Real-Time In Situ Prediction of Ocean Currents

Alexandre Immas [1], Ninh Do [1], Mohammad-Reza Alam

Department of Mechanical Engineering, University of California, Berkeley, California 94720, USA

Abstract

The prediction of ocean currents is essential for the path planning and control of Autonomous Underwater Vehicles. Regional physics-based forecast models provide valid predictions but are too computationally expensive for real-time prediction necessary for AUV navigation. While vehicle sensors allow to measure the spatial evolution of current, temporal prediction remains an open problem as existing data-driven models with real-time capabilities have only been shown to work at locations where data have been used to develop the model. We propose in this paper two predictive tools using deep learning techniques, a Long Short-Term Memory Recurrent Neural Network and a Transformer, to perform real-time in-situ prediction of ocean currents at any location. We use a data set from the National Oceanic and Atmospheric Administration to show that they provide state-of-the-art predictions at various locations across the United States where no data have previously been used to train the models.

Keywords: Ocean currents prediction, Recurrent Neural Network, Attention mechanism, AUV navigation

1. Introduction

Ocean currents prediction is critical for the safe and reliable navigation of Unmanned Underwater Vehicles (UUVs). Path planning of Autonomous Underwater Vehicles (AUVs) use oceanographic predictions from numerical forecast models based on mass and hydrostatic momentum balance [1, 2]

[1]These authors contributed equally

such as the Regional Ocean Model Systems (ROMS) [3], the Navy Coastal Ocean Model (NCOM) [4] and the Global Real-Time Ocean Forecast System (RTOFS-Global) [5]. These predictive tools have computation time of 24 hours and a resolution respectively of 3, 3.7 and 9.3km. They are therefore not suited for trajectory re-planning and control where the spatio-temporal variations of the current must be computed on the vehicle's embedded system with relatively high frequency for precise navigation [6]. In this situation, the current speed vector has been estimated in real-time using Kalman Filtering combined with the vehicle navigation sensors [7] or with acoustic positioning [8]. These techniques however do not provide any information on the spatio-temporal evolution of the ocean current. The spatial evolution can be measured with an on-board Acoustic Doppler Current Profiler (ADCP) that provides an instantaneous current profile along one or multiple lines. This technique has been used for AUV navigation [9, 10] but neglecting the temporal evolution leads to suboptimal trajectories that can turn out catastrophic.

Ocean current temporal prediction has historically been made with the Harmonic Method (HM) [11] that uses sets of harmonic constituents precomputed at a finite number of near-shore sites from current data acquired with ADCPs. The predictions are therefore only valid where the data have been acquired which is obviously not suitable for AUVs navigation. Machine learning techniques such as Support Vector Regression [12, 13] or Genetic Algorithm [14] have been proposed to reduce the amount of data necessary to train the predictive tools and to improve predictions accuracy. Other researchers have proposed the use of artificial neural networks using fully connected layers [15] or recurent LSTM layers [16] with similar motivations. These models have been trained and tested with data coming from the same sites providing no information on their generalization to locations where data have not been used during training. A new predictive tool allowing real-time in-situ prediction of ocean currents at any location is therefore needed for AUV navigation.

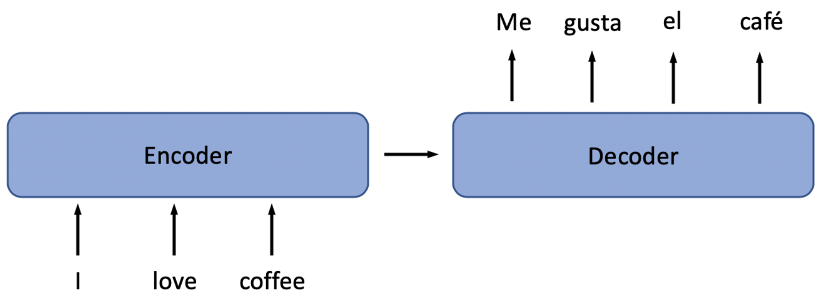
In this paper, we propose a Recurrent Neural Network (RNN) Long-Short Term Memory (LSTM) model [17] and a Transformer model [18] to perform predictions of ocean current speed and direction as described in section 2. These two methods have outperformed fully connected neural networks and other machine learning techniques in Natural Language Processing [19, 20] that has many similarities with ocean currents prediction. RNNs have been the state-of-the-art method in modelling time series data in the last decade. Recently, attention-based methods and in particular the Transformer have however exceeded RNN performance in many Natural Language Processing tasks [20]. We show in section 3 that these deep learning techniques can predict ocean currents at various locations across the United States with a state-of-the-art accuracy and without having been trained with data from these locations.

2. Methodology

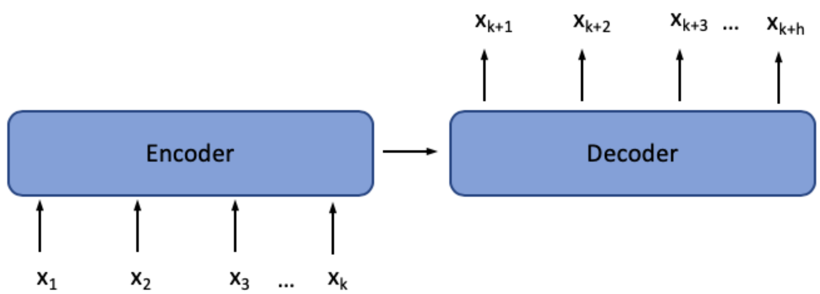
2.1. Deep Auto-Regressive Networks

Our approach is inspired from the latest progress in Natural Language Processing (NLP) and particularly in sequence-to-sequence models. The latter have obtained great results on complex tasks involving time series data such as translation, speech recognition, video captioning, etc. These models are made of two parts, an encoder and a decoder, as shown on Figure 1a for machine translation. On the Encoder side, the neural network collects information from embedded words and propagates the information forward. The Decoder starts with a special token indicating that the model should now predict output words. As words of the output sentence are predicted, additional information is propagated forward in the neural network until a special token is predicted, signifying the end of the sentence. It is straightforward to modify this model for time series and thus, creating a deep auto-regressive network. Vectors with numerical values now replace embedded words in both Encoder and Decoder. The model is therefore able to predict time series values based on previous values as shown on Figure 1b. For the prediction of ocean currents, we propose deep auto-regressive networks that have inputs and outputs of similar length. A network can thus be trained to use 24 hours, 1 week or 1 month of data to respectively predict the next 24 hours, 1 week or 1 month.

Recurrent Neural Network (RNN) Long Short-Term Memory (LSTM) and Transformer are special types of neural networks that are commonly used for sequence-to-sequence models. Loosely speaking, a neural network is a matrix-based algorithm to be used for approximating a process which is, in the scope of this paper, defined as a black-box function that takes as



(a) Machine Translation from English to Spanish.



(b) Deep Auto-regressive Model for time series.

Figure 1: Sequence-to-Sequence Models

input and outputs data. Neural network emerged in recent years as a state-of-the-art approach to data modelling thanks to its universal approximating ability through a sequence of matrix operations and nonlinear activations. Among its core components are its weights that act on and transform input into output. The visualization of a neural network with multi-dimensional input and scalar output is presented in Figure 2 [21]. It is noticed that there is a component σ at each hidden node, called activation function. It plays an important role in adding nonlinearity to the model, thereby enabling the model to approximate the nonlinear processes. Otherwise, a sequence of linear matrix operations can be reduced to a single linear matrix operation, and a neural network without activation is not able to simulate a nonlinear process.

Initially, all the weights are generated randomly, simulating a random process that outputs random results no matter what input it takes. After

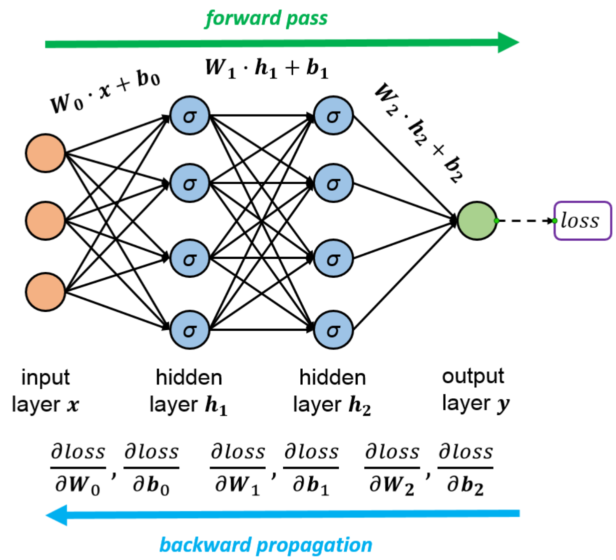
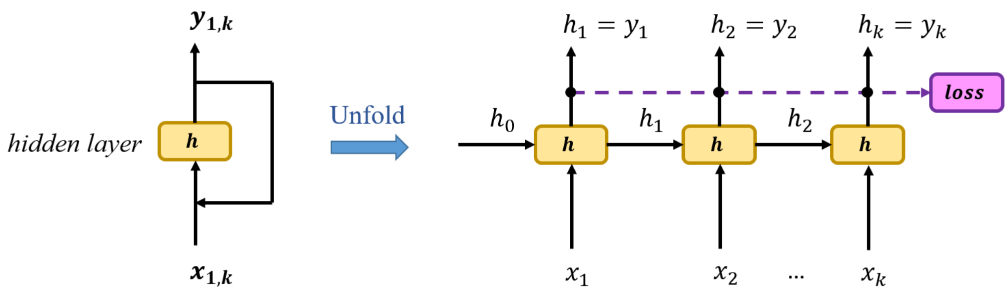


Figure 2: Neural network scheme and its operations.

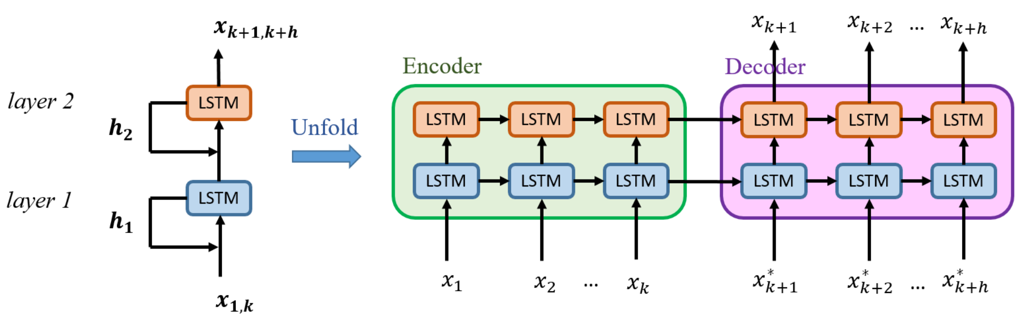
being trained on the collected data from a true process using the backpropagation algorithm, neural network has its weights converging to the state that can closely imitate the behavior of the true process. The backpropagation algorithm involves a forward pass and a backward pass using the chain rule on the pre-defined loss. It then usually uses the stochastic gradient descent to update the weights with a fraction of the loss gradient. The procedure is illustrated in Figure 2 with forward pass and backpropagation.

2.2. Long Short-Term Memory

Recurrent neural network is a recurrent version of neural network in which the input components, along with the hidden outputs, are recurrently fed into the network. Its architecture is similar to that of neural network with a single set of weight and bias per layer, but they differ in the ways they work. Figure 3a illustrates the structure of a typical recurrent neural network and its unfolds in operation. Let us say the input x has n components $x_1, ..., x_n$, i.e. n-dimensional input. First, x_1 and the all-zero hidden state h_0 are fed into the RNN to output the hidden state h_1. Next, x_2 and the hidden state h_1 are fed into the RNN to output the hidden state h_2, and so on until x_n and the hidden state h_{n-1} to output h_n which is also the output y. RNN is



(a) Recurrent neural network and its unfold in operation. LSTM neural network shares the same architecture except that the hidden layer h is replaced with a LSTM layer.



(b) Long Short-Term Memory based model consists of an encoder and a decoder.

Figure 3: The proposed LSTM-based model is the seq2seq based on LSTM.

also trained using backpropagation and stochastic gradient descent similar to a regular neural network.

Long-Short Term Memory (LSTM) is a special type of RNN whose hidden layers have a more complicated architecture with several combinations of nonlinear activations on matrix operations. In specific, a LSTM hidden layer has a cell, an input gate, an output gate and a forget gate [22]. These three gates regulate the flow of information into and out of the cell which in turn selectively memorizes the information throughout the past. This feature makes LSTM particularly suitable for modelling long sequence data such as texts, signals, time series. For it efficiency in solving similar problems, LSTM is attempted as an approach to modelling and predicting the ocean currents.

The proposed LSTM model in this paper consists of two sequentially connected LSTM networks, each of two hidden layers, serving as an encoder and a decoder as depicted in figure 3b. This is a typical seq2seq model based on LSTM. The encoder first encodes the information in the past $\mathbf{x_{1,k}}$, storing the data patterns in the LSTM cells. The encoder output combined with the information query $\mathbf{x^*_{k+1,k+h}}$ are subsequently fed into the decoder to produce the future values. In our problem, $\mathbf{x_{1,k}}$ composes of time, speed and direction of ocean currents while $\mathbf{x^*_{k+1,k+h}}$ includes only time as query for speed and direction.

2.3. Attention

In the simplest words the attention is a mechanism to capture the correlation between difference data sequences. Its output is employed to boost the accuracy of the existing neural network models thanks to the extra information of correlation that is incorporated into the models. The correlation is obtained through the cross-product of two data sequences. Give two random variable vectors $\mathbf{X} = [\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}]^T$ and $\mathbf{Y} = [\mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y_n}]^T$, for instance, the correlation of $\mathbf{X}$ with itself is given by $\mathbf{X} \times \mathbf{X}^T$ and the correlation between $\mathbf{X}$ and $\mathbf{Y}$ is $\mathbf{X} \times \mathbf{Y}^T$. These are called, in the scope of this paper, the self-attention and the cross-attention, respectively, each represented by a correlation matrix.

The correlation mechanism is not new but its application in the neural networks makes a real difference. Each element in a correlation matrix represents the correlation of only two random variables, and thus, of the limited use. In order to capture more information, instead of gaining the correlation of raw input and target sequences the attention obtains the correlation of the encoder and decoder outputs, each of which is the nonlinear combination of different random variables of the raw input and target, respectively. Thus, the captured correlation matrix turns more informative as each of its elements represents correlation of the combined random variables. Since the encoder and decoder are trainable, the attention is trained consequently and converges to the optimal state in which the most correlated information is stored.

The attention output is incorporated as a context into the neural network models, adding more information to the predictions, thereby enhancing the models accuracy. In brief, the attention operations are summarized as follows:

$$
\begin{aligned}
correlations &\leftarrow encoder\ output \times decoder\ output \\
weights &\leftarrow normalized\ correlations \\
context &\leftarrow weights \cdot encoder\ output \\
predicted\ output &\leftarrow decoder\ output + context
\end{aligned}
$$

where, the *weights* are the normalized *correlations*, and the *context* is the weighted *encoderoutput*. The name *attention* comes from the fact that the *predicted output* attend to the *encoder output*, which is the encoded version of input, through the *context*.

Figure 4 visualizes the attention of the predicted output to the input, i.e. the normalized correlations map indicating how much each output value attends to all the input values. The normalized weights sum up to 1 for each column. The red predicted output on the top correlate with the black input on the left through two obvious periodic trends: daily and half-daily, shown by a main diagonal bright strip and two subordinate diagonal less-bright strips, respectively. The trends are more separating towards future values in both the input and output forward directions. This could be because the earlier values rely more on the past while the later values leverages the earlier ones and attend less but more specific points in the past.

2.4. Transformer

The Transformer is a neural network with an encoder-decoder structure that is solely based on attention mechanisms [18]. Originally developed for NLP, both encoder and decoder use attention layers and fully connected feed-forward layers as shown on Figure 5. Both layers include a residual connection and a normalization layer to improve the learning process. The model uses self-attention to encode information about other time steps when processing an input at a specific time. The input representation therefore includes contextual information allowing the model to account for dependencies. While the encoder self-attention layer can access inputs at all time steps, the decoder self-attention can only access inputs up to the time step being processed to respect the auto-regression. Finally, the encoder-decoder attention layer allows the decoder to focus on the most relevant inputs to carry out is predictions.

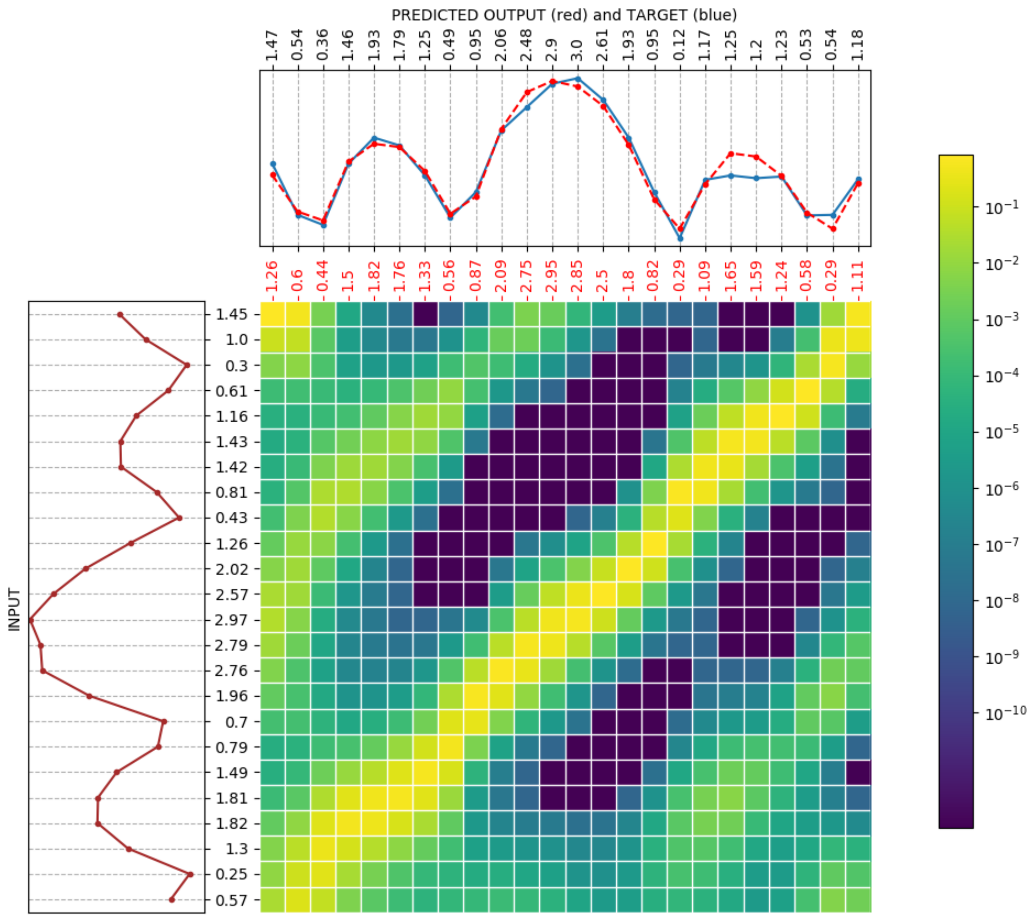


Figure 4: The attention visualization shows the weights or normalized correlations map of the input and output. The range of weights is 0-1 The y axis includes the input values in the chronological order upward. The x axis includes the predicted output (red) and target (blue) values in the chronological order rightward. The brighter color indicates the more correlation.

In this paper, we propose to use a Transformer encoder-decoder structure modified to work for time-series. An embedding layer is used in the original Transformer [18] to map words into vectors. For time series, we replace this layer by a linear fully connected layer to transform input vectors into m-dimensional vectors [23]. A similar fully connected layer is used at the decoder end to convert back predictions into physical quantities and the

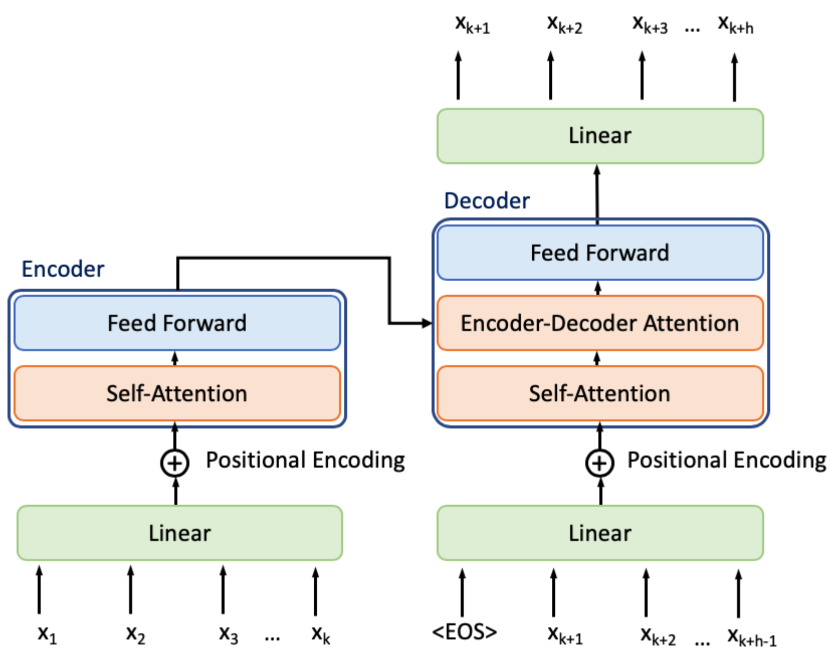


Figure 5: Transformer model architecture adapted for ocean currents prediction.

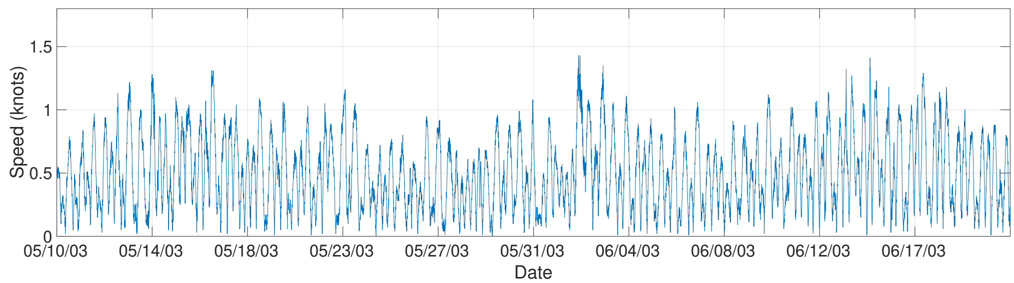
softmax loss is replaced by a L2 loss. Current speed and direction x_1, x_2, ..., x_k are used as input of the Encoder and predicted current speed and direction x_{k+1}, x_{k+2}, ..., x_{k+h} are the output of the Decoder. Decoder input is first an empty time series with a start-of-prediction token in first position. Predictions are made step-by-step and are added to the decoder input to predict current speed and direction at the next time step using attention mechanism to account for dependencies.

2.5. Data

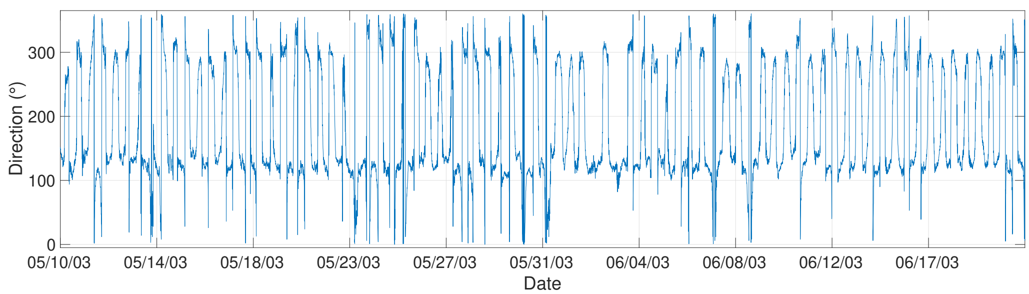
Ocean current data at 831 sites was downloaded from the Historic Stations dataset on the National Oceanic and Atmospheric Admninistration (NOAA) website [24]. Current speed and directions have been acquired with ADCPs that were either bottom-mounted or shore-mounted over measurement campaigns from 1 to 4 months. On this dataset, 222 stations have data length greater than 2 months. Sampling interval of the raw data are either 6 minutes or 10 minutes. We estimated that this was the minimum amount of data necessary to train and test the Transformer and the LSTM and therefore set the maximum time horizon of the models to 1 month. After training, both models can predict ocean current for 1 month based on 1 month of data acquired anywhere in the United States. Note that the length restriction is only due to the data available and that there is no numerical restriction to increase the model prediction length. Depending on the AUV mission, a shorter horizon time may also be more practical as less data is required to start the predictions. The authors decided to present the predictions with the highest time horizon allowed by the database because shorter horizon models mechanically benefit from a higher quantity of data points leading to better predictive capabilities. The performance of the models presented in this paper is therefore a lower bound for models with shorter horizons.

Two data cleaning steps were applied to form the final dataset. First, we observed that currents at some sites had complex directions change that may be due to local specificity or data acquisition error. A typical example is shown on Figure 6. We decided to remove these sites from our dataset as we did not have enough information to conclude on the validity of the data. Second, we observed that missing data points in the raw time series coincided with a non-physical variation of the ocean current speed as shown on Figure 7. We observe a tidal pattern from May 20, 2012 to August 6, 2012 with clear semidiurnal, diurnal and monthly variations and a sudden change of pattern around August 6, 2012. As tides are generated by steady-state periodic gravitational forces of the Sun and the Moon, the authors believe that this is a consequence of a faulty sensor or of data acquisition errors. Corresponding sequences were therefore removed from the time series. Finally, an interpolation was carried out to obtain a uniform sampling interval of 1 hour. This value was set to reduce the size of the models while capturing the important variation of the currents. A dataset including 78 stations was finally selected. 66 stations were used to train the model and 12 stations were used in a test set to show the ability of the models to predict ocean currents at various locations across the United States where no data have previously been used to train the models.

In addition to the aforementioned Historic Stations, NOAA is currently deploying ADCPs over 61 stations that actively collect the ocean current data at various locations around the United States [24]. These active stations are similar to the historic stations in most senses but actively working at the



(a) Ocean current speed.



(b) Ocean current direction.

Figure 6: Ocean current measured at Chesapeake Bay (CHB0303) from May 9, 2003 to August 21, 2003. Depth: 11.7ft.

moment, collecting and sending data in real time. The time granule of data sampling is 6 minutes. The sensors are deployed over every 5-8 months, then recovered for a couple of hours for maintenance and checking purpose, which causes some data gaps. In addition, there are several data gaps within any deployment time span that might probably be caused by the sensing inconsistency of the sensors due to their harsh working environment. Indeed, we found it challenging to obtain any complete two-month data in any active station. We instead looked for the incomplete two-month data with small data gaps and interpolated the missing data points. The data cleansing is conducted in the same manner as that for the historic stations. We completed the test set with 3 active stations to ensure that the models work consistently not only with historic data, but also with data recently acquired without the need of re-training the models

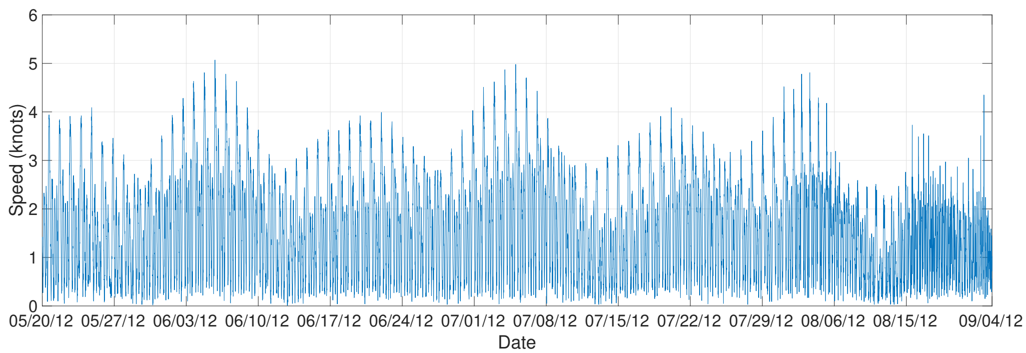


Figure 7: Ocean current speed measured at San Francisco Bay entrance (SFB1201) from May 19, 2012 to September 11, 2012. Depth: 38.6ft.

3. Results and discussion

3.1. Validation with measurements from historic stations

We present in this section the validation of the models on the 12 historic stations, 2 depths for each, set aside in the test set, as described in section 2.5. We used the Adam optimization algorithm [25] to train the Transformer model for 10 epochs and the LSTM model for 2 epochs with a constant learning rate of 10^{-3} and a batch size of 64 [26, 27]. The models were built on the TensorFlow platform supporting the implementation of neural network, and their training was carried out on a GPU NVIDIA TITAN V with a memory of 12 GB. After hyperparameter optimization, we found that the best Transformer architecture has a single layer of encoder and decoder as shown on Figure 5, a model dimension $m = 128$ and a fully connected feed-forward network with a single layer of dimensionality 256 with ReLu action and a dropout of 0.5. On the other hand, the best LSTM model adopts a slightly different architecture with two LSTM layers, each of dimensionality 512, as demonstrated in Figure 3b.

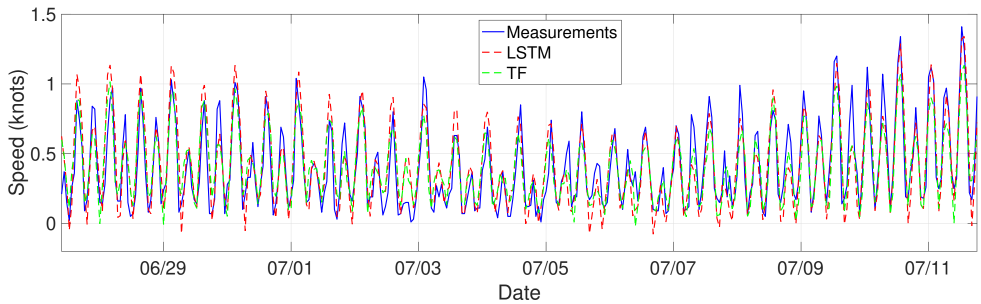
To analyze the performance of the models, we computed the speed Normalized Root Mean Squared Error (NRMSE), defined as the speed RMSE normalized by the maximum speed amplitude, that allows to compare the predictions accuracy at sites with different currents speed magnitude. In addition, we computed the direction Mean Absolute Error (MAE) to evaluate the predictions of the speed vector orientation. While the RMSE has been

Table 1: Speed Normalized Root Mean Squared Error (NRMSE) and Direction Mean Absolute Error (MAE) for the Long Short-Term Memory model (LSTM) and the Transformer (TF) at various locations across the United States.

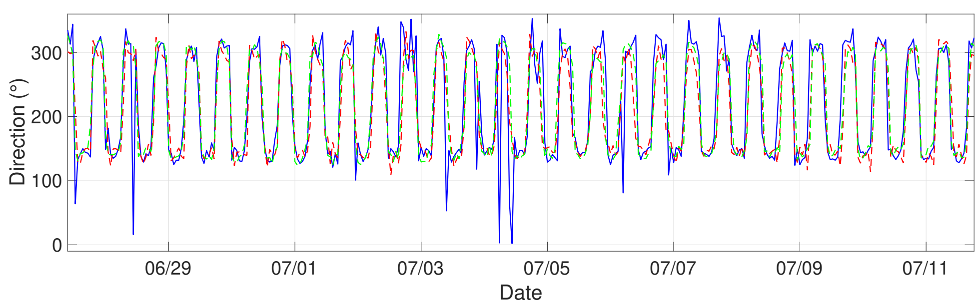
Station ID	Depth (ft)	Speed NRMSE		Direction MAE (°)	
		LSTM	TF	LSTM	TF
CAB1401	9	0.09	0.10	21	22
CAB1401	35	0.10	0.10	27	22
CHB9904	15	0.12	0.13	30	29
CHB9904	42	0.13	0.15	29	28
FPI0903	9	0.11	0.13	22	20
FPI0903	16	0.12	0.14	23	18
HUB0402	15	0.14	0.13	25	24
HUB0402	34	0.14	0.15	25	25
KOD0903	6	0.09	0.09	28	26
KOD0903	16	0.08	0.08	25	25
PEV0901	16	0.12	0.13	19	19
PEV0901	35	0.14	0.16	27	25
PIR0705	8	0.06	0.06	12	9
PIR0705	31	0.06	0.05	12	8
SEA0624	147	0.1	0.12	24	24
SEA0624	252	0.13	0.14	28	27
SFB1202	54	0.08	0.07	19	18
SFB1202	126	0.08	0.08	24	20
SFB1316	17	0.08	0.07	15	13
SFB1316	33	0.09	0.09	20	19
SJR9801	16	0.09	0.09	16	16
SJR9801	30	0.09	0.10	16	15
UNI1010	128	0.09	0.09	21	26
UNI1010	207	0.09	0.08	22	24

commonly used to evaluate the prediction accuracy of current speed, no standard error has been defined to evaluate prediction accuracy of the current direction. The MAE represents well the error observed on temporal comparisons because it is proportional to the error at each time step and is therefore presented here. Both errors are computed on an identical sample for each site and depth with the LSTM and the Transformer and are shown in Table 1 . The LSTM and Transformer have a global averaged NRMSE, defined as the average of the NRMSE for all sites, respectively of 0.100 and 0.105 with a corresponding standard deviation respectively of 0.024 and 0.031. In addition, the direction MAE of the two models remains lower than 30° for all depths and sites. As current direction follows a simple repeating pattern, this error can be easily corrected using a heuristic technique. The similar performance achieved by both models show that the Transformer's attention mechanism and the LSTM's gating mechanism are able to adequately regulate the flow of information for ocean currents data processing.

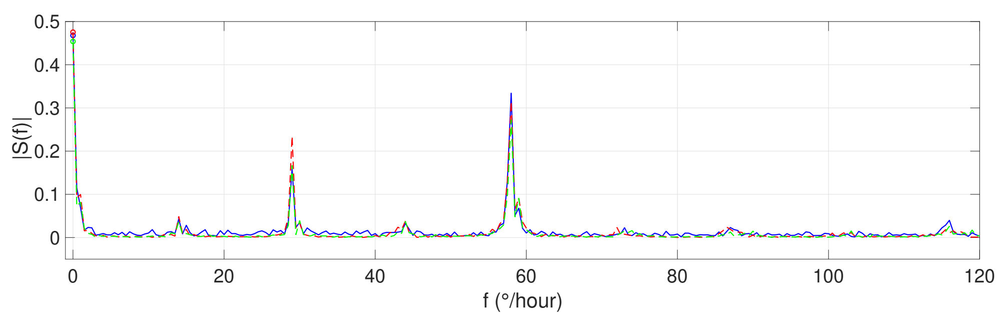
To better understand the predicting capabilities of the Transformer and of the LSTM, we show on Figure 8 the speed and direction as a function of time as well as the speed Fourier transform for station CAB1401 at a depth of 35 ft. Station CAB1401 is located at the harbor entrance of Portland, Maine. We chose this station to illustrate the performance of the models because station CAB1401 has speed NRMSE at a depth of 35 ft equal to the global averaged NRMSE for both models as shown in Table 1 and is therefore representative of the models accuracy. Speed Predictions for station CAB1401 match well with experimental measurements as shown on Figure 8a. The timing of the slack before flood, slack before ebb, maximum flood and ebb currents are well predicted as well as well as the alternating high and low tides but maximum flood and ebb currents speed shows some slight discrepancies. Direction predictions, shown on Figure 8b, match well with measurements and confirm that the timing of the currents variation is well captured. We observe that the regularization of the models, i.e. dropout for the Transformer, dropout and early stopping for the LSTM, prevents them from overfitting as they do not capture the spikes in the measured data that are probably due to data acquisition errors. We observe on the Fourier transform of the currents speed on Figure 8c that ocean currents at station CAB1401 are driven by long-period constituents around 0°/hour, semidiurnal constituents at 30°/hour as well as constituents generated by nonlinear mechanisms in shallow water at 60°/hour. The models are able to capture all the harmonic constituents up to 120°/hour which corresponds to the angular speed range covered by the 37 tidal harmonic constituents typically used in harmonic analysis [28].



(a) Ocean currents speed.



(b) Ocean currents direction.



(c) Spectral analysis of ocean currents speed.

Figure 8: Comparison of Transformer and LSTM predictions with experimental measurements at Portland Harbor Entrance (CAB1401) from June 12 to July 12, 2014. Depth: 35ft. Only the last two weeks of the comparisons are shown on figures (a) and (b) to facilitate reading.

3.2. Comparison with Harmonic Method on measurements from active stations

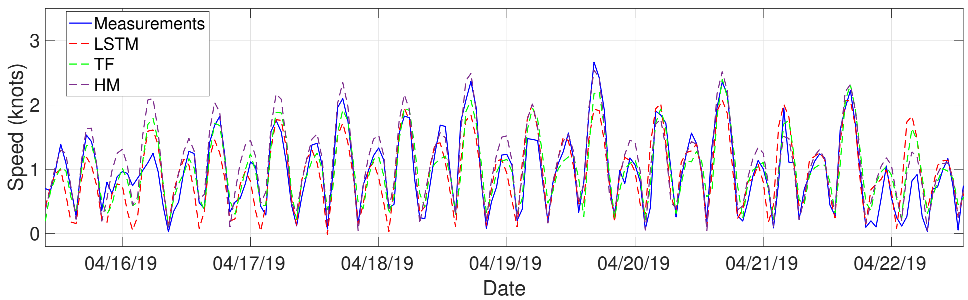
Harmonic method is one of the most popular approaches to modelling time series data, particularly to those whose visualization exposes some pe-

Table 2: Comparison of Speed Prediction Normalized Root Mean Squared Error (NRMSE) in knots for the Long Short-Term Memory model (LSTM), Transformer model and Harmonic Method (HM) in Active Stations.

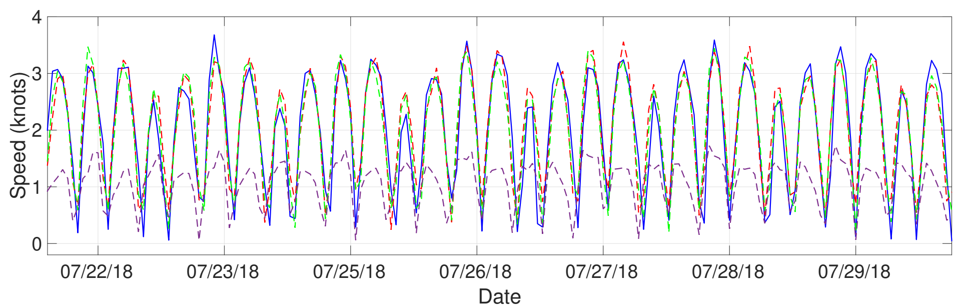
Station ID	Depth (ft)	Speed NRMSE		
		LSTM	TF	HM
cb0701	13.9	0.18	0.18	0.2
hb0401	15.7	0.14	0.12	0.12
jx0302	29.5	0.1	0.09	0.33
jx0701	15	0.09	0.10	0.12

riodic pattern. It is, along with ARIMA models, classified as of the classic methods as opposed to the modern ones using neural network, yet it is still widely used today. The essence of harmonic method is to find harmonic components that can be obtained by applying the Fourier transform to the selected data windows and filtering out the harmonic components with small weights. If the harmonic components are identified beforehand, the method is focused on determining their weights using, for instance, least square fit, and adjusting for the variations due to the non-harmonic factors. In modelling ocean currents, NOAA took the second approach. It is obvious that the harmonic patterns of ocean currents could be identified with specific frequencies since they are governed by the sun and the moon. NOAA applied different procedures, the choice of which is based on the available data volume, in order to determine the harmonic weights. The final models were subsequently adjusted for variations such as de-tiding the time series data.

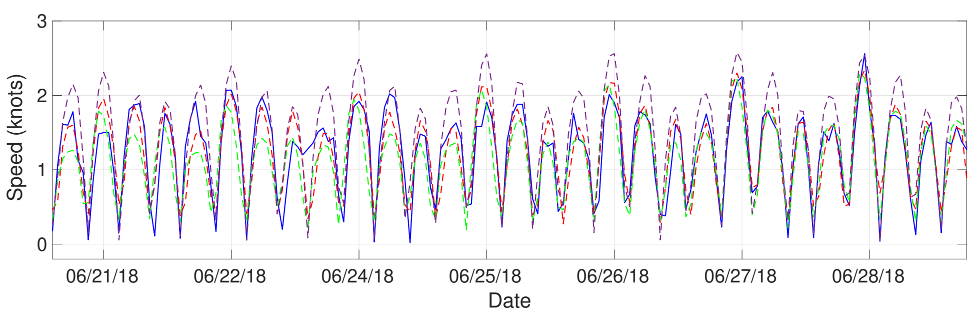
It is essential to note that each time series data requires a model that is a set of harmonic constants with their amplitudes. Thus, the number of models is corresponding to the number of stations in the ocean current modeling problem, which results in a huge workload. In specific, NOAA used the historic stations data to build the separate models for the corresponding sites. Depending on the amount of collected data in the stations, the models were built independently using different procedures. The predictions using these models are called harmonic predictions as they are based directly on the harmonic models, and these station are called reference station. Beside the harmonic prediction, the subordinate predictions are made for the subordinate stations close to the reference station but without the collected



(a) hb0401



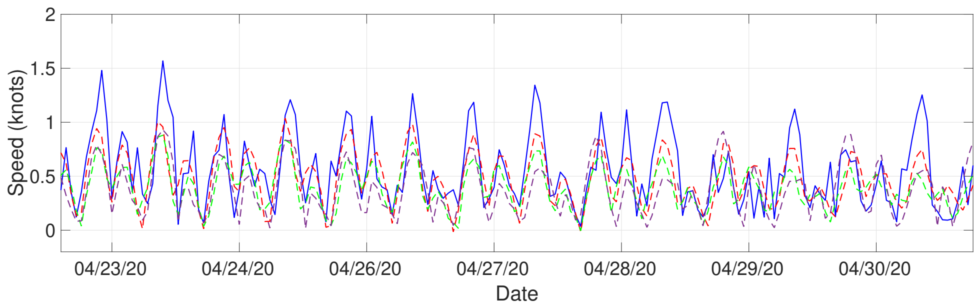
(b) jx0302



(c) jx0701

Figure 9: Comparison of Transformer and LSTM ocean currents speed predictions with experimental measurements and predictions obtained with Harmonic Method at various locations across the United States. Only one week of comparisons is shown to facilitate reading.

data. We are interested in the harmonic predictions as we want to compare our models performance to the harmonic model performance. Table 2 sum-



(d) cb0701

Figure 9: Comparison of Transformer and LSTM ocean currents speed predictions with experimental measurements and predictions obtained with Harmonic Method at various locations across the United States. Only one week of comparisons is shown to facilitate reading.

marizes the speed NRMSEs by LSTM, Transfomer, and harmonic methods, followed by Figure 9 visualizing their performances on speed prediction. The number of tests on active stations are limited since, although there are a total of 61 active stations as of this study's time, few ones have sufficient continuous data for testing purpose.

Table 2 shows that our models LSTM and Transformer outperform the harmonic model in 3 out of 4 active stations. Due to the limitation of active data, this may not be a fair comparison but it shows some interesting points regarding the prediction capacities of these models. As can be seen in Figure 9, all the models do a good job in catching the periodic time points when the flood and ebb currents occur. The harmonic method, however, fails to match the peak values in the stations jx0302 and jx0701 due to its inflexibility. That is, the harmonic method can identify well the current patterns but it is not adapting to the changes of current amplitudes due to its fixed harmonic constituents. The neural network-based methods, on the other hand, take into account the most recent information combined with the memorized patterns to produce the predictions that incorporate the details of long and short time. Thus, the proposed methods are more adapting to the change of currents over time.

It is interesting to notice that the neural network-based models perform better than the harmonic one in cb0701 and worse in hb0401 while the former station looks less patterned than the latter, i.e. more noisy. The authors do not have solid explanation for this. Our conjecture is that the proposed models are more capable of capturing minor details in time series. The reason for that is, when training the network-based models, we do not deliberately select patterns beforehand, the training procedure converges the network weights to embrace all major and minor patterns except the randomness. In the harmonic method, on the other hand, people do determine the harmonic periods before identifying it constituents so the number of patterns is limited.

4. Conclusions

We developed deep auto-regressive networks to predict ocean currents speed and direction at various locations across the United States. A LSTM Recurrent Neural Network and a Transformer, using attention mechanisms, have been modified and trained with the Historic Stations dataset from NOAA. Both models are able to predict ocean currents for one month at any site in the territorial sea of the United States using one month of measured data as input. They have similar performance and their accuracy is equivalent to the well known harmonic method. The LSTM is likely to be a better choice for future deployments because it is easier to train and more widely spread among the scientific and engineering community. These models provide a significant improvement compared to harmonic method where harmonic constituents need to be analyzed and computed for each site before predictions can be carried out. Notably, these models allow the real-time in-situ prediction of ocean currents for AUVs navigation. AUVs are expected to be widely deployed in the near future for seafloor mapping. Both models could, for instance, be used to support the seafloor mapping of the United States Exclusive Economic Zone (EEZ) and of the Alaskan coastline. The extension of the models to other regions in the world remain to be evaluated. Both models have been trained to learn the principal ocean currents variations and should therefore be able to provide, to some extent, predictions at any location in the world. Re-training of the models on a dataset representative of the regional ocean currents distribution should however be considered to fully exploit the predictive power of the neural networks.

5. Acknowledgement

This work was partially supported by the Link OE foundation; and by the National Science Foundation under Grant No. 1932595.

References

[1] R. N. Smith, Y. Chao, P. P. Li, D. A. Caron, B. H. Jones, G. S. Sukhatme, Planning and implementing trajectories for autonomous underwater vehicles to track evolving ocean processes based on predictions from a regional ocean model, The International Journal of Robotics Research 29 (2010) 1475–1497.

[2] Z. Zeng, H. Zhou, L. Lian, Exploiting ocean energy for improved auv persistent presence: path planning based on spatiotemporal current forecasts, Journal of Marine Science and Technology 25 (2020) 26–47.

[3] A. F. Shchepetkin, J. C. McWilliams, The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model, Ocean modelling 9 (2005) 347–404.

[4] P. J. Martin, Description of the navy coastal ocean model version 1.0, Technical Report, Naval Research Lab Stennis Space Center MS, 2000.

[5] Z. Yang, P. H. Richardson, R. A. Schmaltz Jr, Assessment of noaa water level, sea-surface temperature, and salinity guidance from the global real-time ocean forecast system (g-rtofs) and water level guidance from the extratropical storm surge (etss) system in western us coastal waters (2015).

[6] A. Immas, M.-R. Alam, Optimization of a swarm of autonomous underwater vehicles for high-bandwidth underwater wireless communication, in: International Conference on Offshore Mechanics and Arctic Engineering, volume 58851, American Society of Mechanical Engineers, 2019, p. V07BT06A053.

[7] B. Allotta, R. Costanzi, F. Fanelli, N. Monni, L. Paolucci, A. Ridolfi, Sea currents estimation during auv navigation using unscented kalman filter, IFAC-PapersOnLine 50 (2017) 13668–13673.

[8] P. Baccou, B. Jouvencel, V. Creuze, Single beacon acoustic for auv navigation, in: International Conference on Advanced Robotics, 2001, pp. 22–25.

[9] B. Garau, A. Alvarez, G. Oliver, Auv navigation through turbulent ocean environments supported by onboard h-adcp, in: Proceedings

2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006., IEEE, 2006, pp. 3556–3561.

[10] L. Medagoda, S. B. Williams, Model predictive control of an autonomous underwater vehicle in an in situ estimated water current profile, in: 2012 Oceans-Yeosu, IEEE, 2012, pp. 1–8.

[11] R. Pawlowicz, B. Beardsley, S. Lentz, Classical tidal harmonic analysis including error estimates in matlab using t_tide, Computers & Geosciences 28 (2002) 929–937.

[12] Y. Guozhen, W. Haifeng, Q. Hui, F. Jianming, Tidal current short-term prediction based on support vector regression, in: IOP Conference Series: Materials Science and Engineering, volume 199, IOP Publishing, 2017, p. 012024.

[13] A. Kavousi-Fard, W. Su, A combined prognostic model based on machine learning for tidal current prediction, IEEE Transactions on Geoscience and Remote Sensing 55 (2017) 3108–3114.

[14] P. Remya, R. Kumar, S. Basu, Forecasting tidal currents from tidal levels using genetic algorithm, Ocean engineering 40 (2012) 62–68.

[15] S. Dauji, M. C. Deo, K. Bhargava, Prediction of ocean currents with artificial neural networks, ISH Journal of Hydraulic Engineering 21 (2015) 14–27.

[16] C. Bayindir, Predicting the ocean currents using deep learning, arXiv preprint arXiv:1906.08066 (2019).

[17] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need (2017) 5998–6008.

[19] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, S. Valaee, Recent advances in recurrent neural networks, arXiv preprint arXiv:1801.01078 (2017).

[20] R. Al-Rfou, D. Choe, N. Constant, M. Guo, L. Jones, Character-level language modeling with deeper self-attention 33 (2019) 3159–3166.

[21] N. Do, M.-R. Alam, Data-based approach to optimizing the ocean wave energy carpet using deep neural network, Proceedings of the ASME 2020 39th International Conference on Ocean, Offshore and Arctic Engineering OMAE2020 (June 2020).

[22] C. Olah, Understanding LSTM Networks, 2015. URL: http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[23] M. Allard, What is a transformer?, 2019. URL: https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04.

[24] Co-ops current station data, 2020. URL: https://tidesandcurrents.noaa.gov/cdata/.

[25] D. P. Kingma, J. L. Ba, Adam: A method for stochastic optimization, in: ICLR, 2015.

[26] L. Bottou, Online algorithms and stochastic approximations, in: D. Saad (Ed.), Online Learning and Neural Networks, Cambridge University Press, Cambridge, UK, 1998. URL: http://leon.bottou.org/papers/bottou-98x, revised, oct 2012.

[27] L. Bottou, Stochastic gradient learning in neural networks, ATT Bell Laboratories, (1991).

[28] B. B. Parker, Tidal analysis and prediction., NOAA, NOS Center for Operational Oceanographic Products and Services (2007).