COCE 2024

The First International Conference on Technologies for Marine and Coastal Ecosystems

ISBN: 978-1-68558-325-5

November 17th – 21st, 2024

Valencia, Spain

COCE 2024 Editors

Jaime Lloret, Universitat Politècnica de València, Spain

COCE 2024

Forward

The First International Conference on Technologies for Marine and Coastal Ecosystems (COCE 2024), held between November 17th, 2024, and November 21st, 2024, in Valencia, Spain, was a leading event dedicated to exploring and promoting innovative solutions and technologies for oceanic research and conservation. This conference served as a platform for collaboration among researchers, engineers, policymakers, and industry leaders, aiming to showcase cutting-edge advancements in smart ocean monitoring, data analytics, and sustainable ocean engineering. COCE 2024 highlights the transformative potential of technologies such as Artificial Intelligence (AI), Internet of Things (IoT), robotics, and data analytics in addressing critical challenges facing our oceans and marine ecosystems.

We take here the opportunity to warmly thank all the members of the COCE 2024 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to COCE 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the COCE 2024 organizing committee for their help in handling the logistics of this event.

We hope that COCE 2024 was a successful international forum for the exchange of ideas and results between academia and industry for the promotion of progress in the field of marine and coastal ecosystems

COCE 2024 Chairs

COCE 2024 General Chair
Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COCE 2024 Steering Committee
Cesar Bordehore, University of Alicante, Spain
José Tena, Catholic University of Valencia, Spain
Miquel Ardid, Universitat Politecnica de Valencia, Spain
Kiseon Kim, Gwangju Institute of Science and Technology, Korea
Eduardo Jorge Belda Perez, Universitat Politecnica de Valencia, Spain
Mahdi Abdelguerfi, The University of New Orleans, USA

COCE 2024
Committee

COCE 2024 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COCE 2024 Steering Committee

Cesar Bordehore, University of Alicante, Spain
José Tena, Catholic University of Valencia, Spain
Miquel Ardid, Universitat Politecnica de Valencia, Spain
Kiseon Kim, Gwangju Institute of Science and Technology, Korea
Eduardo Jorge Belda Perez, Universitat Politecnica de Valencia, Spain
Mahdi Abdelguerfi, The University of New Orleans, USA

COCE 2024 Technical Program Committee

Mahdi Abdelguerfi, The University of New Orleans, USA
Lavanya Addepally, Universitat Politecnica de Valencia, Spain
Chiguk Ahn, (Former Director and Vice President) Korea Institute of Aids to Navigation, Korea
Christhian N. Aldana, Universidad Nacional de Frontera, Peru
Miquel Ardid, Universitat Politecnica de Valencia, Spain
Salva Ardid, Universitat Politecnica de Valencia, Spain
Eduardo Jorge Belda Perez, Universitat Politecnica de Valencia, Spain
Cesar Bordehore, University of Alicante, Spain
Lismer Andres Caceres Najarro, Chosun University, Korea
Rafael Caparó, Universidad Nacional de Ingeniería, Peru
Bok Kyoung Choi, Korea Institute of Ocean Science and Technology (KIOST), Korea
Edson S. Gomi, University of São Paulo, Brazil
Kiseon Kim, Gwangju Institute of Science and Technology, Korea
Jeong-A Lee, Chosun University, Korea
Alvaro Liebana Carrascosa, Universitat Politecnica de Valencia, Spain
Satish R. Jondhale, Savitribai Phule Pune University, Pune, India
Alberto Ivars Palomares, Universitat Politecnica de Valencia, Spain
Carlos Antonio Puig Mengual, Universitat Politecnica de Valencia, Spain
Amine Rghioui, Université Ibn Zohr, Morocco
Juan A. Montiel-Nelson, University of Las Palmas de Gran Canaria | Institute for Applied Microelectronics, Spain
Minh Tuan Nguyen, Posts and Telecommunications Institute of Technology (PTIT), Vietnam
V. Pandiyaraju, VIT University, Chennai, India
Pujan Pokhrel, Amazon AWS & GulfSCEI, USA
Edwin Saavedra, Universidad Nacional de Frontera, Peru
Fermin Saavedra Cano, Universidad Nacional de Frontera, Peru
Yesenia Saavedra, Universidad Nacional de Frontera, Peru
María-Teresa Sebastiá-Frasquet, Universitat Politecnica de Valencia, Spain
Vinie Lee Silva Alvarado, Universitat Politecnica de Valencia, Spain

José Tena, Catholic University of Valencia, Spain
Norsuzila Ya'acob, Universiti Teknologi MARA, Malaysia
Chan-su Yang, Korea Institute of Ocean Science and Technology (KIOST), Korea
Vincent Xiaochuan Yu, University of New Orleans, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Running Hydraulics Simulations at Scale Using Inductiva Python API

Luís Sarmento, Hugo Penedones, Sérgio Santos, and Paulo Barbosa
Inductiva.AI - Porto, Portugal
e-mail: {sarmento | hpenedones | ssantos, | pbarbosa}@inductiva.ai

***Abstract*—This demo will introduce participants to the Inductiva cloud-based High Performance Computing (HPC) platform, designed for large-scale simulations. Inductiva provides pre-installed ready-to-use simulation models for marine, coastal, and hydraulic projects. Researchers and engineers can run these simulations immediately using Python scripts. In addition to scripting, users can also interact through an intuitive web interface that simplifies data management, resource allocation, and cost control, giving them full oversight of their simulations.**

Keywords-Cloud-based Computing, Numerical Simulation, Python API.

I. Introduction

Despite advancements in simulation technologies, engineers and scientists still face challenges in running numerical models, from configuring simulators and setting up suitable hardware to managing large data volumes.

To help engineers and scientists run large-scale simulations, companies like Rescale [1] and Sabalcore [2] introduced cloud-based solutions that enable simulation jobs on powerful cloud machines. However, these platforms often lack model-specific support, particularly in fields like hydraulics, coastal engineering, and marine sciences, and are primarily optimized for single simulation runs, making them less suitable for workflows requiring hundreds or thousands of simulation variations. Such scalability is essential for exploring design spaces, conducting sensitivity analyses, or generating training data for Physics-Artificial Intelligence (AI) models.

Inductiva offers a cloud-based HPC infrastructure controlled via simple Python scripts and a user-friendly web interface, streamlining complex simulation tasks so experts can focus on their models without concern for underlying infrastructure.

This demo introduces Inductiva's Application Programming Interface (API)-driven infrastructure through several use cases, and demonstrates why Inductiva is an ideal solution for automating complex simulation workflows. In Sections II and III, we will cover how to submit simulations on Inductiva's platform using Python, with practical examples of hydraulics models like Reef3D [3] and SWASH [4], both of which come pre-installed. We will also explore how the platform allows users to generalize simulation parameters and run simulations in parallel to explore different scenarios. Sections IV and V cover Inductiva's task and cost-management features, while Section VI presents supported hydraulics simulators. Section VII concludes the work.

II. The Infinite Laptop Experience

The workflow with Inductiva is simple: users start running their simulation on their local machine and when more computational power is needed—such as for running a longer version of the simulation—users can redirect their workloads to cloud machines equipped with hundreds of central processing units (CPUs), using Inductiva's Python API. Consider the example in **Figure 1**, where a user switches from running a Reef3D simulation on their local machine to a cloud machine with 112 virtual CPUs (vCPUs) provided by Inductiva, using a few lines of Python code:

```
import inductiva

my_machine = inductiva.resources.MachineGroup(
    machine_type="c2d-highcpu-112",
    data_disk_gb=20)
my_machine.start()

reef3d = inductiva.simulators.REEF3D()

simulation_task = reef3d.run(
    input_dir="./3D_Dam_Break_with_Obstacle",
    on=my_machine,
    n_vcpus=112,
    storage_dir="3D_dam_break_with_obstacle")

simulation_task.wait()
simulation_task.download_outputs()

my_machine.terminate()
```

Figure 1. Starting a Reef3D simulation on the cloud from a local machine.

In addition to running simulations on a single large machine, as shown above, users can also set up groups of machines to operate as a Message Passing Interface (MPI) cluster. This allows them to scale their computations even further without worrying about the resource allocation or formalities required by traditional HPC facilities.

III. A Tool for Optimal Design and Physics-Machine Learning (ML)

Inductiva's API-driven computational infrastructure provides a fresh approach for thinking about marine and coastal engineering projects. By orchestrating multiple simulation jobs in parallel with simple Python scripts, users can systematically explore different parameter configurations or scenarios for a single base simulation. With Inductiva's *templating mechanism*, users can transform specific parameters in the simulation configuration files into variables that they can easily adjust programmatically through Python.

As an example, the code below (see **Figure 2**) illustrates how to run five SWASH simulations on five different machines, each testing a different value for the base water level. The API's TemplateManager class is used to automatically adjust the water level in the SWASH simulation configuration

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

files. All simulations are executed in parallel, running in the background to allow for simultaneous processing.

Using a simple “FOR loop", users can launch thousands of variations of a base simulation case. This capability allows users to: **i)** explore different parameters in optimal design projects, **ii)** perform sensitivity analysis on models, or **iii)** generate large-scale synthetic data for training Physics-Machine Learning (ML) models (see one of our tutorials [5] where we explain how to use the API to generate the data required to replicate the work by [6]).

```
import inductiva

# Start 5 preemptible machines
machines = inductiva.resources.MachineGroup(
    machine_type="c2-standard-30",
    num_machines=5,
    spot=True)
machines.start()

swash = inductiva.simulators.SWASH()

water_levels = [3.5, 3.75, 4.0, 4.5, 5.0]

# Launch one simulations for each water level.
for i, water_level in enumerate(water_levels):
    target_dir = f"./my_outputs/swash-sim-{i}"
    inductiva.TemplateManager.render_dir(
        source_dir=template_dir,
        target_dir=target_dir,
        water_level=water_level)

    simulation_task = swash.run(
        input_dir=target_dir,
        sim_config_filename="input.sws",
        on=machines)
```

Figure 2. Running five variations of a SWASH simulation in parallel.

Unlike many providers of cloud resources for simulation that focus on providing resources to run *a single* simulation, Inductiva distinguishes itself by offering an efficient and convenient solution for users who need to run multiple variations of a base simulation case. This capability empowers users to explore different scenarios—something that is typically overlooked by other platforms.

IV. Task and Data Management

One of the key challenges when running simulations at scale is managing all the simulation tasks, especially when it comes to tracking inputs, outputs, and statuses across different projects. To address this, Inductiva provides the **Inductiva Console** [7], a web-based **User Interface (UI)** that enables users to monitor the status of all tasks, organize them by project, download outputs, and track resource usage, including costs. Users can also get an overview of all their activity over time and drill down into specific dates to review the simulation tasks executed on those days.

V. Performance Metrics and Cost Control

Numerical simulation is costly. The performance-to-cost ratio of the computational infrastructure, along with the project’s budget, determines how many and how large simulations can be run. To address this, Inductiva maintains full transparency on all costs, including computation, storage, and auxiliary tasks (e.g., compressing and moving output data).

The platform also helps users save by optimizing resource selection and configuration. A key strategy is using “spot instances”—cloud machines that can be preempted by the provider but are up to 91% cheaper than standard instances [8].

VI. Available Hydraulics Simulators

Table 1 lists the simulation models currently supported by Inductiva for marine, coastal, and hydraulic sciences. New simulators are regularly added. To suggest a specific simulator to be integrated, please e-mail *support@inductiva.ai*.

TABLE I. Ready to use simulators available via Inductiva API

Simulator / Model	Versions
DualSPHysics	5.2.1
Reef3D	24.02
SCHISM	5.11
SPlisHSPlasH	2.13.0
SWAN	41.45
SWASH	9.01A, 10.01
XBeach	1.23, 1.24

VII. Conclusion

Inductiva offers an efficient and scalable solution for engineers and researchers to run simulations. It integrates cloud resources into a Python API and an intuitive web interface. This enables users to focus on analysis and innovation in their respective fields engineering. Future work will focus on enhancing platform stability to support increasingly complex simulations, lowering the barrier of entry by making the platform more user-friendly, and integrating additional simulators.

Free experimental access to Inductiva can be obtained by registering on Inductiva’s website or directly via [7].

VIII. Conclusion

References

[1] Rescale, *Why rescale: Cloud-based high-performance computing*, Accessed: 2024-11-11, 2024.

[2] Sabalcore, *Sabalcore cloud computing solutions*, Accessed: 2024-11-11, 2024.

[3] Reef3D, *Reef3d: Open-source computational fluid dynamics*, Accessed: 2024-11-11, 2024.

[4] SWASH, *Swash: Simulating waves nearshore*, Accessed: 2024-11-11, 2024.

[5] *Generating synthetic data for training physics-ml models*, https://tutorials.inductiva.ai/generating-synthetic-data/synthetic-data-generation-1.html, Accessed: 2024-19-23.

[6] A. Sanchez-Gonzalez *et al.*, “Learning to simulate complex physics with graph networks,” in *International Conference on Machine Learning*, 2020.

[7] *Inductiva web console*, https://console.inductiva.ai/, Accessed: 2024-19-23.

[8] *Spot vms: Affordable compute instances suitable for batch jobs and fault-tolerant workloads*, https://cloud.google.com/spot-vms, Accessed: 2024-19-23.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

A Multiple-Location Modeling Scheme for Physics-Regularized Networks: Recurrent Forecasting of Fixed-Location Buoy Observations

Elias Sandner
CoDiS-Lab ISDS
Graz University of Technology
Graz, Austria
email: sandner@student.tugraz.at

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

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

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

David Dobson
Center for Geospatial Sciences
Naval Research Laboratory
Mississippi, United States
email: david.dobson@nrlssc.navy.mil

Christian Guetl
CoDiS-Lab ISDS
Graz University of Technology
Graz, Austria
email: c.guetl@tugraz.at

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

***Abstract*—Reliable oceanic and climate analysis depend on high-quality sensor readings, yet these systems commonly encounter significant sensor limitations, leading to missing data. Addressing this issue is critical for ensuring accurate forecasts and analyses. In this work, the data gap problem is studied by developing physics-regularized machine learning models with multiple-location modeling to forecast missing sensor data. Utilized are recurrent statistical surrogate models that generate hourly 24-hour forecasts. To train these models, we use a selection of five sensor features collected over three years. Introduced is a multi-location modeling scheme that uniquely combines sensor data from nearby buoys as a novel methodology. This approach allows for more stable and accurate predictions compared to forecasting with single buoy data alone. Our experiments reveal that grouping six buoys yields the best forecasting performance. Furthermore, we improve model accuracy by integrating buoy data with numerical ocean models and applying a physics-regularized loss function. This technique mitigates the impact of missing or erratic data, leading to more dependable 24-hour forecasts. Our findings demonstrate that the combination of multiple-location modeling and physics-based regularization enhances the stability and accuracy of oceanic data forecasting.**

Keywords-Buoy Forecast; Multiple Location Forecast; Physics-Regularized; Numerical Models; Surrogate Models.

I. Introduction

Accurate forecasting of ocean and climate parameters is useful in industry and research. Climate analysis, ocean pollution management, extreme weather event tracking, and marine life monitoring, as examples, benefit from ocean modeling techniques [1][2]. Numerical models use initial data collected by observation sensors through buoys, ships, or satellites as inputs to their underlying physical equations. The result is a full coverage analysis of the physical features used to describe ocean and climate states. The European Centre for Medium-Range Weather Forecasts (ECMWF) research institute provides such forecast models for use in decision-making and analysis problems. These models rely on accurate observations of physical phenomena in two major ways. The first is the initialization values used as initial conditions for the physical equations. Then, after producing an analysis based on those initial conditions, a historical re-analysis of the model is generated by integrating the results with real-world sensor data through Data Assimilation (DA) techniques [3]. ECMWF's fifth reanalysis experiment (ERA5) dataset is a popular example of this and often used in statistical surrogate modeling tasks [4]. One source of observations is the fleet of free-floating ocean buoys anchored to fixed locations which are maintained by the National Data Buoy Center (NDBC). The processes of initialization and DA both require reliable, consistent, and high-quality observed sensor measurements to maintain accurate representations. Technical limitations or poorly calibrated sensors can yield noisy interpretations, and physical damage or scheduled maintenance can completely halt data collection in that location. Ocean and climate analysis or reanalysis, which rely on the steady stream of ocean sensor data provided by the NDBC, might benefit from short to midterm regional and sparse forecasts in this situation. This gives justification to investigate deep learning surrogate models to conduct sparse observation forecasting for data assimilation and other uses.

The sensor data is geographically sparse in the sense that there are collections of buoys within the same region which are separated by up to hundreds of kilometers. A surrogate model might be trained for individual analysis using single buoy data or multiple buoy data [5][6]. In both cases, only a single buoy is modeled at a time, ignoring any spatial complexities between surrounding buoys. Therefore, it is reasonable to investigate methodologies for modeling buoys at multiple locations within a shared context. One method for implementing spatial and temporal frequencies in a deep learning model is by introducing a graph neural network that leverages graph convolutions based on a buoy's spatial relations [7]. However, this approach relies on additional training and space overhead that is undesirable in a lightweight framework. Instead, we focus on less specialized deep learning frameworks. To apply a higher degree of spatial awareness in the model, all buoys in

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

a region of interest are included simultaneously. By expanding the radius to include more buoys, it can be demonstrated whether the machine learning surrogate benefits from the collective information. In addition, the impact on a specific buoy is investigated to determine if contextual clues from surrounding buoys aid in forecasting current conditions at the chosen location.

Surrogate modeling of ocean observations is useful in scenarios where fast and relatively accurate forecasts are needed. One problem with these models is that forecasts are accurate on short leads but lose accuracy as the time horizon increases [8]. For this reason, research into ways to combine the surrogate model with data assimilation, physical equations, or other numerical models have been investigated. The main goal is to reduce surrogate model error by incorporating physics understanding through alternative data sources. For example, machine learning models have been successfully incorporated with DA techniques for improved results [9][10]. Similarly, the concept of the Physics-Informed Neural Network (PINN) uses physical Partial Differential Equations (PDEs) solved in the loss function at training time to improve physical understanding [11]. This is similar to our methodology because both simulations and observations are used in the loss function. In contrast, our work seeks to find the best combinations of the contributions of precalculated models and observations. PINNs notably try to optimize the loss from observations to fit a particular model. The paradigm explored in this work is the combination of full-coverage numerical data and sparsely collected sensor data to produce a more stable model. To this end, NDBC data and ERA5 reanalysis data are combined to train surrogate models to forecast buoy-derived geographically sparse ocean observations.

Noisy data derived from sensors has a significant impact on training forecast models. Missing data is a compounding concern when analyzing individually collected sensor observations. To combat this problem, a methodology to combine sensor derived data with numerically modeled data at near-point locations was proposed in [12]. The surrogate model training procedure uses both numerical data and sensor derived outputs when calculating the loss score. These values are combined using a ratio of the two error scores and then back propagated through the architecture. The previous work is extended by using a modified forecasting methodology and data representation to see if similar increases in model accuracy are achieved. Also, numerically modeled features are removed from the model input to reduce numerical reliance post-training. In this way, the underlying physical calculations of the numerical models regularize the statistical surrogate models at training time alone.

The main considerations of this research are in improving techniques of machine learning with ocean data for the predictive modeling of observed phenomena. Specifically, the goal is to investigate whether a geographically sparse set of data can be structured in such a way that the predictive ability of the model is stabilized over 24-hour forecast cycles. The viability of combining sensor data with numerically derived data at training time is consequently explored to verify if further improvements to the surrogate models can be made. Combining data in this way is a new technique which has only been evaluated in two experimental situations [12][13]. This raises the question of whether the combined approaches are better used in unison or separately. So, our main contributions are listed as follows:

- We describe a novel training scheme that uses multiple buoys and their observed parameters as input into the model;
- Using the physics-regularized loss function, we use a grid search to find the best ratio of data for each combined feature;
- By identifying methodologies to improve surrogate model performance, we give further justification to the use of statistical models in an oceanographic context.

The paper begins with a summary of existing related work in Section II. Section III introduces the specific methodology used. The data sources, chosen data representation, physics-regularized loss function, and the selected architectures for the neural networks are detailed. The section concludes with the setup of the executed experiments. In Section IV, the experimental results are presented. The paper concludes with a summary and suggestions for future work in Section V.

II. Related Work

Research into forecasting ocean parameters using machine learning methodologies is abundant in literature. Reviewing recent innovations in modeling the physical parameters yield methodologies for wind modeling [14], ocean wave height/direction [6][15][16], air temperature [17], and sea surface water temperature [18][19]. Popular models for surrogate ocean forecasting include the Long-Short Term Memory (LSTM) model and the attention head transformer model [20][21]. Their use is highlighted in recent ocean parameter forecasting, so the use of these layers are adopted in this research [6][12][22]–[26].

The availability and quantity of ERA5 data make it a suitable choice for machine learning-based surrogate models. In recent years, surrogate models trained or otherwise supported specifically by ERA5 data are studied for use for regional wave modeling [15], weather forecasting [27][28], earth surface temperature modeling [29], and sea surface temperature forecasting [30][18]. The data is also used when enhancing sensor predictions, for example, in the case of satellite sensor models [31][32]. Notably, the focus is usually on one oceanic feature or phenomenon.

To better analyze real-world conditions at a specific location, ocean buoys are modeled. Traditionally, this can be done by training a statistical model on the data collected from a single location for prediction [5][16]. Otherwise, multiple observation locations are combined into a single time series to train and predict an individual location. This incorporates a sense of spatial awareness into the model [14]. With the advent of deep learning, dozens of locations are used to train a single model for a more generalized approach. The model is

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

trained and used to forecast multiple single-location buoys, for example [6][12]. Specialized architectures, like graph neural networks, are also used to increase spatial awareness [30].

Combining both physically collected and numerically modeled data together in an ML context is a wide-reaching methodology. PINNs combine PDE solutions with collected data at training time [11]. This approach is usually used to solve numerical equations directly [33]. Forecasting physical parameters such as sea surface temperature [34] or physical phenomena such as storm surges [35] are also explored. DA for machine learning models is also becoming a popular topic to improve surrogate results [10][28]. Most similar to our research is the use of both modeled and observed data in the training and/or inference of ocean parameters, as seen in [12]. That work differentiates itself from typical DA by combining the data during the training scheme itself, instead of as a reanalysis step. It is different from typical PINN models through the lack of modeling of differential equations in the training scheme, instead using pre-calculated numerical outputs.

The methodologies proposed in this research are novel with respect to surrogate ocean parameter forecasting because the highlighted works typically combine fewer ocean features for forecasting, while we use all buoy sensors available and combine with ERA5 data. This work is novel in the context of buoy forecasting because a Multiple-Location Model (MLM), as defined in this work, either has not been investigated or is difficult to find in literature. Lastly, this extends the previous work of [12] by examining the proposed loss function in a different context, modeling multiple buoys with a different feature set, instead of single buoys. A significant difference is the use of numerically modeled features only in the training dataset and not as an input into the surrogate at inference time. This better aligns the methodology as a class of DA that combines data observations and models at training time to improve model performance when initial conditions are poor at inference time.

III. Methodology

Introduced are methodologies to verify the following research goals. One aim is to see if it is possible to improve the forecasting result for the collection of selected buoy observations. The MLM technique is expected to improve accuracy when increasing the number of modeled buoys. If so, this will give insights on whether the model weights can internalize connection in spatially sparse regions of interest. The physics-regularized loss function is improved by using modeled features in the training scheme while removing them from the inference input. Thus, it must be investigated if the method still improves model accuracy and whether it can be combined with the new sparse modeling scheme.

A. Sources of Data

The National Data Buoy Center (NDBC) is a United States-based organization and a division of the National Oceanic and Atmospheric Administration (NOAA). NDBC is responsible for collecting real-world observations of a wide range of ocean and climate feature measurements. In total, 1311 stations are deployed with 243 owned and maintained directly by NDBC [36], and they have been collecting the observation data since 1970. Within the framework of this project's experiments, the utilization of standard meteorological historical data (STDMET) is employed. As the data is of real-world origin, not every station provides feature data for every year. Furthermore, the available data may contain missing values [37]. Sensors in real-world settings may experience malfunctions or localized noise that are not helpful in a model training context.

Table 1. Considered NDBC Features and ERA5 Equivalents

Measurement	NDBC Feature	ERA5 Equivalent
Air Temperature (°F)	ATMP	2m temperature
Air Pressure (hPa)	PRES	Mean sea level pressure
Dewpoint Temperature (°F)	DEWP	2m dewpoint temperature
Water Temperature (°F)	WTMP	Sea surface temperature
Wind Speed (kts)	WSPD	10m v-component of wind, 10m u-component of wind

The ERA5 dataset, generated by the ECMWF, represents the latest advancement in a series of global atmospheric reanalysis datasets. Reanalysis entails the assimilation of observational data and model simulations to generate a coherent and extensive dataset. Data assimilation leverages observations from ships, buoys, and satellites with physical laws to ensure historical data is modeled as accurately as possible. Within this project, the ERA5 dataset, specifically identified as "ERA5 hourly data on single levels from 1940 to present", is utilized [4]. Hourly estimates are offered for a diverse set of atmospheric and ocean circulation variables. This dataset consists of a regridded subset extracted from the full ERA5 dataset while preserving its native resolution.

NDBC standard meteorological dataset provides 14 measurements [37] while ERA5 provides 262 variables [4]. Five selected NDBC measurements are also modeled within the ERA5 dataset, and therefore, are collected for further analysis and use within the physics-regularized deep learning task. Combining the five NDBC measurements with the corresponding five ERA5 values yields the collection of data required for the upcoming physics-regularized loss function. These five selected features are the focus of the surrogate forecasting task. The selection collected from NDBC and the ERA5 counterparts are depicted in Table 1. We use approximately three full years of hourly data, from 2020 to 2022.

B. Multi-Location Modeling

Oceans are complex and elaborate systems with interconnected dynamics. Therefore, it is hypothesized that enriching training data with measurements from several locations will empower the neural network to recognize spatial dependencies

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

between nearby buoys. The spatial-temporal data from each location is integrated in a unified dataset following a novel design in which each instance encapsulates measurements of multiple stations at a particular timestamp. That is, at N buoy locations, the features in Table 1 are extracted and combined into a single vector of one-hour increments to create an MLM dataset. Corresponding locations for the buoys in the ERA5 data are combined in the same way. The ERA5 data is reserved separately for use in the specialized loss function described in Subsection C. The example in Figure 1 shows the partitioning process. Any number of buoys may be selected in this generalized approach.

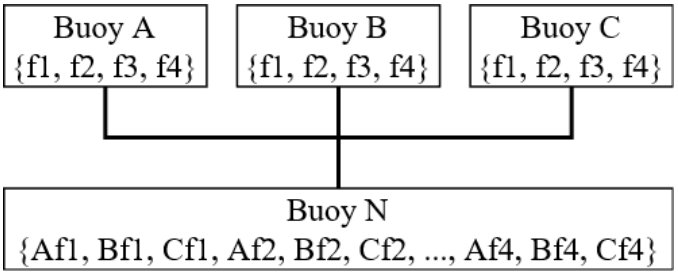

Figure 1. Example of the construction and final representation of the Multi-Location Modeling scheme.

In the figure, buoys A, B and C represent unique locations with a fixed buoy. Features [f1-f4] represent time series vectors of length T. Features are combined in alphabetical order into buoy N, such that sub-sampling a single element t gives the current condition of each buoy. The vector size of buoy N(t) is the number of features multiplied by number of buoys. The geographical locations influence one another implicitly through the neural network's hidden layers. Model weights take advantage of correlations learned at training time. This is contrasted against grid-based numerical models where forward regression is explicitly calculated with respect to adjacent grid cells.

Once the dataset has been generated, preprocessing is completed to make the problem more suitable for the deep learning architecture. First, we take the difference of each feature from t to $t+1$. This simplifies the forecast problem by reducing it to a gradient prediction problem. For each feature, the model only needs to produce an inference representing the rate of change. An inverse integration stage sums the recurrent predictions with the initial condition to get the real-world result. The data is normalized based on the mean and standard deviation seen in the training data alone, and the inverse is completed when investigating the results. Therefore, feature forecasts analyzed in the results section are in their respective scales.

C. Physics-Regularized Loss Function

The loss function of our deep learning models compares the training inference of the MLM vector with an individual observed value and numerically modeled value. By comparing the model inference to the NDBC and ERA5 data, two error scores are generated from the same prediction. Depending on interpolation or noise in the underlying data, one source will better represent the true conditions. So, a ratio of the two error scores is taken, and this error is used for back propagation in the model. The method is first proposed in [12]. The result is that, based on the ratio, the model is trained to approximate either source more strongly.

The disagreement between the datasets is proposed to improve the training procedure in two significant ways. The first is that the disagreement prevents overfitting to either individual source of training data. The second is that interpolated or distorted values are less likely to be frequent in both sources simultaneously. It is assumed that at least one source of data reasonably represents the underlying conditions for a given time step. Ultimately, the multi-location vector of buoy feature predictions is given along with the numerically derived values. Given the NDBC predictions and ERA5 values to be combined in the training procedure, the loss function is defined as follows in (1)-(6).

$$\Delta_1 = |\hat{y}_{\text{NDBC}} - y_{\text{NDBC}}| \tag{1}$$

$$\Delta_2 = |\hat{y}_{\text{NDBC}} - y_{\text{ERA5}}| \tag{2}$$

$$\Omega_{\text{coupled loss}} = (\alpha * \Delta_1) + ((1 - \alpha) * \Delta_2). \tag{3}$$

In (1) and (2), $\hat{y}$ represents the output vector of the surrogate model, while y represents the training ground truth vector. The source of modeling truth is determined by the subscript as y_{obs} or y_{model}. The error for each feature is weighted by coupling term α and represents a mixture of error calculated against two sources, ERA5 and NDBC. Importantly, α is constrained such that $0.0 \leq \alpha \leq 1.0$. When $\alpha = 0.0$, the NDBC term is completely shut off, and the model is only trained by comparing the ERA5 estimations. Otherwise, when $\alpha = 1.0$, only the NDBC data is used in training the model. Additional non-coupled features may be included in the surrogate and are defined as,

$$\Omega_{\text{model loss}} = |\hat{y}_{\text{ERA5}} - y_{\text{ERA5}}| \tag{4}$$

$$\Omega_{\text{observation loss}} = |\hat{y}_{\text{NDBC}} - y_{\text{NDBC}}|. \tag{5}$$

The remaining uncoupled features, as seen in (4) and (5), are used to collect error by comparing the predicted value with the relevant ground truth value. Each piecewise value is summed into the final loss function (6),

$$\Omega_{\text{total loss}} = \Omega_{\text{coupled loss}} + \Omega_{\text{model loss}} + \Omega_{\text{observation loss}}. \tag{6}$$

In this work, only the coupled loss is used. There are no uncoupled model or observation features, so $\Omega_{\text{total loss}} = \Omega_{\text{coupled loss}}$.

D. Deep Learning Architecture

Two deep learning architectures are considered for the forecasting task, given the prior defined multi-location dataset and the physics-coupled loss function. A LSTM unit model and an attention head transformer model are used. The two

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

models are chosen for their complexity and ability to generalize complex and recurrent time series data [38]. LSTM and attention head layers add more parameters to the hidden layers of the relatively shallow networks. The memory unit in the LSTM model particularly excels at storing prior knowledge for improved forecasts. The attention mechanism in the Transformer model is known for statistically weighing the importance of the previous input. The Transformer architecture typically implements more weights in the hidden layers of the model, so comparing this with the smaller LSTM model is often insightful. Both model architectures are implemented using the Python programming language and the TensorFlow machine learning platform [39].

TABLE 2. LSTM MODEL ARCHITECTURE BY LAYER FOR STAGE 2. THE TOTAL NUMBER OF TRAINABLE PARAMETERS IS 146,878.

Layer Type	Shape	Parameters
Input	*Variable*	0
LSTM	(N, 1, 128)	81,408
Dropout	N, 1, 128	0
LSTM	(N, 1, 64)	49,408
Dropout	N, 1, 64	0
LSTM	(N, 1, 32)	12,416
Dropout	N, 1, 32	0
LSTM	(N, 16)	3,136
Dense	(N, *Variable*)	510

TABLE 3. TRANSFORMER MODEL ARCHITECTURE BY LAYER FOR STAGE 2. THE TOTAL NUMBER OF TRAINABLE PARAMETERS IS 602,526.

Layer Type	Shape	Parameters
Input	*Variable*	3,968
Transformer Block	(N, 1, 128)	297,344
Transformer Block	(N, 1, 64)	297,344
Flatten	(N, 128)	0
Dense	(N, *Variable*)	3,870

Both models include dropout layers to prevent model overfitting during the training process. The dropout parameter is set to 0.1. The exact implementation of the LSTM model is found in Table 2 and the transformer model is in Table 3. In each table, N represents the variable batch size when training the model. The *Variable* input and output is equal to the product of the number of features and the number of buoys in each experiment. Therefore, each model's input and output shape is dependent on the number of ocean features and the number of buoys, as described in Figure 1. The Transformer blocks consist of multi head attention layers with four attention heads and feed forward dense layers of size 128 and dropout layers. Each model is trained for 100 epochs using a batch size of 64. The training procedure is set such that the model is given conditions at t and produces a forecast at $t+1$. By setting the input and output shape to the same length, the resulting model can be used to generate recurrent forecasts for any number of consecutive one-hour periods.

E. Experimental Test Case

Experiments using differing numbers of model locations and α ratio values are examined to verify the proposed methodology. The forecast results are centered on a particular buoy with the identification value of 42002. Buoy 42002 is located in the in the Gulf of Mexico. This buoy was chosen because of its central location relative to other buoys and because it has comparatively fewer missing values. Regardless of the number of buoys modeled, focusing on only one buoy assesses the hypothesis that the surrogate is improved when given spatial context. Considering the forecast results of all buoys is impactful, but outside the scope of this investigation.

When increasing the number of modeled buoys, results that are similar or better to the single-buoy model validate the proposed methodology for producing regional inferences. Increasingly better results indicate that the MLM approach either provides spatial context or model regularization through additional data. Tuning the α ratio is proposed to improve model forecast by combining observed and numerical data. So, an α value which produces better results than when $\alpha = 1.0$ (no regularization) is searched for.

For validation, the forecast ability of the surrogate models is examined for 24-hour intervals over the test data. The test data is comprised of the final three months of 2022. A recursive forecast scheme that uses the output of the proceeding inference as the input for the next inference is employed. The temporal resolution of the data is in one-hour increments, so 24 hourly forecasts are made for each interval. The model only uses the real conditions when making the first prediction in a cycle and when the test results are analyzed.

The Root Mean Square Error (RMSE) score (7) is used to determine the accuracy of the 24 hour forecast compared to the actual conditions at the buoy. The RMSE can be examined on a per-time step basis or as an average of all predictions.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

The MLM methodology is tested by increasing the number of buoys to forecast features in Table 1. Buoy 42002 is used to analyze ATMP, PRES, and WSPD, while buoy 42020 is analyzed for DEWP and WTMP. Two separate single-buoy experiments are conducted because no examined buoy contained all five features without extensive missing values. Using the coordinates of the central station, buoy 42002, and the Great Circle Distance [40] method, the distance to the remaining stations in the area was calculated. To evaluate the hypothesis, three datasets were created: one that exclusively covers a central station (Stage 0), one that covers stations within 600 km (Stage 1), and one that covers stations within 900 km (Stage 2). The exact specifications of the three experiments is found in Table 4. The geographical location of each buoy is displayed in Figure 2. To get baseline results, $\alpha = 1.0$ is selected, which means the entire training signal comes from the NDBC buoy dataset and the ERA5 data has no influence on the NDBC predictions.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

TABLE 4. INCLUSION OF BUOYS PER EXPERIMENT. GENERAL LOCATIONS OF EACH BUOY ARE FOUND IN FIGURE 2. STAGE 0 USES TWO SEPARATE INDIVIDUAL BUOYS TO CAPTURE FEATURES MISSING IN EITHER EXPERIMENT.

Stage	Distance (m)	# of Buoys	Buoy List
Stage 0	-	1	42002; 42020
Stage 1	600	4	42002, 42019, 42020, 42035
Stage 2	900	6	42002, 42003, 42019, 42020, 42035, 42040

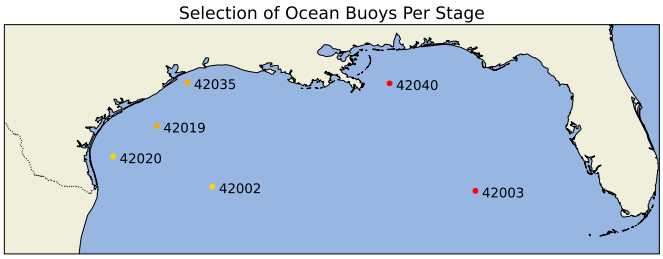

Figure 2. The geographical location of each buoy used in experiments. Stage 0 is labeled yellow, Stage 1 is labeled orange, and Stage 2 is labeled red.

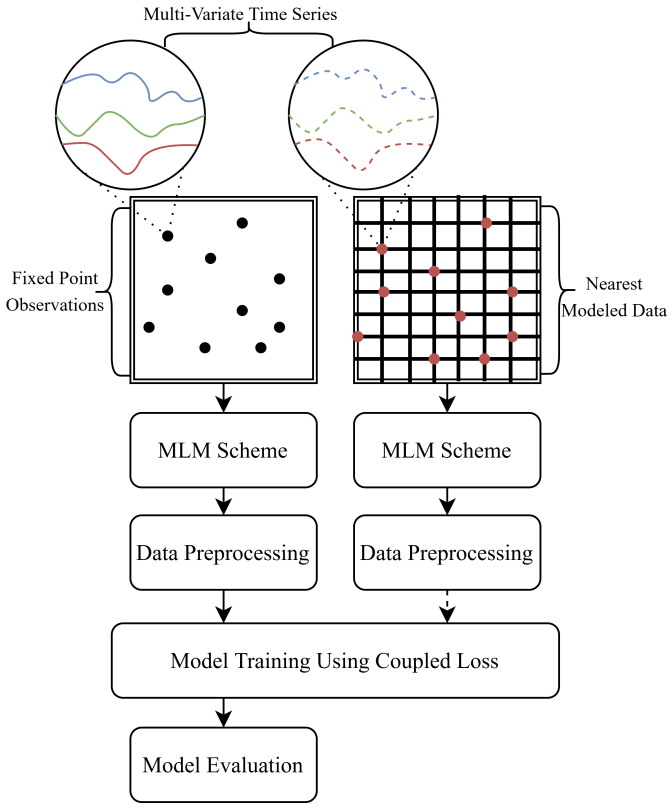

Figure 3. The data collection and processing pipeline.

Once the best number of buoys has been identified, the models re-examined with the physics-regularized loss. Finding the value for hyperparameter α that minimizes the error score is the main goal. A grid search of $\alpha \in [0.0, 1.0]$ with a step size of 0.05 is conducted. The surrogate model is retrained using the same random seed and for each α value. The minimal RMSE score on the test dataset denotes the best performing experimental setup.

For this experiment, the data collection and processing pipeline is divided into four stages, as outlined in Figure 3. First, the time series data is downloaded for each buoy from NOAA. Similarly, the ERA5 data is retrieved after selecting the appropriate geographical region and corresponding time periods. Second, the buoy's latitude, longitude, and time values are used to match the buoy data with the ERA5 data. Since these values may not align perfectly, we select the closest possible location and time points. Third, the MLM scheme is applied to format the forecast features as described in Figure 1. The same process is followed for the ERA5 data, which is stored in a separate vector for use in the coupled loss function (3). Finally, time series data processing is completed for all features. Missing data is interpolated, differencing is applied from t to $t+1$, and each feature is normalized based on the training data.

IV. RESULTS

The results for the MLM data technique are found for the LTSM model in Table 5 and the transformer model in Table 6. Each table displays the model, the examined buoy, and the calculated RMSE rounded to three decimal places for each stage. The average performance of all features is taken to find the overall best performing stage. From the initial stage experiments, a slight improvement is demonstrated for all features when comparing Stage 0 to Stage 1 and/or Stage 2. This implies that the additional data used in the MLM data scheme has an impact on model stability.

One outlier in the experiments is found in water temperature (WTMP), which yields almost the same result in each experiment. We propose the behavior is due to two main factors. The first is that most buoys are missing WTMP data, a differenced forecast of near zero is preferred by the model. The second is that the ERA5 WTMP data, used when $\alpha < 1.0$, is only updated every 24-hours. Therefore, the regularizing data is also biased in the same way. This important lesson shows that poor data, when collected for both sources, produces an uninformed model. Neither the MLM scheme nor the physics-regularized loss function can improve results in this case.

Model improvements when adding additional buoy data is shown in Figure 4. In the figure, consecutive 24H forecasts are generated for the air temperature (ATMP) feature for all three stages. In Stage 0, the prediction is unstable for long

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

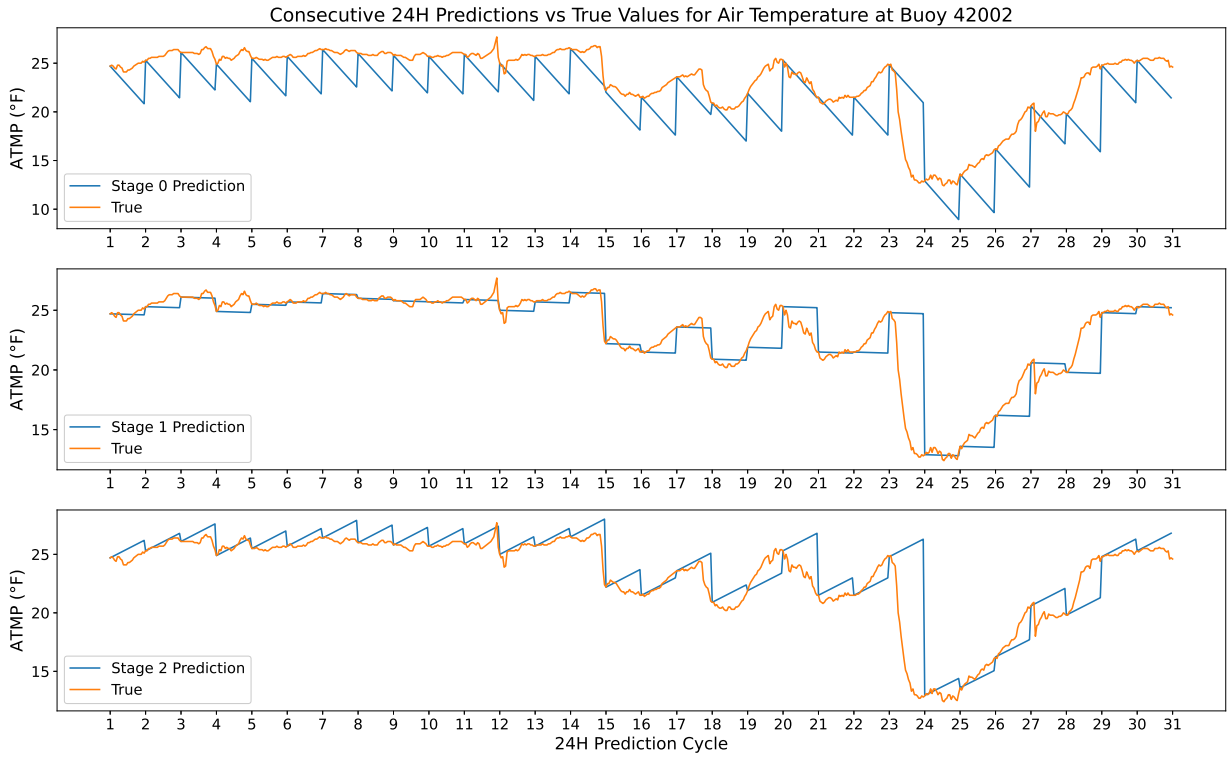

Figure 4. Improvements in Transformer model stability when adding additional buoy data to the prediction vector. Increasing buoy count results in more conservative predictive behavior.

TABLE 5. RESULTS OF CONSECUTIVE 24H FORECASTS USING THE LSTM MODEL FOR EACH STAGE. ATMP, PRES, AND WSPD ARE TESTED USING BUOY 42002 WHILE DEWP AND WTMP ARE COLLECTED FROM BUOY 42020.

LSTM	Buoy	Stage 0	Stage 1	Stage 2
ATMP	42002	1.909	1.736	1.733
PRES	42002	12.683	12.483	12.132
WSPD	42002	2.698	2.513	2.516
DEWP	42020	3.218	3.447	3.164
WTMP	42020	0.625	0.625	0.625
Average		4.227	4.161	**4.034**

TABLE 6. RESULTS OF CONSECUTIVE 24H FORECASTS USING THE TRANSFORMER MODEL FOR EACH STAGE. ATMP, PRES, AND WSPD ARE TESTED USING BUOY 42002 WHILE DEWP AND WTMP ARE COLLECTED FROM BUOY 42020.

Transformer	Buoy	Stage 0	Stage 1	Stage 2
ATMP	42002	2.851	1.733	1.940
PRES	42002	12.677	12.460	12.302
WSPD	42002	2.560	2.595	2.514
DEWP	42020	6.262	3.158	3.160
WTMP	42020	0.625	0.625	0.625
Average		4.995	4.114	**4.108**

forecast horizons. Applying more data via the MLM scheme regularizes the behavior to a more stable outcome. Stage 1 shows results that are very stable with nearly no change. This performs very well in low-change periods. In Stage 2, the addition of more data to forecast yields a less stabilized result overall. The stability is better than Stage 0, and the forecast might be more useful in drastically changing systems.

Following the stage experiments, various α values are iterated over to validate whether the physics-regularized loss function improves the MLM data scheme. The experiment is conducted for both model architectures and all stages. Results when $\alpha = 1.0$ are equivalent to the MLM stage experiments. The average results of these tests can be found in Table 7. In the table, all but one model and stage combination yield

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

TABLE 7. AVERAGE α RESULTS FOR ALL MODELS AND STAGES. THE MOST PERFORMANT EXPERIMENTS ARE SET IN BOLD.

Model: Stage	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
LSTM: 0	5.072	5.037	6.055	4.800	4.647	4.899	5.419	5.063	5.152	**3.956**	4.320
LSTM: 1	5.174	4.921	4.708	4.575	4.652	5.304	8.051	6.33	5.526	4.415	**4.028**
LSTM: 2	4.647	4.412	**3.999**	4.74	5.079	4.66	4.962	6.126	4.473	4.432	5.218
Transformer: 0	5.328	5.187	5.361	5.631	5.761	5.252	5.0140	5.088	4.9780	5.168	4.731
Transformer: 1	5.696	5.604	5.433	5.911	5.273	6.129	5.061	4.964	5.338	5.165	4.938
Transformer: 2	5.066	5.119	5.160	5.251	4.929	4.960	4.773	4.673	4.838	4.597	4.853

Model: Stage	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
LSTM: 0	5.242	6.050	4.670	4.420	4.378	4.576	5.138	4.634	4.818	4.227
LSTM: 1	5.381	6.475	4.688	4.186	4.492	4.971	4.905	4.986	5.257	4.161
LSTM: 2	4.420	4.455	5.522	4.084	4.596	4.878	4.692	4.603	4.244	4.034
Transformer: 0	4.844	4.511	4.543	4.558	4.604	4.317	4.597	**4.088**	4.371	4.995
Transformer: 1	5.210	4.824	4.723	5.008	4.928	4.884	4.605	4.48	4.205	**4.114**
Transformer: 2	4.370	**4.106**	4.495	4.505	4.374	4.364	4.604	4.178	4.416	4.108

TABLE 8. RESULTS OF ITERATIVE α RATIO TESTING USING THE LSTM MODEL AT STAGE 0. WHEN $\alpha = 0.0$, THE MODEL IS TRAINED USING EXCLUSIVELY ERA5 DERIVED DATA. WHEN $\alpha = 1.0$, THE MODEL IS TRAINED USING EXCLUSIVELY NDBC BUOY DATA. FOR $0.0 < \alpha < 1.0$, A MIXED FORMULATION OF (3) IS USED.

Feature	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
ATMP	1.772	1.741	1.979	1.779	1.75	1.81	1.745	1.746	1.863	1.732	1.957
PRES	16.883	15.982	9.825	14.376	13.87	14.713	14.463	14.406	13.879	11.021	12.461
WSPD	2.541	3.311	14.571	3.466	3.453	2.689	2.543	3.061	5.158	2.784	3.355
DEWP	3.539	3.525	3.277	3.755	3.535	4.659	7.720	5.479	4.234	3.617	3.200
WTMP	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.625

Feature	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
ATMP	1.731	1.757	2.027	2.319	1.752	1.789	1.807	2.441	2.127	1.909
PRES	13.09	13.33	13.670	12.738	12.212	11.694	12.877	12.738	13.072	12.683
WSPD	6.168	6.31	3.203	3.007	2.977	3.204	5.803	2.571	2.483	2.698
DEWP	4.597	8.223	3.820	3.413	4.325	5.568	4.577	4.794	5.782	3.823
WTMP	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.625	0.625

results that are superior to the $\alpha = 1.0$ (all NBDC data) model. However, typically the most significant reduction is found for Stage 0. This implies that the stability gained from the MLM modeling is significant enough that the additional regularization gained from the coupled loss function is minimal. In other words, when there is less available geographical context in the data representation, numerical model regularization is more impactful. The results from the LSTM Stage 0 experiment are highlighted in Table 8. We highlight that the combination of data may not be viable for all features at once. That is, the feature yields minimal results at various α values. This suggests that multiple α ratios may be used, one for each feature. This idea is explored further in similar ongoing research [13] but is outside the scope of this work.

To better understand how the error is reduced among the α experiments, the absolute error generated from $\alpha = 1.0$ and $\alpha = 0.9$ for the Transformer Stage 0 experiment is compared in Figure 5. The bottom figure shows the difference between the two errors. When the value is greater than zero, this represents places where $\alpha = 0.9$ is more performant than the original $\alpha = 1.0$ model. In general, the coupled loss function regularizes the model in the same way as the MLM data scheme. Less exaggerated forecasts keep the model stable over longer horizons. In cases not highlighted in figures, it was observed that $\alpha < 1.0$ produced models that were better aligned to general ocean conditions by proactively forecasting changes in the environment. The final RMSE of these forecasts were worse on average but showed the influence of the ERA5 dataset on the models. Considering the results of both experiments, we have improved the forecast error and achieved roughly the best performance possible using the model architecture and data set. The improvements to model performance are capped because of the size of the models and the amount of training data used. Three years of buoy data at only six locations is not enough data to support a statistical model with millions of parameters, hence the smaller model architectures used in this experiment. Increasing the number of model parameters and amount of training data should result in a surrogate model that is more robust to real-world conditions, allowing the RMSE to be reduced further. These changes would provide diminishing returns once the model is large enough. Additional feature engineering for more informative features can also improve baseline results.

When comparing these baseline results to those generally

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

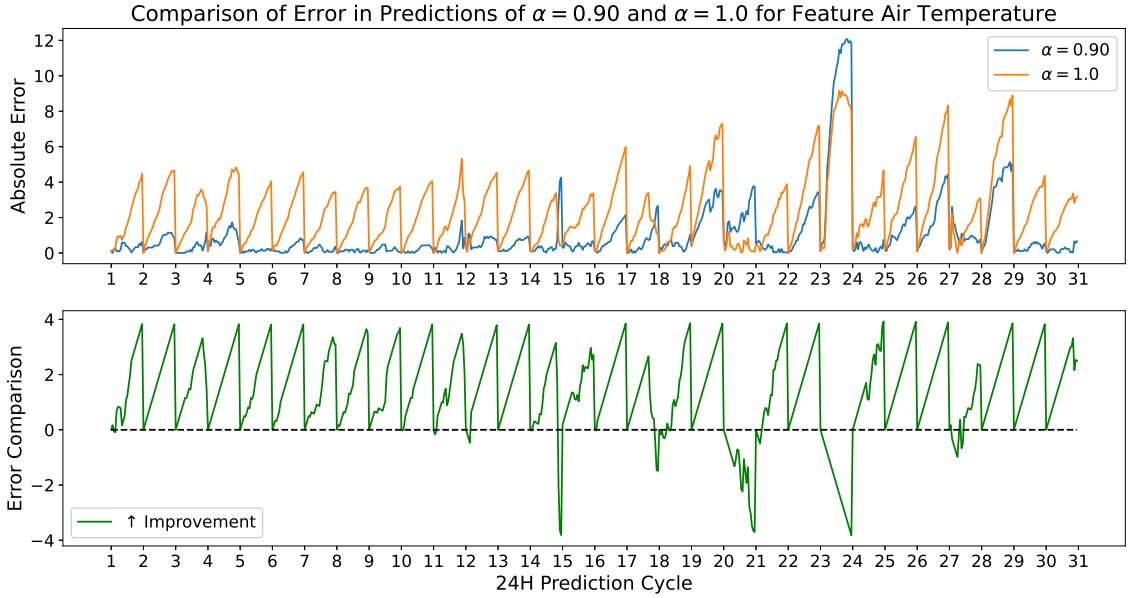

Figure 5. Difference in absolute error of the Transformer Stage 0 forecasts when comparing $\alpha = 0.9$ and $\alpha = 1.0$ (no regularization) for ATMP. The use of model data when training the observation model helps to prevent large swinging changes in the recursive forecast.

achieved in the most similar research to this [12], they are generally worse. This is explained by the smaller model size and dataset. However, the increased number of forecast steps is also a significant factor. The surrogate models in this experiment forecast three times as many periods, 24 instead of the previous eight, which means that general model stability is more important. Also, the previous work considers the combined result of over 100 buoys. Using a more similar experimental setup would likely continue to improve the results seen here. However, increasing the number of modeled buoys would continue to increase the inference vector size to a potentially cumbersome level. Since the input and output vectors of our models is the product of the number of buoys and number of features the required number of trainable parameters would increase considerably. A future investigation might select buckets of nearby buoys to model using the MLM scheme. This would increase the data pool without modeling hundreds of buoys in a single inference.

The MLM scheme improved model accuracy as the number of buoys increased. The distance between the buoys in Stage 1 and Stage 2 is large enough that local conditions are unlikely to affect one another. Therefore, it does not seem like a direct geographical influence is the determining factor of inferences. Instead, it seems most likely that utilizing more data in the input vector improves stability when training the weights of the neural network. To verify this hypothesis, further studies which use the MLM scheme in a variety of ways can be examined. First, using the MLM scheme on very distinct and far away locations should reveal if model behaviors are stabilized from the increased modeling space alone. Then, nearby observation points, which are geographically relevant to one another, can be added one-by-one until diminishing returns are found. This would support that the neural network is internalizing nearby behaviors. Although a more stable output is produced in the context of this research, other mechanisms for adding geospatial context should also be explored.

The physics-regularized loss function displayed the expected results of decreasing RMSE when the best α value is found. This compares to the results in [12], where a similar reduction in error through more stabilized results was seen. Contrary to those results, the error reduction seen is lesser in magnitude, most likely because the testing dataset is significantly larger in that work. Similarly to the previous work, some α values yielded results which were significantly worse than when no regularization is provided. This is typically the case when $\alpha = 0.0$ and most often because of the poor performing PRES feature. When $\alpha = 0.0$ this is equivalent to only using numerical model data to predict observation data. So, this behavior is most likely explained by misalignment between the observation ground truth and the numerically modeled PRES values from ERA5.

Importantly, the ERA5 data was removed from the inference input, and this is novel when comparing to the previous work. This shows the methodology can be used without the use of numerical models as an input parameter after the surrogate has been trained. This was an important contribution because it allows the methodology to be used more flexibly in real-world examples. Numerically modeled data is likely to be available when training a surrogate model, because this is almost always done using historical data. However, numerical models can

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

take hours to run for high-quality analysis. In situations where observation forecasts are needed immediately, this bottleneck would be extremely detrimental.

Finally, both techniques were separately found to regularize the results below the baseline. When combining both MLM and the physics-regularized loss, only a limited performance increase was found. This supports the conjecture that the model has reached a theoretical limit via the data selection and model architecture. Therefore, the use of both methodologies together should be reserved for those cases where more model regularization is required or when the problem is well-suited for both techniques. The MLM technique is recommended for those tasks where a static number of observation points need to be modeled. The physics-regularized loss can be implemented in any problem where multiple sources of data representing the same phenomenon are readily available.

V. Conclusion and Future Work

In this paper, a novel combination of a MLM scheme and a physics-regularized loss function was investigated for deep learning models. Fixed-location ocean buoy observations were used to validate this methodology in a real-world context. The buoy inference model was used to recurrently forecast 24-hour intervals over one month to validate results. Combining multiple buoy locations and the relevant features into a shared inference vector using the proposed MLM scheme improved model performance by stabilizing the inferences over longer time horizons. Multiple locations equaling four and six locations both yielded superior results over the single-buoy model. Improving the physics-regularized loss approach by removing numerical models from the surrogate input was also a success. The grid search for the best α value increased the performance the most in Stage 0 experiments. The combination of both MLM and coupled loss produces more accurate forecasts, but the magnitude of the improvement is lesser than when using either technique alone. This implied we reached the best results for this configuration of model architecture and data set. Although the experimentation was focused on a single buoy, the trends observed are expected to hold for the other forecasted buoys.

All together, it is proposed that the findings demonstrate enhanced stability and accuracy of oceanic data forecasting when using MLM and the physics-regularized loss. This is a practical surrogate for systems where multiple fixed-location observations (e.g., buoys) need to be forecasted simultaneously and in quick succession. Situations where a fleet of buoys have some missing values, due to buoy damage or scheduled maintenance, can benefit from this type of modeling. If only some data is missing, known values can be injected as a model input to support forecasts in regions where data is missing. These approximate observations can be used in place of sensor data while buoy maintenance is conducted. Finally, in this work the ERA5 numerical model data was used when training the model, but no features were used as part of the input during inference time. Practically, this technique allows for more flexibility in real world scenarios while still giving the model some physics-based regularization.

In future work, the same methodology should be compared against other data to verify if the combination of MLM and coupled loss would see more significant increases in model accuracy. Examinations considering all buoys in the forecast model should be conducted. Further comparisons to other geospatial context models, like graph neural networks, and their integration with the coupled loss function would also be extremely relevant. Explorations of the MLM structure and physics-regularized loss should also be adapted for use with other methods of combining physical knowledge or observations. Moreover, more complicated integration schemes like Runge-Kutta can be implemented to further improve performance/stability. Therefore, the use of the proposed methodology with PINNs or data assimilative machine learning is a promising potentiality in the right circumstances.

Acknowledgments

This work was partly supported by the U.S. Department of the Navy, Office of Naval Research (ONR), and Naval Research Laboratory under contracts N0073-16-2-C902 and N00173-20-2-C007, respectively. The work of Austin Schmidt was funded by a SMART (Science, Mathematics and Research for Transformation) Department of Defense (DoD) scholarship for service. The work of Elias Sandner was equally funded by the contract N00173-20-2-C007 and the Austrian Marshall Plan Foundation. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the funding agencies.

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] F. Bouttier and P. Courtier, "Data assimilation concepts and methods march 1999," *Meteorological training course lecture series. ECMWF*, vol. 718, p. 59, 2002.

[4] C3S CDS, *ERA5 hourly data on single levels from 1940 to present*, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), [retrieved: October, 2024], 2023. DOI: 10.24381/cds.adbb2d47.

[5] G. Ibarra-Berastegi *et al.*, "Wave energy forecasting at three coastal buoys in the bay of biscay," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 4, pp. 923–929, 2016.

[6] P. Pokhrel, E. Ioup, J. Simeonov, M. T. Hoque, and M. Abdelguerfi, "A transformer-based regression scheme for forecasting significant wave heights in oceans," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 4, pp. 1010–1023, 2022. DOI: 10.1109/JOE.2022.3173454.

[7] Z. Wu *et al.*, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, Jan. 2021, ISSN: 2162-2388. DOI: 10.1109/tnnls.2020.2978386.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

[8] G. Reikard and W. E. Rogers, "Forecasting ocean waves: Comparing a physics-based model with statistical models," *Coastal Engineering*, vol. 58, no. 5, pp. 409–416, 2011.

[9] M. Bocquet, "Surrogate modeling for the climate sciences dynamics with machine learning and data assimilation," *Frontiers in Applied Mathematics and Statistics*, vol. 9, p. 1 133 226, 2023.

[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] 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.

[12] 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*, pp. 1–20, 2024. DOI: 10.1109/JOE.2024.3378408.

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

[14] Y.-Y. Hong, C. L. P. P. Rioflorido, and W. Zhang, "Hybrid deep learning and quantum-inspired neural network for day-ahead spatiotemporal wind speed forecasting," *Expert Systems with Applications*, vol. 241, p. 122 645, 2024.

[15] L. Huang, Y. Jing, H. Chen, L. Zhang, and Y. Liu, "A regional wind wave prediction surrogate model based on cnn deep learning network," *Applied Ocean Research*, vol. 126, p. 103 287, 2022.

[16] S. Londhe and V. Panchang, "One-day wave forecasts using buoy data and artificial neural networks," in *Proceedings of OCEANS 2005 MTS/IEEE*, IEEE, 2005, pp. 2119–2123.

[17] Q. F. Qian, X. J. Jia, and H. Lin, "Machine learning models for the seasonal forecast of winter surface air temperature in north america," *Earth and Space Science*, vol. 7, no. 8, e2020EA001140, 2020.

[18] X. Yu *et al.*, "A novel method for sea surface temperature prediction based on deep learning," *Mathematical Problems in Engineering*, vol. 2020, no. 1, p. 6 387 173, 2020.

[19] S. Wolff, F. O'Donncha, and B. Chen, "Statistical and machine learning ensemble modelling to forecast sea surface temperature," *Journal of Marine Systems*, vol. 208, p. 103 347, 2020.

[20] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132 306, 2020.

[21] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[22] K.-S. Kim, J.-B. Lee, M.-I. Roh, K.-M. Han, and G.-H. Lee, "Prediction of ocean weather based on denoising autoencoder and convolutional lstm," *Journal of Marine Science and Engineering*, vol. 8, no. 10, p. 805, 2020.

[23] Y. Liu *et al.*, "Spatiotemporal wave forecast with transformer-based network: A case study for the northwestern pacific ocean," *Ocean Modelling*, p. 102 323, 2024.

[24] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE geoscience and remote sensing letters*, vol. 14, no. 10, pp. 1745–1749, 2017.

[25] W. Chang *et al.*, "Real-time prediction of ocean observation data based on transformer model," in *Proceedings of the 2021 ACM International Conference on Intelligent Computing and its Emerging Applications*, 2021, pp. 83–88.

[26] X. Liu, T. Wilson, P.-N. Tan, and L. Luo, "Hierarchical lstm framework for long-term sea surface temperature forecasting," in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, 2019, pp. 41–50.

[27] A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, E. Bach, and K. Kashinath, "Towards physics-inspired data-driven weather forecasting: Integrating data assimilation with a deep spatial-transformer-based u-net in a case study with era5," *Geoscientific Model Development*, vol. 15, no. 5, pp. 2221–2237, 2022.

[28] M. Adrian, D. Sanz-Alonso, and R. Willett, "Data assimilation with machine learning surrogate models: A case study with fourcastnet," *arXiv preprint arXiv:2405.13180*, 2024.

[29] R. Niu *et al.*, "Multi-fidelity residual neural processes for scalable surrogate modeling," *arXiv preprint arXiv:2402.18846*, 2024.

[30] J. Kim, T. Kim, J.-G. Ryu, and J. Kim, "Spatiotemporal graph neural network for multivariate multi-step ahead time-series forecasting of sea temperature," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106 854, 2023.

[31] B. Kesavakumar, P. Shanmugam, and R. Venkatesan, "Enhanced sea surface salinity estimates using machine-learning algorithm with smap and high-resolution buoy data," *IEEE Access*, vol. 10, pp. 74 304–74 317, 2022.

[32] R. Zhang, Q. Liu, R. Hang, and G. Liu, "Predicting tropical cyclogenesis using a deep learning method from gridded satellite and era5 reanalysis data in the western north pacific basin," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2021.

[33] T. de Wolff, H. Carrillo, L. Marti, and N. Sanchez-Pi, "Towards optimally weighted physics-informed neural networks in ocean modelling," *arXiv preprint arXiv:2106.08747*, 2021.

[34] T. Yuan *et al.*, "A space-time partial differential equation based physics-guided neural network for sea surface temperature prediction," *Remote Sensing*, vol. 15, no. 14, p. 3498, 2023.

[35] C. Fu, J. Xiong, and F. Yu, "Storm surge forecasting based on physics-informed neural networks in the bohai sea," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 2718, 2024, p. 012 057.

[36] NDBC, *National Data Buoy Center*, https://www.ndbc.noaa.gov/, [retrieved: October, 2024].

[37] NDBC, *Measurement Descriptions and Units*, https://www.ndbc.noaa.gov/faq/measdes.shtml, [retrieved: October, 2024].

[38] M. Haghbin, A. Sharafati, D. Motta, N. Al-Ansari, and M. H. M. Noghani, "Applications of soft computing models for predicting sea surface temperature: A comprehensive review and assessment," *Progress in earth and planetary science*, vol. 8, no. 1, pp. 1–19, 2021.

[39] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[40] R. Bullock, "Great circle distances and bearings between two locations," *MDT, June*, vol. 5, pp. 1–3, 2007.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

Physics-Regularized Buoy Forecasts: A Multi-Hyperparameter Approach Using Bounded Random Search

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

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

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

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

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

David Dobson
Center for Geospatial Sciences
Naval Research Laboratory
Mississippi, United States
email: david.dobson@nrlssc.navy.mil

***Abstract*—One challenge in oceanographic analysis is the need for accurate initial conditions collected from physical buoys. Temporary sensor outages or noisy conditions can hinder the data collection process. Machine learning surrogate models offer short-term coverage during outages. This study presents a methodology for regularizing machine learning models that predict buoy observations by utilizing multiple data sources. A previous work introduced a ratio-coupling hyperparameter to combine numerically modeled data and ocean observations when calculating training loss. However, applying one ratio across all features failed to capture the unique characteristics of different data sources. To overcome this limitation, this work investigates a multiple-hyperparameter loss function to independently manage the contribution of each data source per feature. A bounded random grid search explores the hyperparameter space to find ratios which produce superior results compared to the single-ratio approach. Surrogate models are validated at the same 88 fixed locations as the previous paper for a direct comparison. The experimental results suggest that this multi-ratio methodology can offer more reliable forecasts over a 24-hour period by applying the correct weight for each pairing of observed feature and numerical model source.**

Keywords-Surrogate; HYCOM; ERA5; Deep Learning; Buoy Forecasting.

I. INTRODUCTION

Modeling ocean and climate conditions is very important in industry settings and oceanographic research. Tasks like climate modeling, marine life population surveys, and tsunami monitoring, all rely on accurate understandings of ocean conditions [1][2][3]. Whether directly or indirectly, each of these tasks depends on the accurate initial conditions gathered from physical sensors. For that reason, this work focuses on machine learning modeling of sensor-derived data to produce short-term forecasts during temporary outages. The resulting forecasts can be used in place of observations for direct analysis, initial conditions for numerical models, or as data assimilation when performing a reanalysis. The types of conditions directly considered in this work include Sea Surface Temperature (SST), air pressure, and gust strength. Anomalies in SST can significantly affect accurate weather prediction [4]. Air pressure predictions are helpful for forecasting energy gain in photovoltaic systems [5] and intelligent weather forecasting systems [6]. Strong gusts cause severe damage in thunderstorms and are a forecasting target in machine learning tasks such as [7]. So, accurate predictions of these interconnected phenomena are highly relevant.

Whether considering sensor-derived observations or a carefully derived numerical solutions, there are often multiple ways to represent ground truth in a physical system. The numerical features used to describe our oceans and atmosphere are simply approximations of the underlying conditions. Systematic errors in data collection, physical errors in sensors, and spatiotemporal gaps in availability make observations unreliable by themselves [8][9]. Likewise, numerically modeled data have spatial and temporal discretization errors or miscalculations from strongly nonlinear interactions [10]. Imperfect approximations always exist, so combining various data sources becomes a worthwhile endeavor to reduce the inherent biases of each individual source. Traditionally, the use of data assimilation systems to improve models has seen great success. Reanalysis of numerical models with 4D variational data assimilation and Kalman filters improve historical model data to high accuracy [11][12]. This process yields high-quality training data for statistical surrogate models. However, data assimilation methods can only be used retroactively or when observation data is readily available. They also do not typically address errors in the underlying numerical model. From a machine learning context, multiple data sources can be combined as part of the training process instead. Due to multiple representations of truth, there is potentially more than one source of relevant training data. For example, SST can be represented by either a numerical model or by sensor observations. Therefore, improving forecasts by selecting the best source of truth for the training signal is a valuable goal.

To experiment with machine learning solutions for ocean

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

modeling problems, statistical surrogate models trained on a mixture of observed data and numerically modeled data are employed. A surrogate model is any model which is an approximation of a system without being numerically derived. This includes pure data driven approaches and also hybrid-physics approaches, like Physics-Informed Neural Networks (PINNs) [13]. Regardless of the method, surrogate model are trained to approximate generalized behaviors of the underlying system. To improve surrogate model performance, data is combined from fixed observation sensors and numerical models for a richer feature set. The combination can be formulated as a mixture of data assimilation and machine learning [14]. Conversely, the entire physical phenomenon can be modeled together by directly training a surrogate model with numerical outputs and sensor data [15][16]. This work follows the latter paradigm where the physical phenomenon is directly modeled. Specifically, the experimental design follows the problem domain presented in [16].

In [16], a specialized loss function was introduced which coupled noisy observation data and imperfect numerically modeled data. The buoys for observation of the ocean in fixed locations collect sensor data from around the coast of the United States, and surrogate models were used to forecast their observed features. To improve the regional surrogate model stability, historical ocean modeling data from the same regions are added to the training set. The Hybrid Circulation Ocean Model (HYCOM) [17] and the fifth reanalysis experiment of the European Center for Medium-Range Weather Forecasts (ERA5) [18] were chosen for their selection of global climate and weather features. Features that are available in both observed and modeled sources were coupled together in the loss function of the training procedure. A performant ratio of the loss signals was identified by balancing error between the surrogate inference and the two sources of training data per coupled feature. That is, each coupled feature has both an observed value and a corresponding numerically modeled value from either HYCOM or ERA5. A limitation in this methodology was the use of a single hyperparameter to control the ratio of all coupled features. To improve the identified limitation, the single hyperparameter is redefined as a vector of N hyperparameters. Consider that the ocean features are combined with multiple numerical sources. One numerical model may be well tuned to the underlying conditions of one feature and necessitate a stronger contribution to the training signal. If the other numerical solution does not align as well with the ground truth, an independent hyperparameter allows the training signal of that source to be reduced, while the other remains a major contributor in the loss calculation. Consequently, the main contributions of this paper are as follows:

- A surrogate training scheme is defined and validated that uses a physics-regularized loss function to independently combine two sources of data for the characteristics of the ocean K.
- Showing improvements in surrogate performance justifies the use of statistical models in oceanographic analysis.
- Finding improvements in combining two data sources for ocean features promotes continued exploration of data combination techniques during model training time.

The remainder of the paper is organized into the following sections. In Section II, the related work identifies similar research and contrasts them to this one. The main research goal is identified in relation to the previously identified work. In Section III the methodology is specified. The experimental dataset details are outlined, the improvements for the physics-regularized loss are detailed, and the deep learning architecture used is described. The experimental setup used for validation is provided for reproduction. Subsequently, Section IV is the Results section where the experimental findings and their impacts on the methodology are described. Finally, in Section V, the major contributions are reiterated and future considerations are identified.

II. Related Work

Buoy forecasting is investigated in some statistical learning contexts similar to this work. Models are trained using one buoy [19][20] or multiple buoys [21] for a region of interest. Most often, buoy observation forecasting focuses on the analysis of a single buoy, instead of many buoys in a variety of conditions. In one work, a collection of buoys are integrated into the input and output vector [22], but the rigid design of the architecture requires less flexible batch forecasting. Comparatively, there do exist works where a deep learning model is used for generalized buoy forecasting [15][16]. This research follows the scheme of generalized deep learning models to forecast a wide range of buoys, given their initial conditions.

The numerical models HYCOM and ERA5 are used in machine learning-based projects as training data for surrogate modeling tasks. In the case of HYCOM, the modeled data is used in machine learning forecasting tasks for ocean conditions [16] and sea surface salinity [23]. HYCOM data is also used to combine observations with modeled data in a machine learning context to parameterize typhoon-ocean interactions [24]. The ERA5 data is used more commonly, most likely due to its ease of access and high number of modeled parameters. It is used as training data for regional wave modeling [25], weather forecasting [26][27], earth surface temperature modeling [28], and sea surface temperature forecasting [29][30]. Numerically modeled data is also used when enhancing sensor predictions, for example, in the case of satellite sensor models [31][32]. Usually only one oceanic feature is forecasted, in contrast to this work. Also, when enhancing sensor forecasts with numerically modeled data, it is less common to combine more than one numerical model.

Recurrent Neural Networks (RNN) and attention models are classically used in time series-based modeling problems, making them a natural choice for oceanic forecasting. The Gated Recurrent Unit (GRU) is one type of RNN unit that employs an update and reset gate as part of the architecture for improved

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

temporal learning [33]. Oceanographic modeling that uses a GRU-based architecture has been used for ocean current prediction [34] and chlorophyll concentration forecasting [35]. One work similarly focuses on buoy sensor SST forecasting using GRU architectures [36]. However, the methodology differs from this work in significant ways including the model architecture, number of features forecasted, and the use of numerical models.

The physics-regularized loss function for training surrogate models has been examined in two experimental papers [16][22] and one theoretical analysis [37]. The methodology proposed in this work bridges the gap between two of those papers. In the original paper, it was proposed in the concluding section that separating the ratio-controlling hyperparameter would continue to improve results [16]. The research in this paper directly extends this previous work by testing that hypothesis using the same experimental setup. Although [37] uses a similar multiple-parameter scheme we propose here, the work does not highlight this fact. There is an assumption separating the hyperparameters is an improvement on the methodology, but no formal study was ever undertaken. Therefore, this is the first study using the physics-regularized loss function that investigates whether the use of multiple hyperparameters improves the physics-regularized model by directly comparing against the original implementation.

III. Methodology

The presence of multiple, potentially biased representations of truth within a domain presents a challenge for optimizing machine learning models. A loss function that effectively leverages these diverse sources of truth can minimize test error. One approach is to use a loss function that balances the contributions of different data sources using a coupling hyperparameter, $\lambda \in [0, 1]$, which determines the ratio of each source's influence [16]. The following subsections describe the methodologies used to extend the previous paper and answer the main research question. That is, whether splitting the single coupled hyperparameter into a vector of independent coupled hyperparameters improves the result. Specifically, this study investigates whether this proposed modification results in continued improvement under the same experimental conditions.

A. Dataset Details

Buoy observations are collected from the United States funded National Oceanic and Atmospheric Administration (NOAA) public data center. Although there exist many types of sensor payloads, we limit the scope to buoys with the Self Contained Ocean Observing Payload (SCOOP) [38]. Observations recorded on SCOOP buoys are transmitted via satellite to NOAA data servers for immediate access. Exactly 88 buoys are chosen for data extraction from a wide area of fixed locations that encompasses coastlines around the United States. Buoy sensors may be damaged, taken down for maintenance, or experience noisy local conditions. Each buoy measures multiple features per location, but uses individual sensors for each, leading to situations where only one feature may be missing. Missing values from the observations are interpolated, adding noise to the potential training data.

The numerically modeled ocean and climate models used are the HYCOM and ERA5 models, respectively [17][18]. The HYCOM and ERA5 data used are selected by finding the closest geographic and temporal resolutions. Both numerical models are grid-aligned, unlike the fixed locations of buoys, so the spatiotemporal alignment is not perfect. HYCOM is a higher resolution than the ERA5 data and typically fits the spatial position more closely as a result. Imperfect spatiotemporal alignment also introduces noise into the training pipeline. All data is combined to create a set of coupled and non-coupled training features. The complete set of features can be found in Table 1, for reference.

Table 1. Selected ocean features for training surrogate models. In bold are the numerical model features coupled with observations.

Feature Name	Feature Units	Feature Source
SST	°C	Buoy
Gust Strength	m/s	Buoy
Air Pressure	hPa	Buoy
SST	°C	**HYCOM**
Salinity	psu	HYCOM
Surf Elevation	m	HYCOM
Water Eastern Flow (U)	m/s	HYCOM
Water Northern Flow (V)	m/s	HYCOM
Wind Eastern Flow (U)	m/s	ERA5
Wind Northern Flow (V)	m/s	ERA5
Evaporation	m of w.e.	ERA5
Gust Strength	m/s	**ERA5**
Mean evaporation Rate	kg/($m^{-2}s^{-1}$)	ERA5
Mean Runoff Rate	kg/($m^{-2}s^{-1}$)	ERA5
Sea-Ice Cover (%)	[0-1]	ERA5
Air Pressure	hPa	**ERA5**
Cloud Cover	[0-1]	ERA5
Precipitation	m	ERA5

The data collected are from January 1, 2011, to December 31, 2011, and are taken in three-hour increments. The data is arranged into training, validation, and testing datasets by date. Training data are chosen from January 1 to September 13, the validation data is from September 13 to October 20, and the testing data includes the remainder of the year. Although one year of data is temporally small selection, this is exactly the same as what was used in the compared work [16]. The data is normalized based on the mean and standard deviation seen in the training data alone. The inverse is transformation upon model inference to investigate the results. Feature forecasts analyzed in the results section are in their respective scales.

B. Multiple-λ Physics-Regularized Loss

The physics-regularized loss function measures the surrogate inference error generated when comparing an observation value and the corresponding numerically modeled value. By evaluating the model inference against both sources of data, two error scores are produced for each observation and model pair. A ratio of the two error scores is taken, and this

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

error score is used for back propagation. The method is first proposed in [16]. The result is that, based on the ratio, the model is trained to approximate either source more strongly. The combination is determined by the hyperparameter λ, which selects a ratio of the errors to use. For example, in a forecasting task with features derived from sensor observations and numerical models, each feature's error is weighted by the singular λ value before being summed. This is proposed to improve the model by reducing the impact of interpolated or distorted values in either source.

Using a single λ value for all features is not optimal, as certain data sources could be more informative for specific features. To address this limitation, a more flexible loss function is explored in this work. Each feature is assigned its own independent λ value represented by a vector. This extension allows the model to assign different weights to each feature depending on the specific data sources being considered. The weighted errors accumulated from all features are then combined to calculate the total loss. Thus, features which display wildly different best-λ values are no longer required to use the same value. The subsequent piece-wise cost function can be calculated as follows in (1)-(6).

$$\Delta_{k,1} = |\hat{y}_{obs}^{(k)} - y_{obs}^{(k)}| \tag{1}$$

$$\Delta_{k,2} = |\hat{y}_{obs}^{(k)} - y_{model}^{(k)}| \tag{2}$$

$$\Omega_{\text{coupled loss}} = \sum_{k=1}^{K}[(\lambda_k * \Delta_{k,1}) + ((1 - \lambda_k) * \Delta_{k,2})]. \tag{3}$$

In (1) and (2), $\hat{y}$ represents the output of the surrogate model, while y represents the training ground truth and k represents an individual coupled feature. The source of modeling truth is determined by the subscript as y_{obs} or y_{model}. The error for each feature is weighted by λ_k before summing for the total coupled loss value. In this implementation, $K = 3$, which implies three coupled feature are included. Additional non-coupled features may be included in the surrogate and are defined as,

$$\Omega_{\text{model loss}} = |\hat{y}_{model} - y_{model}| \tag{4}$$

$$\Omega_{\text{observation loss}} = |\hat{y}_{\text{obs}} - y_{obs}|. \tag{5}$$

The remaining uncoupled features, as seen in (4) and (5), are used to collect error by comparing the predicted value with the relevant ground truth value. Additional numerical model features were added in this formulation, but no non-coupled observation features are included in the selection. Therefore, $\Omega_{\text{observation loss}} = 0$ in this implementation. Each piecewise value is summed into the final loss function (6),

$$\Omega_{\text{total loss}} = \Omega_{\text{coupled loss}} + \Omega_{\text{model loss}} + \Omega_{\text{observation loss}}. \tag{6}$$

This formulation of the loss function is similarly explored in an ongoing work [37]. However, the multiple-λ aspect is not explicitly explored, and the full impact is not seen. Also, the problem domain is significantly different, and does not compare to the original work. This paper expands on both papers by directly highlighting the multiple-λ physics-regularized loss approach and by comparing these results directly to the most related work.

C. Deep Learning Architecture

The prior work, [16], examines three neural network formulations, but the scope is limited in this paper to just one. Out of the three proposed architectures, the GRU model is chosen for further examination. Like the LSTM unit, a GRU-based model has weights which learn to store important context from the prior timestep. However, a GRU has one less internal gating signal and fewer weights as a result [33]. This yields a smaller model that is faster to train, which is important in this experimental setup, because many combinations of the λ vector must be iterated upon. Also, the GRU architecture was the one which benefited the most after using the coupled loss function in the previous paper, opening the possibility of continued improvement.

TABLE 2. GRU MODEL ARCHITECTURE. THERE ARE 24 TOTAL LAYERS WITH 1,827,306 TRAINABLE PARAMETERS. N REPRESENTS A VARIABLE BATCH SIZE.

Layer Type	Output Shape	Param #	Activation
Input Layer	(N, 18, 1)	0	None
Reshape	(N, 1, 18)	0	None
Dense	(N, 1, 256)	4,864	Tanh
Batch Normalization	(N, 1, 256)	1,024	None
Dropout	(N, 1, 256)	0	None
GRU	(N, 1, 256)	394,752	Tanh
Dropout	(N, 1, 256)	0	None
GRU	(N, 1, 256)	394,752	Tanh
Dense	(N, 1, 256)	65,792	Tanh
Batch Normalization	(N, 1, 256)	1,024	None
Dropout	(N, 1, 256)	0	None
GRU	(N, 1, 256)	394,752	Tanh
Dropout	(N, 1, 256)	0	None
GRU	(N, 256)	394,752	Tanh
Dropout	(N, 256)	0	None
Dense	(N, 200)	51,400	Tanh
Dropout	(N, 200)	0	None
Dense	(N, 200)	40,200	Tanh
Dropout	(N, 200)	0	None
Dense	(N, 200)	40,200	Tanh
Dropout	(N, 200)	0	None
Dense	(N, 200)	40,200	Tanh
Dropout	(N, 200)	0	None
Dense	(N, 18)	3,618	Tanh

The exact architecture of the surrogate model is found in Table 2. Dropout and batch normalization layers are added to prevent exploding or vanishing gradients during the training procedure. The Hyperbolic Tangent (Tanh) activation function is used for each layer. The model is trained for 100 epochs with a batch size of 256. The input and output vectors are the same shape to allow for a recurrent forecast style, where the forecast for time $t + 1$ uses the prior forecast from time t. So, only the first forecast is based on initial conditions. The model inference vector corresponds directly to the Table 1.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

However, the only features considered in the ultimate analysis are those which are collected through buoy sensors, i.e., SST, gust strength, and air pressure. The GRU model conducts eight consecutive forecasts of three-hour steps to produce a daily 24-hour forecast for analysis. The model is trained using many buoy locations and is meant to be used as a generalized forecasting model for any buoy, although it only forecasts them individually.

D. Experimental Setup

To compare the proposed methodology with the existing approach, the same GRU network architecture, training procedure, and testing methodology as described in the literature is used [16]. The findings of the experimental setup are compared directly to the prior results. The best λ value was previously found through an extensive grid search to find the best singular hyperparameter value. This was described as a time-consuming process, and the time complexity is worsened by introducing three λ values instead. So, a linear grid search over the entire search space is impractical for this work. As a method to quickly validate the research question, a basic random search scheme is implemented.

The random search scheme is implemented by randomly generating λ combinations to control the coupling ratios in the physics-regularized loss function. This was completed in two phases. In the first phase, completely random values were used, and features were randomly chosen between 0.0 and 1.0 with a step size of 0.001. To this end 24 random permutations are evaluated. Secondly, the search space is narrowed for each feature such that SST is bounded between [0.500,0.990] and both gust strength and air pressure are bounded between [0.800, 0.999]. These values correspond to the regions where the previous paper saw performant results. A further 85 trials are run in this way. Each of the 109 trials are executed with the same random seed. Since the trials are randomly chosen, this is not guaranteed to explore a full range of possible λ combinations. However, the random ranges identified are sufficient for exploring whether a multiple-λ setup can produce more performant results, especially in the case of the secondary bounded search. To compare directly against the original results, the multiple-λ results are compared against the single-valued results between [0.0,1.0] with a step size of 0.1. The best single-λ values recorded for each feature are also compared. Comparing the multiple-λ and single-λ experimental results yields a total of 122 total comparative test cases.

It is notable that this technique does not scale well to problems without prior knowledge of the system, such as the one studied in this case. In the future, an efficient mechanism for discovering the best λ should be explored and ongoing research has for this task has already started [37]. However, justification that the proposed loss function is an improvement should still be given. Therefore, the aim is to show that some set of λ values exists which performs better than the previously found single λ value.

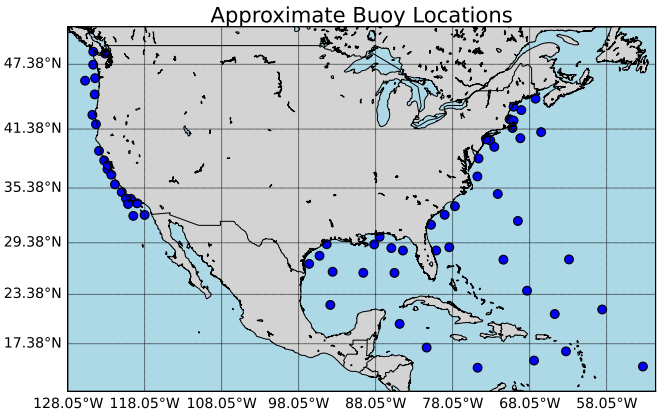

Figure 1. Approximate locations of the 88 buoys used in the testing dataset.

For evaluation of the proposed methodology, a testing dataset is composed of 48,039 instances taken from 88 independent buoys with an 8-step rolling horizon window. This window represents 24-hour forecasts. Each of these forecast windows is then evaluated and aggregated together. Evaluating the models on many buoys means the best surrogate is the one which is most accurate for a wide range of conditions. The approximate locations of each buoy is given in Figure 1, showing the diverse testing conditions. Given the forecasts, the Root Mean Square Error (RMSE) is taken for each coupled feature separately. The RMSE is defined in (7) as,

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}. \tag{7}$$

The parameter N is the number of test samples, y is the ground truth, and $\hat{y}$ is the prediction vector. Once the RMSE has been calculated for each feature, scores are summed together to determine which combination of λ values produces the smallest value.

IV. Results

An analysis of the most impactful results is illustrated in Figure 2 and further detailed in Table 3. Error scores are accumulated through eight forecast steps over a 24-hour horizon. Instead of displaying all experiments, only the top 25 results of the 122 λ combinations are highlighted. Among the top 25, only two are from single-λ experiments. The top two of those are ranked sixth and 23rd, respectively. Some single-λ experiments display minimal RMSE values for a single feature to the detriment of others, yielding a high summed RMSE. For example, the 45th ranked result is the best ever recorded RMSE for gust strength, while the SST RMSE is comparatively very poor. This shows that some selections of λ can minimize the test error of a single feature at the detriment of others. Although multiple λ values can still exhibit this behavior, increasing the hyperparameter search space allows more flexibility to choose λ values which

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

TABLE 3. TOP 10 PERFORMING COMBINATION OF λ VALUES ALONGSIDE EVERY SINGLE-λ BENCHMARK (BOLDED). THE RESULTS ARE SORTED BY THE SUM OF THE RMSE AND THEIR TOTAL RANK OUT OF THE 122 TEST CASES IS DISPLAYED.

Rank	SST λ	Gust λ	Air Pressure λ	SST RMSE	Gust RMSE	Air Pressure RMSE	Sum of RMSE
1	0.569	0.992	0.995	1.844	3.944	5.088	10.877
2	0.573	0.997	0.990	1.604	4.236	5.103	10.943
3	0.894	0.820	0.957	1.826	4.165	4.985	10.976
4	0.837	0.966	0.942	1.925	3.963	5.109	10.997
5	0.518	0.971	0.960	1.698	4.107	5.200	11.005
6	**0.900**	**0.900**	**0.900**	1.607	4.081	5.349	11.037
7	0.870	0.959	0.948	1.635	4.140	5.306	11.080
8	0.670	0.944	0.940	1.847	4.074	5.167	11.087
9	0.848	0.909	0.995	1.704	4.145	5.249	11.098
10	0.900	0.863	0.922	1.748	4.203	5.156	11.108
23	**0.960**	**0.960**	**0.960**	2.126	4.017	5.154	11.296
45	**0.840**	**0.840**	**0.840**	2.262	3.894	5.388	11.544
57	**0.800**	**0.800**	**0.800**	1.801	4.388	5.420	11.609
81	**0.700**	**0.700**	**0.700**	2.238	3.947	5.754	11.938
92	**1.000**	**1.000**	**1.000**	1.970	4.055	6.051	12.076
102	**0.600**	**0.600**	**0.600**	1.785	4.182	6.401	12.368
107	**0.500**	**0.500**	**0.500**	1.757	4.387	7.402	13.545
118	**0.100**	**0.100**	**0.100**	1.907	4.501	8.175	14.583
119	**0.300**	**0.300**	**0.300**	2.029	4.176	8.419	14.624
120	**0.000**	**0.000**	**0.000**	2.138	4.738	8.202	15.079
121	**0.400**	**0.400**	**0.400**	2.045	4.348	8.713	15.106
122	**0.200**	**0.200**	**0.200**	1.850	4.560	9.081	15.492

reduce error on average. This is shown in the fifth ranked result where individual feature do not perform better than in a corresponding single-λ setup, but overall improvements are seen. This is the main benefit of using multiple λ values instead of the previous methodology.

The top five results are all found from using a multiple-λ setup. This surpasses all previous outcomes found in the original findings. This suggests that using multiple λ values can enhance the hyperparameter space for improved test outcomes. The magnitude of improvement between the best single- and multi-λ setups is minimal overall, as displayed in Figure 2. The difference in error is more significant when compared to less optimal single-λ results. Most importantly, a set of independently selected λ values that yields better performance was found, satisfying the main goal. The consistency in the best performing λ configurations indicates that prior domain knowledge of the best-performing λ values is advantageous when conducting a random search. For example, although SST gives minimal RMSE results for a wide range of λ values, SST and Gust Strength both prefer a smaller range. Differences in numerical models influence the best selection of λ, for example, the spatial resolution is different in the HYCOM and ERA5 models. Specifically, ERA5 is a lower resolution than HYCOM, so individual grid points may be further away from the actual buoy locations.

In Table 3, the top ten results are compared with all single-λ results. Each row displays the rank out of all tests, the selection of λ for each feature, and the RMSE scores of each feature. One notable observation is that most single-λ setups are ranked worse than the top 25 results. The values for λ tend to be somewhat similar in the best performing results, depending on the feature. This behavior is in part due to constraining the random selections within previously found performant regions of λ for each feature. The λ choice for air pressure and gust strength benefit the most from prior knowledge of the optimal λ region. Specifically, for air pressure, the RMSE tends to be higher when $\lambda < 0.9$. Interestingly, the SST forecast achieves high performing results for a wider range of values. This lends credibility to the use of a random search setup when there is prior understanding of what λ values might be most effective. Also, this implies that the coupled numerical model highly influences the selection of best-λ value.

Although the sum of the RMSE is reduced when analyzing a multiple-λ experiment, individual feature results should still be considered. The top three single-λ results, are those which previously produced minimal error scores for one feature. Certain multiple-λ combinations yield lower feature-specific RMSE than those prior best results. For example, the third ranked λ configuration yields the lowest RMSE score for air pressure ever recorded using the demonstrated methodology. The second ranked result showed the best performance for SST ever recorded. However, a lower individual gust strength RMSE was never found, compared to the best performing single-λ result. The lowest sum of RMSE did not yield any best-result individual forecasts but had consistently low RMSE across all features. It is notable that a set of λ values which finds most performant forecasts for all features simultaneously was not found. This means that no single feature was optimized to the detriment of the other features. This suggests that using multiple λ values that are specific to each numerical model can overcome bias. This is because a single λ value is not allowing a biased numerical model to be more influential in the training process. Specifically, ERA5 has a lower resolution than the HYCOM data, which tends to mean that the HYCOM data is well fitted to the observations across all values. This describes why a broad selection of λ values work well for HYCOM, but not the ERA5 data. By using separate λ values for each feature neither numerical model source is forced to provide a suboptimal combination of data.

In Figure 3, absolute forecast error is highlighted. The error ranges over consecutive 24-hour cycles and is calculated based on the Mean Absolute Error (MAE) between the buoy derived observation and the predicted value. The numerical model (HYCOM/ERA5) is given as a baseline to compare against. Error is calculated from forecasts of a single buoy with the identification code 41009. The segment of forecasts analyzed are taken from period 40 to 120, demonstrating the error found in 11 forecast cycles. Compared are the best multiple-λ model and the best single-λ model, outlined in Table 3. Compared against both is the numeric error generated when comparing the buoy error to the numerical models HYCOM and ERA5. It is observed that the best performing model does not outperform the single-λ setup in all cases. In some situations, the stability of either model might be superior. The multiple-λ model is more stable on average and tends to experience less extreme fluctuations in the forecast. Occasionally, either surrogate model can outperform the numerical model, but

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

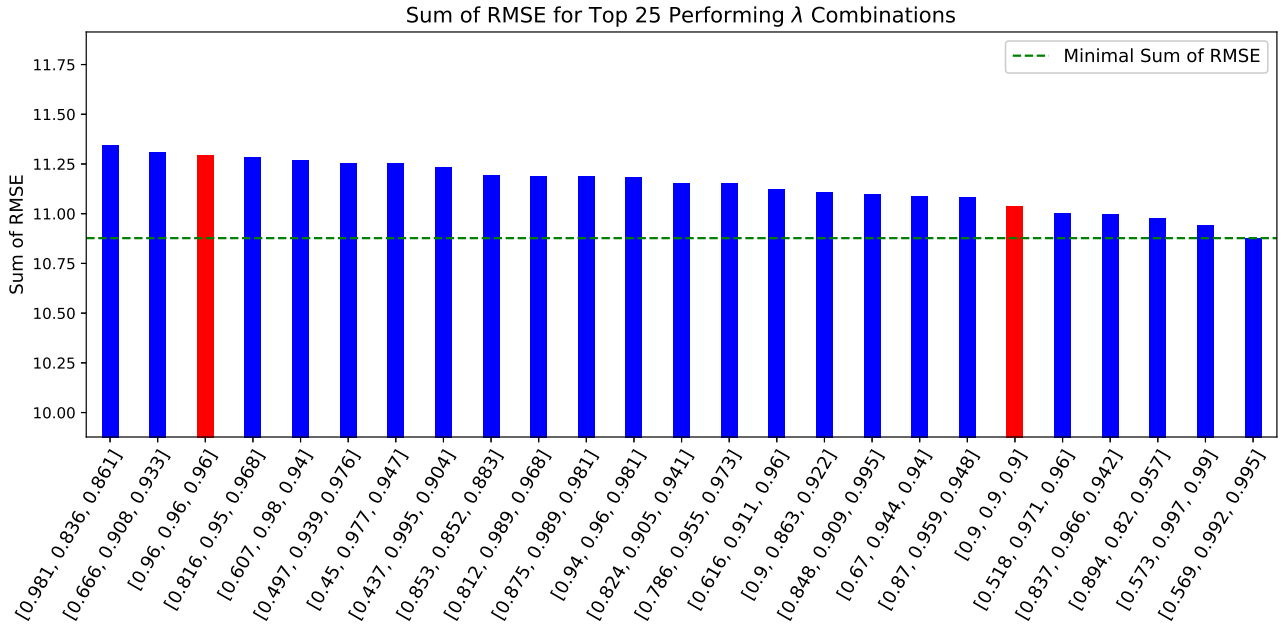

Figure 2. Top 25 performing λ combinations and their summed feature RMSE test scores. The green dotted line shows the separation of values between the most performant result and all others. Multiple-λ combinations are in blue, while any single-λ results are in red.

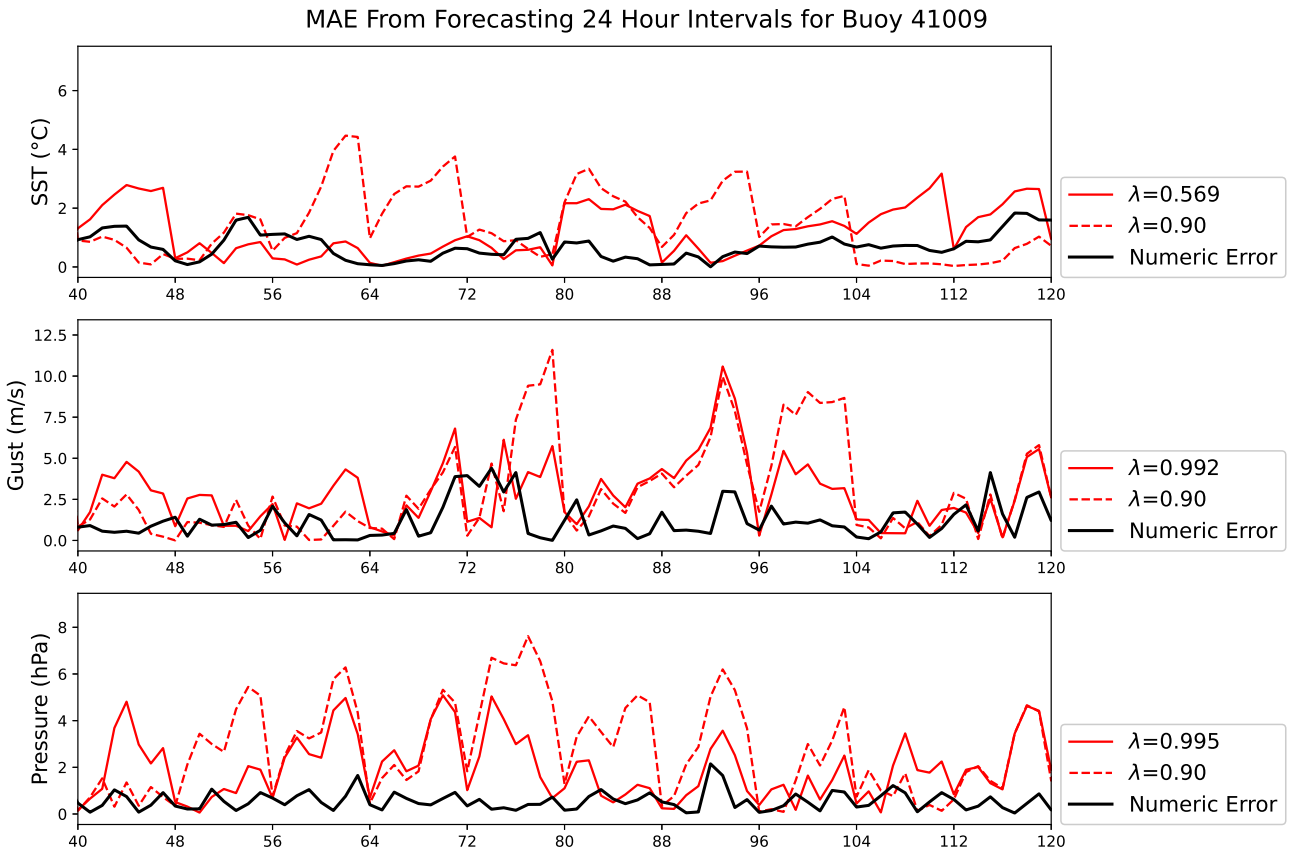

Figure 3. Shown are the forecast errors of the most performant combination of λ values compared against the most performant single-λ for a single buoy.

on average HYCOM and ERA5 show reduced error. This is expected, because statistical models are well-known to be less accurate than numerical models.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

The trade-off in favor of the numerical models is the speed of the forecasts generated. Overall, the findings suggest that an independent selection of λ does improve the methodology by consistently reducing error across all features. Applying the correct weight for each ratio was shown to provide more a stable forecast on average. Finding continued improvements compared to the original research gives justification to the proposed methodology. The best λ values are highly dependent on the secondary data source (i.e., the numerical model) and, to a lesser degree, the selection of domain feature. Although a set of λ values which minimized the RMSE for all features was not found, that does not mean that a configuration does not exist. Exploration of the parameter space is the main limitation of these experimental results. More specialized search techniques should be implemented to efficiently find the best selection of hyperparameters. One further limitation of the methodology is the need for two sets of good-quality data. The benefits of the physics-regularized loss are directly dependent on the ability of the second source of data to be informative when the primary data, i.e., the observations, faces physical constraints.

V. Conclusion and Future Work

A previous methodology improved the forecasting of fixed-location ocean buoy observations by combining observation data with numerically modeled data. In the work, it was found that the selection of the ratio-determining hyperparameter, λ, varied depending on the numerical source and ocean feature. It was hypothesized that the results could be further improved if each feature was independently combined with numerical data. To address the proposed research question, the methodology was modified to include multiple independently selected λ values. The physics-regularized loss function was updated to combine features with numerical models in a less constrained way, which increased the potential hyperparameter search space. Then, a bounded random search was employed to generate random λ selections which produced superior results.

The updated technique was directly validated against the publicly available prior experiments. The outcome was a surrogate model that generated more accurate forecasts overall compared to the single-λ approach. The use of multiple λ values is particularly beneficial when multiple numerical models contribute to the feature set. For example, in this work both global HYCOM and ERA5 reanalysis models were used to improve overall results. A selection of λ values which reduced each individual feature's error below the best recorded value simultaneously was not found using the random search, but average error was improved for five combinations of λ. Such a combination of values may exist, even if the random search did not yield these results. It is acknowledged that the use of random or grid search to find the best parameter combination is time-consuming without prior domain knowledge and does not guarantee optimal results. However, the results justify the further use of multiple λ values, instead of a single value for all features.

Future work should validate the methodology using a wider range of real-world and theoretical datasets. Testing the combining technique with different combinations of input and output data would be very insightful. Different model architectures should be explored to assess the effectiveness of coupling data with more generalizable models. The physics-regularized loss is not reliant on the model architecture and should be attempted with more specialized architectures to see if similar improvements are found. Grid search and random search are not efficient enough, so developing methods for approximating or selecting λ values is a primary focus of future research.

Acknowledgments

This work was partly supported by the U.S. Department of the Navy, Office of Naval Research (ONR), and Naval Research Laboratory under contracts N0073-16-2-C902 and N00173-20-2-C007, respectively. 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] M. Angove *et al.*, "Ocean observations required to minimize uncertainty in global tsunami forecasts, warnings, and emergency response," *Frontiers in Marine Science*, vol. 6, p. 350, 2019.

[4] B.-T. Jong, M. Ting, R. Seager, N. Henderson, and D. E. Lee, "Role of equatorial pacific sst forecast error in the late winter california precipitation forecast for the 2015/16 el niño," *Journal of Climate*, vol. 31, no. 2, pp. 839–852, 2018.

[5] A. Bugała *et al.*, "Short-term forecast of generation of electric energy in photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 306–312, 2018.

[6] L. L. Lai *et al.*, "Intelligent weather forecast," in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826)*, IEEE, vol. 7, 2004, pp. 4216–4221.

[7] H. Wang, Y.-M. Zhang, J.-X. Mao, and H.-P. Wan, "A probabilistic approach for short-term prediction of wind gust speed using ensemble learning," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 202, p. 104 198, 2020.

[8] W. A. Lahoz and P. Schneider, "Data assimilation: Making sense of earth observation," *Frontiers in Environmental Science*, vol. 2, p. 16, 2014.

[9] H. Y. Teh, A. W. Kempa-Liehr, and K. I.-K. Wang, "Sensor data quality: A systematic review," *Journal of Big Data*, vol. 7, no. 1, p. 11, 2020.

[10] W. L. Oberkampf, S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin, "Error and uncertainty in modeling and simulation," *Reliability Engineering & System Safety*, vol. 75, no. 3, pp. 333–357, 2002.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

[11] Y. Tr'emolet, "Accounting for an imperfect model in 4d-var," *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, vol. 132, no. 621, pp. 2483–2504, 2006.

[12] G. Evensen *et al.*, *Data assimilation: the ensemble Kalman filter*. Springer, 2009, vol. 2.

[13] 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.

[14] 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.

[15] P. Pokhrel, E. Ioup, J. Simeonov, M. T. Hoque, and M. Abdelguerfi, "A transformer-based regression scheme for forecasting significant wave heights in oceans," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 4, pp. 1010–1023, 2022. DOI: 10.1109/JOE.2022.3173454.

[16] 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*, pp. 1–20, 2024. DOI: 10.1109/JOE.2024.3378408.

[17] R. Bleck, "An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates," *Ocean modelling*, vol. 4, no. 1, pp. 55–88, 2002.

[18] H. Hersbach *et al.*, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[19] G. Ibarra-Berastegi *et al.*, "Wave energy forecasting at three coastal buoys in the bay of biscay," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 4, pp. 923–929, 2016.

[20] S. Londhe and V. Panchang, "One-day wave forecasts using buoy data and artificial neural networks," in *Proceedings of OCEANS 2005 MTS/IEEE*, IEEE, 2005, pp. 2119–2123.

[21] Y.-Y. Hong, C. L. P. P. Rioflorido, and W. Zhang, "Hybrid deep learning and quantum-inspired neural network for day-ahead spatiotemporal wind speed forecasting," *Expert Systems with Applications*, vol. 241, p. 122 645, 2024.

[22] E. Sandner *et al.*, "A multiple-location modeling scheme for physics-regularized networks: Recurrent forecasting of fixed-location buoy observations," *TechRxiv*, Aug. 2024. DOI: 10.36227/techrxiv.172469936.64312665/v1.

[23] E. Jang, Y. J. Kim, J. Im, Y.-G. Park, and T. Sung, "Global sea surface salinity via the synergistic use of smap satellite and hycom data based on machine learning," *Remote sensing of environment*, vol. 273, p. 112 980, 2022.

[24] G.-Q. Jiang, J. Xu, and J. Wei, "A deep learning algorithm of neural network for the parameterization of typhoon-ocean feedback in typhoon forecast models," *Geophysical Research Letters*, vol. 45, no. 8, pp. 3706–3716, 2018.

[25] L. Huang, Y. Jing, H. Chen, L. Zhang, and Y. Liu, "A regional wind wave prediction surrogate model based on cnn deep learning network," *Applied Ocean Research*, vol. 126, p. 103 287, 2022.

[26] A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, E. Bach, and K. Kashinath, "Towards physics-inspired data-driven weather forecasting: Integrating data assimilation with a deep spatial-transformer-based u-net in a case study with era5," *Geoscientific Model Development*, vol. 15, no. 5, pp. 2221–2237, 2022.

[27] M. Adrian, D. Sanz-Alonso, and R. Willett, "Data assimilation with machine learning surrogate models: A case study with fourcastnet," *arXiv preprint arXiv:2405.13180*, 2024.

[28] R. Niu *et al.*, "Multi-fidelity residual neural processes for scalable surrogate modeling," Proceedings of Machine Learning Research, vol. 235, R. Salakhutdinov *et al.*, Eds., pp. 38 381–38 394, 21–27 Jul 2024.

[29] J. Kim, T. Kim, J.-G. Ryu, and J. Kim, "Spatiotemporal graph neural network for multivariate multi-step ahead time-series forecasting of sea temperature," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106 854, 2023.

[30] X. Yu *et al.*, "A novel method for sea surface temperature prediction based on deep learning," *Mathematical Problems in Engineering*, vol. 2020, no. 1, p. 6 387 173, 2020.

[31] B. Kesavakumar, P. Shanmugam, and R. Venkatesan, "Enhanced sea surface salinity estimates using machine-learning algorithm with smap and high-resolution buoy data," *IEEE Access*, vol. 10, pp. 74 304–74 317, 2022.

[32] R. Zhang, Q. Liu, R. Hang, and G. Liu, "Predicting tropical cyclogenesis using a deep learning method from gridded satellite and era5 reanalysis data in the western north pacific basin," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–10, 2021.

[33] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (gru) neural networks," in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, IEEE, 2017, pp. 1597–1600.

[34] N. Thongniran, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, "Combining attentional cnn and gru networks for ocean current prediction based on hf radar observations," in *Proceedings of the 2019 8th international conference on computing and pattern recognition*, 2019, pp. 440–446.

[35] S. Wu, Z. Du, F. Zhang, Y. Zhou, and R. Liu, "Time-series forecasting of chlorophyll-a in coastal areas using lstm, gru and attention-based rnn models.," *Journal of Environmental Informatics*, vol. 41, no. 2, pp. 104–117, 2023.

[36] Y. Jiang *et al.*, "Prediction of sea temperature using temporal convolutional network and lstm-gru network," *Complex Engineering Systems*, vol. 1, no. 2, p. 1, 2021.

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

[38] P. C. Kohler, L. LeBlanc, and J. Elliott, "Scoop-ndbc's new ocean observing system," in *OCEANS 2015-MTS/IEEE Washington*, IEEE, 2015, pp. 1–5.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

Development of an Acoustic Antenna for Event Triggering in KM3NeT

Miguel Ardid, Salva Ardid, Víctor Espinosa, Alicia Herrero, Carlos D. Llorens, Carlos Quiroz-Rangel
Universitat Politècnica de València, Inst. de Investigación para la Gestión Integrada de Zonas Costeras,
Gandía, Spain.
emails: mardid@fis.upv.es, sardid@upv.es, vespinos@fis.upv.es, aherrero@mat.upv.es, cdavid@upv.es, carquira@upv.es

Manuel Bou-Cabo, Guillermo Lara
Instituto Español de Oceanografía
Murcia, Spain.
emails: manuel.bou@ieo.csic.es, guillermo.lara@ieo.csic.es

***Abstract*— The Cubic Kilometer Neutrino Telescope (KM3NeT), located deep in the Mediterranean Sea, is designed for detecting neutrinos through optical Cherenkov radiation. However, the installation of an acoustic detection system can provide complementary capabilities, extending the detection to ultra-high energies. This paper proposes the development of an acoustic antenna composed of an array of four hydrophones, separated by approximately one meter. This antenna will be used as a trigger for detecting relevant acoustic events, which could include signals from neutrino interactions or other underwater acoustic phenomena. The system is intended to work in conjunction with the existing KM3NeT hydrophone network, enhancing both the detection of neutrinos through acoustic channels and contributing to sea acoustic monitoring.**

Keywords-acoustic detection; deep-sea neutrino telescope; sea acoustic monitoring; hydrophone array.

I. INTRODUCTION

The Cubic Kilometer Neutrino Telescope (KM3NeT) is a neutrino observatory located at depths of up to 3500 meters in the Mediterranean Sea, primarily designed to detect high-energy neutrinos by observing Cherenkov light produced when neutrinos interact with water [1]. In addition to optical detection, there is growing interest in utilizing acoustic techniques to identify neutrino interactions, which may produce characteristic pressure waves in the water and allowing the extension of neutrino detection to ultra-high energies [2]. Such an approach also enables environmental monitoring, offering insight into marine biodiversity (e.g. mammals), seismic activity, and anthropogenic noise pollution.

The integration of an acoustic antenna system that can detect these neutrino-induced acoustic signals or other significant acoustic phenomena is a natural extension of KM3NeT's capabilities. This paper proposes a design for a 4-hydrophone acoustic antenna array that can trigger event detection based on specific acoustic patterns, which complements the current array of hydrophones used in KM3NeT. With this antenna, the feasibility of the acoustic detection of neutrinos will be assessed. In parallel, long-term acoustic studies on marine mammals and on anthropogenic noise are planned.

Notice as well that this kind of infrastructure is very unique, allowing an ideal place for deep-sea science studies and providing the infrastructure the required energy for the equipment to be operated and a broadband connection that allows continuous real-time monitoring with high-frequency sampling.

The paper is structured as follows: in Section II, the conceptual design is discussed. In Section III, the integration of the antenna in KM3NeT is treated. The applications of the antenna beyond acoustic neutrino detection are presented in Section IV. Finally, the main conclusions are highlighted in Section V.

II. CONCEPTUAL DESIGN OF THE ACOUSTIC ANTENNA

The proposed acoustic antenna consists of four hydrophones arranged in a 3D array with a separation of approximately one meter between each hydrophone. It will follow a similar scheme of the acoustic floors in the AMADEUS/ANTARES, or ONDE detection systems [3][4]. The goal is to leverage the spatial configuration and array processing techniques to locate the origin of incoming acoustic signals and assess their characteristics, distinguishing neutrino events from background noise. Fig. 1 shows a schematic view of the conceptual design of the array mounted in a recoverable autonomous structure. There are the four hydrophones of the array fixed with vertical bars at different heights. They are on a structure that contains the electronics of the system and the connector to link it to the KM3NeT infrastructure, thus allowing the control and data capturing communication as well as the needed electrical power. The structure, with sufficient buoyancy, may be automatically recovered using an acoustic release.

Some considerations that have to be taken into account in the design:

- **Hydrophone Spacing:** The separation of one meter is chosen to balance the ability to capture acoustic signals across a range of frequencies, while maintaining a compact structure suitable for deployment in deep-sea conditions. At this separation, it complements the larger distances between acoustic sensors existing in KM3NeT and the array is sensitive to wavelengths in the range of typical acoustic signals generated by high-energy

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

particle interactions or the acoustic phenomena to be studied.

- **Frequency Range:** The hydrophone array will be sensitive to the expected acoustic signature of neutrino interactions, which is in the range of 1–50 kHz. This frequency range also allows for the detection of other underwater acoustic phenomena, including marine mammal vocalizations, oceanographic noise, and seismic signals.

- **Signal Processing and Triggering:** The array will employ beamforming and signal correlation techniques to identify the direction of arrival of acoustic signals and trigger data recording when certain criteria are met [5][6]. This triggering mechanism will be designed to identify sharp, broadband acoustic pulses, such as those expected from neutrino interactions. Advanced algorithms for pattern recognition will be developed to differentiate between neutrino-induced acoustic signals and noise from marine life, ships, or environmental factors.

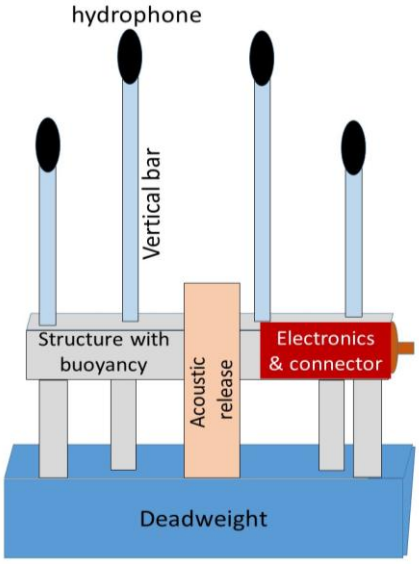

Figure 1. Schematic view of the acoustic antenna.

III. INTEGRATION WITH KM3NET

The KM3NeT neutrino telescope is already equipped with a large array of hydrophones and acoustic sensors primarily used for positioning and calibration of the optical modules [7]. The proposed 4-hydrophone array will be deployed alongside this system, but with a focus on event detection and triggering relevant events. By complementing the existing infrastructure, this acoustic antenna can enhance KM3NeT's ability to monitor and record transient acoustic events, as well as various other natural and anthropogenic sources of underwater sound.

- **Data Synchronization:** The new hydrophone array will be synchronized with the KM3NeT timing system, ensuring that acoustic data can be correlated with the other acoustic data from the observatory, as well as from the optical detections or other neutrino candidate events.

- **Deployment Strategy:** The compact size of the array allows for flexible deployment within the KM3NeT grid. It can either be mounted on the base of the detector strings, in the calibration base or instrumentation line, or deployed as a standalone unit with dedicated infrastructure.

IV. APPLICATIONS BEYOND NEUTRINO DETECTION

In addition to enhancing the detection of neutrino interactions, the proposed hydrophone array will contribute to acoustic studies in the deep sea. The Mediterranean Sea is a rich environment for studying natural and anthropogenic sounds. The proposed system could provide valuable data for:

- **Marine Biology:** Detection of vocalizations from marine mammals and fish populations [8], contributing to biodiversity assessments and behavioral studies.

- **Seismic Monitoring:** Monitoring underwater earthquakes or sub-sea volcanic activity by detecting associated acoustic waves. It could also be part of a warning system for marine hazard, e.g., tsunamis.

- **Anthropogenic Noise Studies**: Long-term monitoring of human-generated noise pollution, such as from shipping or underwater construction, and its impact on marine ecosystems.

V. CONCLUSIONS

The proposed 4-hydrophone acoustic antenna represents a novel addition to the KM3NeT neutrino telescope, offering enhanced capabilities for detecting acoustic signals from neutrino interactions and providing valuable data for oceanographic and environmental monitoring. The small array size, signal processing capabilities, and compatibility with the existing KM3NeT infrastructure make it an ideal tool for complementing physics and environmental research objectives in the deep sea. As future work, the design, construction and installation of the acoustic antenna will be finalized. It will be tested initially in the laboratory and at sea at shallow depths. Once verified that it is working properly, it will be integrated into the KM3NeT deep sea neutrino telescope.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

ACKNOWLEDGMENT

The authors acknowledge the financial support of the Spanish Ministerio de Ciencia, Innovación y Universidad, for PID2021-124591NB-C42 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", for ASFAE/2022/014 with funding from the EU NextGenerationEU (PRTR-C17.I01) and Generalitat Valenciana, as well as for the grant CIDEGENT/2019/043, from Generalitat Valenciana.

REFERENCES

[1] S. Adrián-Martínez et al., "Letter of intent for KM3NeT 2.0" J. Phys. G: Nucl. Part. Phys. 43 084001, 2016.

[2] D. Diego-Tortosa, "Positioning System and Acoustic Studies for the KM3NeT deep-sea neutrino telescope", Ph. D. thesis, Universitat Politècnica de València, 2022.

[3] J. A. Aguilar et al., "AMADEUS - The Acoustic Neutrino Detection Test System of the ANTARES Deep-Sea Neutrino Telescope", Nucl. Instrum. Meth. A 626-627 pp. 128-143, 2011.

[4] G. Riccobene, "Long-term measurements of acoustic background noise in very deep sea", Nucl. Instrum. Meth. A 604 pp. S149-S157, 2009.

[5] M. Ardid, Calibration in acoustic detection of neutrinos, ", Nucl. Instrum. Meth. A 604 pp. S203-S207, 2009.

[6] S. Adrián-Martínez et al., "Acoustic Signal Detection Through the Cross-Correlation Method in Experiments with Different Signal to Noise Ratio and Reverberation Conditions". In: Ad-hoc Networks and Wireless. ADHOC-NOW 2014. Lecture Notes in Computer Science, vol 8629. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-46338-3_7.

[7] S. Viola et al., "Acoustic positioning system for KM3NeT", PoS(ICRC2015)1169, 2016. https://doi.org/10.22323/1.236.1169.

[8] C. Guidi, M. Bou-Cabo, and G. Lara, "Passive acoustic monitoring of cetaceans with KM3NeT acoustic receivers". *JINST* 16 C10004, 2021.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org