

Loop Current SSH Forecasting: A New Domain Partitioning Approach for a Machine Learning Model

Justin L. Wang ¹, Hanqi Zhuang ^{2,*}, Laurent Chérubin ³, Ali Muhamed Ali ² and Ali Ibrahim ²

¹ Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA; jlwang5@illinois.edu

² CEECS Department, Florida Atlantic University, Boca Raton, FL 33431, USA; amuhamedali2014@fau.edu (A.M.A.); aibrahim2014@fau.edu (A.I.)

³ Harbor Branch Oceanographic Institute, Florida Atlantic University, Boca Raton, FL 33431, USA; lcherubin@fau.edu

* Correspondence: zhuang@fau.edu

Abstract: A divide-and-conquer (DAC) machine learning approach was first proposed by Wang et al. to forecast the sea surface height (SSH) of the Loop Current System (LCS) in the Gulf of Mexico. In this DAC approach, the forecast domain was divided into non-overlapping partitions, each of which had their own prediction model. The full domain SSH prediction was recovered by interpolating the SSH across each partition boundaries. Although the original DAC model was able to predict the LCS evolution and eddy shedding more than two months and three months in advance, respectively, growing errors at the partition boundaries negatively affected the model forecasting skills. In the study herein, a new partitioning method, which consists of overlapping partitions is presented. The region of interest is divided into 50%-overlapping partitions. At each prediction step, the SSH value at each point is computed from overlapping partitions, which significantly reduces the occurrence of unrealistic SSH features at partition boundaries. This new approach led to a significant improvement of the overall model performance both in terms of features prediction such as the location of the LC eddy SSH contours but also in terms of event prediction, such as the LC ring separation. We observed an approximate 12% decrease in error over a 10-week prediction, and also show that this method can approximate the location and shedding of eddy Cameron better than the original DAC method.

Keywords: Gulf of Mexico; loop current; SSH; forecasting; eddy shedding; machine learning; divide-and-conquer; partitioning



Citation: Wang, J.L.; Zhuang, H.; Chérubin, L.; Muhamed Ali, A.; Ibrahim, A. Loop Current SSH Forecasting: A New Domain Partitioning Approach for a Machine Learning Model. *Forecasting* **2021**, *3*, 570–579. <https://doi.org/10.3390/forecast3030036>

Academic Editor: Sonia Leva

Received: 14 May 2021

Accepted: 11 August 2021

Published: 16 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Gulf of Mexico (GoM) is a semi-enclosed basin whose circulation is dominated by the Loop Current (LC), which sheds large anticyclonic eddies called Loop Current eddies (LCE) at varying time intervals [1]. These eddies drift westward and dissipate near the western boundary of the GoM. Predicting the LC and its eddy shedding is fundamental to many aspects of marine life and human activities in the GoM region, including natural disaster responses, short-term weather anomaly predictions, offshore oil and gas operations, and ecosystem services. The importance of this issue is stressed by a call to action in a report prepared by the National Academies of Sciences, Engineering, and Medicine [2]. The report recommends three tasks for researchers, one of which is to improve predictive skill in forecasting an eddy shedding event from an extended LC out to a forecast period of approximately three months.

In an effort to address this challenge, a deep learning approach to forecast the sea surface height (SSH) of the LC system was proposed in [3,4]. The problem was formulated as time sequence regression and prediction [5], and solved with a Recurrent Neural Network (RNN). RNN is a suitable tool that can learn from recurrent patterns and therefore realize predictions. Long Short-Term Memory (LSTM) networks, a type of RNN, were chosen

because of their capability of handling long-term dependencies contained in time series [6]. In order to reduce the computational costs associated with the handling of large datasets, the region of interest was partitioned into smaller non-overlapping sub-regions. In such an approach, the local features associated with the LC evolution and eddy shedding can be resolved. For each partition, an empirical orthogonal function (EOF) decomposition was applied to the SSH in order to reduce its dimension. At this point, each time series of principle components from each partition was independently predicted with local expert LSTM networks. After each prediction, these principle component time series were reconstructed into the original subregions, and the partitions were pieced back together to produce an overall prediction of the region's SSH. With this Divide and Conquer (DAC) method, Wang et al. [4] predicted the LC evolution and eddy shedding more than two and three months in advance, respectively. At every prediction step, the predicted SSH across the neighboring partitions were smoothed using an interpolation function described in [4]. While the DAC method was relatively effective in forecasting the eddy frontal positions, merging the SSH predictions across the partitions contributed to error propagation.

In this article, a new method to reduce partition boundary errors is presented. In digital signal processing, signal time windows are often chosen to overlap one another to improve the signal-to-noise ratio of digital filters. Similarly, in this study, the region of interest is partitioned into 50% overlapped sub-regions. The SSH value at each prediction point is obtained by a weighted average of the SSH values of the same point in the overlapping region of the partitions. This procedure avoids the progressive smoothing process across partition boundaries, which was implemented in the original DAC method.

The details of the methodology are given in Section 2. Results from the methods are presented in Section 3, and concluding remarks follow in Section 4.

2. Methodology

2.1. Dataset

The SSH used in this study was obtained from the HYCOM+CFRS 1/25 GoM 54-year Experiment (here- after GoM-HYCOM), which consists of 18 years of simulated SSH from 1992 to 2009 [7]. The dataset was split into approximately 15-, 1-, and 2-year periods for training, validation, and testing, respectively. Forecast experiments were conducted for approximately eighty 20-week sliding windows over the testing period.

2.2. Recurrent Learning

The prediction model used in this study is an RNN, which was developed to represent a time sequence (see Wang et al. [4] for details). RNNs work by feeding the output of each neuron, along with a new input, back into itself, forming loops within its architecture. In an RNN network, a simple RNN neuron or hidden unit's output behavior can be modeled by a recursion shown in Equation (1), where W_{g_i} represents the weights of the input, W_{g_s} the weights of the recurrence, f is an activation function, and x_n and s_n are the input and state at time n , respectively. Note that the output can be obtained from the state whenever it is needed.

$$s_n = f(W_{g_i}x_n + W_{g_s}s_{n-1}) \quad (1)$$

A problem associated with the RNNs given in Equation (1) is its gradient vanishing problem. This is because RNNs are typically trained with a gradient descent algorithm, and the gradients may vanish for a multi-layer RNN due to the chain rule in differentiation [8]. Long Short-Term Memory (LSTM) neural networks were designed to deal with this issue [6,9,10], in which a memory unit m_n was added to avoid the disappearance of gradients. Let α and β be constants and $\oplus$ denote an element-wise multiplication, then the memory unit is updated by the following rule:

$$m_n = \alpha \oplus m_{n-1} + \beta \oplus f(W_{g_i}x_{n-1} + W_{g_s}s_{n-1}) \quad (2)$$

The state is then related to the memory unit with an activation function. In this way, the derivatives will not vanish due to the additive relationship described in Equation (2). The overall prediction scheme is the same as the one used in [4], in which the Long Short-Term Memory (LSTM) network was paired with EOF decomposition to predict the evolution of the LC system using SSH data. A brief overview of the prediction method is given here (a detailed description can be found in Wang et al. [4]).

In each sub-domain, the EOF+LSTM model predicts SSH of the future weeks using all available data (see Figure 1). Let us denote the SSH data for week 1 to week n by $p_1, p_2, \dots, p_n$. To predict $\hat{p}_{n+1}$, thus the PCs of the SSH at week $n+1$, $p_1, p_2, \dots, p_n$ are first used to train the prediction model sequentially, so that it “learns” the evolution of the PCs. Then, the model can predict the PCs for week $n+1$, or $\hat{p}_{n+1}$. To predict the PCs for week $n+2$, or $\hat{p}_{n+2}$, $p_1, p_2, \dots, p_n, \hat{p}_{n+1}$ are then used to retrain the neural network. With the retrained neural network, the PCs for week $n+2$, thus $\hat{p}_{n+2}$ are predicted. This recursive process is thus continued for as long as the prediction errors remain within the range set for the forecast. In practice, new SSH measurements in the forecasting period can be added progressively to retrain the prediction model.

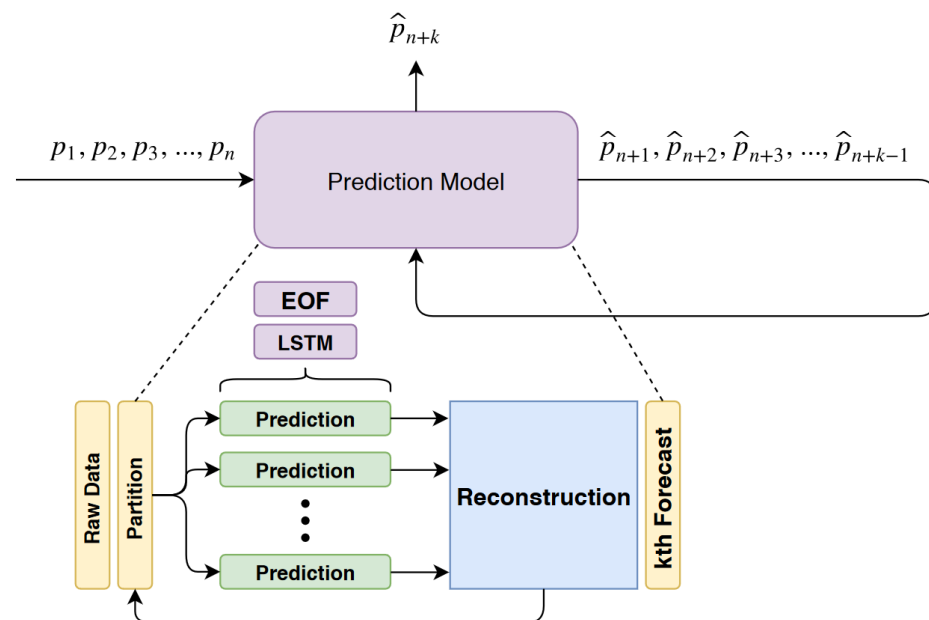


Figure 1. Flow chart of the recurrent learning algorithm used [4]. The figure above consists of two parts: the overarching abstract model and the implementation of the model with partitioning. The abstract model is essentially a dynamic prediction model, recursively predicting k steps ahead. The implementation detail shows the core components of the algorithm.

2.3. Divide and Conquer Prediction Model of the Gom SSH

A key step in the algorithm is to decompose the SSH data into temporal and spatial parts using the EOF method. Only the temporal series is forecasted at each time step by the prediction model. The method, however, requires an eigen value decomposition of the SSH matrices. Given a SSH two-dimensional field time series, it can be represented by a matrix X such that the rows are a representation of the temporal direction, and the columns are that of the spatial direction. The singular value decomposition of X is given below.

$$X = U\Sigma V^T \quad (3)$$

Here, matrix $U\Sigma$ represents the temporal principal components (PCs) and matrix V^T contains the spatial patterns (EOFs).

When the SSH dataset is large, the above matrix decomposition may result in memory overflow errors. In order to reduce the computational burden and to exploit the high

resolution of the SSH data, a divide-and-conquer strategy was proposed in [4]. The region of interest was first partitioned into a number of “non-overlapping” sub-regions. At each prediction step and for each partition, a “local” prediction is obtained from the EOF/LSTM algorithm. To reconstruct the SSH field over the whole prediction domain, a progressively weighted average was applied to the boundaries of each partition to smooth errors due to discontinuities across boundaries (Figure 1). However, these partition boundary errors contributed to the increase of model forecast errors over time as shown hereafter. For further details on the DAC approach, readers are referred to Wang et al. [4].

2.4. Sinusoidal-Weighted Overlapped Partitioning

In the original DAC method, a progressively weighted averaging procedure was used to smooth the SSH differences across every partition’s boundaries in an effort to remove discontinuities between partitions. However, this smoothing approach may interfere with the SSH dynamics across the region by removing bumps or troughs that would grow across the boundaries. The method proposed herein relies on removing discontinuities across boundaries that now overlap. SSH in the overlapping areas will be calculated as the combined solution of all overlapping partitions. This new approach is shown to preserve event prediction integrity and reduce error propagation.

For simplicity, we group 50% overlapped partitions into two groups such that no partition in the same group overlaps one another. These groups are referenced as the first- and second-level partitions, respectively. The first-level partitions have equal areas, cover the entire domain, and are denoted by f (blue partitions in Figure 2). The second-level partitions are superimposed on the joint boundary of the previous ones (yellow partitions in Figure 2) and denoted by g . The predicted SSH of overlapping partitions is averaged at each point with weights computed with a sinusoidal function of the normalized distance between the particular point and the center of the partitions f and g , respectively, labeled as $(C_{l,x,y})$. l represents the input layer (f layer or g layer) and (x, y) represents the index of the partition. Partitions h contain the final SSH state after merging of partitions f and g using the formula shown in Equation (5), which depends on the weights r_f and r_g , which are the distances calculated in Equation (4).

$$\forall (l, i, j) \in \text{Partition}_{l,x,y}, r_l = \text{dis}((i, j), C_{l,x,y}) \quad (4)$$

$$\begin{aligned} \forall (i, j) \in h, h(i, j) = & \left(1 - \sin \left(\frac{r_f}{r_f + r_g} \right) \right) * f(i, j) \\ & + \left(\sin \left(\frac{r_g}{r_f + r_g} \right) \right) * g(i, j) \end{aligned} \quad (5)$$

where $h(i, j)$ is the resulting SSH, and $f(i, j)$ and $g(i, j)$ are the respective SSH values in each overlapping partition.

At last, a median filter with a window size of 3×3 is applied to the h partitions to remove outliers. One may use a scheme with either more than or less than 50% overlapping. The advantage of using 50% overlapping is that we can construct two non-overlapping planes, as shown in Figure 2, so that for any point within the second (g) plane, two and only two values (one from each of the planes f and g) are obtained for reconstruction of each point in the output plane h .

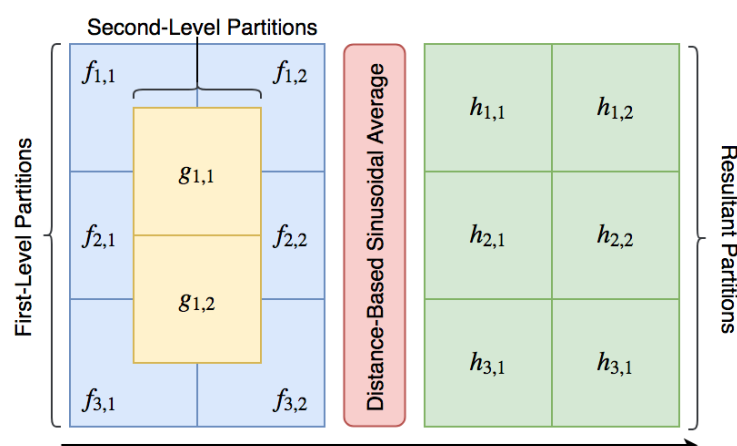


Figure 2. Graphical concept of overlapping partitions where l is the length of each partition. f and g denote the partition levels. h is the reconstructed partition from the application Equation (5) to f and g partitions.

3. LSTM Forecasting of the Gom SSH with Overlapping Partitions

The LSTM model was applied to the forecasting of the SSH during the 2-year testing period. Eighty 20-week forecasts were analyzed by using performance measure of the model skills defined in [4]. The forecasts of the overlapping partition method were compared to the forecasts made with no overlapping partitions. The effect of discontinuities between partitions of the SSH forecast can be seen in Figures 3 and 4.

We present a 3D plot (to stress the effect of the discontinuities on smooth fields) as well as a 2D plot which shows the affect the discontinuity may have on the LCS. Both these plots are snapshots of predictions 10 weeks ahead, so that the error propagation is large enough to see. We can clearly see in Figure 4 that the discontinuity alters not only the shape of the LC, but also alters its path. While the true LC and the overlapping model predict the current veering to the West, the non-overlapping partitions predict a pinch in the LC without spin. This shows that providing a more fine-grained partitioning method that allows for error corrections on the boundaries of the original non-overlapping partitions can aid long-term forecasting of eddy movement and shedding.

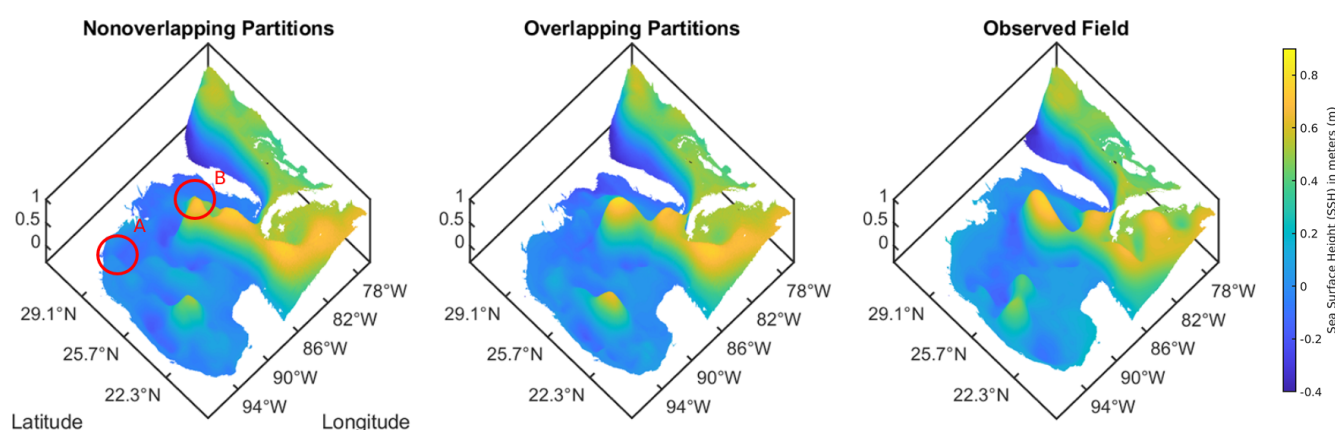


Figure 3. Example of discontinuity error effect on a 10-week SSH forecast. The left panel shows the SSH (m) prediction of the DAC model with no overlapping between partitions. The center panel shows the results of the DAC model with overlapping partitions. The right panel shows the HYCOM reference SSH. Red circles in the left panel shows two areas of interest, a trough (A) and a bump (B) where discontinuities have grown.

The Root Mean Square Error of the distances from the eddy front to seven reference points between the observed (HYCOM) and predicted reference field was calculated as

$$RMSE_{dF} = \sqrt{\sum_{i=1}^k \left(\frac{(dM_{i,n} - dO_{i,n})^2}{k} \right)} \quad (6)$$

where $dM_{i,n}$ and $dO_{i,n}$ denote the distance from the i th reference point at week n for the predicted and observed SSH, respectively. For more details on the selection of these reference points, the selection of contours, and the calculation of this metric, the reader is referred to Wang et al. [4] and Oey et al. [1]. For consistency with previous studies, the 0.45 m SSH contour was used [1]. Figure 5 shows a decrease by approximately 2.5 km in the first 5 weeks of the $RMSE_{dF}$ followed by a 5 km decrease in the following weeks. The overlapping partition model significantly improved the SSH prediction over the non-overlapping one along with the forecast duration. Overall, we can observe a 12% decrease in $RMSE_{dF}$ over a 10 week prediction.

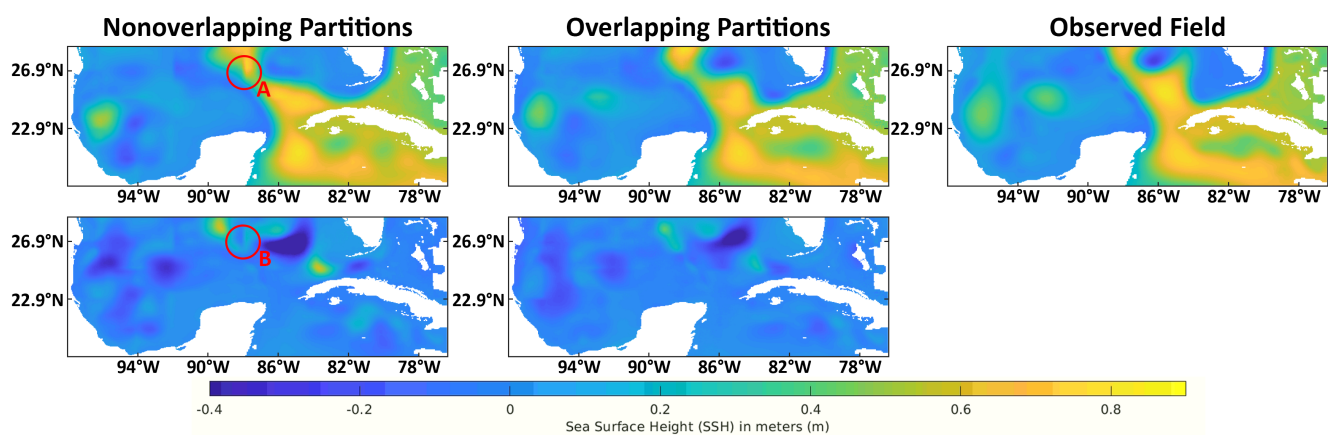


Figure 4. Another 10-week SSH forecast which shows that a discontinuity may contribute to misidentified shedding or significantly alter the predicted path of an eddy. The bottom row of the figure shows the difference plot between the prediction and ground truth observation. Note that in (A) we can clearly see that the information from the right-partition was unable to be properly relayed the left. As a result, in the error plot (B), we can see that there is a sharp valley signally that the discontinuity prevented what could have been an eddy formation event and significantly altering its future path.

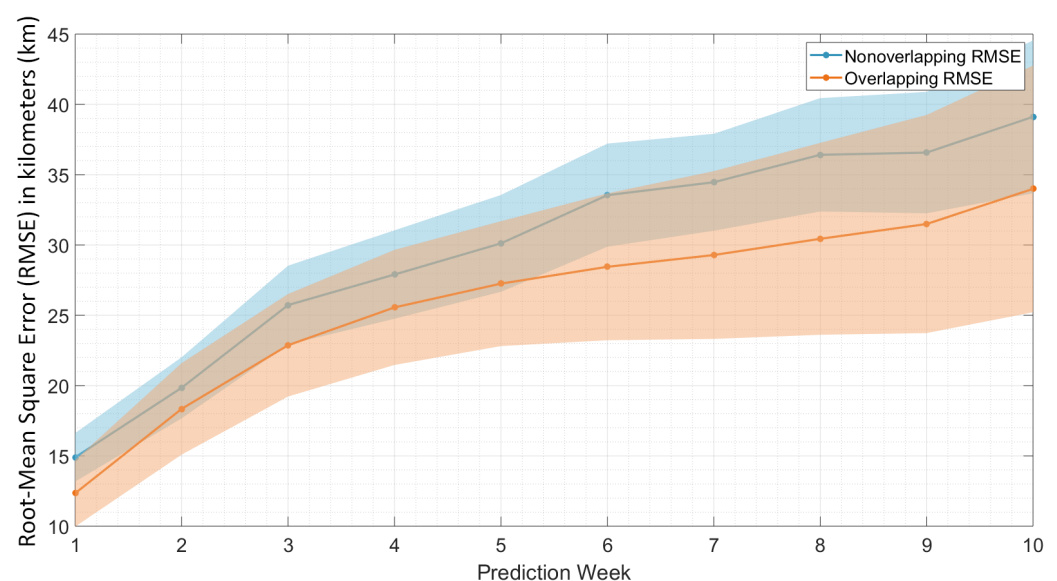


Figure 5. $RMSE_{dF}$ mean (solid line) and standard deviation (shaded area) for both the non-overlapping model (blue) and the overlapping (red) partition models. Note that while the resolution of our data is 1/25 or 4 km, a 5 km difference is still significant as it is a 12% improvement on average.

This improvement is noticeable in the SSH field of the LC shown in Figure 6, particularly for the forecast of eddy detachment and re-attachment. Week 7 and 10 SSH show the first detachment of eddy Cameron in May 2008. Based on the SSH only, the detachment is better predicted by the model with overlapping partitions than the original DAC model.

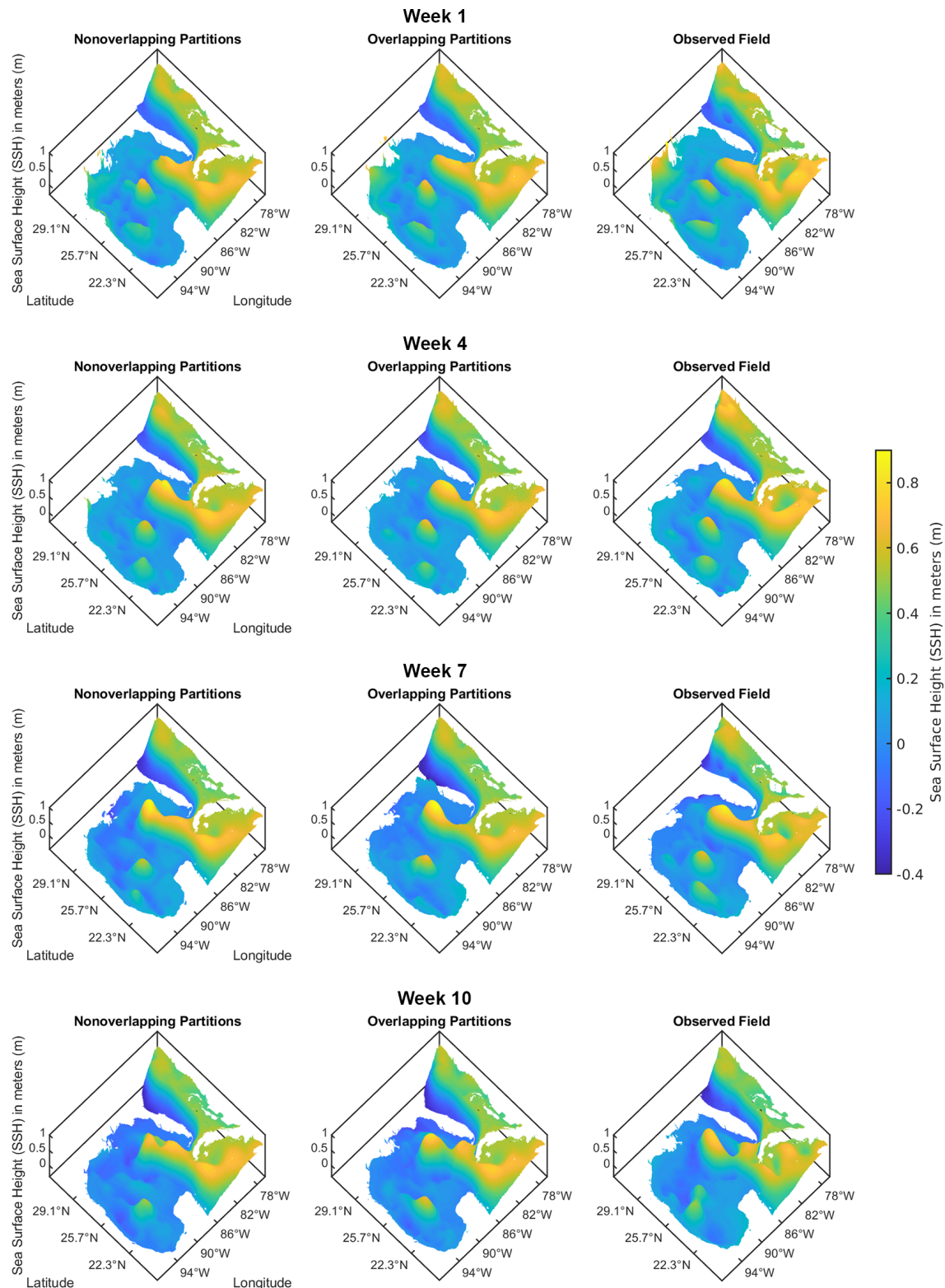


Figure 6. SSH weekly snapshots in meters of a 10-week prediction of the LC system SSH spanning March to May of 2008. Predicted SSH on Week 1, 4, 7, and 10 are shown for the model with no overlapping boundaries (left column), with overlapping boundaries (center column), and the observed (HYCOM) SSH is shown for the same weeks in the right column.

To concretely show the advantage of using this improved partitioning scheme, we revisit one of the original experiments presented in Wang et al. [4], specifically the prediction of eddy Cameron eight weeks in advance. On this eight-week horizon, the non-overlapping scheme works to perfectly capture and predict the timing/dynamics of eddy Cameron. However, on a longer horizon, its performance decays significantly. While eddy Darwin was able to be predicted 10 weeks in advance, the dynamics of Cameron were shown to be more complex. Thus, to show the decay of performance as well as the strength of the overlapping partitions scheme, we present Figure 7. Originally, Wang et al. found a 10-week prediction of eddy Cameron to be too erroneous, and thus presented an eight week prediction instead. With this new overlapping method, we can clearly extend the predictability of eddy Cameron a couple of weeks ahead.

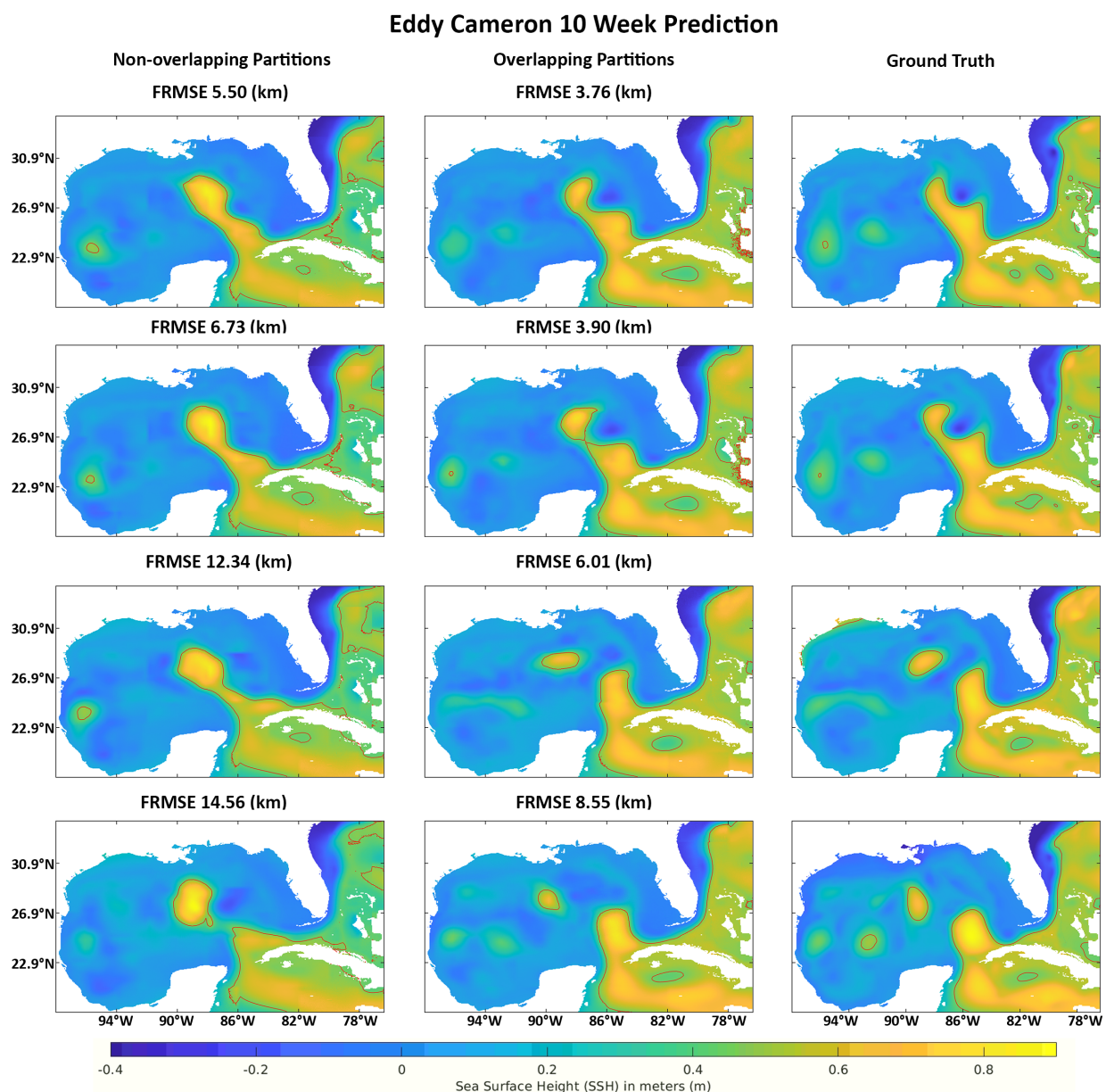


Figure 7. Presents the 10 week prediction of eddy Cameron with the non-overlapping DAC algorithm (**left**), overlapping DAC algorithm (**middle**), and ground truth (**right**). These results show that the original DAC algorithm, in a 10-week-ahead prediction does not capture the proper timing of the eddy shedding event of eddy Cameron. Notice the significant discontinuities that arise on the field as well, showing that perhaps errors in the pinched area may be responsible for misinforming the timing of the shedding. We also provide a 0.45m contour to better show the location of the eddy front.

4. Conclusions

In this study, we applied a RNN model with a memory unit and an activation function, called LSTM, to the prediction of the LC system's SSH. In order to forecast the GoM domain, a DAC approach was used. The prediction domain was initially divided into multiple non-overlapping partitions and a simple smoothing function was used to enforce continuity of the SSH forecast at the partition boundaries. This model, implemented in [4], exhibited significant error growth at partition boundaries. In order to mitigate this DAC model error, we proposed in the study herein a new prediction domain partition approach in which the partitions are overlapping. The SSH in the overlapping regions was calculated as a sinusoidal weighted function of the SSH in each overlapping partitions. This new approach led to a significant improvement of the overall model performance both in terms of features prediction, such as the location of the LC eddy SSH contours, but also in terms of event prediction, such as the LC ring separation as shown in Figure 6. Although the overlapping method can be simply implemented, it is slightly more computationally expensive than its non-overlapping counterpart. Moreover, the sinusoidal weight function can also be swapped for other types of weight functions that could further improve the SSH prediction.

Author Contributions: Conceptualization, J.L.W. and H.Z.; methodology, J.L.W., H.Z., A.M.A. and A.I.; software, J.L.W.; validation, J.L.W.; formal analysis, J.L.W.; investigation, J.L.W.; resources, H.Z.; data curation, A.I.; writing—original draft preparation, J.L.W.; writing—review and editing, J.L.W., H.Z. and L.C.; visualization, J.L.W.; supervision, H.Z.; project administration, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the NSF MRI grant #1828181, which provided computational equipment for this research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this research was downloaded from HYCOM: <https://www.hycom.org/> (accessed on 14 May 2021).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DAC	Divide and Conquer
SSH	Sea Surface Height
LC	Loop Current
LCS	Loop Current System
GoM	Gulf of Mexico
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory Network
PC	Principal Components
EOF	Empirical Orthogonal Functions
RMSE	Root Mean Squared Error

References

1. Oey, L.Y.; Ezer, T.; Forristall, G.; Cooper, C.; DiMarco, S.; Fan, S. An exercise in forecasting loop current and eddy frontal positions in the Gulf of Mexico. *Geophys. Res. Lett.* **2005**, *32*. [[CrossRef](#)]
2. National Academies of Sciences Engineering, Medicine. *Understanding and Predicting the Gulf of Mexico Loop Current: Critical Gaps and Recommendations*; The National Academies Press: Washington, DC, USA, 2018. [[CrossRef](#)]
3. Ali, A.M.; Zhuang, H.; Ibrahim, A.K.; Wang, J.L. Preliminary results of forecasting of the loop current system in Gulf of Mexico using robust principal component analysis. In Proceedings of the 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, 6 December 2018; pp. 1–5.

4. Wang, J.L.; Zhuang, H.; Chérubin, L.M.; Ibrahim, A.K.; Muhamed Ali, A. Medium-Term Forecasting of Loop Current Eddy Cameron and Eddy Darwin Formation in the Gulf of Mexico With a Divide-and-Conquer Machine Learning Approach. *J. Geophys. Res. Ocean.* **2019**, *124*, 5586–5606. [[CrossRef](#)]
5. Zeng, X.; Li, Y.; He, R. Predictability of the Loop Current Variation and Eddy Shedding Process in the Gulf of Mexico Using an Artificial Neural Network Approach. *J. Atmos. Ocean. Technol.* **2015**, *32*, 1098–1111. [[CrossRef](#)]
6. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
7. Dukhovskoy, D.S.; Leben, R.R.; Chassignet, E.P.; Hall, C.A.; Morey, S.L.; Nedbor-Gross, R. Characterization of the uncertainty of loop current metrics using a multidecadal numerical simulation and altimeter observations. *Deep. Sea Res. Part Oceanogr. Res. Pap.* **2015**, *100*, 140–158. [[CrossRef](#)]
8. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [[CrossRef](#)]
9. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: continual prediction with LSTM. In Proceedings of the 1999 Ninth International Conference on Artificial Neural Networks ICANN 99, Edinburgh, UK, 7–10 September 1999; Volume 2, pp. 850–855. [[CrossRef](#)]
10. Pulver, A.; Lyu, S. LSTM with working memory. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14 May 2017; pp. 845–851. [[CrossRef](#)]