A spatio-temporal LSTM model to forecast across multiple temporal and spatial scales

Yihao Hu,[1] **Fearghal O'Donncha,**[2] **Paulito Palmes,**[2]
Meredith Burke,[3] **Ramon Filgueira,**[3] **Jon Grant.**[3]

[1]University of Notre Dame
[2]IBM Research Europe
[3]Dalhousie University
yhu5@nd.edu, feardonn@ie.ibm.com, paulpalmes@ie.ibm.com,
meredith.burke@dal.ca, ramon.filgueira@dal.ca, jon.grant@dal.ca

Abstract

This paper presents a novel spatio-temporal LSTM (SPATIAL) architecture for time series forecasting applied to environmental datasets. The framework was evaluated across multiple sensors and for three different oceanic variables: current speed, temperature, and dissolved oxygen. Network implementation proceeded in two directions that are nominally separated but connected as part of a natural environmental system – across the spatial (between individual sensors) and temporal components of the sensor data. Data from four sensors sampling current speed, and eight measuring both temperature and dissolved oxygen evaluated the framework. Results were compared against RF and XGB baseline models that learned on the temporal signal of each sensor independently by extracting the date-time features together with the past history of data using sliding window matrix. Results demonstrated ability to accurately replicate complex signals and provide comparable performance to state-of-the-art benchmarks. Notably, the novel framework provided a simpler pre-processing and training pipeline that handles missing values via a simple masking layer. Enabling learning across the spatial and temporal directions, this paper addresses two fundamental challenges of ML applications to environmental science: 1) data sparsity and the challenges and costs of collecting measurements of environmental conditions such as ocean dynamics, and 2) environmental datasets are inherently connected in the spatial and temporal directions while classical ML approaches only consider one of these directions. Furthermore, sharing of parameters across all input steps makes SPATIAL a fast, scalable, and easily-parameterized forecasting framework.

Introduction

The coastal ocean represents perhaps the most complex modeling challenge, intimately connected as it is to the deep ocean and the atmosphere (Song and Haidvogel 1994). Irregular coastlines, and steep highly variable bottom topography can generate highly complex patterns of flow; wind forcings produce both surface and internal waves and contributes to surface flows directly through wind drift and Ekman transport; tidal forcing in addition to barotropic flow processes induce internal tides; while freshwater discharges

add buoyancy fluxes, which further complicate local water motion (O'Donncha et al. 2015). Consequently, the accurate quantification of flow processes in the ocean is a complex, nonlinear task requiring a holistic consideration of dynamics and driving forces in both space and time. Further, understanding, quantifying, and forecasting ocean dynamics is of major importance for societal and economic factors: we look to the oceans for food, water, and energy, while ocean-atmosphere dynamics are critical to understanding weather and climate.

Traditionally, forecasting of ocean processes rely on physics-based approaches that resolve a set of governing equations such as Navier-Stokes (Constantin and Foias 1988) or convective-diffusion equations (Stockie 2011). An extensive body of literature exists on this subject with early investigations on the topic described in Bryan (1969) and a useful overview provided in Ji (2017). The primary challenge of physics-based approaches are the immense computational expense to deploy at high resolution across broad spatial and temporal scales (generally requiring high performance computing facilities), and the complexity of configuring and parameterizing the model that typically requires an expert user (DeYoung et al. 2004).

Machine-learning (ML) approaches to forecast ocean processes are in a nascent stage and have traditionally been limited by the challenge of data sparsity. To circumvent this limitation, recent years have seen interest in using ML to develop low cost approximates or surrogates of physic-based models (e.g. Lary, Müller, and Mussa (2004); Ashkezari et al. (2016). In this paradigm, there is the luxury of being able to run the model as many times as necessary to develop a sufficient data set to train the ML model. A key issue with this approach is that one is limited by the performance of the physics-based model; as an example James, Zhang, and O'Donncha (2018) have recently observed a factor of 12,000 improvement in the speed of obtaining results comparable to that of the physics model, but because the model itself is utilized to train the deep-learning network, the accuracy and spatial extents are limited to the levels dictated by the traditional approaches.

Progress in satellite technology generates large volumes of remotely sensed observations providing long-term global measurements at varying spatial and temporal resolution.

These datasets provide a natural fit for ML-based approaches and there is a wide body of literature related to mining and forecasting ocean processes from satellite observations. Deep Learning (DL) models have been adopted to increase the resolution of satellite imagery through downscaling techniques (Ducournau and Fablet 2016), while a number of studies have presented data-mining approaches that extract pertinent events from measured data such as detection of harmful algal blooms (Gokaraju et al. 2011). Some caution is necessary here since satellite estimates of ocean processes are invariably contaminated by external influence on the optical properties (Holt et al. 2009; Smyth et al. 2006), while one is naturally limited to sampling surface processes and cannot penetrate the ocean depth.

This paper investigates a framework that extends model accuracy by explicitly learning across the spatial and temporal components of a disparate – but related – time-series signal. We refer to our framework as *Spatial-LSTM* or SPATIAL.

SPATIAL is applied to three real-world datasets, namely, ocean current speeds, temperatures, and dissolved oxygen. Incorporating both physical and biogeochemical datasets, the performance of the model with different temporal scales and process memory are evaluated: ocean current patterns are influenced by external physical drivers such as tidal effects and surface winds stress together with density driven flows generated by changes in temperature and salinity; variations in ocean temperature are explained by multiple exogenous processes such as solar radiation, air temperature, sea bed heat transfer, and external heat flux from rivers and the open ocean (Pidgeon and Winant 2005); oceanic oxygen content, on the other hand is a biogeochemical process influenced by hydrodynamics (horizontal and vertical mixing, residence times, etc.), weather (temperature reducing the solubility of oxygen), and nutrient loads (both anthropogenic and organic enrichment) (Caballero-Alfonso, Carstensen, and Conley 2015). The combination of nonlinear response and sensitivity to multiple, opaque variables makes the capabilities of DL an appealing solution for these prediction-requirements.

Related Work

There is a large number of application papers on using ML for time-series forecasting. For classical (linear) time series when all relevant information about the next time step is included in the recent past, multilayer perceptron (MLP) is typically sufficient to model the short-term temporal dependence (Gers and Schmidhuber 2001). Compared to MLPs, convolutional neural networks (CNN) automatically identify and extract features from the input time series. A convolution can be seen as a sliding window over the data, similar to the moving average (MA) model but with nonlinear transformation. Different from MLPs and CNNs, recurrent neural networks (RNN) explicitly allow long-term dependence in the data to persist over time, which clearly is a desirable feature for modeling time series. A review of many of these approaches for time series-based forecasting is provided in Esling and Agon (2012); Gamboa (2017); Fawaz et al. (2019).

Forecasting of environmental variables relies heavily on physics-based approaches and a well-established literature exists related to hydro-environmental modeling and applications to numerous case studies. An excellent operational forecasting product are the Copernicus Marine Service ocean model outputs which provide daily forecasts of variables such as current speed, temperature, and dissolved oxygen (DO) at a 10-day forecasting window (EU Copernicus 2020). Data are provided at varying horizontal resolution from approximately $10\,\mathrm{km}$ globally, to regional models at resolution of up to $3\,\mathrm{km}$. These datasets serve as a valuable forecasting tool but are generally too coarse for coastal process studies and for most practical applications (e.g. predicting water quality properties within a bay) are rescaled to higher resolution using local regional models (e.g. Gan and Allen (2005).

Due to the heavy computational overhead of physical models, there is an increasing trend to apply data-driven DL or ML methods to model physical phenomena (de Bezenac, Pajot, and Gallinari 2017; Wiewel, Becher, and Thuerey 2018). A number of studies have investigated data-driven approaches to provide computationally cheaper surrogate models, applied to such things as wave forecasting (James, Zhang, and O'Donncha 2018), viscoelastic earthquake simulation (DeVries, Thompson, and Meade 2017), and water-quality investigation (Arandia et al. 2018). In all examples, the ML has been used to perform non-linear regression between the inputs and outputs.

An alternative approach aims to embed information from physics or heuristic knowledge within the network. Physics-informed DL is a novel approach for resolving information from physics. The philosophy behind it is to approximate the quantity of interest (e.g., governing equation variables) by a deep neural network (DNN) and embed the physical law to regularize the network. To this end, training the network is equivalent to minimization of a well-designed loss function that contains the PDE residuals and initial/boundary conditions (Rao, Sun, and Liu 2020).

A further stream of related work has been started by Chen et al. (2018), who presented a novel approach to approximate the discrete series of layers between the input and output state by acting on the derivative of the hidden units. At each stage, the output of the network is computed using a black-box differential equation solver which evaluates the hidden unit dynamics to determine the solution with the desired accuracy. In effect, the parameters of the hidden unit dynamics are defined as a continuous function, which may provide greater memory efficiency and balancing of model cost against problem complexity. The approach aims to achieve comparable performance to existing state-of-the-art with far fewer parameters, and suggests potential advantages for time series modeling. For follow-up work in this stream, see Grathwohl et al. (2019); Gulgec et al. (2019).
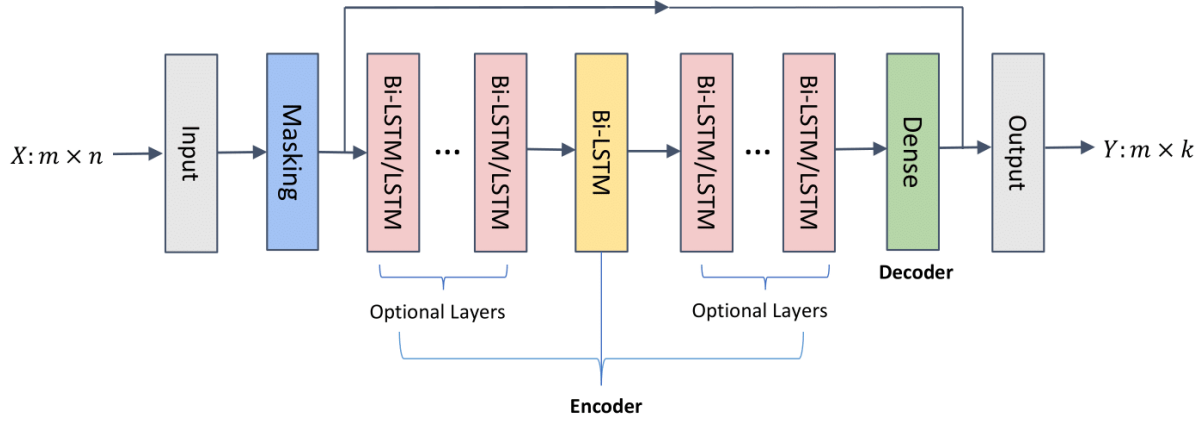
Figure 1: SPATIAL allows multidimensional input, which enables the deep neural network to extract features from different sensors and exploit the information learned to predict the time series for every sensor. That's being said for each sensor, the prediction is not only based on its previous time series, but also the information from other sensors.

(a) Classical LSTM Model for single sensor i

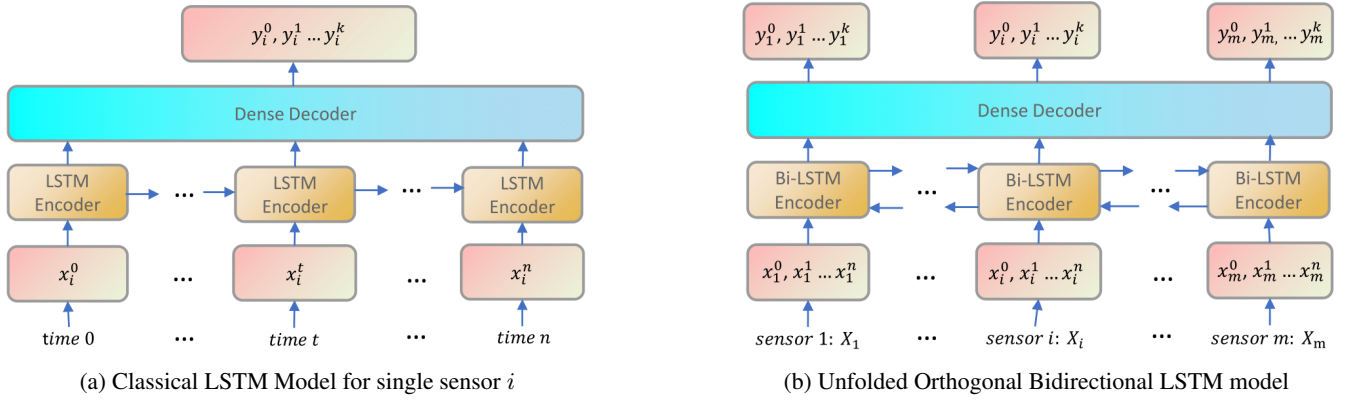(b) Unfolded Orthogonal Bidirectional LSTM model

Figure 2: Left figure (a) presents a schematic of a classical LSTM architecture to predict a single time series signal while (b) presents our SPATIAL implementation in which the spatial (sensor-to-sensor) and temporal patterns are explicitly learned by the network. In this representation, **m** is the number of sensors, **n** is the training time period, and **k** is the time period we wish to predict.

Methodology

SPATIAL is a first attempt at using a bidirectional LSTM model on both the spatial and temporal directions of a time-series signal. More specifically, based on geographical proximity or domain expertise, it is known that the signals that we wish to predict (ocean currents, temperature, and dissolved oxygen) have some dependency on neighbouring signals. We implement a bidirectional LSTM model across the *spatial* direction of pertinent sensors and train the model to learn the spatial and temporal structure. An overview of SPATIAL is provided in Figure 1, and described in detail in the following section.

Deep Neural Networks

RNN and its variants are widely studied for time series forecasting (Hochreiter and Schmidhuber 1997). A fundamental extension of RNNs compared to ANN approaches is param-

eter sharing across different parts of the model. This has natural applicability to the forecasting of time-series variables with historical dependency. A schematic representation is provided in Figure 2(a) illustrating the neural network mapping framework from an input time series, $(x_1, x_2, \cdots, x_t)$, to output label vector, $(y_{t+1}, y_{t+2}, \cdots, y_T)$. In effect, the RNN has two inputs, the present state and the past.

Standard RNN approaches have been shown to fail when lags between response and explanatory variables exceed 5–10 discrete timesteps (Gers 1999). Repeated applications of the same parameters can give rise to vanishing, or exploding gradients leading to model stagnation or instability. A number of approaches have been proposed in the literature to address this, with the most popular being LSTM (Goodfellow, Bengio, and Courville 2016).

An extension of the parameter sharing enabled by recurrent networks is bidirectional LSTM which processes se-

#	ADCP name	Long	Lat	Dist. (km)
1	E39_A_Sulafjorden	6.18	62.56	00.0
2	E39_B_Sulafjorden	6.35	62.67	22.4
3	E39_D_Breisundet	6.09	62.60	11.1
4	E39_F_Vartalsfjorden	5.74	62.16	65.2

Table 1: The geographical longitude and latitude coordinates, and reference name (complying with the naming convention used on the OpeNDAP portal (Institute 2020) of the four ADCP datasets used in this study, together with the distance between each sensor and the first one listed

quence data in backward and forward directions (Schuster and Paliwal 1997). This has natural application to sequential or time series data, and has demonstrated improved performance against unidirectional LSTM in areas such as language processing (Wang et al. 2016), and speech recognition (Graves, Jaitly, and Mohamed 2013). While bidirectional LSTM has previously been applied to time series forecasting, this paper is the first that applies across the spatial as well as the temporal components. Figure 2(b) provides a schematic of our SPATIAL implementation. A unidirectional LSTM is applied across the time direction of each sensor, and between individual sensors a series of stacked bidirectional layers support learning between distinct time series. The number of stacked layers is a hyperparameter selected during training. The features to the network consist of an $\mathbf{m} \times \mathbf{n}$ array (where $\mathbf{m}$ is number of sensors and $\mathbf{n}$ is number of time points) and labels consist a corresponding $\mathbf{m} \times \mathbf{k}$ array (where $\mathbf{k}$ is the prediction window).

The data

In this study, SPATIAL is applied to the following datasets: ocean current speed or velocity collected along the Norwegian coast, ocean temperature, and dissolved oxygen sampled from a high-density network of sensors at an aquaculture site in Atlantic Canada.

Acoustic Doppler Current Profiler (ADCP) is used to measure water current velocities over a depth range, using the Doppler effect of sound waves scattered back from particles within the water column. Four ADCPs were deployed at individual locations along the Norwegian coast. The sensors collect 1-minute interval measurements at vertical resolution or bins of between 5 and $15\,\mathrm{m}$ over a 1-year period (January 1st – December 31st 2018). Observations were averaged over the depth to generate a single time series for each ADCP.

Table 1 provides summary metrics of the ADCP deployments including longitude, latitude, and distance between ADCPs (relative to arbitrarily selected, ADCP 1). The data were downloaded from the OPeNDAP server provisioned by the Norwegian Meteorological Institute (Institute 2020). While these sensors are geographically distant (up to $60\,\mathrm{km}$ apart), theory imparted from transfer learning (Pan and Yang 2009) suggests that knowledge of the dynamics at one location could be used to guide training at others.

Temperature and dissolved oxygen observations were collected within an aquaculture farm of sea cages cultivat-

ing Atlantic salmon. Each cage was $100\,\mathrm{m}$ in circumference, $12\,\mathrm{m}$ deep, with around $25{,}000$ fish. The cages were equipped with an array of sensors sampling both temperature and dissolved oxygen. We gathered information from eight sensors distributed across a cage: 4 at $2\,\mathrm{m}$, and 4 at $8\,\mathrm{m}$ (North, South, East, West configuration).

These sensors sample a complex system described by natural conditions (natural ocean variation due to dynamics) and the effects of fish farming (modified temperature stratification patterns, consumption of oxygen by fish, etc.). Furthermore, the geographical proximity of cages exerted an influence: modification of upstream dynamics will be amplified at downstream sensors as effects accumulate (e.g. oxygen depletion) (O'Donncha, Hartnett, and Nash 2013). Tidally driven flows have the most significant effect on dissolved oxygen levels, with its influence a function of cage location within the farm. As waters flow from one end of the farm to the other, fish behaviour and physiology, as well as flow restriction from cage infrastructure reduce oxygen levels. This results in significant oxygen variations throughout the farm with high oxygen levels at one end of the farm and low oxygen levels at the other. Since farmers use low oxygen as an indicator as to whether to feed, it is important to implement robust predictive models that reflect spatial variations at high resolutions to enable informed decisions that impact fish growth and welfare (Burke et al. 2020).

The DO and temperature data was collected over a four-month period from August 16th – December 21st 2018 over relatively small geographical extents, characterised by significant dynamics and aquaculture-industry activities. Additional details on the data collection and application to directing aquaculture operations are provided in Burke et al. (2020).

Data Preprocessing

Anomalous data and gaps in the sensor data are common in ocean monitoring. Table 2 presents summary statistics including information on the frequency of occurrence of missing data. The *BMean* column quantifies the mean length of contiguous blocks of missing data for each sensor dataset (averaged across all sensors). The ADCP data in particular reports large data gaps which complicates learning and requires bespoke data cleaning routines to adjust to the characteristics of the models.

Each individual sensor data consists of a one-dimensional time series at 1-minute interval. We resampled each dataset to 30-minute intervals to avoid spurious fluctuations while maintaining the fidelity of the data. Data processing proceeded according to the requirements of the particular ML pipeline.

SPATIAL has the simplest data preparation requirement with each time series loaded into a single array of $[N_{sensors}, N_{timepoints}]$ for each of the three variables.

The AutoAI and AMLP pipelines follow the tabular ML workflow of data preparation. We used the TSML (Palmes, Ploennigs, and Brady 2020) which is an open-source package to automatically convert the time series data into matrix form by *sliding window*. Each row represents the autoregressive features together with the date-time features (year,

Dataset	Mean	σ	Skewness	Kurtosis	BMean
ADCP	20.22	7.54	2.51	9.11	216.61
Temp	12.21	0.85	-0.30	-1.34	4.36
DO	08.50	0.55	-0.13	-0.66	4.44

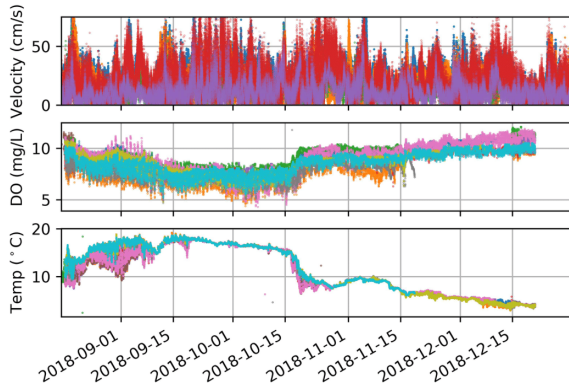Table 2: Summary statistics of the three datasets.

Figure 3: Time series plot of the three datasets. Top figure represents measurements of ocean current speed from four locations (summarised in Table XXX), while middle and bottom figure denotes dissolved oxygen and temperature observations respectively at eight locations.

month, day hour, day of week, etc), time-aligned with the corresponding label at the desired prediction window. We setup the problem as a 24-hour ahead prediction using half hour stride with 6-hour autoregressive window size predetermined based on the initial experiments.

For the AutoAI and AMLP approaches, missing data were flagged and any rows that contained missing data were dropped in the sliding window matrix generated. SPATIAL on the other hand simply applied a masking layer that acted on the data flagged as missing and automatically ignored them during learning.

!ignore the missing data, in starima to make it more consisten

The complete source code of SPATIAL has been released publicly under Apache license at https://github.com/IBM/spatial-lstm, whereby further details of the implementation can be studied in detail. The ADCP data we used is freely available and the open source code contains scripts to download and preprocess it. We currently do not have permission to release the oxygen and temperature data as it pertains to commercial aquaculture operations.

Results and Discussion

To evaluate the relative performance of SPATIAL with respect to existing approaches, we included two optimal baseline models based on the AutoML (Drori et al. 2018) technology:

- IBM AutoAI (IBM 2020): a technology that is directed at automating the end-to-end AI Lifecycle, from data clean-

ing, to algorithm selection, and to model deployment and monitoring in the ML workflow (Wang et al. 2020).

- AutoMLPipeline (AMLP) (Palmes 2020): a toolbox that provides a semi-automated ML model generation and forecasting.

Models generated by the above AutoML technology were used to benchmark the predictive skill of the SPATIAL framework, provide additional insight into data characteristics and pattern extraction of environmental datasets, and assess the ease of deployment of the different model frameworks.

Figure 3 presents all sensor data illustrating a number of characteristics:

- The current speed is highly volatile and reports rapid variations over a relatively large range between $0 - 75\,\mathrm{cm/s}$.

- Temperature data exhibits a pronounced seasonal pattern, reducing from a peak of about $18\,^{\circ}\mathrm{C}$ to $3\,^{\circ}\mathrm{C}$ in latter half of December.

- DO is exposed to relatively large fluctuations that are influenced by exogenous, non-seasonal factors (particularly aquaculture-driven impacts such as supplemental oxygenation and fish metabolism). Values drop rapidly from a peak of about $10\,\mathrm{mg/L}$ in late August to a minimum of about $7\,\mathrm{mg/L}$ in mid-October before increasing back to approximately $10\,\mathrm{mg/L}$.

These pronounced patterns encapsulated a wide-ranging array of time-series behaviours including volatility, seasonality, and exogenous drivers. Of particular interest are seasonal patterns in temperature data and the capabilities to learn seasonally-varying patterns from short, sub-seasonal datasets (both temperature and DO contain 17 weeks of observation data). These constitute a challenging time-series forecasting problem and early analysis of standard approaches such as GAM and ARIMA modeling reported substandard predictive skill.

	SPATIAL		AMLP-RF		AutoAI-XGB	
Dataset	MAE	σ	MAE	σ	MAE	σ
Temp	0.86	0.54	0.25	0.04	0.21	0.03
DO	0.32	0.25	0.27	0.04	0.28	0.04
ADCP	4.67	5.40	3.80	1.68	4.51	1.94

Table 3: The MAE and standard deviation of reported MAE across all three datasets for a five-fold cross-validation interrogation.

We compared our SPATIAL model against the two baseline implementations described previously. Tuning of model performance considered effects of appropriate number of time lags and hyperparameter selection, and adopted a 5-fold cross-validation for all models. The baseline models relied on a robust automated hyperparameter optimisation routine, while hyperparameter tuning of the SPATIAL adopted a greedy grid-search approach.
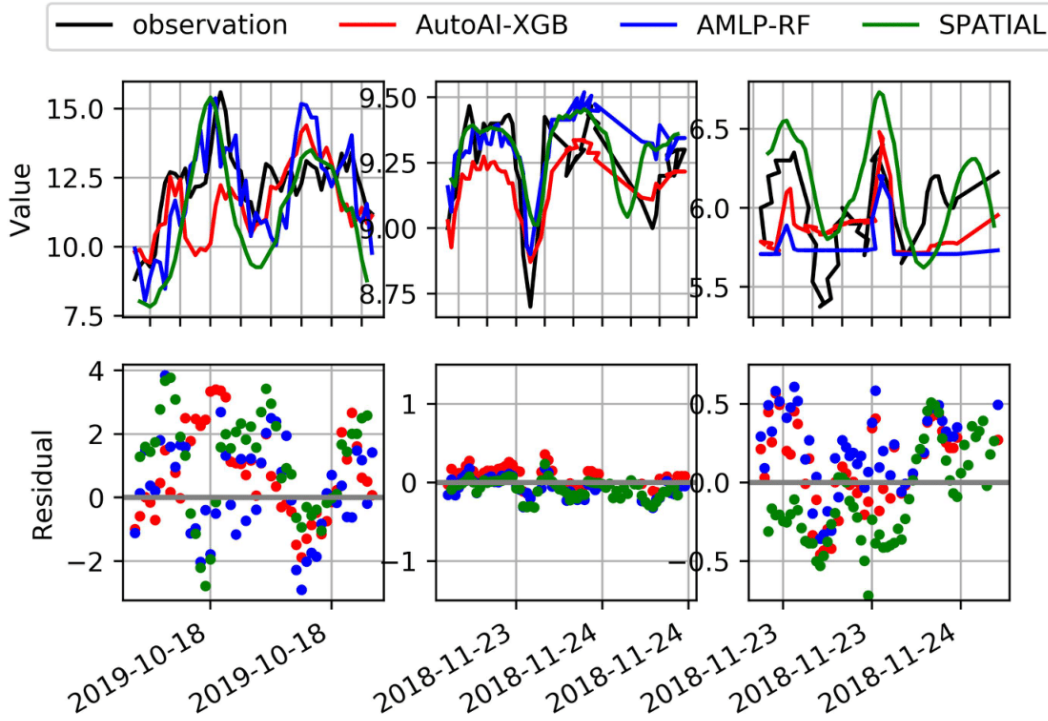
Figure 4: Comparison of SPATIAL performance against two benchmark models: An XGBoost model configured and deployed by AutoAI (red line), and a Random Forest model generated by using the AMLP pipeline (blue line). Black line indicates observations. The top row compares each model against observations while the bottom row presents the residual plots between observation and prediction. Each model predicts a 24-hour ahead forecast over 24 hours with a stride of 30 minutes (new forecast generated every 30 minutes). Data is presented for ADCP (left column), Dissolved Oxygen (middle column), and Temperature (right column).

Table 3 summarises model performance in terms of MAE and the standard deviation of the error across series of cross-validation tests. The baseline models that demonstrated best performance across all datasets were XGBoost and Random Forest, for AutoAI (AutoAI-XGB) and AMLP (AMLP-RF), respectively. We see strong performance of the different models across all datasets, despite the high complexity of the data. ADCP predictions reported significantly higher MAE for all models, which was largely a result of larger volatility and range in the data. Table 2 presents summary statistics of the three datasets indicating that the standard deviation varies across the datasets. Considering the reported MAE relative to the standard deviation of the data, we see comparable predictive skill across the three datasets.

Figure 4 provides a time series plot comparing the three models against observations (we selected one sensor from each of the dataset to illustrate). All models closely capture both the short-scale fluctuations and seasonal patterns of the data. This is particularly true for data such as temperature which has a pronounced seasonal pattern and can be challenging for ML models to learn with less than one-year of data (in our case, our training data was 17 weeks).

Performance comparisons demonstrated that existing algorithms give strong predictive skill on these datasets. This

performance is broadly replicated by our SPATIAL model. Namely, adopting a simpler data preprocessing, and model generation and implementation pipeline (fewer models to support by grouping sensor data), we get performance comparable to existing state-of-the-art. Our SPATIAL model has a number of practical advantages against existing algorithms, particularly:

1. The data preprocessing pipeline is simplified as the SPATIAL pipeline loads all data into a single array that is fed to the networks (the model does not rely on exogenous variables or feature transformations). Missing data is handled readily with masking layer. This compares with the the benchmark models that requires a bespoke model for each sensor signal and transforming into appropriate matrix structure dependent on the desired prediction window, historic features and stride forward step length (Palmes, Ploennigs, and Brady 2020).

2. By feeding data from multiple sensors that are geographically in proximity or share certain characteristics, the network can capture certain physical relationships that exist in nature and potentially serves to regularize the model.

3. Treating all sensors simultaneously provides a large increase in computational efficiency – only one model is

trained rather than a model for each signal.

Points 1 and 2 above are closely connected. In forecasting, we wish to use the simplest model that most closely mirrors what we know about the system. The proposed SPATIAL framework simplifies the model training and deployment process for environmental modeling by:

- requiring only one model be trained and maintained;

- eliminating need to implement transformations or matrix manipulations of the input data;

- naturally generating a time series sequence for the desired forecasting period rather than a single value for each direct model (although classical approaches can adopt an iterative method to generate a time-sequence prediction via multi-step forecasting, predictive skill is generally limited (Hamzaçebi, Akay, and Kutay 2009)).

An important point is that the SPATIAL model processes information from multiple sensors that are known to be connected (based on the spatiotemporal nature of the dataset). There is a large volume of literature related to applying physics-informed constraints to ML which is discussed in the Related Work section. The objective in many of those studies is to augment the models with data external to the training data via aspects such as modified loss functions, data augmentation, or specifying consensus filters to guide disparate models or data towards convergence (Haehnel et al. 2020). The proposed methodology presents a natural framework to ingest information external to the time series signal to the predictive model positing the opportunity to enhance learning. Within a time series forecasting paradigm, the nominal architectures that deliver performance are well-established, and often the challenge reduces to appropriate data conditioning and feature engineering. Missing data, anomalous data, and sensor drift are major obstacles, particularly for ocean-related deployments. SPATIAL presents a natural framework to reduce the impacts of individual sensor error or bias by enabling parameter sharing and pattern learning across multiple sensors that can serve to regularize individual deviations.

Further, we investigated how interdependence of different variables can inform learning. Dissolved oxygen has a nonlinear dependency on temperature and both data were sampled at the same location. In a classical sense, one could adopt temperature as a feature to a ML framework. We explored a combined oxygen and temperature SPATIAL that fed data from both simultaneously to the network. Results demonstrated reduced MAE compared to treating both independently indicating the ability to learn from multiple scales. The combined SPATIAL reported MAE of $0.3\,\mathrm{mg/L}$, and $0.65\,^\circ\mathrm{C}$ for oxygen and temperature, respectively, when SPATIAL was applied across both datasets simultaneously, while the corresponding results for the individual models were 0.32 and 0.86 for oxygen and temperature.

Ocean variables such as dissolved oxygen and temperature are intrinsically connected and it is intuitive to apply the capabilities of DL to extract and exploit these relationships to achieve uplift in predictive skill. Traditionally, this is a part of the feature engineering phase. SPATIAL however, provides a natural framework to readily explore these relationships via a simple DL framework that extracts these spatiotemporal patterns.

Finally, this paper compares SPATIAL against two baseline frameworks. Importantly, the baseline models required training on each dataset individually, representing a total of 20 model training iterations compared to three for the SPATIAL. The computational saving is approximately similar (six-fold reduction), although in reality the savings are larger, as the SPATIAL architecture is relatively simple and required minimal hyperparameter tuning compared to a full algorithm search implemented by AutoML approaches. These computational savings are particularly attractive for edge-computing applications and is a major area of interest for marine industry applications (O'Donncha and Grant 2020). Many ocean sensor deployments consist of multiple sensors connected to a low-power compute device to manage storage, data processing, and connectivity. Extending these with a lightweight predictive framework that can forecast based only on the data collected from the sensors can allow for more intelligent data processing and enable adaptive sampling for sensor networks (Jain and Chang 2004).

Conclusions

This paper presents a DL framework based on bidirectional LSTM to extend learning of time series signals. The resulting model provides a scalable forecasting system that adjusts naturally to spatiotemporal patterns. The cost of model training was reduced by an order of magnitude as we trained a single model for each *dataset* rather than each *signal*. SPATIAL adjusts naturally to missing data through a simple masking layer and the random distribution of missing data between sensors can be leveraged to lessen the effect of data gaps on learning (i.e. since typically sensors exhibit data gaps at different times, combining information from multiple sensors in a single framework can improve learning on noisy and error-prone data). Finally extending the recurrent structure of LSTM-type approaches for time series applications to the spatiotemporal direction has potential to translate to many different industry applications such as weather, transport, and epidemiology.

The ML approach is extremely useful in forecasting oceanographic conditions of water quality (e.g. oxygen), especially as it assimilates realtime forcing variables such as temperature from sources including sensor networks or remote sensing. As additional big data layers become available from instruments such as underwater autonomous vehicles, this approach can be extended to 3D and 4D geospatial data to map nearshore dynamics of oxygen or other variables. These forecasts are essential in management decisions regarding coastal resources including aquaculture, fisheries, and pollution discharge.

References

Arandia, E.; O'Donncha, F.; McKenna, S.; Tirupathi, S.; and Ragnoli, E. 2018. Surrogate modeling and risk-based analysis for solute transport simulations. *Stochastic Environmental Research and Risk Assessment* 1–15. ISSN

1436-3240. doi:10.1007/s00477-018-1549-6. URL http://link.springer.com/10.1007/s00477-018-1549-6.

Ashkezari, M. D.; Hill, C. N.; Follett, C. N.; Forget, G.; and Follows, M. J. 2016. Oceanic eddy detection and lifetime forecast using machine learning methods. *Geophysical Research Letters* 43(23): 12–234.

Bryan, K. 1969. Climate and the ocean circulation. *Mon. Weather Rev* 97(11): 806–827.

Burke, M.; Grant, J.; Filgueira, R.; and Stone, T. 2020. Oceanographic processes control dissolved oxygen variability at an Atlantic salmon farm: Application of a real-time sensor network. *Aquaculture* .

Caballero-Alfonso, A. M.; Carstensen, J.; and Conley, D. J. 2015. Biogeochemical and environmental drivers of coastal hypoxia. *Journal of Marine Systems* 141: 190–199.

Chen, T. Q.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural Ordinary Differential Equations. In Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 31*, 6572–6583. Curran Associates, Inc. URL http://papers.nips.cc/paper/7892-neural-ordinary-differential-equations.pdf.

Constantin, P.; and Foias, C. 1988. *Navier-stokes equations*. University of Chicago Press.

de Bezenac, E.; Pajot, A.; and Gallinari, P. 2017. Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge. *arxiv preprint arXiv:1711.07970* URL http://arxiv.org/abs/1711.07970.

DeVries, P. M. R.; Thompson, T. B.; and Meade, B. J. 2017. Enabling large-scale viscoelastic calculations via neural network acceleration. *Geophysical Research Letters* 44(6): 2662–2669. ISSN 00948276. doi:10.1002/2017GL072716. URL http://doi.wiley.com/10.1002/2017GL072716.

DeYoung, B.; Heath, M.; Werner, F.; Chai, F.; Megrey, B.; and Monfray, P. 2004. Challenges of modeling ocean basin ecosystems. *Science* 304(5676): 1463–1466.

Drori, I.; Krishnamurthy, Y.; Rampin, R.; de Paula Lourenco, R.; Ono, J.; Cho, K.; Silva, C.; and Freire, J. 2018. Alphad3m: Machine learning pipeline synthesis. In *Workshop on AutoML (ICML)*.

Ducournau, A.; and Fablet, R. 2016. Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived SST data. In *2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing (PRRS)*, 1–6. IEEE. ISBN 978-1-5090-5041-3. doi:10.1109/PRRS.2016.7867019. URL http://ieeexplore.ieee.org/document/7867019/.

Esling, P.; and Agon, C. 2012. Time-series data mining. *ACM Computing Surveys (CSUR)* 45(1): 1–34.

EU Copernicus. 2020. Copernicus - Marine environment monitoring service. URL https://marine.copernicus.eu/.

Fawaz, H. I.; Forestier, G.; Weber, J.; Idoumghar, L.; and Muller, P.-A. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33(4): 917–963.

Gamboa, J. C. B. 2017. Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887* .

Gan, J.; and Allen, J. S. 2005. On open boundary conditions for a limited-area coastal model off Oregon. Part 1: Response to idealized wind forcing. *Ocean Modelling* 8(1-2): 115–133.

Gers, F. 1999. Learning to forget: continual prediction with LSTM. In *9th International Conference on Artificial Neural Networks: ICANN '99*, volume 1999, 850–855. IEE. ISBN 0 85296 721 7. doi:10.1049/cp:19991218. URL http://digital-library.theiet.org/content/conferences/10.1049/cp{_}19991218.

Gers, F. A.; and Schmidhuber, E. 2001. LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE Transactions on Neural Networks* 12(6): 1333–1340.

Gokaraju, B.; Durbha, S. S.; King, R. L.; and Younan, N. H. 2011. A Machine Learning Based Spatio-Temporal Data Mining Approach for Detection of Harmful Algal Blooms in the Gulf of Mexico. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4(3): 710–720. ISSN 1939-1404. doi:10.1109/JSTARS.2010.2103927. URL http://ieeexplore.ieee.org/document/5701668/.

Goodfellow, I. J.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. Cambridge, MA, USA: MIT Press. http://www.deeplearningbook.org.

Grathwohl, W.; Chen, R. T. Q.; Bettencourt, J.; and Duvenaud, D. 2019. Scalable Reversible Generative Models with Free-form Continuous Dynamics. In *International Conference on Learning Representations*.

Graves, A.; Jaitly, N.; and Mohamed, A.-r. 2013. Hybrid speech recognition with deep bidirectional LSTM. In *2013 IEEE workshop on automatic speech recognition and understanding*, 273–278. IEEE.

Gulgec, N. S.; Shi, Z.; Deshmukh, N.; Pakzad, S.; and Takáč, M. 2019. FD-Net with Auxiliary Time Steps: Fast Prediction of PDEs using Hessian-Free Trust-Region Methods. In *NeurIPS 2019 Workshop (Beyond First Order Methods in ML)*.

Haehnel, P.; Mareček, J.; Monteil, J.; and O'Donncha, F. 2020. Using deep learning to extend the range of air pollution monitoring and forecasting. *Journal of Computational Physics* 408: 109278.

Hamzaçebi, C.; Akay, D.; and Kutay, F. 2009. Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert Systems with Applications* 36(2): 3839–3844.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Holt, J.; Harle, J.; Proctor, R.; Michel, S.; Ashworth, M.; Batstone, C.; Allen, I.; Holmes, R.; Smyth, T.; Haines, K.; et al. 2009. Modelling the global coastal ocean. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367(1890): 939–951.

IBM. 2020. IBM Watson Studio - AutoAI. URL https://www.ibm.com/ie-en/cloud/watson-studio/autoai.

Institute, N. M. 2020. Norkyst Model. URL https://thredds.met.no/thredds/catalog.html.

Jain, A.; and Chang, E. Y. 2004. Adaptive sampling for sensor networks. In *Proceeedings of the 1st international workshop on Data management for sensor networks: in conjunction with VLDB 2004*, 10–16.

James, S.; Zhang, Y.; and O'Donncha, F. 2018. A machine learning framework to forecast wave conditions. *Coastal Engineering* 137. ISSN 03783839. doi:10.1016/j.coastaleng.2018.03.004.

Ji, Z.-G. 2017. *Hydrodynamics and water quality: modeling rivers, lakes, and estuaries*. John Wiley & Sons.

Lary, D. J.; Müller, M. D.; and Mussa, H. Y. 2004. Using neural networks to describe tracer correlations. *Atmospheric Chemistry and Physics* 4(1): 143–146.

O'Donncha, F.; and Grant, J. 2020. Precision Aquaculture. *IEEE Internet of Things Magazine* 2(4): 26–30. ISSN 2576-3180. doi:10.1109/iotm.0001.1900033.

O'Donncha, F.; Hartnett, M.; and Nash, S. 2013. Physical and numerical investigation of the hydrodynamic implications of aquaculture farms. *Aquacultural Engineering* 52: 14–26.

O'Donncha, F.; Hartnett, M.; Nash, S.; Ren, L.; and Ragnoli, E. 2015. Characterizing observed circulation patterns within a bay using HF radar and numerical model simulations. *Journal of Marine Systems* 142: 96–110.

Palmes, P. 2020. AutoMLPipeline: A Toolbox for Building ML Pipelines. doi:10.5281/zenodo.3980593. https://doi.org/10.5281/zenodo.3980593.

Palmes, P.; Ploennigs, J.; and Brady, N. 2020. TSML (Time Series Machine Learning). *Proceedings of the JuliaCon Conferences* 1(1): 51. doi:10.21105/jcon.00051. https://doi.org/10.21105/jcon.00051.

Pan, S. J.; and Yang, Q. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22(10): 1345–1359.

Pidgeon, E. J.; and Winant, C. D. 2005. Diurnal variability in currents and temperature on the continental shelf between central and southern California. *Journal of Geophysical Research: Oceans* 110(C3).

Rao, C.; Sun, H.; and Liu, Y. 2020. Physics informed deep learning for computational elastodynamics without labeled data. *arXiv preprint arXiv:2006.08472* .

Schuster, M.; and Paliwal, K. K. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 45(11): 2673–2681.

Smyth, T. J.; Moore, G. F.; Hirata, T.; and Aiken, J. 2006. Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment. *Applied Optics* 45(31): 8116–8131.

Song, Y.; and Haidvogel, D. 1994. A semi-implicit ocean circulation model using a generalized topography-following coordinate system. *Journal of Computational Physics* 115(1): 228–244.

Stockie, J. M. 2011. The mathematics of atmospheric dispersion modeling. *SIAM Review* 53(2): 349–372.

Wang, C.; Yang, H.; Bartz, C.; and Meinel, C. 2016. Image captioning with deep bidirectional LSTMs. In *Proceedings of the 24th ACM international conference on Multimedia*, 988–997.

Wang, D.; Ram, P.; Weidele, D. K. I.; Liu, S.; Muller, M.; Weisz, J. D.; Valente, A.; Chaudhary, A.; Torres, D.; Samulowitz, H.; et al. 2020. AutoAI: Automating the End-to-End AI Lifecycle with Humans-in-the-Loop. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, 77–78.

Wiewel, S.; Becher, M.; and Thuerey, N. 2018. Latent-space Physics: Towards Learning the Temporal Evolution of Fluid Flow. *arxiv preprint arXiv:1802.10123* URL http://arxiv.org/abs/1802.10123.