Multi-Source Constrained Machine Learning for Oceanic Parameters Forecasting

Pujan Pokhrel
GulfSCEI,
University of New Orleans
New Orleans, United States
email: ppokhre1@uno.edu

Austin B. Schmidt
GulfSCEI,
University of New Orleans
New Orleans, United States
email: sbaustin@uno.edu

Md Meftahul Ferdaus
GulfSCEI,
University of New Orleans
New Orleans, United States
email: mferdaus@uno.edu

Elias Ioup
Center for Geospatial Sciences
Naval Research Laboratory
Mississippi, United States
email: elias.z.ioup.civ@us.navy.mil

Julian Simeonov
Sediment Dynamics Section
Naval Research Laboratory
Mississippi, United States
email: julian.a.simeonov.civ@us.navy.mil

Mahdi Abdelguerfi
GulfSCEI,
University of New Orleans
New Orleans, United States
email: gulfsceidirector@uno.edu

***Abstract*—This study evaluates multi-source fusion techniques for environmental forecasting, focusing on their effectiveness in predicting oceanic and atmospheric variables. Three neural network architectures are examined: a baseline Long Short Term Memory (LSTM) model, a Softmax Fusion model, and a Lagrangian Fusion model. A central component of the approach is the incorporation of physics-based constraints during training to ensure physically consistent predictions. Results based on Root Mean Squared Error (RMSE) indicate that fusion-based models consistently outperform the baseline for wave-related and thermodynamic variables such as air and water temperature. RMSE reductions for these variables range from approximately 5% to over 40%, driven by the models' ability to enforce spatiotemporal smoothness and reduce spatial variability. In contrast, wind components show higher RMSE in fusion models, highlighting a trade-off between global physical consistency and the accurate representation of localized, high-variance wind phenomena. These findings demonstrate the advantages of fusion architectures for improving buoy-based wave and thermodynamic forecasts, while suggesting that future work on wind predictions may benefit from adaptive regularization or hybrid loss functions to capture both global coherence and local detail better.**

Keywords-multi-source fusion, physics-informed neural networks, constrained machine learning.

I. INTRODUCTION

Accurate short-term forecasting of oceanic and atmospheric variables is important for a wide range of applications, including maritime navigation, coastal hazard mitigation, and climate monitoring [1][2]. Traditional Numerical Weather Prediction (NWP) uses partial differential equations based on physics [3], which provide physically consistent forecasts. However, these models require substantial computation and exhibit sensitivity to initial conditions, restricting their real-time utility. In contrast, data-driven methods based on Long Short Term Memory Networks (LSTMs) have demonstrated strong performance in forecasting buoy-observed quantities such as gush speed, wave height, pressure, and sea surface temperature [4]–[8]. These methods offer faster inference and ease of deployment, but frequently rely on single-source input, not exploiting auxiliary information from satellite retrievals or reanalysis datasets that can offer broader spatial and temporal perspectives.

In practice, buoy measurements are often noisy and sparse, and simulations may not capture the full dynamics at relevant scales. These limitations motivate the study of fusion architectures, which aim to produce well-calibrated initial states that physical laws can later constrain during forecasting. Multisource fusion combines diverse observations, such as buoy recordings, ECMWF Reanalysis v5 (ERA5) reanalysis, and NOAA Global Forecast System (GFS) analysis data to generate more accurate estimates of surface wind components (u_{10}, v_{10}). Adaptive fusion models, especially those that use softmax-based weight mechanisms, dynamically allocate trust among sources in response to measurement discrepancies. However, the model outputs are not always physically plausible. The weights can be negative or do not sum up to one, and the forecasts may violate the principles of energy or mass conservation [9]. Physics-Informed Neural Networks (PINNs) address these issues by embedding PDEs into loss functions. Still, they often require explicit physical equations and incur heavy training costs, which limit their applicability in noisy or incomplete data regimes [9].

Schmidt *et al.* [6][7] and Pokhrel *et al.* [10][11] studied fusion models in the context of physics-based forecasting. In these models, numerical model predictions are integrated with observations in a weighting scheme to generate fused predictions. Specifically, Pokhrel *et al.* extended this to multiple data sources in [12]. In this study, we extend the previous studies by using differentiable constraints to obtain physically consistent predictions. This study focuses on architectures that both fuse multisource data effectively and enforce physical constraints on learned outputs. This study frames fusion as a method for generating physically meaningful initial estimates that remain valid when subjected to downstream constraints. Our primary contribution is the Lagrangian Fusion Model, which employs an augmented Lagrangian formulation to constrain fusion weights (ensuring non-negativity and unit sum) to enforce consistency between predicted wind vectors and observed Gust

Speed (GST $\approx \sqrt{u_{10}^2 + v_{10}^2}$), without requiring explicit PDE formulas.

To validate our approach, three end-to-end architectures are compared:

1) **Baseline Model**: a sequence-to-sequence LSTM that forecasts Gust Speed (GST), Wave Heights (WVHT), Pressure (PRES), Water Temperature (WTMP), and wind components (u_{10} and v_{10}) using buoy data alone. This model benchmarks single-source performance under noisy conditions.
2) **Softmax Fusion Model**: adds softmax-weighted fusion of multisource inputs for u_{10} and v_{10}, while maintaining direct forecasting for other variables. It demonstrates the benefit of fusing information, along with enforcing the physical constraints. The weights of different sources are generated implicitly using the softmax layer in this approach.
3) **Lagrangian Fusion Model**: enhances the softmax approach by incorporating augmented Lagrangian penalties to enforce physically meaningful fusion weights and outputs, ensuring meteorological consistency and robustness in the presence of noisy measurements.

All fusion models are trained using buoy-observed targets and multi-source wind input. Physics-inspired loss terms, including constraint penalties and temporal smoothing, are integrated to improve the forecast fidelity. Evaluated on three years of data (2021–2023) with an 80/20 train/test split and rolling-horizon validation, the Lagrangian Fusion Model achieves substantial improvements while maintaining physically consistent behaviors. Our study advances the state-of-the-art by presenting a robust and interpretable fusion strategy that combines data-driven flexibility with physical realism, generating reliable initial states for downstream constrained forecasting or data assimilation frameworks.

This paper is organized as follows. Section 2 reviews related work. Section 3 details the dataset, model architectures, and training procedures. Section 4 presents the experimental results and their analysis. Section 5 discusses the limitations of the study and outlines directions for future research. Finally, Section 6 concludes the study.

II. Related Work

A. Neural Forecasting for Marine Variables

LSTMs have become important in marine environmental forecasting because of their ability to model temporal dependencies. For example, Bonino *et al.* combined buoy and reanalysis inputs in an LSTM framework and achieved a significant error reduction in sea surface temperature prediction compared to purely physical models [4]. Park *et al.* and Kim *et al.* showed comparable improvements in wave height and wind speed forecasting using buoy-only LSTM [13]. These models demonstrate the benefits of data-driven modeling but remain constrained by single-source inputs and lack of explicit physical constraints, which our work overcomes through multi-source fusion and constrained optimization.

B. Multisource Fusion Techniques

Optimal interpolation and classical data assimilation methods combine multiple sources of observational data by modeling error covariances [14]. Recent deep learning-based fusion approaches learn weighting schemes directly from data. Scher and Messori fused satellite and reanalysis data using convolutional architectures for global weather forecasts, reporting improved accuracy in spatial detail [15]. Shaw *et al.* used Spatiotemporal Dynamic Graph (SDG) networks to fuse multiple features to improve prediction accuracy for significant wave heights prediction [16]. However, neither approach enforces explicit physical constraints on fusion outputs. Our Softmax Fusion Model extends these works by enabling component-specific weighting for wind vectors, while our Lagrangian Fusion Model further ensures weight validity and meteorological consistency.

C. Physics-Informed Learning and Constrained Optimization

PINNs integrate physical laws into deep networks via loss-term penalties, enabling PDE-informed prediction capabilities [9]. While effective, they require explicit PDE formulations and tend to be computationally expensive, particularly in noisy or incomplete-data regimes [9]. Constrained optimization frameworks, such as those reviewed by Kotari *et al.*, embed structural constraints into network parameters using Lagrange multipliers, without relying on full physical equations [17]. Gramacy *et al.* demonstrated this approach in fluid dynamics by enforcing mass conservation using augmented Lagrangian methods [18]. Our Lagrangian Fusion Model adopts this strategy, applying constraints to both fusion weights and output predictions, ensuring that noisy, fused initial estimates remain physically consistent and usable for downstream data assimilation or forecasting frameworks.

Furthermore, we frame fusion as a preconditioning step that enhances subsequent constraint enforcement and data-assimilation efforts. This is supported by assimilation literature, which highlights the importance of good-quality initial states for robust real-time forecasts and correction of model deficiencies [19]–[21].

III. Methodology

A. Dataset

Table 1 presents the essential summary statistics (count, mean, standard deviation, minimum, and maximum) for each variable. Note that the fusion models are weighted using ERA-5 and NOAA wind fields in the cost function and not used as a part of input to the models.

Table 1 displays the count, mean, standard deviation, minimum, and maximum for each feature. In particular, the buoy-measured wind direction and wind speed, each with 29,778 samples, have means of 223.27° (std 96.07°) and 6.06 m/s (std 3.01 m/s), respectively. The zonal and meridional wind components at 10 meters ($u10\_buoy$ and $v10\_buoy$) were derived from the gust speed and wind direction. These wind-related variables are fused in the loss function using either fixed Lagrangian weights or learned weights via a Softmax

TABLE 1. BASIC SUMMARY STATISTICS FOR ALL VARIABLES

Variable	Meaning (Units)	Mean	Std	Min	Max
WDIR	Wind Direction (degrees)	223.27	96.07	2.44	360.00
WSPD	Wind Speed (m/s)	6.48	2.94	0.00	21.00
GST	Gust Speed (m/s)	8.38	3.61	0.21	26.61
WVHT	Wave Height (m)	2.12	1.03	0.35	10.08
DPD	Dominant Wave Period (s)	10.42	2.83	3.67	22.29
APD	Average Wave Period (s)	6.66	1.27	3.29	14.11
MWD	Mean Wave Direction (radians)	3.45	1.61	0.02	6.28
WTMP	Water Temperature (°C)	18.82	7.29	2.90	30.39
DEWP	Dew Point Temperature (°C)	15.08	6.74	-9.40	28.52
ATMP	Air Temperature (°C)	17.99	7.11	-3.26	30.84

layer. The data set comprises a total of 94,442 observations that span the period from January 1, 2021, to December 31, 2023 and the points are sampled every 3 hours. We utilize global predictions for ERA-5 and NOAA datasets.

B. Data Preprocessing

The raw time-series data is preprocessed through several key steps to ensure temporal continuity and improve model performance. First, the features that are redundant, irrelevant, or represent metadata not useful for forecasting are removed. To maintain the integrity of temporal sequences, rows with missing values are discarded instead of applying imputation, which can introduce artifacts. All continuous variables are standardized to have zero mean and unit variance, which facilitates faster convergence and stabilizes training in recurrent architectures. We then segment the data using 80/20 time-based split for training and testing. This setup supports multi-step forecasting aligned with short-term marine prediction needs.

Since the buoy observations lack corresponding numerical simulations, their values are predicted directly. The fusion targets in this study are the 10-meter wind components, u_{10} and v_{10}, estimated using multi-source integration of buoy measurements, ERA5 reanalysis, and NOAA data products. Although this work focuses on a single variable pair, the approach is extendable to other features, provided suitable numerical models are available for integration.

C. Model Architectures

This study evaluates three neural network architectures built on a sequence-to-sequence (Seq2Seq) LSTM backbone for multi-step environmental forecasting. All models share a common structure and incorporate physics-based constraints during training.

The architecture consists of an LSTM encoder and decoder, with an optional attention layer to improve temporal representation. The overall design is:

$$\begin{aligned} \text{Encoder:} \quad & \mathbf{h}_t, (\mathbf{h}_T, \mathbf{s}_T) = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}) \\ \text{Attention:} \quad & \mathbf{z}_t = \text{Attention}(\{\mathbf{h}_t\}) \\ \text{Decoder:} \quad & \mathbf{d}_t, (\mathbf{d}_*, \mathbf{s}_*) = \text{LSTM}(\mathbf{z}_t, \mathbf{d}_{t-1}) \end{aligned} \tag{1}$$

Here, $\mathbf{x}_t \in \mathbb{R}^{D_{in}}$ is the input at time t, and the encoder summarizes the input sequence into hidden states, and s represents the state at the last layer. The attention layer (when used) reweights the encoder output to form a context vector $\mathbf{z}_t$, which helps the decoder focus on relevant time steps during prediction. The decoder then produces latent outputs for the forecast window.

To encourage physically meaningful predictions, a physics-based loss term is added in the fusion models:

$$\mathcal{L}_{\text{physics}} = \sum_{c \in \mathcal{C}} w_c \cdot \phi_c(\hat{\mathbf{Y}}) \tag{2}$$

Each constraint $c \in \mathcal{C}$ has a weight w_c and penalty function ϕ_c that measures violations based on model outputs $\hat{\mathbf{Y}}$. Details are provided in Section III-D.

1) Baseline Model: The Baseline Model employs a straightforward prediction head applied to the decoder's latent representation:

$$\hat{\mathbf{Y}} = \text{Linear}(\text{ReLU}(\text{Linear}(\mathbf{d}))) \tag{3}$$

In this model, the constraints of physics are strictly enforced through the $\mathcal{L}_{\text{physics}}$ term, acting as a soft regularization.

2) Softmax Fusion Model: The Softmax Fusion Model extends the Baseline by incorporating a dynamic, attention-based fusion mechanism for wind component prediction. This model learns softmax-normalized weights to combine wind estimates from multiple input data sources (e.g., buoy, ERA5and NOAA inputs). The architecture integrates source-aware fusion with physics-constrained training:

$$\begin{aligned} \mathbf{\Lambda}_u &= \text{Softmax}(\text{MLP}_u(\mathbf{d})) \\ \mathbf{\Lambda}_v &= \text{Softmax}(\text{MLP}_v(\mathbf{d})) \\ \hat{\mathbf{Y}}_{\text{direct}} &= g_d(\mathbf{d}) \\ \hat{\mathbf{Y}}_{\text{wind}} &= g_w(\mathbf{d}) \\ \hat{u}_{\text{fused}} &= \sum_s \lambda_u^{(s)} u_s \\ \hat{v}_{\text{fused}} &= \sum_s \lambda_v^{(s)} v_s \\ \mathcal{L}_{\text{total}} &= \mathcal{L}_{\text{MSE}} + \alpha \mathcal{L}_{\text{physics}} \end{aligned} \tag{4}$$

where $\mathbf{\Lambda}_u$ and $\mathbf{\Lambda}_v$ are the learned softmax weights for the wind components u and v of the sources s, respectively. u_s and v_s represent the wind components from source s. g_d and g_w are multi-layer perceptrons (MLPs) that predict

direct variables and wind residuals, respectively. The softmax mechanism facilitates the differentiable and interpretable weighting of each modality in each forecast timestep, thereby improving the robustness to noise and variations in data quality between different sources. Figure 1 displays the architecture of the fusion models. Note that the buoy data is used to make predictions and the other datasets are only used during training.

3) Lagrangian Fusion Model: While the Softmax Fusion approach implicitly ensures non-negativity and sum-to-one constraints for fusion weights, it may limit the model's expressiveness by tightly coupling all weights through a single normalization. To address this, we adopt an augmented Lagrangian formulation that allows for explicit control over constraints, enabling the model to learn fusion weights more flexibly.

The Lagrangian Fusion Model learns raw (unnormalized) weights $\mathbf{\Lambda}_u$ and $\mathbf{\Lambda}_v$ for wind component fusion. The fusion loss is then defined as:

$$\begin{aligned}
\mathbf{\Lambda}_u &= \text{Linear}_u(\mathbf{d}) \\
\mathbf{\Lambda}_v &= \text{Linear}_v(\mathbf{d}) \\
\mathcal{L}_{\text{constraints}} =& \mu_u \left(\sum_s \lambda_u^{(s)} - 1 \right) + \frac{\rho}{2} \left(\sum_s \lambda_u^{(s)} - 1 \right)^2 \\
&+ \sum_s \nu_u^{(s)} (-\lambda_u^{(s)}) + \frac{\rho}{2} \sum_s (-\lambda_u^{(s)})^2 \\
&+ \mu_v \left(\sum_s \lambda_v^{(s)} - 1 \right) + \frac{\rho}{2} \left(\sum_s \lambda_v^{(s)} - 1 \right)^2 \\
&+ \sum_s \nu_v^{(s)} (-\lambda_v^{(s)}) + \frac{\rho}{2} \sum_s (-\lambda_v^{(s)})^2 \\
\mathcal{L}_{\text{total}} &= \mathcal{L}_{\text{MSE}} + \alpha \mathcal{L}_{\text{physics}} + \beta \mathcal{L}_{\text{constraints}}
\end{aligned} \tag{5}$$

Here, μ_u, μ_v are Lagrange multipliers for the sum-to-one equality constraint, and $\nu_u^{(s)}, \nu_v^{(s)}$ are Lagrange multipliers for the non-negativity inequality constraints. ρ is the penalty parameter of the augmented Lagrangian. This formulation allows for decoupled optimization of weights while explicitly enforcing physical constraints.

D. Physics-Based Constraints

To promote physically consistent forecasts, the fusion models incorporate a set of differentiable physics-based constraints within the loss term $\mathcal{L}_{\text{physics}}$. These constraints are grounded in well-established relationships from wave physics and are summarized in Table 2.

E. Constraint Integration

Each constraint is implemented as a differentiable component of the total loss, allowing for gradient-based optimization. Denote each constraint as $\phi_c(\hat{\mathbf{Y}})$, then the total physics-informed penalty is:

$$\mathcal{L}_{\text{physics}} = \sum_c w_c \cdot \phi_c(\hat{\mathbf{Y}}) \tag{6}$$

where w_c is a fixed weight determined through empirical tuning. The wave development constraint and wave steepness limit are assigned lower weights (e.g., $w_c = 0.1$), reflecting their supportive role in shaping output realism without overwhelming the primary forecasting loss. The gravitational constant $g = 9.8\,\text{m/s}^2$ is used in the wave steepness constraint, derived from classical wave theory.

This constraint formulation ensures that model predictions remain physically plausible, especially for wave-related parameters where empirical structure is well understood.

F. Training Framework

All models were trained using a consistent framework to ensure a fair comparison of their architectural merits. The key configuration parameters are as follows:

- **Optimizer:** Adam optimizer with an initial learning rate (η) of 5×10^{-4}. For the Lagrangian model, separate Adam optimizers were used for main model parameters and Lagrangian multipliers, both with an initial learning rate of 5×10^{-4}.
- **Learning Rate Schedule:** A step-down learning rate scheduler was applied, reducing the learning rate by a factor of 0.5 (for Baseline and Softmax) or 0.05 (for Lagrangian) every 25 epochs.
- **Regularization:** Gradient clipping with a maximum L2 norm of 1.0 was applied to prevent exploding gradients during training.
- **Batch Size:** Training was performed using a batch size of 64 sequences.
- **Physics Weight:** The weighting factor α for the $\mathcal{L}_{\text{physics}}$ term was set to 10^{-4} for all models. For the Lagrangian model, an additional weight $\beta = 0.1$ was applied to the $\mathcal{L}_{\text{constraints}}$ term.
- **Epochs:** Each model was trained for 50 epochs.

IV. Results and Discussion

This section presents a comprehensive analysis of the proposed models' performance, focusing on quantitative evaluation of first-step prediction (3 hours ahead) RMSE and a discussion of the observed trends and their implications.

Table 3 shows the first-step Root Mean Squared Error (RMSE) for the Baseline, Softmax, and Lagrangian models across a suite of oceanic and atmospheric variables. A clear trend of improvement is observed for the fusion-based models compared to the Baseline, especially for wave-related and temperature variables.

The most substantial relative improvements occurred in the prediction of air temperature (ATMP), sea surface temperature (WTMP), and average wave period (APD). For example, the RMSE for WTMP dropped from 0.2286 to 0.1793 with the Softmax model. These gains can be attributed to:

1) **Spatiotemporal Smoothness:** Variables like temperature evolve more smoothly in space and time, aligning well with the inductive bias of fusion models that leverage weighted averaging and regularization.

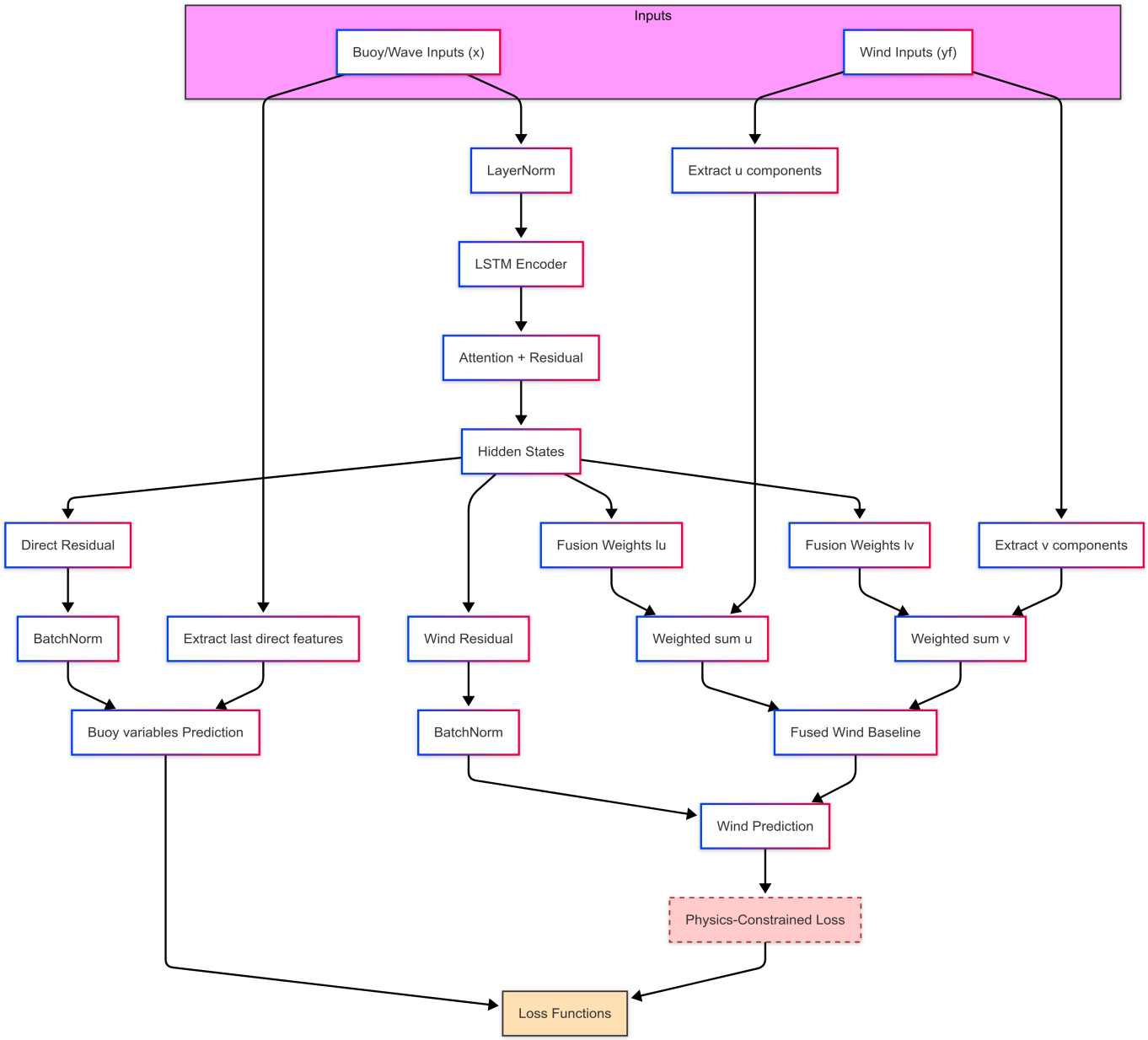

Figure 1. Model Architecture. Note that the wind inputs from multiple sources and fusion weights training is used only during training in the loss function calculation. The testing loop proceeds as normal.

2) **Low Spatial Variability:** ATMP and WTMP are governed by large-scale synoptic systems, making them more amenable to fusion from multiple sources. Fusion reduces noise and integrates consistent large-scale patterns effectively.

These variables also tend to have higher signal-to-noise ratios, enhancing the ability of models to generalize from reanalysis and auxiliary data sources.

Moderate improvements are also seen in wave-related variables (WVHT, DPD, APD) and wind speed (WSPD), suggesting that fusion helps smooth out some of the high-frequency noise or discrepancies among sources.

Interestingly, wind vector components (u_{10} and v_{10}) show degraded performance under fusion. These variables are subject to rapid and localized changes influenced by turbulence, topography, and mesoscale systems. Pointwise buoy observations may not align with gridded inputs (like ERA5), and fusion could oversmooth sharp transitions or localized bursts, reducing apparent accuracy.

Fusion models, especially the Lagrangian variant, demonstrate strong performance for coherent variables like temperature and wave metrics. Their inductive bias toward smoothness and large-scale consistency aligns with the physics and statistical characteristics of these variables. However, for high-variance, small-scale phenomena like wind vectors, fusion may inadvertently degrade performance. This does not imply

TABLE 2. PHYSICS CONSTRAINTS APPLIED TO FUSION MODELS

Constraint	Formula	Basis	Weight
Wave Development Limit	$\mathrm{ReLU}(\mathrm{WVHT} - 1.3 \times 0.02 \times \mathrm{GST}^2)$	Pierson–Moskowitz spectrum	0.3
Wave Period Alignment	$\mathrm{ReLU}(\lvert\mathrm{DPD} - 0.55 \times \mathrm{GST}\rvert - 1.5)$	JONSWAP spectrum relation	0.4
Wave Steepness Limit	$\mathrm{ReLU}(\mathrm{WVHT} - 0.1 \times \frac{g \cdot \mathrm{DPD}^2}{2\pi})$	Non-breaking wave criterion	0.3

TABLE 3. FIRST-STEP PREDICTION (3 HOURS AHEAD) ROOT MEAN SQUARED ERROR (RMSE) COMPARISON (BEST VALUES IN **BOLD**)

Variable	Baseline	Softmax	Lagrangian
GST	**1.1134**	1.1217	1.1196
WVHT	0.1988	0.1961	**0.1945**
DPD	0.9134	0.9038	**0.9032**
APD	0.3305	0.2996	**0.2964**
MWD	0.4713	0.4362	**0.4353**
ATMP	0.3783	0.3767	**0.3691**
WTMP	0.2286	**0.1793**	0.1799
WSPD	0.9521	**0.9477**	0.9510
DEWP	0.5302	**0.5268**	0.5304
u_{10}	**6.5590**	9.9120	9.8987
v_{10}	**5.8451**	14.8446	14.8560

model failure but highlights a resolution mismatch and the challenges of noisy labels in wind observations. To address this, future work should explore hybrid or multi-scale loss structures that balance smoothness with local adaptability. Adaptive regularization or uncertainty-aware weighting may help retain fidelity in turbulent regimes while maintaining the benefits of fusion elsewhere.

A. Additional Physics Constraints

Beyond wind speed consistency and temporal smoothness, the following physics-based penalties, computed from the GFS or ERA5 fields, can further improve forecasts. Each constraint is tagged with the approach it benefits most and a brief rationale:

- **Boundary-Layer Momentum Balance (Turbulence-driven).** In the lowest $\sim$100 m the Coriolis force is negligible, so we can enforce a non-rotating, turbulent momentum balance forced by the wind at the layer top. Let τ be the turbulent stress tensor and $\mathbf{F}_{\mathrm{turb}}$ the wind-driven forcing. We then penalize deviations from
$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho\,(\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla p + \nabla\cdot\tau + \rho\,\mathbf{F}_{\mathrm{turb}},$$
via
$$\mathcal{L}_{\mathrm{BL}} = \mathbb{E}\Big[\big\|\rho\,\partial_t\hat{\mathbf{u}} + \rho\,(\hat{\mathbf{u}}\cdot\nabla)\hat{\mathbf{u}} + \nabla\hat{p} - \nabla\cdot\hat{\tau} - \rho\,\hat{\mathbf{F}}_{\mathrm{turb}}\big\|^2\Big].$$
(Best addressed via high-frequency observations + turbulence parameterizations.) This leverages detailed measurements to capture rapid, subgrid-scale fluctuations that a coarse grid would miss.
- **Hydrostatic Balance (Grid-resolved).** Using pressures at multiple levels (e.g. 925 hPa, 850 hPa) we can approximate
$$\frac{\Delta\hat{p}}{\Delta z} = -\,\hat{\rho}\,g, \quad \hat{\rho} = \frac{\hat{p}}{R_d\,\hat{T}}.$$
Anchoring $\hat{p}_{\mathrm{surf}}$ with buoy measurements, we can penalize any violation via
$$\mathcal{L}_{\mathrm{hydro}} = \mathbb{E}\Big[\big(\tfrac{\Delta\hat{p}}{\Delta z} + \hat{\rho}\,g\big)^2\Big].$$
(Resolved accurately on a typical numerical grid.) Large-scale pressure gradients vary smoothly over kilometers, so a grid model naturally enforces hydrostatic balance without needing high-frequency corrections.
- **Moisture Consistency (Hybrid).** We can enforce Clausius–Clapeyron and relative-humidity relations from ERA5 dewpoint and air temperature:
$$e_s(T) = 6.112\exp\Big(\tfrac{17.67\,T}{T+243.5}\Big), \quad \mathrm{RH} \approx 100\,\frac{e(\mathrm{DEWP})}{e_s(\mathrm{ATMP})}.$$
The loss penalizes squared errors between $(\hat{T}, \hat{T}_d, \hat{\mathrm{RH}})$ and these thermodynamic constraints:
$$\mathcal{L}_{\mathrm{moist}} = \mathbb{E}\Big[\big\|\big(\hat{T}, \hat{T}_d, \hat{\mathrm{RH}}\big) - \mathrm{thermo}\big\|^2\Big].$$
(Combines observations for humidity with grid-model thermodynamics.) Observations provide accurate moisture content but grid models capture large-scale transport, so a hybrid enforces both.
- **Surface Energy Budget (Hybrid).** Conservation of heat at the surface implies
$$\rho\,c_p\,\frac{\partial T}{\partial t} \approx H_s + H_l + R_n - Q_h,$$
where R_n is net radiation and H_s, H_l, Q_h are turbulent fluxes. We penalize
$$\mathcal{L}_{\mathrm{energy}} = \mathbb{E}\Big[\big(\rho\,c_p\,\partial_t\hat{T} - (H_s + H_l + R_n - Q_h)\big)^2\Big].$$
(Requires both grid-scale radiative fields and observations of fluxes.) Radiative inputs are well resolved on

the grid, but surface fluxes fluctuate rapidly and need observational constraints.

- **Mass Continuity (Grid-resolved).** Assuming incompressible horizontal flow, any nonzero divergence is penalized:

$$\mathcal{L}_{\text{mass}} = \mathbb{E}\big[(\nabla \cdot \hat{\mathbf{u}})^2\big].$$

Horizontal winds can be taken directly from ERA5 or buoy observations, but this balance is naturally enforced on a resolved grid. *(Grid models handle continuity inherently over large scales.)*

B. Limitations and Future Work

The current model enforces only wind speed and smoothness physics, leaving out important couplings such as moisture, energy, and large-scale balance. It is a single-location model without spatial gradients, which limits its ability to conserve mass or energy across a region. Moreover, the coarser resolution of ERA5 and GFS fields compared to buoy data can introduce alignment errors. In such cases, it is better to handle numerical data and observations separately, as is common in data assimilation. Moreover, since buoy data represent turbulence scale phenomena which is not incorporated in global/regional models, physics loss can be independently applied to the fusion of numerical data, apart from observations.

From the experimental setup described in the paper, the method succeeds when the true physical parameter lies within the span of the available data sources. A natural concern is what happens if the optimal value falls outside this range, i.e., if all data sources are biased. In such scenarios, incorporating the governing PDEs as an additional "data source" within the framework ensures that predictions remain physically consistent.

Future work includes incorporating spatial embeddings using convolutional or graph modules to enforce conservation of mass, momentum, and energy over regions. We also plan to apply multiscale constraints that address synoptic-scale pressure and hydrostatic balance as well as subdaily energy and moisture budgets. Adaptive scheduling of Lagrangian multipliers will help gradually increase physics penalties during training to avoid conflicts with data-driven loss. Additional extensions include modeling ocean-atmosphere coupling with a mixed-layer sea surface temperature model and improving uncertainty quantification through ensembles or Bayesian networks calibrated with ERA5 error statistics.

V. Conclusion

This study developed and evaluated multi-source fusion architectures for accurate short-term forecasting of oceanic and atmospheric variables, addressing key limitations of existing data-driven and traditional numerical weather prediction models. Integrating heterogeneous data sources (buoy, ERA5, NOAA) through fusion models consistently improved predictive performance compared to a Baseline LSTM model relying on single-source observations. Multi-source fusion architectures offer a compelling improvement over single-source models for short-term forecasting of oceanic and atmospheric variables. Across a diverse set of targets, fusion consistently improves precision for parameters characterized by smooth, large-scale dynamics, most notably wave metrics and temperature fields, by leveraging complementary information from reanalysis, and observational products (buoys). This trend underscores the ability of models to take advantage of spatio-temporal coherence and reduce noise when integrating heterogeneous inputs.

Although fusion models showed a trade-off in performance for wind components (u_{10}, v_{10}), highlighting the challenge of balancing global physical consistency with capturing localized, high-variance wind phenomena, the overall benefits for buoy forecasting systems are clear. These findings emphasize the value of physics-informed constrained machine learning in enhancing the reliability and interpretability of environmental forecasts. Future work will extend the constraint set to include large-scale balances (boundary layer momentum, hydrostatic) and energy budgets, and will explore dynamic fusion weights and spatial embeddings to better reconcile global coherence with local fidelity. This research lays the groundwork for robust, physically consistent data-driven forecasting systems vital to maritime and climate-related applications.

Acknowledgment

This work was partly supported by the U.S. Department of the Navy, Naval Research Laboratory under contract N00173-20-2-C007. The work of Austin Schmidt was funded by a SMART (Science, Mathematics and Research for Transformation) Department of Defense (DoD) scholarship for service.

References

[1] M. J. Kaiser and A. G. Pulsipher, "The impact of weather and ocean forecasting on hydrocarbon production and pollution management in the gulf of mexico," *Energy policy*, vol. 35, no. 2, pp. 966–983, 2007.

[2] A. J. Hobday, C. M. Spillman, J. Paige Eveson, and J. R. Hartog, "Seasonal forecasting for decision support in marine fisheries and aquaculture," *Fisheries Oceanography*, vol. 25, pp. 45–56, 2016.

[3] P. Bauer, "What if? numerical weather prediction at the crossroads," *Journal of the European Meteorological Society*, vol. 1, p. 100 002, 2024, ISSN: 2950-6301. DOI: https://doi.org/10.1016/j.jemets.2024.100002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2950630124000024.

[4] G. Bonino, G. Galimberti, S. Masina, R. McAdam, and E. Clementi, "Machine learning methods to predict sea surface temperature and marine heatwave occurrence: A case study of the mediterranean sea," 2, vol. 20, 2024, pp. 417–432. DOI: 10.5194/os-20-417-2024. [Online]. Available: https://os.copernicus.org/articles/20/417/2024/.

[5] S. Scher and G. Messori, "Weather and climate forecasting with neural networks: Using general circulation models (gcms) as a study ground," *Geoscientific Model Development*, vol. 12, pp. 2797–2809, 2019. DOI: 10.5194/gmd-12-2797-2019.

[6] A. B. Schmidt, P. Pokhrel, E. Ioup, D. Dobson, and M. Abdelguerfi, "An algorithm for modelling differential processes utilising a ratio-coupled loss," *Authorea Preprints*, 2024.

[7] A. B. Schmidt, P. Pokhrel, M. Abdelguerfi, E. Ioup, and D. Dobson, "Forecasting buoy observations using physics-informed neural networks," *IEEE Journal of Oceanic Engineering*, 2024.

[8] A. B. Schmidt *et al.*, "Evaluating hyperparameter selection techniques for the ratio-coupled loss function," *Authorea Preprints*, 2025.

[9] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[10] P. Pokhrel, M. Abdelguerfi, and E. Ioup, "A machine-learning and data assimilation forecasting framework for surface waves," *Quarterly Journal of the Royal Meteorological Society*, vol. 150, no. 759, pp. 958–975, 2024.

[11] P. Pokhrel, E. Ioup, M. T. Hoque, M. Abdelguerfi, and J. Simeonov, "Forecasting rogue waves in oceanic waters," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2020, pp. 634–640.

[12] P. Pokhrel, A. Schmidt, M. Abdelguerfi, and E. Ioup, "Generalized multi-source assimilation: A framework for cross-modal integration and source optimization," *Authorea Preprints*, 2025.

[13] F. C. Minuzzi and L. Farina, "A deep learning approach to predict significant wave height using long short-term memory," *Ocean Modelling*, vol. 181, p. 102 151, 2023, ISSN: 1463-5003. DOI: https://doi.org/10.1016/j.ocemod.2022.102151. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1463500322001652.

[14] T. Archambault, A. Filoche, A. Charantonis, D. Bereziat, and S. Thiria, "Learning of sea surface height interpolation from multi-variate simulated satellite observations," *arXiv*, 2024. arXiv: 2310.07626 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2310.07626.

[15] X. Sun *et al.*, "Fuxi weather: A data-to-forecast machine learning system for global weather," Aug. 2024. DOI: 10.48550/arXiv.2408.05472.

[16] S. Yin, T. Wu, J. Yu, X. He, and L. Xu, "A significant wave height prediction method with ocean characteristics fusion and spatiotemporal dynamic graph modeling," *Acta Oceanologica Sinica*, vol. 43, pp. 13–33, Mar. 2025. DOI: 10.1007/s13131-024-2450-4.

[17] J. Kotary, F. Fioretto, P. Van Hentenryck, and B. Wilder, "End-to-end constrained optimization learning: A survey," *arXiv preprint arXiv:2103.16378*, 2021.

[18] R. B. Gramacy *et al.*, "Modeling an augmented lagrangian for blackbox constrained optimization," *Technometrics*, vol. 58, no. 1, pp. 1–11, 2016.

[19] F. Zhang, S. Zhang, and S. Yang, "Data assimilation in numerical weather and climate models," *Advances in Meteorology*, vol. 2015, p. 626 893, 2015. DOI: 10.1155/2015/626893.

[20] E. Kalnay, "Atmospheric modeling, data assimilation and predictability," in *Cambridge University Press*, Cambridge, UK, 2003.

[21] M. A. Balmaseda and D. L. T. Anderson, "Impact of initialization strategies and observations on seasonal forecast skill," *Geophysical Research Letters*, vol. 36, no. 1, p. L01702, 2009. DOI: 10.1029/2008GL035561.