLoader
- Trains ConvLSTM
- Saves best-performing weights (`predictor_best.pt`)

Outcome

The trained model learns to simulate the **forward evolution** of oil thickness maps from historical states.

---

3. Full Pipeline Summary (One Diagram)

[Literature-derived parameters]

↓

Physics / Chemistry / Biology step models

↓

Synthetic spill simulator → train_sequences.npz

↓

PyTorch Dataset → ConvLSTM training

↓

Trained model (predict future oil spread)

↓

(Next) Web dashboard / HTML visualization

↓

Apply to real scenarios (e.g., Taean oil spill)

---

4. Key Features of the Model

1. Physics-informed modeling

Not purely data-driven — incorporates real scientific parameters:

- diffusion coefficients
- biodegradation constants
- photooxidation half-lives
- viscosity–temperature equations

This ensures **physically consistent synthetic data**.

2. Synthetic data solves real-world scarcity

Most oil spill datasets are not publicly available.
Our approach bypasses this limitation by generating **high-quality, literature-based simulations**.

3. Spatio-Temporal Forecasting

ConvLSTM predicts **future oil movement maps** in a fully autoregressive manner.

4. Modular & Extensible

New features can be added:

- wave height
- salinity
- real SAR imagery
- real current forecasts (HYCOM, CMEMS)

---

5. Current Completion Status

✔ Codebase fully constructed

- Mini-simulator modules (physics/chemistry/biology)
- Synthetic data generator
- ConvLSTM model
- Training pipeline

✔ GitHub project initialized

You can now push updates and deploy versions easily.

✖ Synthetic data not yet generated

Once `train_sequences.npz` is created successfully, training will begin.

---

6. One-sentence summary (for presentations)

> **We developed a physics-informed synthetic spill simulator and trained a ConvLSTM network to forecast marine oil spill evolution in space and time.**