



Physics-Informed Tensor-Train ConvLSTM for Volumetric Velocity Forecasting of the Loop Current

Yu Huang¹, Yufei Tang¹, Hanqi Zhuang^{1*}, James VanZwieten² and Laurent Cherubin³

¹ Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, United States,

² Department of Civil, Environmental and Geomatics Engineering, Florida Atlantic University, Boca Raton, FL, United States,

³ Ocean Dynamics and Modeling, Harbor Branch Oceanographic Institute, Florida Atlantic University, Fort Pierce, FL, United States

OPEN ACCESS

Edited by:

Ognjen Arandjelovic,
University of St Andrews, United
Kingdom

Reviewed by:

Qian Du,
GNS Healthcare, United States
Yuxin He,
Shenzhen Technology University,
China

*Correspondence:

Hanqi Zhuang
zhuang@fau.edu

Specialty section:

This article was submitted to
Machine Learning and Artificial
Intelligence,
a section of the journal
Frontiers in Artificial Intelligence

Received: 20 September 2012

Accepted: 06 December 2021

Published: 24 December 2021

Citation:

Huang Y, Tang Y, Zhuang H,
VanZwieten J and Cherubin L (2021)
Physics-Informed Tensor-Train
ConvLSTM for Volumetric Velocity
Forecasting of the Loop Current.
Front. Artif. Intell. 4:780271.
doi: 10.3389/frai.2021.780271

According to the National Academies, a week long forecast of velocity, vertical structure, and duration of the Loop Current (LC) and its eddies at a given location is a critical step toward understanding their effects on the gulf ecosystems as well as toward anticipating and mitigating the outcomes of anthropogenic and natural disasters in the Gulf of Mexico (GoM). However, creating such a forecast has remained a challenging problem since LC behavior is dominated by dynamic processes across multiple time and spatial scales not resolved at once by conventional numerical models. In this paper, building on the foundation of spatiotemporal predictive learning in video prediction, we develop a physics informed deep learning based prediction model called—Physics-informed Tensor-train ConvLSTM (PITT-ConvLSTM)—for forecasting 3D geo-spatiotemporal sequences. Specifically, we propose (1) a novel 4D higher-order recurrent neural network with empirical orthogonal function analysis to capture the hidden uncorrelated patterns of each hierarchy, (2) a convolutional tensor-train decomposition to capture higher-order space-time correlations, and (3) a mechanism that incorporates prior physics from domain experts by informing the learning in latent space. The advantage of our proposed approach is clear: constrained by the law of physics, the prediction model simultaneously learns good representations for frame dependencies (both short-term and long-term high-level dependency) and inter-hierarchical relations within each time frame. Experiments on geo-spatiotemporal data collected from the GoM demonstrate that the PITT-ConvLSTM model can successfully forecast the volumetric velocity of the LC and its eddies for a period greater than 1 week.

Keywords: physics-informed deep learning, time series forecasting, spatiotemporal predictive modeling, loop current, ocean current modeling, volumetric velocity prediction

1. INTRODUCTION

As part of the North Atlantic western boundary current system, the Loop Current (LC) originates as the Yucatan Current, flows northward into the Gulf of Mexico (GoM), and veers east to exit through the Straits of Florida where it becomes the Florida Current, which then transitions into the Gulf Stream system. The LC dynamics is characterized by the shedding of clockwise (anticyclonic) rotating eddies that move westward further into the GoM. The LC and its eddies together are called the Loop Current System (LCS). On April 20, 2010, the oil drilling rig Deepwater Horizon,

operating in the Macondo Prospect in the GoM, exploded and sank, resulting in 4 million barrels of oil gushing uncontrolled into the Gulf over an 87-day period, before it was finally capped on July 15, 2010 (The Commission, 2011). This disaster, also known as the GoM Oil Spill, threatened livelihoods, precious habitats, and even a unique way of life in the GoM region. Since the disaster, a large amount of research reported in the literature has improved scientific understanding and forecasting of the LC with aspirations for alleviating impacts on ecosystems and for future oil-spill prevention. The ability to predict the LC evolution is critical and fundamental to almost all aspects of the GoM community, including (1) anthropogenic and natural disaster response, (2) the prediction of short-term weather anomalies, and hurricane intensity and trajectories, (3) national security and safety, and (4) ecosystem services (Walker et al., 2009). Due to a reinforcing interaction between seasonal hurricanes and the LCS, long-term prediction of the LCS states is of particular interest and becoming increasingly relevant for mitigating potential environmental and ecological threats.

Modeling and forecasting the dynamics of the LC in the GoM remains an open scientific problem. The National Academies of Science, Engineering and Medicine (NASEM) published a report (National Academies Press, 2018) specifically calling for the development of models that are capable of forecasting (a) current speed, vertical structure, and duration of the LC and its eddies at a given location one week in advance, (b) LC propagation one month ahead, and (c) eddy shedding events up to 13 weeks ahead. Current prediction models fall mostly into one category which consists of numerical model-based methods. They primarily use finite-differences or finite-element techniques to discretize partial differential equations (PDEs) of momentum and tracer transport, and conservation. Although the various models used differ in their details, features such as the LC eddy shedding periods, eddy propagation, flow transport, and vertical structure in the LC and the formation of LC frontal cyclonic eddies are relatively similar. Some of these models use data assimilation techniques to constrain their model solution with remote sensing and *in-situ* observations (see Wang et al., 2019, 2021 for a literature review of the models). However, despite the data constraint the reliable prediction window for surface and subsurface currents of the LCS is from a few hours to 2 days at most (Wang et al., 2021).

Therefore, the long-term forecasting of the volumetric velocity has remained an unattainable goal despite the large amount of research focused on the LC dynamics. The major challenges of the LC velocity field prediction are found in the nature of the LC current dynamics which is dominated by instabilities and multi-scale interactions in the temporal, spatial and frequency domain (Yang et al., 2020). These interactions occur not only in the horizontal but also in the vertical dimension, and are controlled by the vertical discretization scheme chosen in the numerical ocean model. The choice of the vertical coordinate system has a significant impact of the physics and processes resolved by the model, which results in diverging solution between models, among other factors (Halliwell, 2004; Bruciaferri et al., 2018).

In this paper, we develop a novel technique to predict the evolution of this type of geophysical dynamical system, named the Physics-informed Tensor-train ConvLSTM

(PITT-ConvLSTM) model, that enables 4D spatiotemporal LC prediction, namely the prediction of a continuous sequence of 3D flow fields (see section 4.1). We address the problem of modeling the LCS based on convolutional Long Short Term Memory (LSTM) networks (ConvLSTM) by incorporating prior domain knowledge described as non-linear differential dynamic equations. Our PITT-ConvLSTM model can leverage required physics explicitly, and learn implicit patterns from data. Moreover, unlike most of the first-order ConvLSTM-based approaches, we propose a higher-order generalization to ConvLSTM with convolutional tensor-train decomposition (Su et al., 2020) to effectively learn the long-term spatiotemporal structure in the LCS.

The remainder of this paper is organized as follow. In section 2, we discuss the motivation of proposing physics informed neural network for LC forecasting. Section 3 presents a gallery of related works. Section 4 describes the dataset and its pre-processing method. Section 5 describes the proposed framework and our main results including the baselines are presented in section 6. Finally, concluding remarks are made in section 7.

2. MOTIVATION

In the past decades, tremendous progress has been made in understanding multiscale physics in diverse geophysics applications, by numerically solving partial differential equations (PDEs) using finite differences, finite elements, spectral, etc. Such numerical model-based approaches still cannot seamlessly incorporate noisy data into existing algorithms, mesh generation remains complex, and high-dimensional problems governed by parameterized PDEs cannot be tackled (Karniadakis et al., 2021), showing limitation in solving real-world physical problem. Machine learning has emerged as a promising alternative. They are purposed to discover complex patterns from large volumes of data. However, training deep neural networks requires big data, while the volume of useful experimental/real-world data for complex physical systems is limited. For instance, in our case, gaps remain in oceanographic observations necessary to further understanding and prediction of the LCS, as the volumetric velocity data of the LCS is inadequate compare to other data domains (Hamilton et al., 2016).

Most machine-learning-based approaches are currently unable to extract interpretable information and knowledge from data. Moreover, machine-learning-based approaches, as a purely data-driven method in nature, are easy to be overfitting. Predictions may be physically inconsistent or implausible, owing to extrapolation or observational biases that may lead to poor generalization performance (Karniadakis et al., 2021). While numerical model-based approaches are generally limited by their high complexity and incompleteness, they do not require large amounts of data and retain interpretation. To this end, incorporating prior physics knowledge into machine learning can help enhance generalization and interpretability, and reduce the need for massive amounts of training data. In general, physics can be used as “teachers” to guide the discovery of meaningful and explainable machine learning models. It is hypothesized

that the combination of machine learning and physics could lead to performance gains by leveraging the advantages of both sides.

Machine learning models can be trained from additional information obtained by enforcing the law of physics (Raissi et al., 2019). By doing so, such physics-informed neural networks can (1) provide theoretical guidance that the model is required to follow and thus, facilitate the model training with fewer labeled data; and (2) improve the model robustness for unseen data and prevent over-fitting. Considering the difficulty in identifying which observations are following the physics explicitly and how much in accordance to it, the law of physics will not be directly applied to the raw observations, but rather to the latent space for representation learning (e.g., the hidden states of the deep neural networks). Therefore, the neural networks would remain flexible to extract unknown patterns as well, instead of being strictly constrained to the knowledge of physics.

3. RELATED WORK

3.1. Physics-Informed Neural Networks

Physics is one of the fundamental pillars that can explain how nature behaves. Because it is unable to describe all the rules governing real-world events, machine learning is desired to bridge the known physics and real-world observations. Physics-informed machine learning integrates data and mathematical models, even in partially understood, uncertain, and high-dimensional contexts (Karniadakis et al., 2021). In addition, (Raissi et al., 2017; Raissi, 2018; Raissi and Karniadakis, 2018) demonstrated that neural networks are capable of finding solutions to partial differential equations (PDEs), which enables us to obtain fully differentiable physics-informed models with respect to all input coordinates and free parameters. (de Bezenac et al., 2019) demonstrated how fluid physics could be incorporated for forecasting sea surface temperature, and such a method not only captures the dominant physics but also infers unknown patterns by CNNs. Further, a number of physics-informed models were developed, assuming that neural networks can learn complex dynamic interactions and simulate unseen dynamics based on current states (Kipf et al., 2018; Sanchez-Gonzalez et al., 2018). Unlike those works that implicitly extract latent patterns from data only, the physics-informed graph networks proposed in Seo and Liu (2019) allow incorporating known physics and simultaneously extracting latent patterns in data that is unable to be captured by numerical model-based approaches.

3.2. Recurrent Neural Network With Tensor-Train Decomposition

Tensor-train decomposition (TTD) factorizes model parameters into smaller tensors (Oseledets, 2011) in order to reduce the space dimension. This decomposition is especially beneficial in multi-relational data analysis, and has been widely used in machine learning, including CNNs (Su et al., 2018, 2020; Kolbeinsson et al., 2019), RNNs (Yang et al., 2017; Su et al., 2020), and transformers (Ma et al., 2019). A multiplicative

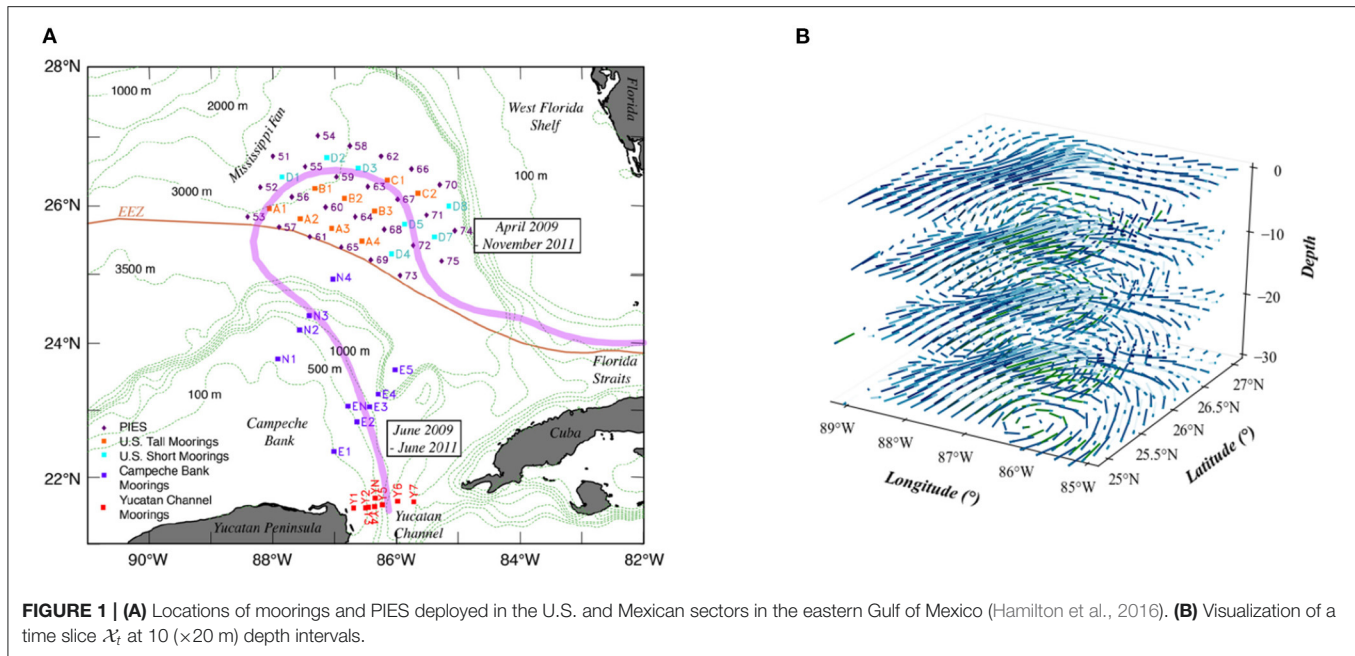
RNN with factorized (weights) tensors was proposed for input-state interactions (Sutskever et al., 2011). Further, (Yang et al., 2017) factorized the input-to-hidden weights within each cell by TTD, and showed improvement in video classification. Different from the first-order RNNs mentioned above, higher-order RNNs (Soltani and Jiang, 2016) require many more parameters since they introduce connections across multiple previous time steps for better long-term dynamics learning. This was improved by TTD-RNNs (Yu et al., 2018), whose higher-order structures within each cell were compressed by TTD, which was shown to improve the model performance in video classification. Recently, (Su et al., 2020) introduced a new convolutional TTD and constructed Convolutional Tensor-Train LSTM that was able to capture higher-order spatiotemporal correlations.

4. DATASET

4.1. Data Description

The dataset used in this study was produced by a funded observational program called "Dynamics of the Loop Current in US Waters Study" (Hamilton et al., 2016) and consisted of a high-density array of moored instruments over a two-and-a-half-year period, beginning in April 2009. The observational array (**Figure 1A**) collected LC velocity and density structure over the region where the LC extended northward and, more importantly, where eddy shedding events occurred most often. This sensor array consisted of nine tall-moorings, seven short-moorings, 25 pressure-equipped inverted echo sounders (PIES), and remote sensors. Each of the nine tall-moorings recorded water velocities using an upward facing 75 kHz acoustic Doppler current profiler (ADCP) deployed at an approximate depth of 450 m. Additionally, point current meters were attached to each tall mooring at five depths ranging from around 600 to 3,000 m. Collectively, these instruments provided high resolution water velocity data from about 60 to 440 m depths and at much lower resolution below this. Each of the seven short-moorings provided near bottom water velocity data using a single point current sensor deployed 100 m above the sea floor. PIES sensors measured both pressure and acoustic round trip travel times from the sea floor to the sea surface. Vertical profiles of density, salinity, and temperature were created from these data. The PIES pressure records were also combined with estimated horizontal density gradients to calculate geostrophic water velocities. The complete procedure for producing these mapped velocity fields was described in Donohue et al. (2016b) and Hamilton et al. (2016).

The utilized dataset, summarized in **Table 1**, contains velocity data gathered from April 2009 to June 2011. The southwest edge of this array was just to the US side of the exclusive economic zone (EEZ) from 25°50.2'N; 88°24.5'W to 24°59.2'N; 85°56.8'W, with the northeast edge of this array ~160 km to the US side of the EEZ. The horizontal separation between moorings was around 50–80 km and between the PIES sensors was around 40–50 km. Sampling frequency from the multiple sensors varied from minutes to hours. In this study, we utilize a time series which was processed with a fourth order Butterworth filter and sub-sampled at 12-h intervals. Consequently, the dataset contains a total of

**TABLE 1 |** Statistics of ocean current dataset.

Time range	Sampling rate	Spatial range	Dimensions				
			Depth	Length	Width	Channel	Steps (T)
April 2009 to June 2011	12 h	25°50.2'N, 88°24.5'W to 24°59.2'N, 85°56.8'W	30 (D)	29 (H)	36 (W)	2 (C)	1810 (T)

1,810 records, covering 905 days. At each time step t , as shown in **Figure 1B**, the data $\mathcal{X}_t$ were formatted as $\mathcal{X}_t \subseteq \mathbb{R}^{D \times H \times W \times C}$, where each dimension represents *Depth* (30), *Longitude* (29), *Latitude* (36), and *Channel* (2), respectively. *Channel* stands for the geostrophic velocity vector (u, v) . The first 80% (June 2009 to December 2010) of the dataset was reserved for training and the remaining 20% (January 2011 to June 2011) for prediction and validation (Shi et al., 2010).

4.2. Data Preprocessing

Empirical Orthogonal Function (EOF) analysis has been extensively used in oceanic and atmospheric sciences. In addition to its ability to decompose a time series into its temporal and spatial components, EOF can drastically reduce the dimension while preserving data integrity. Considering the non-linear dynamics and high dimension characteristics of oceanic phenomena in this work, the EOF analysis is employed to represent spatial patterns (*EOFs*) and temporal components (*Principal Components, PCs*) of the LCS in the GoM region. Specifically, the EOF analysis was conducted to extract the *PCs* of zonal and meridional velocity of ocean current. The *PCs* may contain otherwise hidden and medium-term uncorrelated

patterns (Navarra and Simoncini, 2010). Further, *PCs* are used to train the PITT-ConvLSTM prediction model, with details given in section 5.

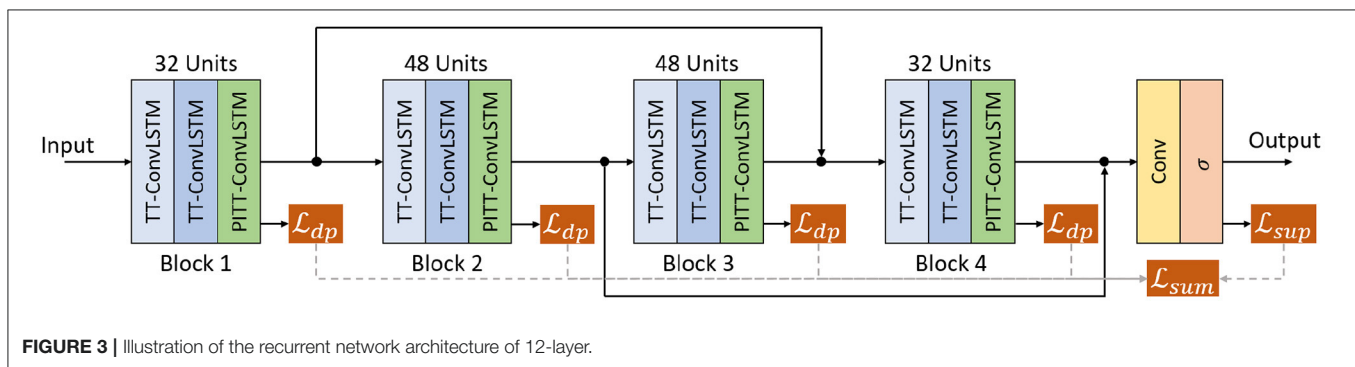
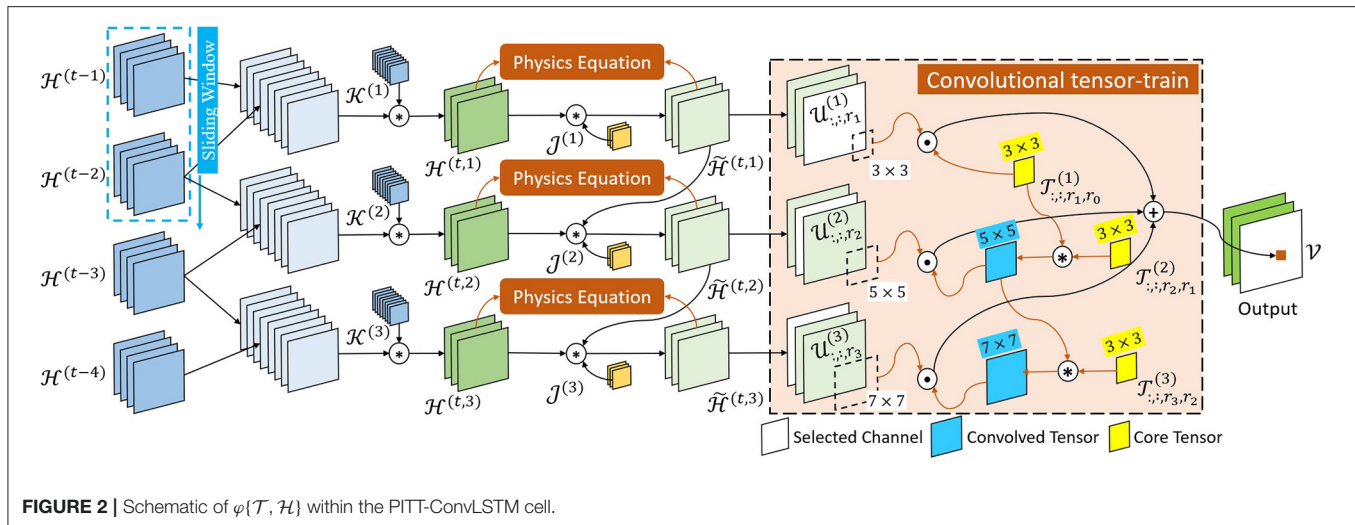
Singular Value Decomposition (SVD) is a mathematical tool to determine both the *EOFs* and *PCs* simultaneously. For the convenience of ocean modeling, the ocean is sliced into layers in the depth direction. Here, SVD is applied to each depth slice S to obtain two unitary orthogonal matrices (U and V) and one diagonal eigenvalue matrix Σ . Consequently, $U\Sigma$ represents the temporal *PCs* and V^T the spatial *EOFs*. Thus, S can be represented as:

$$S = U\Sigma V^T$$

$$\Leftrightarrow \begin{bmatrix} s_{1,1} & \cdots & s_{1,m} \\ \vdots & \ddots & \vdots \\ s_{n,1} & \cdots & s_{n,m} \end{bmatrix} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,m} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,m} \end{bmatrix} \begin{bmatrix} e_{1,1} & \cdots & e_{1,m} \\ \vdots & \ddots & \vdots \\ e_{m,1} & \cdots & e_{m,m} \end{bmatrix} \quad (1)$$

$\underbrace{\hspace{10em}}_S \qquad \underbrace{\hspace{10em}}_{PCs} \qquad \underbrace{\hspace{10em}}_{EOFs}$

After the EOF analysis, we concatenate the flattened *PCs* along the depth dimension. In this way, the raw data $\mathcal{X} \subseteq \mathbb{R}^{T \times D \times H \times W \times C}$ is compressed to $\mathcal{X}' \subseteq \mathbb{R}^{T \times D \times P \times C}$, where P is the number of selected elements in *PCs* (we utilized the first 50 *PCs* that contained over 95% of the energy). The 3D volume sequence prediction $\{\mathcal{X}^{(t)}\}_{t=1}^T$, in consequence, is compressed to a 2D matrix sequence prediction $\{\mathcal{X}'^{(t)}\}_{t=1}^T$. Each row of the 2D matrix $\mathcal{X}'^{(t)}(i, :)$ (flattened *PCs*) represents a depth slice of the original 3D volume $\mathcal{X}^{(t)}(i, :, :)$. The model in section 5 will be trained to predict future *PCs*, which could also be used to reconstruct predicted 3D volume by simply multiplying them by the known *EOFs* (constant) as shown in Equation (1).



5. METHOD

5.1. Method Overview

Convolutional LSTM network (ConvLSTM), a basic building block for sequence-to-sequence prediction, demonstrated promising performance in video forecasting. In ConvLSTM, the spatial information is encoded explicitly as tensors in the LSTM cells, where each cell is a first-order Markovian model (i.e., the hidden states are updated based on their adjacent step). Since ConvLSTM is successful in modeling complex behaviors and extracting abstract features through real-world data (Xingjian et al., 2015), it is natural to explore how such a predictive model can be used to solve practical problems in geo-spatiotemporal modeling domains with higher-order dynamics.

In this paper, we propose a novel model Physics-informed Tensor-train ConvLSTM (PITT-ConvLSTM), a ConvLSTM network under physics constraints. The ConvLSTM is integrated with convolutional tensor-train to model higher-order spatiotemporal correlations explicitly, with its hidden states incorporated with prior physics. The proposed model is illustrated in **Figures 2, 3**.

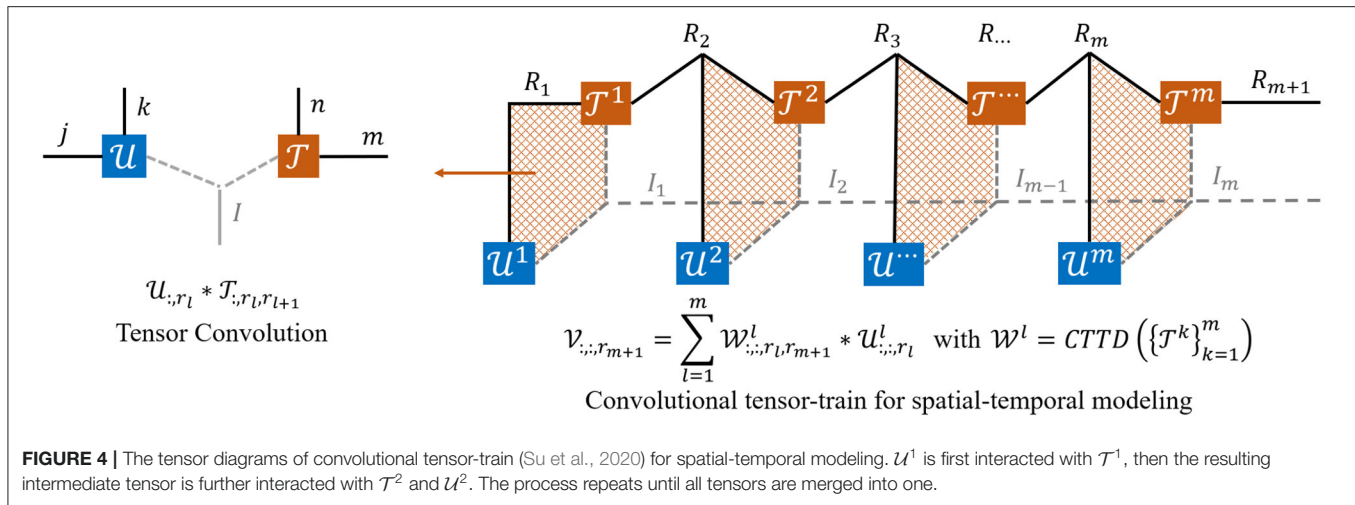
5.2. ConvLSTM With Convolutional Tensor-Train Decomposition

Different from the conventional LSTM, the ConvLSTM introduces convolution tensor operation into input-to-state and state-to-state transitions within each recurrent cell (Xingjian et al., 2015). In ConvLSTM, all features are encoded as third-order tensors with dimensions $(H \times W \times C)$. At each time step t , a ConvLSTM cell updates its hidden states $\mathcal{H}^{(t)}$ based on the previous $\mathcal{H}^{(t-1)}$ and the current input $\mathcal{X}^{(t)}$:

$$[\mathcal{I}^{(t)}, \mathcal{F}^{(t)}, \tilde{\mathcal{C}}^{(t)}, \mathcal{O}^{(t)}] = \sigma \left(\mathcal{W}_{input} * \mathcal{X}^{(t)} + \mathcal{T}_{hidden} * \mathcal{H}^{(t-1)} \right) \quad (2)$$

where $\sigma(\cdot)$ applies sigmoid on the input gate $\mathcal{I}^{(t)}$, forget gate $\mathcal{F}^{(t)}$, and output gate $\mathcal{O}^{(t)}$, and $\tanh(\cdot)$ on memory cell $\tilde{\mathcal{C}}^{(t)}$. The $\mathcal{I}^{(t)}, \mathcal{F}^{(t)}, \mathcal{O}^{(t)}, \tilde{\mathcal{C}}^{(t)} \subseteq \mathbb{R}^{H \times W \times C}$, and the parameters are characterized by two 4th order tensors $\mathcal{W}_{input} \subseteq \mathbb{R}^{K \times K \times S \times 4C}$ and $\mathcal{T}_{hidden} \subseteq \mathbb{R}^{K \times K \times C \times 4C}$, where K is the kernel size, and S and C are the numbers of channels of $\mathcal{X}$ and $\mathcal{H}$ separately.

To capture multi-steps spatiotemporal correlations in ConvLSTM, we introduce a higher-order recurrent unit, where the hidden state $\mathcal{H}^{(t)}$ is updated based on the current input



$\mathcal{X}^{(t)}$ and its n previous steps $\{\mathcal{H}^{(t-l)}\}_{l=1}^n$ with an m -order convolutional tensor-train decomposition $\varphi(\{\mathcal{T}^{(l)}\}_{l=1}^m, \cdot)$ as follows:

$$\{\tilde{\mathcal{H}}^{(t,o)}\} = f(\mathcal{J}^{(o)}, \mathcal{K}^{(o)}, \{\mathcal{H}^{(t-l)}\}_{l=1}^n), \forall o \subseteq [m] \quad (3)$$

$$[\mathcal{I}^{(t)}, \mathcal{F}^{(t)}, \tilde{\mathcal{C}}^{(t)}, \mathcal{O}^{(t)}] = \sigma(\mathcal{W}_{input} * \mathcal{X}^{(t)} + \varphi(\{\mathcal{T}^{(o)}\}_{o=1}^m, \{\tilde{\mathcal{H}}^{(t,o)}\}_{o=1}^m)) \quad (4)$$

where the m -order convolutional tensor-train $\varphi(\{\mathcal{T}^{(l)}\}_{l=1}^m, \cdot)$ are parameterized by m core tensors $\{\mathcal{T}^{(l)}\}_{l=1}^m$. Considering the consistency constraint, the previous n steps are mapped into m intermediate tensors $\tilde{\mathcal{H}}^{(t,o)}$ by $f(\mathcal{J}, \mathcal{K}, \cdot)$ in Equation (3) at first, where $\mathcal{K}$ and $\mathcal{J}$ are 3D convolutional kernels. Note that $f(\mathcal{J}, \mathcal{K}, \cdot)$ is a mapping function with dynamic physics constraints, which will be discussed in section 5.3.

The convolutional tensor-train decomposition, first proposed by Su et al. (2020), is a counterpart of tensor-train decomposition (TTD) which aims to represent a higher-order tensor $\mathcal{T} \subseteq \mathbb{R}^{I_1 \times \dots \times I_m}$ in a set of smaller and lower-order core tensors $\{\mathcal{T}^{(l)}\}_{l=1}^m$ with $\mathcal{T}^{(l)} \subseteq \mathbb{R}^{I_l \times R_l \times R_{l-1}}$. The ranks $\{R^{(l)}\}_{l=1}^m$ here control the number of parameters in the tensor-train format. In this way, the original $\mathcal{T}$ of size $\prod_{l=1}^m I_l$ is compressed to $\sum_{l=1}^m I_l R_{l-1} R_l$, i.e., the complexity only grows linearly with the order m (assuming R_l 's are constants). Similar to TTD, φ is designed to significantly reduce both parameters and operations of higher-order spatio-temporal recurrent models by factorizing a large convolutional kernel into a chain of smaller kernels. The details of convolutional tensor-train decomposition $\mathcal{T} = CTTD\{\mathcal{T}^{(l)}\}_{l=1}^m$ is formulated as:

$$\mathcal{T}_{:,r_1,r_{m+1}} \triangleq \sum_{r_2=1}^{R_2} \dots \sum_{r_m=1}^{R_m} \mathcal{T}_{:,r_1,r_2} * \dots * \mathcal{T}_{:,r_m,r_{m+1}}^m \quad (5)$$

and the convolutional tensor-train for spatiotemporal modeling (Figure 4) can be equivalently stated as:

$$\mathcal{V}_{:,r_{l+1}}^{l+1} = \sum_{r_l=1}^{R_l} \mathcal{T}_{:,r_l,r_{l+1}}^l * (\mathcal{V}_{:,r_l}^l + \mathcal{U}_{:,r_l}^l) \quad (6)$$

where $\mathcal{U}$ is the input feature corresponding to $\tilde{\mathcal{H}}$. $\{\mathcal{V}^l\}_{l=1}^m$ are intermediate results, in which $\mathcal{V}^l \subseteq \mathbb{R}^{H \times W \times R_l}$ for $l > 1$, and $\mathcal{V}^1 \subseteq \mathbb{R}^{H \times W \times R_m}$ is initialized as all zeros and final prediction is returned as $\mathcal{V} = \mathcal{V}^{m+1}$.

Proof of Equation (6) The concrete mathematical expression of convolutional tensor-train decomposition $CTTD\{\mathcal{T}^{(l)}\}_{l=1}^m$ is formulated in Equation (5) and illustrated in Figure 4. Consider a model that predicts a feature $\mathcal{V} \in \mathbb{R}^{I_m \times R_{m+1}}$ based on inputs $\mathcal{U}$. R_l is the number of channel in $\mathcal{U}^l$. The proof is inspired by the work of Su et al. (2020). Since $\mathcal{V}^1$ is initialized as an all zero tensor, the Equation (6) can be rewritten as:

$$\mathcal{V}_{:,r_{m+1}} \quad (7a)$$

$$= \sum_{l=1}^m \sum_{r_l=1}^{R_l} CTTD(\{\mathcal{T}_{t=1}^m\}_{:,r_l,r_{m+1}}) * \mathcal{U}_{:,r_l}^l + \sum_{r_l=1}^{R_l} CTTD(\{\mathcal{T}_{t=1}^m\}_{:,r_l,r_{m+1}}) * \mathcal{V}_{:,r_l}^1 \quad (7b)$$

$$= \sum_{l=2}^m \sum_{r_l=1}^{R_l} CTTD(\{\mathcal{T}_{t=1}^m\}_{:,r_l,r_{m+1}}) * \mathcal{U}_{:,r_l}^l + \sum_{r_l=1}^{R_l} CTTD(\{\mathcal{T}_{t=1}^m\}_{:,r_l,r_{m+1}}) * (\mathcal{U}_{:,r_1}^1 + \mathcal{V}_{:,r_1}^1) \quad (7c)$$

where,

$$\sum_{r_l=1}^{R_l} \mathcal{CTTD}(\{\mathcal{T}\}_{l=1}^m)_{:,r_l,r_{m+1}} * (\mathcal{U}_{:,r_l}^1 + \mathcal{V}_{:,r_l}^1) \quad (8a)$$

$$= \sum_{r_1=1}^{R_1} \left(\sum_{r_2=1}^{R_2} \cdots \sum_{r_m=1}^{R_m} \mathcal{T}_{:,r_1,r_2}^1 * \cdots * \mathcal{T}_{:,r_m,r_{m+1}}^m \right) * (\mathcal{U}_{:,r_1}^1 + \mathcal{V}_{:,r_1}^1) \quad (8b)$$

$$= \sum_{r_2=1}^{R_2} \left(\sum_{r_3=1}^{R_3} \cdots \sum_{r_m=1}^{R_m} \mathcal{T}_{:,r_2,r_3}^1 * \cdots * \mathcal{T}_{:,r_m,r_{m+1}}^m \right) * \left(\sum_{r_m=1}^{R_m} \mathcal{T}_{:,r_1,r_2}^1 * (\mathcal{U}_{:,r_1}^1 + \mathcal{V}_{:,r_1}^1) \right) \quad (8c)$$

$$= \sum_{r_2=1}^{R_2} \mathcal{CTTD}(\{\mathcal{T}\}_{l=2}^{m-1})_{:,r_2,r_{m+1}} * \mathcal{V}_{:,r_2}^2 \quad (8d)$$

and by plugging Equation (8d) into Equation (7c), we reduce Equation (7c) back to the form of Equation (6), which completes the induction:

$$\mathcal{V}_{:,r_{m+1}} = \sum_{l=2}^m \sum_{r_l=1}^{R_l} \mathcal{CTTD}(\{\mathcal{T}\}_{l=1}^m)_{:,r_l,r_{m+1}} * \mathcal{U}_{:,r_l}^l + \sum_{r_2=1}^{R_2} \mathcal{CTTD}(\{\mathcal{T}\}_{l=2}^{m-1})_{:,r_2,r_{m+1}} * \mathcal{V}_{:,r_2}^2 \quad (9)$$

5.3. Physics Constraints

In this section, we provide how the law of physics that describes the physical processes (e.g., the LCS) can be incorporated into the deep learning framework.

5.3.1. Physics Equations

Physical processes are often modeled by a set of nonlinear partial differential equations (PDEs), which describe how a physical quantity is changed in a given region over time. Such dynamic equations are usually written as a relation between time derivatives and spatial derivatives such as:

$$\sum_{i=1}^P b_i(u, t) \frac{\partial^i u}{\partial t^i} = \sum_{i=1}^Q c_i(u, x) \frac{\partial^i u}{\partial x^i} \quad (10)$$

where u is a physical quantity, say, velocity, and x is its spatial coordinate. Furthermore, coefficients b_i , c_i are functions of u and x . Finally, P and Q denote the highest order of time derivatives and spatial derivatives, respectively.

There are multiple ways to incorporate the law of physics into deep neural networks. In this paper, inspired by Seo and Liu (2019), we modeled the law of physics by replacing the updating functions in neural networks with corresponding operators. Considering the characteristics of ocean current, we adopt two dynamic equations, i.e., the diffusion equation and the wave equation, shown in **Table 2**. The diffusion

TABLE 2 | Ocean current related dynamic equations.

Updating function	Dynamic equation
$\mathcal{V}_i' = \mathcal{V}_i + \alpha(\Delta \mathcal{V}_i)$	Diffusion: $\dot{u} = \alpha \nabla^2 u$
$\mathcal{V}_i'' = 2\mathcal{V}_i' - \mathcal{V}_i + c^2(\Delta \mathcal{V}_i)$	Wave: $\ddot{u} = c^2 \nabla^2 u$

equation describes the behavior of continuous physical quantities (macroscopic behavior of many micro-particles in Brownian motion or turbulence in oceanic flows) resulting from the random movement. The wave equation is a second-order PDE for the description of waves (e.g., water, sound, or seismic waves), which are present in the current signal of the LCS through the propagation of wide range of perturbations (Donohue et al., 2016a,b).

As an illustrative example, let us consider time derivatives of the following liner wave equation:

$$\frac{\partial^2 u(x, t)}{\partial t^2} = c^2 \frac{\partial^2 u(x, t)}{\partial x^2} \quad (11)$$

A discrete analog of this equation is usually considered in the form of a second order finite difference equation, which is consistent with the updating function $\mathcal{V}_i'' - 2\mathcal{V}_i' + \mathcal{V}_i$ in **Table 2**. In the continualization method (Tarasov, 2017), it is assumed that the continuous displacement $u(x, t)$ equals the lattice displacement $u_n(t)$ at particle n by $u_n(t) = u(nh, t)$, where the particle spacing is denoted by h . Thus, Equation (11) can then be written as:

$$\frac{\partial^2 u_n(t)}{\partial t^2} = \frac{c^2}{h^2} (u_{n+1}(t) - 2u_n(t) + u_{n-1}(t)). \quad (12)$$

This can be shown as follows. Applying the Fourier series transform $\hat{f}(k) = \sum_{n=-\infty}^{+\infty} f[n] e^{-iknh} = \mathfrak{F}_{h,\Delta}\{f[n]\}$ to Equation (12) gives:

$$\frac{\partial^2 \hat{u}_n(k, t)}{\partial t^2} = -\frac{2c^2}{h^2} \sum_{m=1}^{+\infty} \frac{(-1)^m}{(2m)!} (kh)^{2m} \hat{u}(k, t) \quad (13)$$

and applying the inverse Fourier integral transform $f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk \hat{f}(k) e^{ikx} = \mathfrak{F}^{-1}\{\hat{f}(k)\}$ to the latter gives:

$$\frac{\partial^2 u(x, t)}{\partial t^2} = \frac{2c^2}{h^2} \sum_{m=1}^{+\infty} \frac{(h)^{2m}}{(2m)!} \frac{\partial^{2m} u(x, t)}{\partial x^{2m}} \quad (14)$$

Equation (14) becomes the wave Equation (11) in the limit $h \rightarrow 0$, since,

$$\lim_{h \rightarrow 0} \frac{2}{h^2} \sum_{m=1}^{+\infty} \frac{(h)^{2m}}{(2m)!} \frac{\partial^{2m} u(x, t)}{\partial x^{2m}} = \frac{\partial^2 u(x, t)}{\partial x^2} \quad (15)$$

As a result, the discrete Equation (12) can be considered as an approximation of the wave Equation (11).

5.3.2. Physics-Informed Learning

In Equation (3), a sliding window strategy was adopted. As shown in **Figure 2**, sliding subsets of $\{\mathcal{H}^l\}$ were concatenated and then transformed into m (i.e., the number of core tensors in φ) intermediate hidden tensors $\{\mathcal{H}^{(t,o)}\}$. The concrete process is as follows:

$$\mathcal{H}^{(t,o)} = \mathcal{K}^o * [\mathcal{H}^{(t-n+m-l)}; \dots; \mathcal{H}^{(t-l)}] \quad (16)$$

where n is the sliding window size. The $\{\mathcal{H}\}_{l=1}^n$ were first concatenated into m tensors $\{\mathcal{H}_{cat}\}_{l=1}^m$ ($\mathcal{H}_{cat} \subseteq \mathbb{R}^{H \times W \times (n-m+1) \times C}$) along the time axis, which were thereafter mapped to $\mathcal{H}^{(t,o)} \subseteq \mathbb{R}^{H \times W \times R}$ by the 3D convolutional kernel $\mathcal{K}^o \subseteq \mathbb{R}^{K \times K \times (n-m+1) \times R}$. Then, the intermediate hidden tensor $\{\mathcal{H}^{(t,o)}\}$ was updated to $\{\tilde{\mathcal{H}}^{(t,o)}\}$ with physics constraints.

It is hardly practical to simulate a complicated real-world system with the operators alone unless all physics equations governing the observed phenomena are explicitly known. To date, the LCS is not fully understood, making it almost infeasible to employ all related fluid equations to model the LC and its eddies. Consequently, it is necessary to utilize the learn-able parameters in neural networks to extract latent representations $\mathcal{H}$ in LC observations, which is then further incorporated with domain physics (i.e., constrained by physics equations). The physics-informed part within each PITT-ConvLSTM cell is shown in **Algorithm 1**.

Finally, the sequentially updated intermediate hidden tensors were transformed to the output space by convolutional tensor-train decomposition described in section 5.2. It is notable that the law of physics is not directly constrained by the raw observations, but rather by the latent representations (i.e., the hidden states of the ConvLSTM in this work). This is a desired configuration considering the difficulty in identifying which observations are following the law of physics explicitly and how much in accordance. Consequently, instead of individually applying the equation to each observation, we found that it is more efficient to introduce the constraints on the latent representations.

Specifically, as illustrated in **Figure 2**, $\mathcal{J}$ is the learn-able parameter vector while the orange blocks ("Physics Equations") are objective functions related to physics constraints. First, we defined the physics constraints between the previous and updated states based on the known/assumed knowledge as:

$$\mathcal{L}_{dp}^{(t,o)} = g(\mathcal{H}^{(t,o)}, \tilde{\mathcal{H}}^{(t,o)}) \quad (17)$$

where g is case-specific. In particular, if we are aware/assume that the observations should have a diffusive property, the diffusion equation can be used as the physics-informed constraint as:

$$g_{dif} = \|\nu' - \nu - \alpha \nabla^2 \nu\|^2 \quad (18)$$

where ν (or ν') is the corresponding element of $\mathcal{H}$ (or $\tilde{\mathcal{H}}$). Finally, the physics-informed objective function of the total sequence is defined as:

$$\mathcal{L}_{dp} = \sum_t \sum_{o=1}^m g(\mathcal{H}^{(t,o)}, \tilde{\mathcal{H}}^{(t,o)}) \quad (19)$$

Algorithm 1: Intermediate Hidden Tensor Updating with Physics Constraints

Input : Hidden tensor $\{\mathcal{H}^{(t,o)}\}_{o=0}^m$
Output: Updated hidden tensor $\{\tilde{\mathcal{H}}^{(t,o)}\}_{o=0}^m$, physics loss $\mathcal{L}_{dp}$
Initialize
 Physics equation $pde = wave$ (or *diffusion*);
 Convolutional Kernel $\mathcal{J}^{(o)}$;
 Initial states $\mathcal{H}_{init}$;
 Coefficient $coef$;

Time derivatives $\Delta t = []$; Spatial derivatives $\Delta x = []$;
 Current states $\mathcal{H}_{cur} = \mathcal{H}_{init}$; Previous states $\mathcal{H}_{pre} = None$;
 Physics loss $\mathcal{L}_{dp} = []$;
for $\mathcal{H}^{(t,o)}$ in $\{\mathcal{H}^{(t,o)}\}_{o=0}^m$ **do**
 $\mathcal{H}_{next} = \mathcal{J} * [\mathcal{H}^{(t,o)} \oplus \mathcal{H}_{cur}]$;
if $pde = wave$ **then**
 | $\Delta t = \mathcal{H}_{next} - 2\mathcal{H}_{cur} + \mathcal{H}_{prev}$;
else
 | $\Delta t = \mathcal{H}_{next} - \mathcal{H}_{cur}$;
end
 $\Delta x = -coef \cdot [\nabla^2 \mathcal{H}_{cur}]$;
 $\mathcal{H}_{pre} = \mathcal{H}_{cur}$;
 $\mathcal{H}_{cur} = \mathcal{H}_{next}$;
 $\tilde{\mathcal{H}}^{(t,o)} = \mathcal{H}_{cur}$;
 $\mathcal{L}_{dp}.append((\Delta t - \Delta x)^2)$;
end

and the overall objective function is the sum of the supervised loss and physical loss:

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \lambda \mathcal{L}_{dp} \quad (20)$$

where $\mathcal{L}_1$ is the mean absolute error, $\mathcal{L}_2$ is the mean squared error, and λ controls the importance of the physics term.

6. RESULTS AND DISCUSSION

6.1. Model Implementation

All experiments, shown in **Figure 3**, used a stack of 4 blocks (3 stacked layers of TT-ConvLSTM or PITT-ConvLSTM per block) and two skip connections added between the 1st and 3rd, 2nd and 4th blocks that perform concatenation over channels (Byeon et al., 2018). The channels are set to 32 for the 1st and 4th blocks and 48 for the middle blocks. A convolutional layer is applied on top of all the recurrent layers to compute the predicted frames. The PITT-ConvLSTM is of order 3, sliding window step size 3, rank 8 (the hyper-parameter tuning is based on the validation dataset. we considered an order of the decomposition from [1, 2, 3, 5], tensor ranks from [4, 8, 16] and number of hidden states in sliding window from [1, 3, 5] Su et al., 2020). All models were trained with the ADAM optimizer (Kingma and Ba, 2014) with objective function in Equation (20). Once the model did not improve after 20 epochs (in terms of validation loss), scheduled sampling (Bengio et al., 2015) was then activated to ease the

TABLE 3 | Comparisons of 10, 20, and 30 time-steps prediction (equal to 5, 10, and 15 days ahead forecast), where lower MSE (in 10^{-3} magnitude) or higher SSIM indicates better model performance.

Methods	10 → 10		10 → 20		10 → 30		Parms
	MSE	SSIM	MSE	SSIM	MSE	SSIM	
ConvLSTM (Xingjian et al., 2015)	9.257	0.623	18.584	0.465	26.725	0.374	4.94 m
PredRNN (Wang et al., 2017)	8.752	0.648	19.229	0.490	27.512	0.351	8.84 m
TT-ConvLSTM (Su et al., 2020)	8.575	0.656	17.445	0.498	23.802	0.420	0.97 m
PITT-ConvLSTM (diffusion)	7.429	0.686	15.835	0.527	22.426	0.439	0.99 m
PITT-ConvLSTM (wave)	6.971	0.687	14.097	0.543	19.638	0.463	0.99 m

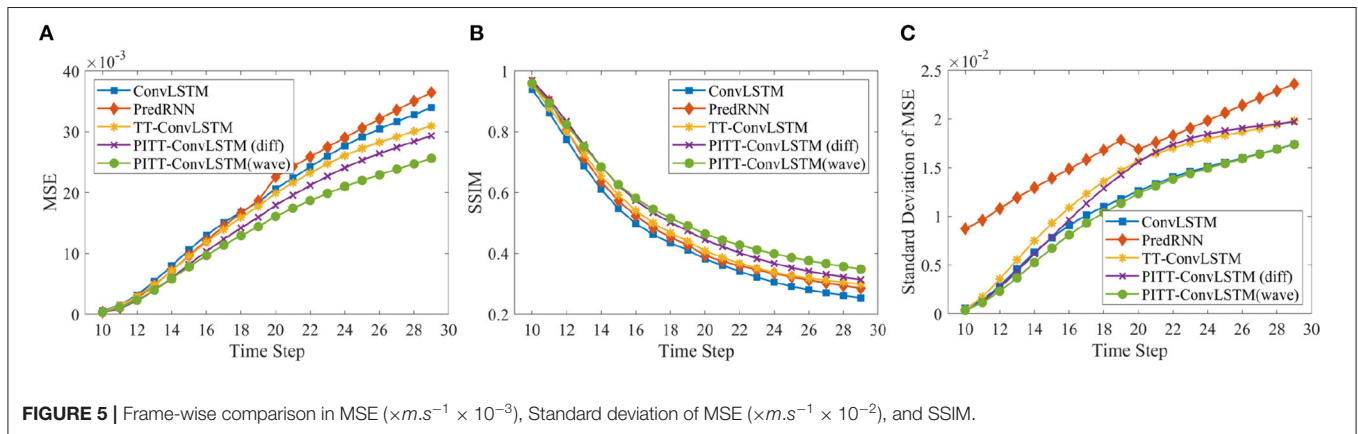


FIGURE 5 | Frame-wise comparison in MSE ($\times m.s^{-1} \times 10^{-3}$), Standard deviation of MSE ($\times m.s^{-1} \times 10^{-2}$), and SSIM.

training with linearly decreased sampling ratio from 1 to 0. Learning rate decay was further activated if the loss did not drop after 20 epochs, and the rate was decreased exponentially by 0.98 every 5 epochs. The initial learning rate is set to 10^{-4} . All models were trained to predict 10, 20, 30 frames given 10 input frames, where we use the teacher forcing mechanism (Su et al., 2020; Zhuang and Ibrahim, 2021), when a step is within the number of input steps, it's direct multi-step forecast strategy, when the step exceeds the number of input steps, it adopts recursive multi-step forecast strategy. These baseline models are listed below:

1. ConvLSTM is first proposed by Xingjian et al. (2015) for precipitation nowcasting by extending the fully connected LSTM (FC-LSTM) to have convolutional structures in both the input-to-state and state-to-state transitions. The spatio-temporal information is encoded in each cell.
2. PredRNN proposed in Wang et al. (2017), is a recurrent network for spatiotemporal predictive learning. The core of PredRNN is a new Spatio-temporal LSTM unit that extracts and memorizes spatial and temporal representations simultaneously.
3. TT-ConvLSTM in Su et al. (2020) is a higher-order generalization to ConvLSTM, which is able to learn long-term spatio-temporal structure in videos.

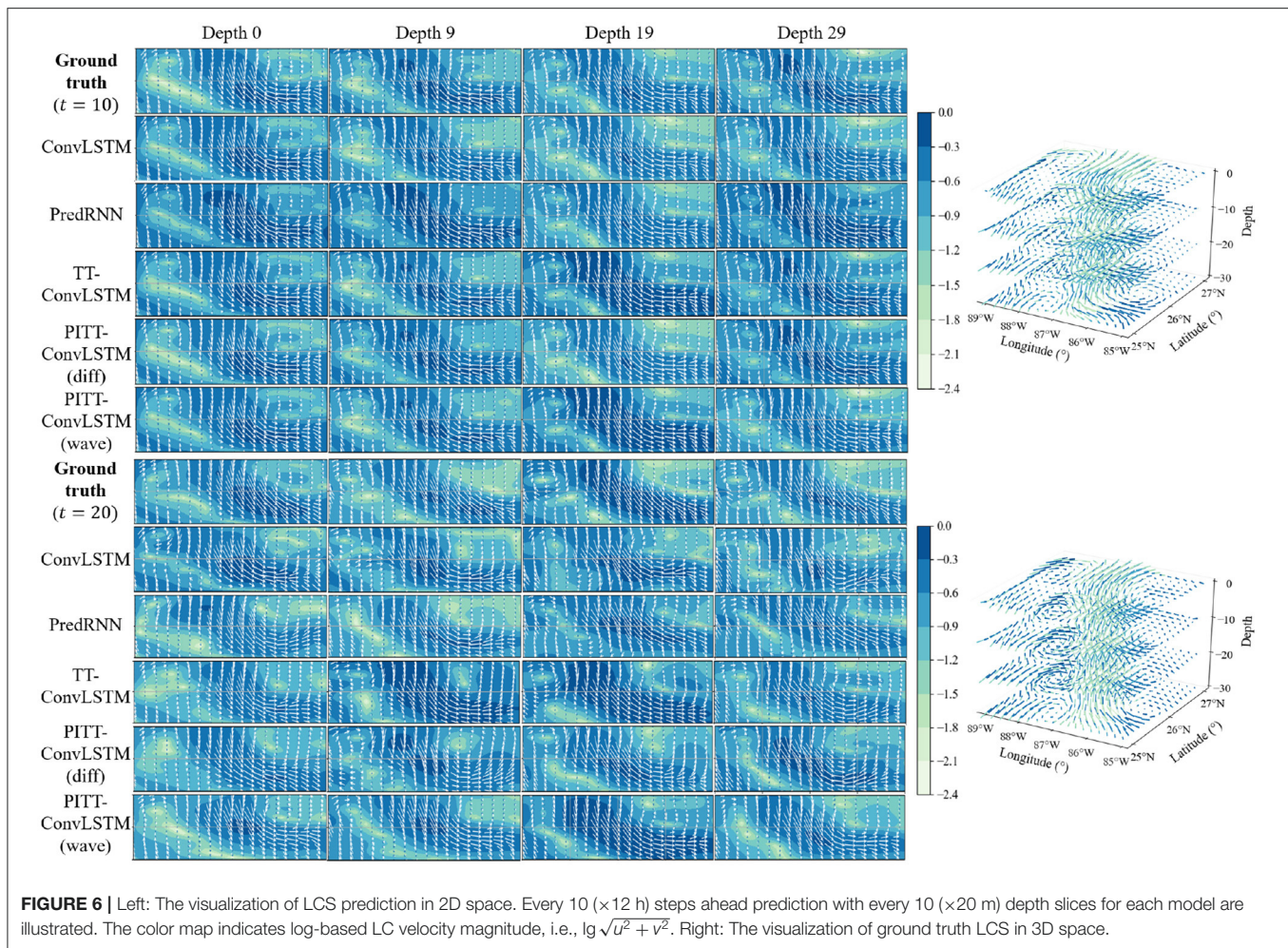
6.2. Results

Two metrics were adopted to evaluate the performance and provide frame-wise quantitative comparisons. Specifically, the mean squared error (MSE) was used for element-wise differences, while the structural similarity index measure (SSIM) (Wang et al.,

2004), which ranges between -1 and 1 , was used for perceptual similarity. Here, we report the averaged MSE and SSIM over 10, 20, 30 predict horizons.

The average evaluation statistics, i.e., MSE and SSIM scores, over 10, 20, and 30 frames prediction is given in **Table 3**. **Figure 5** shows the comparisons of the predictions by ConvLSTM, PredRNN, TT-ConvLSTM and proposed PITT-ConvLSTM informed by *wave* and *diffusion* equations. The ConvLSTM model generates blurred future frames (lowest SSIM), since it fails to memorize the detailed spatial representations. The reason lies in that the ConvLSTM is a first-order Markovian model in nature, which means the hidden state is updated based on its previous step. Thus, it can not efficiently capture higher-order temporal correlations. ConvLSTM is more effective in short-term forecasting, focusing on next or first few frames prediction. By contrast, as shown in **Table 3** the methods using convolutional tensor-train decomposition, TT-ConvLSTM and PITT-ConvLSTM, achieve better performance in long-term prediction with fewer parameters.

The proposed PITT-ConvLSTM models showed superior skills for both measures over the other models. It validates that the convolutional tensor-train, combined with physics constraints, is capable of modeling higher-order spatio-temporal correlations. The results showed how the physics constraints helped improve the volumetric velocity prediction of the LCS. Among the convolutional tensor-train decomposition (CCTD) based models, PITT-ConvLSTM realized the smallest MSEs and highest SSIMs. It validates the effectiveness of incorporating physical rules in latent representation learning, since knowing



physics-constrained neighboring information is helpful to infer its own states. Specifically, the wave-informed PITT-ConvLSTM consistently outperformed all baseline models and showed a superior predicting capability both spatially and temporally. The LC prediction comparison among ConvLSTM, PredRNN, and TT-ConvLSTM is visualized in **Figure 6**.

7. CONCLUSION

Understanding the dynamics of the LCS is fundamental to understanding the Gulf of Mexico's full oceanographic system, and vice versa. Hurricane intensity, offshore safety, oil spill response, fisheries, and the Gulf Coast economy are all affected by the position, strength, and structure of the LC and associated eddies. The LC's position varies greatly from its retracted state in the Yucatan Channel, directly east of the Florida Straits, to its extended state into the far northern and western Gulf. Of particular interest to the modeling community is the prediction of the velocity field and the duration of the LC circulation at a given location.

In this paper, we proposed a physics-informed tensor-train ConvLSTM that is capable of effectively capturing long-term spatiotemporal correlations in a temporal sequence of volumetric

data. Within the PITT-ConvLSTM cell, a large convolutional kernel was factorized into a set of smaller core tensors through convolutional tensor-train decomposition. In the learnt latent space, physical domain knowledge in the form of PDEs over time and space were incorporated to facilitate learning. The performance of our proposed PITT-ConvLSTM model was demonstrated on volumetric velocity forecasting of the LCS in the GoM, and the results verified that our model can produce superior results compared to published state-of-the-art deep learning models, including ConvLSTM, PredRNN, and TT-ConvLSTM. The results of this study suggest that it is now possible to create reliable current forecasts for a period longer than 5 days. The MSE of the PITT-ConvLSTM was less than $0.03(m.s^{-1})$ 15 days ahead, which is a significant improvement over the current state of the art ocean numerical models (Cooper et al., 2016). Note that although the proposed model is a sequence-to-sequence prediction, a sequence of arbitrary length can be generated in a recursive manner. In a future study, we aim to design an effective prediction model that can capture simultaneously both fast changing local patterns and slow varying global trends. Furthermore, as suggested in Zhang et al. (2018), multiple complex factors should be considered or integrated in modeling, including spatial correlation (nearby and

distant), various types of temporal correlation (closeness, period, trend), and external conditions (e.g., weather).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YH, HZ, and YT contributed to the conception and design of the study. JV organized the database. YH performed the statistical analysis and wrote the first draft of the manuscript. YH, YT, HZ, and LC wrote sections of the manuscript. All

authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was partially supported by an Understanding Gulf Ocean Systems Phase II grant (#NAS/GRP 2000011052) from the National Academies of Sciences, Engineering, and Medicine (NASEM), and by an MRI grant (#1828181) from National Science Foundation (NSF).

ACKNOWLEDGMENTS

We would like to thank Kathleen Donohue and her team for providing the Dynloop dataset, which made this study possible.

REFERENCES

- Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. (2015). "Scheduled sampling for sequence prediction with recurrent neural networks," in *Advances in Neural Information Processing Systems*, Montreal, 1171–1179.
- Bruciaferri, D., Shapiro, G. I., and Wobus, F. (2018). A multi-envelope vertical coordinate system for numerical ocean modelling. *Ocean Dyn.* 68, 1239–1258. doi: 10.1007/s10236-018-1189-x
- Byeon, W., Wang, Q., Kumar Srivastava, R., and Koumoutsakos, P. (2018). "Contextvp: fully context-aware video prediction," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 753–769.
- Cooper, C., Danmeier, D., Frolov, S., Stuart, G., Zuckerman, S., Anderson, S., et al. (2016). "Real Time Observing and Forecasting of Loop Currents in 2015," in *OTC Offshore Technology Conference*. Houston, TX.
- de Bezenac, E., Pajot, A., and Gallinari, P. (2019). Deep learning for physical processes: incorporating prior scientific knowledge. *J. Stat. Mech. Theor. Exp.* 2019:124009. doi: 10.1088/1742-5468/ab3195
- Donohue, K., Watts, D., Hamilton, P., Leben, R., Kennelly, M., and Lugo-Fernández, A. (2016a). Gulf of Mexico loop current path variability. *Dyn. Atmos. Oceans* 76, 174–194. doi: 10.1016/j.dynatmoce.2015.12.003
- Donohue, K. A., Watts, D., Hamilton, P., Leben, R., and Kennelly, M. (2016b). Loop current eddy formation and baroclinic instability. *Dyn. Atmos. Oceans* 76, 195–216. doi: 10.1016/j.dynatmoce.2016.01.004
- Halliwel, G. R. (2004). Evaluation of vertical coordinate and vertical mixing algorithms in the hybrid-coordinate ocean model (hycom). *Ocean Model.* 7, 285–322. doi: 10.1016/j.ocemod.2003.10.002
- Hamilton, P., Lugo-Fernández, A., and Sheinbaum, J. (2016). A loop current experiment: Field and remote measurements. *Dyn. Atmos. Oceans* 76, 156–173. doi: 10.1016/j.dynatmoce.2016.01.005
- Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nat. Rev. Phys.* 3, 422–440. doi: 10.1038/s42254-021-00314-5
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kipf, T., Fetaya, E., Wang, K.-C., Welling, M., and Zemel, R. (2018). Neural relational inference for interacting systems. *arXiv preprint arXiv:1802.04687*.
- Kolbeinson, A., Kossafi, J., Panagakis, Y., Bulat, A., Anandkumar, A., Tzoulaki, I., and Matthews, P. (2019). Robust deep networks with randomized tensor regression layers. *arXiv preprint arXiv:1902.10758*.
- Ma, X., Zhang, P., Zhang, S., Duan, N., Hou, Y., Zhou, M., et al. (2019). "A tensorized transformer for language modeling," in *Advances in Neural Information Processing Systems*, Vancouver, 2229–2239.
- National Academies Press (2018). *Understanding and Predicting the Gulf of Mexico Loop Current: Critical Gaps and Recommendations*. Washington, DC: National Academies Press.
- Navarra, A., and Simoncini, V. (2010). *A Guide to Empirical Orthogonal Functions for Climate Data Analysis*. Dordrecht: Springer Science & Business Media.
- Oseledets, I. V. (2011). Tensor-train decomposition. *SIAM J. Sci. Comput.* 33, 2295–2317. doi: 10.1137/090752286
- Raissi, M. (2018). Deep hidden physics models: deep learning of nonlinear partial differential equations. *J. Mach. Learn. Res.* 19, 932–955. <https://arxiv.org/abs/1801.06637>
- Raissi, M., and Karniadakis, G. E. (2018). Hidden physics models: machine learning of nonlinear partial differential equations. *J. Comput. Phys.* 357, 125–141. doi: 10.1016/j.jcp.2017.11.039
- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2017). Physics informed deep learning (part i): data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*.
- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707. doi: 10.1016/j.jcp.2018.10.045
- Sanchez-Gonzalez, A., Heess, N., Springenberg, J. T., Merel, J., Riedmiller, M., Hadsell, R., et al. (2018). Graph networks as learnable physics engines for inference and control. *arXiv preprint arXiv:1806.01242*.
- Seo, S., and Liu, Y. (2019). Differentiable physics-informed graph networks. *arXiv preprint arXiv:1902.02950*.
- Shi, L., Campbell, G., Jones, W. D., Campagne, F., Wen, Z., Walker, S. J., et al. (2010). The microarray quality control (maqc)-ii study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.* 28, 827–838. doi: 10.1038/nbt.1665
- Soltani, R., and Jiang, H. (2016). Higher order recurrent neural networks. *arXiv preprint arXiv:1605.00064*.
- Su, J., Byeon, W., Huang, F., Kautz, J., and Anandkumar, A. (2020). Convolutional tensor-train lstm for spatio-temporal learning. *arXiv preprint arXiv:2002.09131*.
- Su, J., Li, J., Bhattacharjee, B., and Huang, F. (2018). Tensorized spectrum preserving compression for neural networks. *arXiv preprint arXiv:1805.10352*.
- Sutskever, I., Martens, J., and Hinton, G. E. (2011). "Generating text with recurrent neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, Bellevue, 1017–1024.
- Tarasov, V. (2017). Discrete wave equation with infinite differences. *Appl. Math. Inform. Sci. Lett.* 5, 41–44. doi: 10.18576/amisl/050201
- The Commission (2011). *Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling: Report to the President*. Washington, DC: The Commission.
- Walker, N., Leben, R., Anderson, S., Feeney, J., Coholan, P., and Sharma, N. (2009). *Loop Current Frontal Eddies Based on Satellite Remote-Sensing and Drifter Data*. United States: Department of the Interior.
- Wang, J. L., Zhuang, H., Chérubin, L., Muhamed Ali, A., and Ibrahim, A. (2021). Loop current ssh forecasting: a new domain partitioning approach for a machine learning model. *Forecasting* 3, 570–579. doi: 10.3390/forecast3030036

- Wang, J. L., Zhuang, H., Chérubin, L. M., Ibrahim, A. K., and Muhamed Ali, A. (2019). 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. Oceans* 124, 5586–5606. doi: 10.1029/2019JC015172
- Wang, Y., Long, M., Wang, J., Gao, Z., and Philip, S. Y. (2017). “Predrnn: recurrent neural networks for predictive learning using spatiotemporal lstms,” in *Advances in Neural Information Processing Systems*, Long Beach, CA, 879–888.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861
- Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c. (2015). “Convolutional lstm network: a machine learning approach for precipitation nowcasting,” in *Advances in Neural Information Processing Systems*, Montreal, 802–810.
- Yang, Y., Krompass, D., and Tresp, V. (2017). “Tensor-train recurrent neural networks for video classification,” in *Proceedings of the 34th International Conference on Machine Learning-Volume 70* JMLR. org, Sydney, 3891–3900.
- Yang, Y., Weisberg, R. H., Liu, Y., and Liang, X. S. (2020). Instabilities and multiscale interactions underlying the loop current eddy shedding in the Gulf of Mexico. *J. Phys. Oceanogr.* 50, 1289–1317. doi: 10.1175/JPO-D-19-0202.1
- Yu, R., Zheng, S., Anandkumar, A., and Yue, Y. (2018). “Long-term forecasting using tensor-train rnns. h ps,” in *Openreview. net/forum*. Pasadena, CA.
- Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., and Li, T. (2018). Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artif. Intell.* 259, 147–166. doi: 10.1016/j.artint.2018.03.002
- Zhuang, D., and Ibrahim, A. K. (2021). Deep learning for drug discovery: a study of identifying high efficacy drug compounds using a cascade transfer learning approach. *Appl. Sci.* 11, 7772. doi: 10.3390/app1117772

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Huang, Tang, Zhuang, VanZwieten and Cherubin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.