This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140398, IEEE Transactions on Geoscience and Remote Sensing

An information flow-based sea surface height reconstruction through machine learning

Yineng Rong, X. San Liang

***Abstract*—The advent of satellite altimetry datasets of sea surface height (SSH) is a major advance in oceanography and other earth system sciences. But, while the along-track data coverage is dense, the relatively poor resolution between tracks poses a challenge to the reconstruction of those processes such as mesoscale and sub-mesoscale eddies. This study proposes a machine learning algorithm based on a causal inference tool, i.e., the Liang-Kleeman information flow (L-K IF) analysis, to address the challenge. For a region in the South China Sea where eddies frequently appear but unobserved, it is shown that the algorithm can reconstruct the desired mesoscale eddies in a remarkably successful way in geometry, orientation, strength, etc., while with the objective analysis interpolation or the traditional neural network technique the results are not satisfactory. This study provides prospects for developing the next generation of SSH products with the available altimetry data.**

***Index Terms*—Sea surface height; causal inference; Liang-Kleeman information flow; machine learning.**

I. Introduction

PERHAPS one of the most significant achievements in satellite observation is the successful global reconstruction of a spatiotemporal field of sea surface height (SSH), an important ocean variable which allows for an accurate estimation of in-situ large-scale ocean circulation (see [1] for a review). This not just sets a milestone in oceanography, but also aids to advance atmospheric and other earth sciences in the fields of, say, Madden-Julian Oscillation(e.g., [2]), El Niño-Southern Oscillation (ENSO) (e.g., [3]–[6]), Indian Ocean Dipole (IOD) (e.g., [7], [8]), global change (e.g., [9]–[11]), to name but a few. Indeed, during the past 30 years, the satellite e altimeters have gradually become a major observing technique in geoscience due to their dense resolution in space and time, as well as the simultaneous coverage on a global scale. For example, Fig. 1 shows the track of the Jason-2 satellite altimeters over the South China Sea. The along-track SSH can resolve waves with a wavelength as small as 30–50 km, with a root mean square error (RMSE) of only 2 cm.

With the densely covered data along the track, the resolution between the tracks, however, is in general not satisfactory. As a result, the global mesoscale resolution is rather limited, mainly due to the spatial and temporal gaps between the altimeter-derived SSH profiles. Currently, the mapping of the SSH observations onto the two-dimensional (2D) grid is mainly through the widely adopted global suboptimal space-time objective analysis (OA) [12] together with optimal interpolation (OI) (e.g., [13]–[15]). The spatial gap can be reduced by using the multi-satellite altimeter data [15]–[18], and the temporal resolution can be improved with dynamic interpolation (DI) [19]. Nonetheless, so far as of today, mesoscale and synoptic variabilities of SSH are still poorly reconstructed [20].

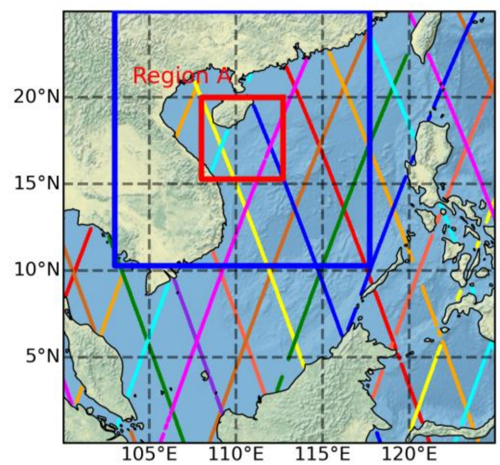

Fig. 1. A scatter of the Jason-2 trajectory over the South China Sea from January 1 to January 10, 2011. Different colors represent the trajectories on of different days. The red box between Hainan Island and Vietnam is the region A as referred to in the text; the blue one is the observation region for region A.

The widely used OA method employs a statistical estimation under a linear assumption [13]; it has difficulty in revealing the nonlinear relationship between the along-track data and the grid data. This is unfortunate, as nonlinear processes are ubiquitous, particularly in mesoscale processes. A conspicuous example is that, when between two tracks there exists an eddy, with a local extremum SSH (local high or local low) lying in between, then a linear interpolation scheme will never be able to reconstruct such a pattern---In fact, this is a benchmark which we will be examining in this study. Besides,

Manuscript received November 20, 2020. This work was supported in part by National Science Foundation of China under Grant 41975064 and the 2015 Jiangsu Program for Innovation Research and Entrepreneurship Groups.
(*Corresponding author: X. San Liang.*)

Yineng Rong is with the School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China. (e-mail: yinengrong@foxmail.com).

X. San Liang is with the Department of Atmospheric and Oceanic Sciences, Fudan University, Shanghai 200433, China (e-mail: x.san.liang@gmail.com).

OA, like many other interpolation algorithms, is based on a given empirical model to calculate the interpolation field corresponding to the observation. Whether such an empirical model could really approximate the mapping between observations and grid data is still a problem. Obviously, there is still much room for improvement for the satellite-derived SSH field reconstruction.

On the other hand, if we can find out a mapping which does not depend on the given model but, instead, depends on the available samples composed of the observations and the gridded ground truth, then the grid data between the tracks can be well reconstructed, provided that the new set of observations have the same format as the samples. This is the so-called "learning from samples" in machine learning [21]. Different from the traditional OA, which “learns from instruction", this method does not rely on a model; it draws from the training samples the needed functional relation. In this regard, the methods of neural networks (NN) [22] and NN-based deep learning [23], [24] have attracted enormous interest in recent decades. NNs are developed from the multilayer perceptron [25] and generally consist of three layers, namely, the input, output, and hidden layers. They do not depend on a particular model, but work with linear or nonlinear transformations (called “activation function”) in hidden layers to convert inputs (from input layer) into outputs (to output layer), and adjusting the structure of the NNs in order to minimize the error or loss function between outputs and the ground truth. Such examples include [26], who used an NN to interpolate SSH to a grid in a way of “pixel to pixel.” Their results are generally acceptable (the RMSE is about 4.7 cm), but, still, at some spots the bias can exceed 10 cm.

In most of the current algorithms all available data are put into an NN model for training. The adequacy is questionable, as the quality of data determines the effect of the NN model of concern. Indeed, it is a very important problem on how to determine the covariates for the outputs. Apart from this, while the "Pixel to Pixel" approach can allow one to map from the observations to each “Pixel”, for 2D or 3D interpolation problems (such as SSH reconstruction), it may fail in dealing with the spatially distributed data. To address the first problem, in this study we will use a newly rigorously developed causal inference tool (Liang-Kleeman information flow, or IF for short; see Section II.A) to eliminate noncausal observations and hence achieve the data quality. For the second one, the principal component analysis (PCA), an efficient method for reducing the dimension of datasets, is used to extract the main spatial features of the data for NN. After these preprocessings a multilayer NN (Section II.C) to extract the mapping from the along-track SSH to the gridded SSH. For demonstration, we will focus on a region in the South China Sea where an eddy appears between two trajectories and traditional approaches fail. We first introduce the data in Section III, and the IF-based neural network in Section IV. The power of this new approach will be demonstrated using a benchmark problem (Section V). This study is summarized in Section VI.

II. Methods

A. Information Flow

Since Liang and Kleeman[27], a series of studies have been conducted for a rigorous formalism of information flow, which now has been established from first principles in physics [28]. A causal inference technique is henceforth developed [29], validated and applied with success in problems in different disciplines. Hereafter is a brief introduction of some material needed for this study.

Consider, for an example, a 2D stochastic dynamical system:

$$d\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, t)dt + \boldsymbol{B}(\boldsymbol{x}, t)d\boldsymbol{w}, \tag{1}$$

where $\mathbf{F} = (F_1, F_2)$ is the deterministic components, $\mathrm{x} = (x_1, x_2) \in \mathbb{R}^2$ is the state variables, $\mathbf{w} = (w_1, w_2)$ is a standard 2D Wiener process and $\boldsymbol{B} = (b_{ij})$ is the matrix of perturbation amplitude. Liang[30] showed that the transfer of Shannon entropy, or information flow (IF), from x_2 to x_1 is:

$$T_{2\to1} = -E\left(\frac{1}{\rho_1}\frac{\partial F_1\rho_1}{\partial x_1}\right) + \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right), \tag{2}$$

where $\rho(t; x_1, x_2)$ is the joint probability density function, $\rho_1(t; x_1) = \int_{\mathbb{R}} \rho\, dx_2$ is the marginal density of x1, $g_{11} = \sum_{k=1}^{2} b_{1k}^2$, and E is the expectation with respect to ρ. The so-obtained IF, i.e., Eq. (2), has several nice properties, one being the “principle of nil causality” [28]: an event is not causal to another event if the evolution of the latter does not depend on the former. This is a principle that all other formalisms try to verify in applications, while in the above rigorous formalism, this is a proven theorem: if both F_1 (the deterministic component of the system) and g_{11} (the stochastic component) are independent of x_2, then $T_{2\to1} = 0$. Moreover, Liang [31] proved that the so-obtained causality is **invariant upon any nonlinear coordinate transformation**, indicating that (2) is an intrinsic physical property, in contrast to those empirical formalisms.

The IF formula has been validated in many highly chaotic systems, such as baker transformation, Hénon map, Kaplan-Yorke map, Langevin equation, etc. [28], [32]. Under a linearity assumption, Liang[29] further established that it can be estimated from two time series, say, x_1 and x_2, and the resulting maximum likelihood estimator is remarkably simple in form, involving only covariances between the time series:

$$T_{2\to1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11}C_{12}^2} \tag{3}$$

where C_{ij} is the covariance between x_i and x_j, $C_{i,dj}$ the covariance between x_i and $\dot{x}_j$, and $\dot{x}_j$ the difference approximation of dx_j/dt using the Euler forward scheme. In terms of correlation coefficient, (3) becomes

$$T_{2\to1} = \frac{r}{1-r^2}\left(r'_{2,d1} - r r'_{1,d1}\right) \tag{4}$$

where $r = \frac{C_{11}}{\sqrt{C_{11}C_{22}}}$ is the sample correlation coefficient between x_1 and x_2, and $r'_{i,dj} = \frac{C_{i,dj}}{\sqrt{C_{11}C_{22}}}$ the “correlation coefficient” between x_i and $\dot{x}_j$. Obviously, two uncorrelated events ($r = 0$) must be noncausal ($T_{2\to1} = 0$); the converse, however, does not hold; That is to say, **causation implies**

correlation, but correlation does not imply causation. In a simple mathematical equation, (4) fixes the long-standing debate over causation versus correlation in philosophy. So far as of today, the L-K IF has been widely applied in the diverse problems such as global warming [33], [34], El Niño and Indian Ocean Dipole [29], typhoon genesis prediction [35], space weather [36], chlorophyll variability[37], soil moisture versus precipitation [38], Finance [39], [40], neuroscience [41], to name a few.

B. Objective Analysis (OA)

OA is an interpolation algorithm widely used in satellite remote sensing, who aims to estimate the value θ_x of a scalar variable θ at a point $\boldsymbol{x}$ from measurements φ_r at a limited number of data points $\boldsymbol{x}_r (r = 1, \dots, N)$[12]. Here θ is one realization out of a homogeneous statistical ensemble, which has a zero mean and some given covariance. Under this assumption, the least square linear estimator for θ is:

$$\hat{\theta}_x = \sum_{r=1}^{N}\left[C_{xr}\left(\sum_{s=1}^{N} A_{rs}^{-1}\left(\varphi_s - \frac{1}{N}\sum_{p=1}^{N}\varphi_p\right)\right) + \frac{1}{N}\varphi_r\right], \tag{5}$$

where

$$A_{rs} = \overline{\varphi_r \varphi_s} = F(\boldsymbol{x}_r - \boldsymbol{x}_s) + \mathfrak{E}\delta_{rs},$$

is the matrix of covariance between all pairs of observations, $\mathfrak{E}$ is the variance of the errors, δ_{rs} is the Dirichlet function. Here we have used A_{rs}^{-1} to indicate the entries of A^{-1}, i.e., the inverse of A.

$$C_{xr} = \overline{\theta_x \varphi_r} = F(\boldsymbol{x} - \boldsymbol{x}_r)$$

is the covariance between the quantity θ_x to be estimated and the rth measurement.

Equation (5) shows that such a calculation requires only the location of the data points and a knowledge of the covariance function $F(\xi)$.Thus for different realizations of the field θ_x, the estimate θ_x depends linearly on the observations φ_s. In other words, θ_x is a linear estimator.

C. Neural Network (NN)

The traditional methods of interpolating the along-track data to the gridded data are mostly based on some empirical given models. These empirical models are often based on our understanding of the system. Compared to large-scale ocean processes, up to now, the understanding of mesoscale or submesoscale processes is still limited. It is therefore necessary to have a model free method to map the data along the orbit and grid points. Neural network (NN) is one of the most popular methods.

Unlike OA, the NN offers a way to do nonlinear transformations. In this study, a series of NNs are designed to realize the regression of the output label to the multi-dimensional inputs. Specifically, in one NN, there are n training samples, each sample containing an N-dimensional input variables and a one-dimensional output label. Then the input and output of the ith ($i \in \mathbb{Z}[1, n]$) sample is $\boldsymbol{X}[i] = (x_{i1}, \dots, x_{iN})$ and $\boldsymbol{Y}[i] = y_i$ respectively. Fig. 2 shows the structure of the NN model. Input and output layers are

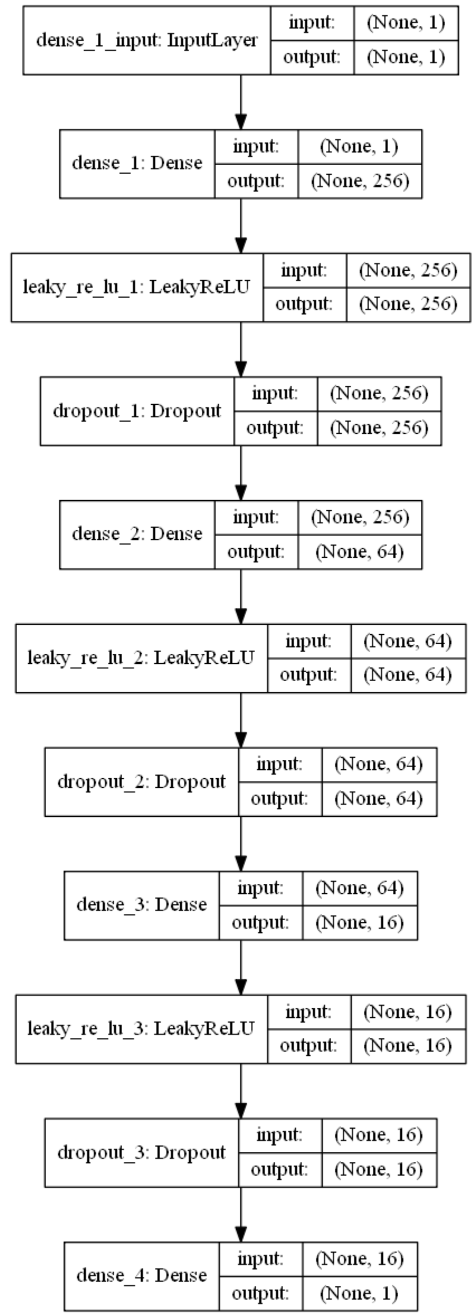

Fig. 2. Architecture diagram for the NN algorithm.

connected by three fully connected hidden layers, consisting of, respectively, 256, 64, and 16 neurons. After each fully connected layer, a nonlinear transformation and a regularization are applied to accelerate the convergence. The former realized by an activation function enables the NN to learn the nonlinear relations between different layers. We use the leaky rectified linear unit (LeakyReLU [42]; parameter $\alpha = 0.2$) as the activation function. The regularization is a dropout method ([43]; parameter drop rates: 0.1). It works by randomly dismissing parts of the weight parameters, having the same effect as data augmentation[44]. For the NN, the loss function is chosen as the mean square error (MSE), and the optimizer is the adaptive moment estimation [45] (Adam; parameter $lr = 0.0002$ and $\beta = 0.5$).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140398, IEEE Transactions on Geoscience and Remote Sensing

III. Datasets

We use the Hybrid-Coordinate Ocean Model (HYCOM) dataset and the satellite altimetry of Jason-2 within the South China Sea region. HYCOM is a model with a hybrid coordinate in the vertical direction such that, in shallow waters, vertical grid points are geometrically constrained to remain at a fixed depth, while in deep oceans an isopycnic coordinate is adopted [46]. This allows its simulation to be relatively accurate along the coast. In this study, we use the sea surface height derived from "HYCOM + NCODA Global 1/12° GOFS 3.1 41-layer Analysis (GLBy0.08 Experiment 9X.X)". It has a horizontal resolution of 0.08° × 0.08° and a temporal resolution of 24 hour (daily snapshot at 0Z, start from 1994/01/01). The Jason-2 satellite altimetry data (including latitude, longitude, and date) [47] has a resolution of 1-second (1.0786 s) along the orbit and repeats at regular intervals in 10 days (9.9156 days, to be precise) (Fig. 1).

IV. Information Flow-based Neural Network

A. Preprocessing of Input Data

In order to minimize the error from the satellite altimetry data, this study assumes that the HYCOM data is the ground truth, and interpolates the HYCOM grid data onto the trajectory of Jason-2. Considering that the temporal sampling interval of Jason-2 (around 1 second) is much larger than the ground truth (once a day at UST 0:00) and the spatial sampling interval (about 0.137^o) is of the same order as the ground truth (0.08° × 0.08°), we do not interpolate the HYCOM grid data in time, but instead divide the 10-day (a period of Jason-2) orbital data into 10 groups according to the day. More specifically, the sampling time of Jason-2 is rounded to integer days, and then we put the same unit digit of sampling time into a group. For regional problem, the global trajectory data may be redundant [48], [49]. We hence discard the orbits far away from the South China Sea ($[100^oE{-}125^oE,\ 0-24^oN]$), and finally obtain 10 sets of observations, each set containing a label (from 0 to 9) and the information of coordinates (longitude and latitude).

Since a time interval of 20 days (i.e., two periods) is commonly used as an e-fold time in OA [16], in this paper, we also use this characteristic time to preprocess the along-track data. As shown in the following Algorithm I, the along-track SSH used for interpolating the grid data at the date t_i (t is the time series with the unit of day) can be expressed as a label k (reflects the category of the track at t_i.) and a 2D array (one dimension is the input characteristics (latitude, longitude, t_i, SSH) ; the other one is made of the observations t_{it-20} to t_{it+19}).

B. Construction of the Model

In order to avoid the interference from the irrelevant data, the IF analysis is first applied. In Algorithm II, its application eliminates the along-track observations which are noncausal to the principal components of the corresponding grid data. Considering the huge dimension of the gridded data and the difficulty for the "pixel to pixel" model to learn the spatial structure from a global perspective, principal component analysis [50] (PCA) is used to simplify the problem from learning the mapping from the along-track data to gridded data into a mapping to the time series of the principal components (PCs), and hence greatly reduce the dimension of the NN model. Besides, PCA could also help to remove the impacts of extreme values on the whole datasets, and improve the generalization ability of the model[51]. This is the information flow and PCA-based neural network (IF-PCA-NN). The workflow is summarized in the following algorithm.

Algorithm I Data Preprocessing

Input: Grid data according to grouped latitudes and longitudes from the orbital data, time instants i for interpolation.

Step 1: Let $j = i - 20$. Interpolate linearly the j^{th} date of the grid data into the k^{th} group of the orbital data, save the interpolated data into the K^{th} cell, where

$$k = \mathrm{mod}(j, 10),$$

$K = \mathrm{mod}(j, 10) + [\mathrm{floor}(j/10) - (\mathrm{floor}((i-20)/10))]\mathrm{X}10,$

with mod and floor being the remainder function and round-down function, respectively.

Step 2: If $j < i + 19$

  $j = i + 1$ and go back to Step 1;

 else

  if $K > 39$

   $j = 1$ and go to Step 3;

  else

   Output

 end if

end if

Step 3: While $j < 5$

  take the $(10j)^{\text{th}}$ - $[(K-39)+10j]^{\text{th}}$ cell to the$[10j-10]^{\text{th}}$-$[(K-39)+10j-11]^{\text{th}}$

  increment j

 end while

Step 4: Reshape the data in cells into a vector $\left(\boldsymbol{o}_{N_{obs}}\right)_i$, where N_{obs} is the number of the orbital data.

Output: $\left(\boldsymbol{o}_{N_{obs}}\right)_i$ with a label $k(=\mathrm{mod}(i, 10))$

Algorithm II provides the details of the IF-PCA-NN). When such a model is built with training samples, the SSH reconstruction now becomes easy. With the new along-track data coming in, we just need to use the output parameters to generate the desired gridded data. The procedure is as follows. First, preprocess the input data using Algorithm I. Second, use the causal indices $\boldsymbol{ind}$ to eliminate the non-causal observations. Third, perform PCA for the observations and obtain the time series of $\boldsymbol{PC_O}$. Fourth, put the time coefficients of the causal observation PCs into the network sets respectively and calculate the counterpart of the grid data. Finally, the grid data are reconstructed with the $\boldsymbol{PC_G}$ (PCs of grid data in the training set) and the corresponding time series $\left[\widehat{\boldsymbol{A}}_G\right]_{M_1 \times n_1}$ estimated with IF-PCA-NN.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140398, IEEE Transactions on Geoscience and Remote Sensing

Algorithm II: IF-PCA-NN Training

Input: The grid data $\boldsymbol{G}_{N_G\times n}$, along-track data $\boldsymbol{O}_{N_{obs}\times n}=\left\{\left(\boldsymbol{o}_{N_{obs}}\right)_i\right\}(i=1,\ldots,n)$, where N_G /N_{obs} is the grid/observation number and n the sample size

Step 1: Perform PCA for $\boldsymbol{G}$, get the matrix of the first M_G PCs $[\boldsymbol{PC_G}]_{N_G\times M_1}$, where the variance of the M_G PCs is greater than or equal to 99% of the total variance, and the matrix of the time series coefficients of the M_G PCs $[\boldsymbol{A_G}]_{M_G\times n}=\left[(\boldsymbol{a_G})_{\boldsymbol{g}}\right]_{M_1}$, where $(\boldsymbol{a_G})_{\boldsymbol{g}}$ is the gth PC's time series.

Step 2: Start with $g=1$

Step 3: Using **Eq. (3)**, calculate the Liang-Kleeman IF from the along-track data $\boldsymbol{O}$ to the g^{th} grid data PC's time series $(\boldsymbol{a_G})_{\boldsymbol{g}}$, perform significance test (at a 0.01 level), and select N_g along-track data $[\boldsymbol{O_g}]_{N_g\times n}(\in \boldsymbol{O}_{N_{obs}\times n})$ that are significantly causal, and their index $\boldsymbol{ind}_g$.

Step 4: Same as **Step 1**, but with along-track data $\boldsymbol{O_g}$. Obtain the matrix of the first $M_{o,g}$ PCs$[\boldsymbol{PC_{O,g}}]_{N_{obs}\times M_{O,g}}$, where the $M_{o,g}$ PCs contribute 99% of the total variance, and the matrix of the time series coefficients of the $M_{o,g}$ PCs $[\boldsymbol{A_{o,g}}]_{M_{o,g}\times n}$, where $(\boldsymbol{A_{o,g}})_{M_{o,g}}$ is the $M_{o,g}{}^{\text{th}}$ PC's time series of $\boldsymbol{O_g}$.

Step 5: Put $[\boldsymbol{A_{o,g}}]$ and $(\boldsymbol{a_G})_g$ into the neural network model in Section II.B, save the best network structure and weights $BNN_g=\{bnn\}_g$.

Step 6: If $g\le M_g$ then
  $g++$ (increment g)
  return **Step 3**
 else
  Output

Output: PCs of the grid data $\boldsymbol{PC_G}$, causal indices $\boldsymbol{ind}=\{\boldsymbol{ind}_g\}(g=1,\ldots,M_1)$, groups of the observation PCs $\boldsymbol{PC_O}=\{\boldsymbol{PC_{O,g}}\}(g=1,\ldots,M_1)$ (where the $\boldsymbol{PC_{O_g}}$ is the PC matrix of the set consisting of the along-track data which are causal to the g^{th} PC's time series $(\boldsymbol{a_G})_{\boldsymbol{g}}$ of the grid data $\boldsymbol{G}$) and of the grid data PCs $\boldsymbol{PC_G}==\{\boldsymbol{PC_g}\}(g=1,\ldots,M_G)$, and the neural network sets BNN.

V. Results

A. OAed map

As Section II.B shows that, the OA method only needs the locations of gridded data and observation points, as well as the observed values. In Section IV.A, we assume that the HYCOM outputs as ground truth and construct the along-track data by the outputs. So $\mathfrak{E}$ is equal to 0 here. The covariance function is hence:

$$C(dx,dt)=e^{-dx/a}e^{-dt^2/T^2}$$

where dx and dt are the spatial and temporal distance respectively, the parameter $a=110\ km$ and $T=20\ day$. The estimation of the gridded SSH is realized with (5).

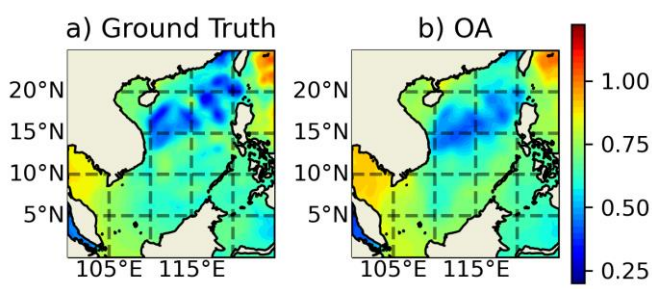

Fig. 3. SSH distribution in the South China Sea on March 26, 2020. a) Ground truth; b) OA results.

Fig. 3 shows the HYCOM ground truth (Fig. 3a) and the OAed map (Fig. 3b) of SSH in the South China Sea on March 26, 2020. For the South China Sea, we take the whole along-track data in the region $[0-25°N, 98-130°E]$, which contains 12792 observations in 40 days (from March 7 to April 15, 2020), with the OA method to map the grid data. From the ground truth, there are obvious low SSH zones between the area southwest of Luzon Strait and the region off Vietnam, and, particularly, four mesoscale cyclonic eddies with different sizes. The OA map has roughly captured the large feature, but with only two eddies with limited amplitude are reconstructed. In particular, the strength of vortex off Vietnam is by far underestimated.

B. Information Flow-Based Neural Network

Region A $[15.28-20°N, 108-112.72°E]$ in Fig. 1 is the northern branch of the South China Sea where the west boundary current (SCSWBC) dominates. Just like the two famous boundary currents namely Kuroshio and Gulf Stream, the SCSWBC is associated with a lot mesoscale eddies, with significant seasonal variabilities (e.g., [51]–[53]). This makes the SSH reconstruction for this region a big challenge.

TABLE I
PERIODS COVERED BY THE TRAINING, VALIDATION, AND TESTING DATASETS.

Datasets	Beginning date	Ending date	Sample number
Training	01/01/1995	12/31/2016	8036
Validation	01/01/2017	12/31/2018	730
Testing	01/01/2019	12/31/2020	731

For region A, we expand the interpolation region by 5 degrees and call it the "observation region" (blue block in Fig. 1). For each sample if we are targeted to map at time step t_0, there are 3812 along-track observations during t_{-19} to t_{20}.As Table 1 shows, the samples are divided by date into training sets (8036 groups from January 1, 1995 to December 31, 2016), validation sets (730 groups from January 1, 2017 to December 31, 2018) and testing sets (731 groups from January 1, 2019 to December 31, 2020). Input these above sets into the model, and iterate until the validation loss reaches its minimum without updating through the next consecutive 100 iterations. Then, the model with the minimum validation loss is taken as the optimal one.

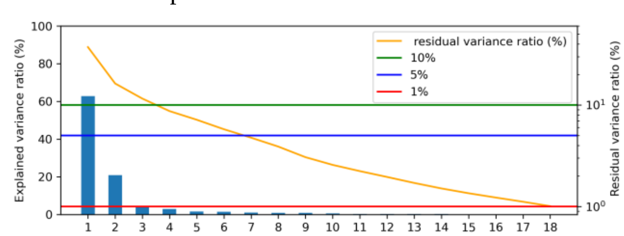

Fig. 4. The explained (bar, left ordinate) and residual (yellow line, right ordinate in logarithm) variance ratios for the first 18 PCs of the grid data in the training set. The green, blue, and red lines denote the 10%, 5%, and 1% of the residual variance ratios (right axis), respectively.

In order to reduce the dimension of the model outputs, we first apply PCA to the grid data in the training set. As shown in Fig. 4, the eigenvectors of the first four PCs have been able to capture the basic features in region A (91.49% by variance). We choose the first 18 PCs, which make 99% of the total

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140398, IEEE Transactions on Geoscience and Remote Sensing

variance. Next, AlgorithmII is used to train the IF-PCA-NN model, and the mapping from the along-track data to the different PCs series are obtained successively. Take the first PC training process of IF-PCA-NN model as an example (Fig. 5). No matter which set it is with (training/ verification/ test set), the loss function generally decreases with iteration. The loss with the verification set does not reach a minimum (within the first 157 echoes) until the training is over 257 echoes. The deviations of the loss functions with the testing set and that of the verification set are roughly the same, implying that the samples of the training and verification sets have embedded all the variabilities in region A.

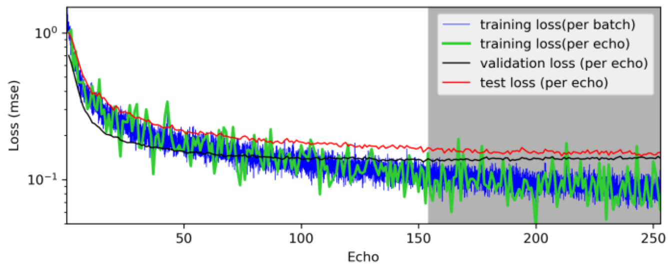

Fig. 5. The evolution of the loss functions for the first PC with different sets. The blue line indicates the training set (per batch), while the green, red, and black lines are for is the training, validation, and testing sets, respectively (per echo).

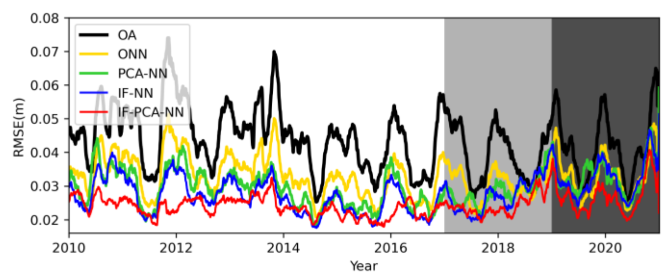

Fig. 6. The time series of the spatial average over region A of the RMSEs of SSH among the three experiments (unit: m; OA method, black line; ONN method, yellow line; PCA-NN method, green line; IF-NN method, blue line; IF-PCA-NN, red line). Lightly shaded is the period over which the validation set is formed, and the period with dark shading is for test set.

To illustrate the role of IF and PCA in the IF-PCA-NN model, we also design three NN-related experiments: PCA-NN, without IF (skip **Step 3** in Algorithm 2); IF-NN, without PCA (skip **Steps 1** and **4** in Algorithm 2); ONN, original NN, without IF and PCA (skip **Steps 1**,**3**, and **4** in Algorithm 2). Fig. 6 is the time series of the spatial average of RMSE of the SSH among the five experiments (black--OA; yellow--ONN; green--PCA-NN; blue--IF-NN; red—IF-PCA-NN). It can be seen that in the training set, the RMSEs of all four NN-related results are significantly smaller than that of the OA result (4.49 cm). Their RMSEs, from largest to smallest, are: 3.34 cm (ONN), 2.92 cm (PCA-NN), 2.65 cm (IF-NN) and 2.33 cm (IF-PCA-NN). For the validation set and the test set, the RMSE with the OA method is still around 4 cm (3.97 and 4.20 cm), which is consistent with the training set by a 99% F-test. For the NN-related models, their RMSEs increase slightly. For example, the best performed IF-PCA-NN increases to 2.44 (validation set) and 2.69 cm (test set). But anyhow, these RMSEs are still significantly lower than that with the OA method.

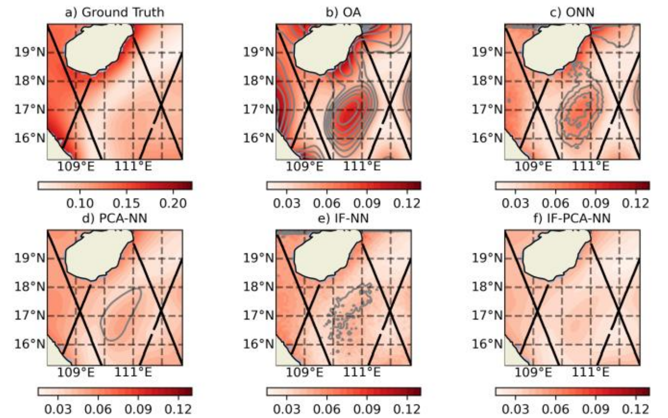

Fig. 7. The spatial distribution of the RMSE over test set (2019-2020). a) The standard deviation of the HYCOM ground truth; the RMSEs (shaded, unit: m) and the normalization (>50% are marked by contour lines) of b) OA; c) ONN; d) PCA-NN; e) IF-NN; f) IF-PCA-NN. The black lines mark the trajectory of Jason-2.

To better show how the difference of the RMSEs with the five methods arises, we calculate the spatial distribution of the RMSE during the test set period (Fig. 7). Fig. 7a is the standard deviation for the HYCOM ground truth (Fig. 7a). It can be found that in the west of the region, the along-track data is relatively sparse and the standard deviation of SSH is large. These account for the large RMSEs (shaded) in this region for the results with all the five methods (Fig. 7b-f). Another observation is that there is a maximum region between the area from Hainan Island to the offshore of Vietnam. It is located at gap of the Jason-2's orbits, and hence makes sense. Compared to the OA result, the maxima on all the NN-related maps are significantly weaker. Clearly NN-related models can effectively improve the problem caused by the gap. Among these NNs, the ONN has the maximum RMSE and obvious noise in distribution, while PCA-NN and IF-NN have their RMSEs significantly reduced in the gap area. Most notably, the IF-PCA-NN shows a very excellent performance, whose RMSE in the whole region can be further

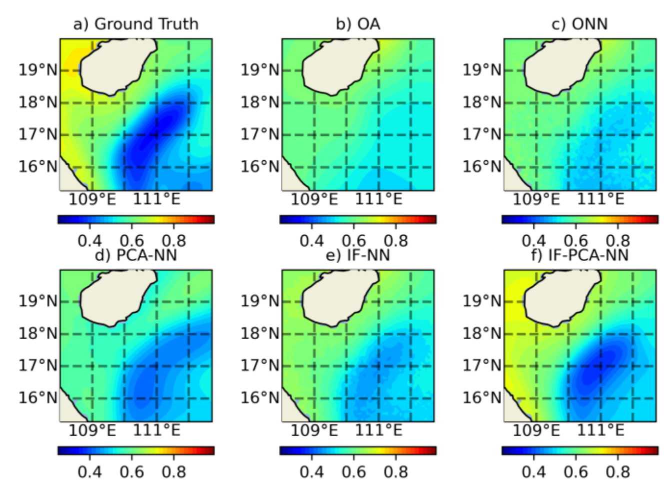

Fig. 8. The sea surface height (unit: m) for region A on March 26, 2020 where a) is the HYCOM original field, b-f) are the OA, ONN, PCA-NN, IF-NN, and IF-PCA-NN results, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140398, IEEE Transactions on Geoscience and Remote Sensing

reduced.

In order to better understand the reason for the performance differences, we choose an eddy event occurring on March 26, 2020, for demonstration. As shown in Fig. 8, overall, the SSH has a pattern with an increasing trend overall from east to west. A conspicuous feature is an isolated cyclonic eddy southeast off Hainan Island. This eddy lies between the tracks (cf. Fig. 1), and hence poses a great challenge for interpolation. For this reason, the SSH reconstruction for region A makes an ideal testing example for our proposed algorithms.

Figs. 8b-f are the reconstructed results with the five methods. They all can have the general trend reconstructed. However, when going to details, OA and ONN are far less accurate than the other three methods. For example, the OA and ONN reconstructions obviously underestimate the positive anomaly (especially OA) on the western side of Hainan Island. Moreover, as expected, OA fails to reveal the cyclonic eddy. By comparison, ONN performs slightly better; it captures the fragmentary low values, though with noise. The PCA-NN and IF-NN do capture the eddy, but the orientation and geometry of the eddy are not as those in Fig. 8a. The pattern resulting from the IF-NN result is closer to the ground truth but is noisier than that from the PCA-NN. In contrast, the IF-PCA-NN method yields a rather appealing result, with the strength, size, shape, and orientation rather satisfactorily reconstructed. If compared quantitatively, the mean deviations and RMSEs with OA, ONN, PCA-NN, IF-NN, and IF-PCA-NN with the test set are, respectively, 1.69, -1.29, 0.95, 0.79, 0.72 cm and 8.22, 6.88. 6.64, 5.13, 2.50 cm. As the distribution of the deviation shows in Fig. 9, the isolated cyclonic eddy is exactly located within the gap of the Jason-2's orbits, where the maximum OA error results exist. As expected, OA cannot reconstruct mesoscale structure(s) lying in between the surrounding observations. In contrast, the deviations of the NN-related results within the gap, especially the causality-based NN (IF-NN and IF-PCA-NN), are much smaller.

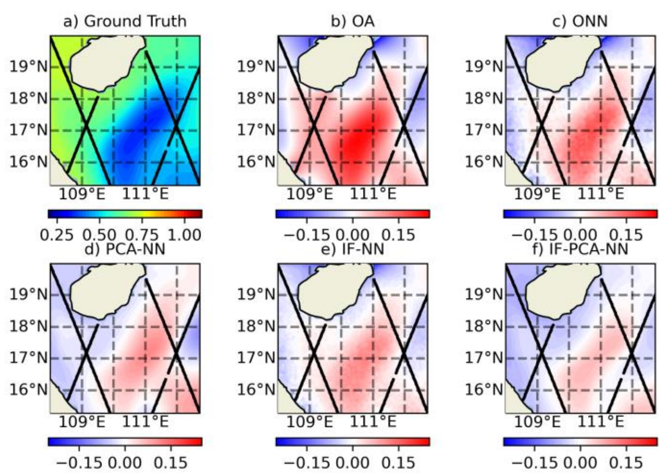

Fig. 9. Same as Fig. 8 but b-f) are the deviations of the OA, ONN, PCA-NN, IF-NN, and IF-PCA-NN reconstructions from the ground truth. The black lines mark the trajectory of Jason-2.

Fig. 10 tracks the evolution of the eddy southeast off Hainan Island. It can be clearly seen that the eddy propagates from east to west and moves southward toward the area off Vietnam. This process is observed successively by different Janson-2 tracks within 20 days around March 26, 2020 (from March 16 to April 10, 2020). Although there are no observations in the area where the eddy appears on March 26, 2020, IF-NN can reconstruct the mapping from observations to the grid data by learning the intrinsic dynamic relationships between the observations at different times, and thus reveal to us the eddy activity which would otherwise unobserved.

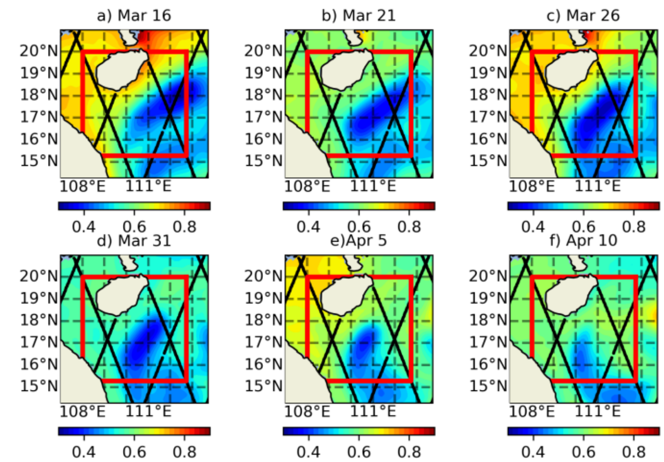

Fig. 10. The evolution of the eddy southeast off Hainan Island from March 16 to April 10, 2020.

VI. Conclusion

The advent of satellite altimetry datasets of sea surface height (SSH) has set a milestone in the advancement of oceanography and other earth system sciences. But, while the along-track data coverage is dense, the relatively poor resolution between tracks poses a great challenge to the reconstruction of those processes such as mesoscale and sub-mesoscale eddies. In this study, a machine learning algorithm has been developed to address this change, with a neural network model combined with a causal inference technique based on the information flow (IF) analysis[28], [29]. By discarding the redundant observations and reducing the dimension in the sample, the IF analysis and principal component analysis are used respectively in order to construct a simple but nonlinear mapping from the along track-data to the gridded data. It is shown that, with such an IF-based neural network model, the characteristics of the grid data can be basically restored from the along-track data. As a demonstration, we picked an area in the South China Sea, where mesoscale and submesoscale eddies frequently appear but without satellite observations. By training the model as proposed above using the HYCOM data as ground truth, we immediately had the desired eddies reconstructed, which agree remarkably with the ground truth in strength, geometry, and orientation.

This study provides a new way of thinking for remote sensing data reconstruction. Traditional algorithms are based on a mapping method with some specific, preset model, while

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TGRS.2022.3140398, IEEE Transactions on Geoscience and Remote Sensing

the neural network method allows for the computer to find the intrinsic relationship between inputs (observations) and outputs (gridded data), so as to build a mapping without relying on any model given a priori. By helping remove irrelevant factors, causal inference can speed up the process and improve the performance of the neural network to search for the best functional form for such a mapping. This technique to identify covariates is expected to play an important role in machine learning and artificial intelligence in the future.

It should be noted that the model training of the above method relies on 2D gridded fields, while usually we have only along-track data. Nonetheless, with the accumulation of more and more observational and modeling data, more accurate historical reanalysis gridded data will be made available for the training purpose and, henceforth, the method is expected to show its power.

Acknowledgment

We thank HYCOM and AVISO for making the data available, and the NUIST High Performance Computing Center for providing computational resources.

References

[1] J. Isern-Fontanet, J. Ballabrera-Poy, A. Turiel, and E. García-Ladona, “Remote sensing of ocean surface currents: a review of what is being observed and what is being assimilated,” *Nonlinear Process. Geophys.*, vol. 24, no. 4, pp. 613–643, Oct. 2017, doi: 10.5194/npg-24-613-2017.

[2] L. A. Edwards, R. E. Houseago-Stokes, and P. Cipollini, “Altimeter observations of the MJO/ENSO connection through Kelvin waves,” *Int. J. Remote Sens.*, vol. 27, no. 6, pp. 1193–1203, Mar. 2006, doi: 10.1080/01431160500380182.

[3] K. S. Casey and D. Adamec, “Sea surface temperature and sea surface height variability in the North Pacific Ocean from 1993 to 1999,” *J. Geophys. Res. Oceans*, vol. 107, no. C8, pp. 14-1-14–12, 2002, doi: 10.1029/2001JC001060.

[4] C.-R. Wu, “Interannual modulation of the Pacific Decadal Oscillation (PDO) on the low-latitude western North Pacific,” *Prog. Oceanogr.*, vol. 110, pp. 49–58, Mar. 2013, doi: 10.1016/j.pocean.2012.12.001.

[5] M. Saraceno, C. G. Simionato, and L. A. Ruiz-Etcheverry, “Sea surface height trend and variability at seasonal and interannual time scales in the Southeastern South American continental shelf between 27°S and 40°S,” *Cont. Shelf Res.*, vol. 91, pp. 82–94, Dec. 2014, doi: 10.1016/j.csr.2014.09.002.

[6] J. Song, C. Li, and W. Zhou, “High and low latitude types of the downstream influences of the North Atlantic Oscillation,” *Clim. Dyn.*, vol. 42, no. 3, pp. 1097–1111, Feb. 2014, doi: 10.1007/s00382-013-1844-3.

[7] S. K. Behera *et al.*, “Paramount impact of the Indian Ocean dipole on the east african short rains: a CGCM study,” *J. Clim.*, vol. 18, no. 21, pp. 4514–4530, Nov. 2005, doi: 10.1175/JCLI3541.1.

[8] M. J. McPhaden and M. Nagura, “Indian Ocean dipole interpreted in terms of recharge oscillator theory,” *Clim. Dyn.*, vol. 42, no. 5, pp. 1569–1586, Mar. 2014, doi: 10.1007/s00382-013-1765-1.

[9] R. Knutti and T. F. Stocker, “Influence of the thermohaline circulation on projected sea level rise,” *J. Clim.*, vol. 13, no. 12, pp. 1997–2001, Jun. 2000, doi: 10.1175/1520-0442(2000)013<1997:IOTTCO>2.0.CO;2.

[10] G. A. Meehl *et al.*, “How much more global warming and sea level rise?,” *Science*, 2005, Accessed: Oct. 31, 2020. [Online]. Available: https://agris.fao.org/agris-search/search.do?recordID=US201300982452

[11] A. S. Gardner *et al.*, “A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009,” *Science*, vol. 340, no. 6134, pp. 852–857, May 2013, doi: 10.1126/science.1234532.

[12] F. P. Bretherton, R. E. Davis, and C. B. Fandry, “A technique for objective analysis and design of oceanographic experiments applied to MODE-73,” *Deep Sea Res. Oceanogr. Abstr.*, vol. 23, no. 7, pp. 559–582, Jul. 1976, doi: 10.1016/0011-7471(76)90001-2.

[13] T. Kuragano and A. Shibata, “Sea surface dynamic height of the pacific ocean derived from TOPEX/POSEIDON altimeter data: Calculation method and accuracy,” *J. Oceanogr.*, vol. 53, no. 6, Art. no. 6, 1997.

[14] N. Ducet, P. Y. L. Traon, and G. Reverdin, “Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2,” *J. Geophys. Res. Oceans*, vol. 105, no. C8, pp. 19477–19498, 2000, doi: 10.1029/2000JC900063.

[15] R. Morrow and P.-Y. Le Traon, “Recent advances in observing mesoscale ocean dynamics with satellite altimetry,” *Adv. Space Res.*, vol. 50, no. 8, pp. 1062–1076, Oct. 2012, doi: 10.1016/j.asr.2011.09.033.

[16] P. Y. Le Traon, F. Nadal, and N. Ducet, “An improved mapping method of multisatellite altimeter data,” *J. Atmospheric Ocean. Technol.*, vol. 15, no. 2, pp. 522–534, Apr. 1998, doi: 10.1175/1520-0426(1998)015<0522:AIMMOM>2.0.CO;2.

[17] P. Y. L. Traon and G. Dibarboure, “Mesoscale mapping capabilities of multiple-satellite altimeter missions,” *J. Atmospheric Ocean. Technol.*, vol. 16, no. 9, pp. 1208–1223, Sep. 1999, doi: 10.1175/1520-0426(1999)016<1208:MMCOMS>2.0.CO;2.

[18] J. Verron *et al.*, “The benefits of the ka-band as evidenced from the SARAL/AltiKa altimetric mission: scientific applications,” *Remote Sens.*, vol. 10, no. 2, Art. no. 2, Feb. 2018, doi: 10.3390/rs10020163.

[19] C. Ubelmann, P. Klein, and L. Fu, “Dynamic interpolation of sea surface height and potential applications for future high-resolution altimetry mapping,” *J. ATMOSPHERIC Ocean. Technol.*, vol. 32, pp. 177–184, 2014.

[20] C. A. S. Tanajura, L. N. Lima, and K. Belyaev, “Impact on oceanic dynamics from assimilation of satellite

surface height anomaly data into the Hybrid Coordinate Ocean Model Ocean Model (HYCOM) over the Atlantic Ocean," *Oceanology*, vol. 56, no. 4, pp. 509–514, Jul. 2016, doi: 10.1134/S000143701603022X.

[21] J. G. Carbonell, J. Anderson, R. S. Michalski, R. Banerjee, T. M. Mitchell, and J. Mostow, *Machine learning: an artificial intelligence approach*. Palo Alto, CA: Tioga, 1983.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, Art. no. 6088, Oct. 1986, doi: 10.1038/323533a0.

[23] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[25] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, Nov. 1958, doi: 10.1037/h0042519.

[26] V. Krasnopolsky, C. J. Lozano, D. Spindler, I. Rivin, and D. B. Rao, "Using neural network to enhance assimilating sea surface height data into an ocean model," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, Jul. 2006, pp. 4299–4301. doi: 10.1109/IJCNN.2006.247004.

[27] X. S. Liang and R. Kleeman, "Information transfer between dynamical system components," *Phys. Rev. Lett.*, vol. 95, no. 24, p. 244101, Dec. 2005, doi: 10.1103/PhysRevLett.95.244101.

[28] X. S. Liang, "Information flow and causality as rigorous notions ab initio," *Phys. Rev. E*, vol. 94, no. 5, p. 052201, Nov. 2016, doi: 10.1103/PhysRevE.94.052201.

[29] X. S. Liang, "Unraveling the cause-effect relation between time series," *Phys. Rev. E*, vol. 90, no. 5, p. 052150, Nov. 2014, doi: 10.1103/PhysRevE.90.052150.

[30] X. S. Liang, "Information flow within stochastic dynamical systems," *Phys. Rev. E*, vol. 78, no. 3, p. 031113, Sep. 2008, doi: 10.1103/PhysRevE.78.031113.

[31] X. S. Liang, "Causation and information flow with respect to relative entropy," *Chaos Interdiscip. J. Nonlinear Sci.*, vol. 28, no. 7, p. 075311, Jul. 2018, doi: 10.1063/1.5010253.

[32] X. S. Liang, "The Liang-Kleeman information flow: theory and applications," *Entropy*, vol. 15, no. 1, pp. 327–360, Jan. 2013, doi: 10.3390/e15010327.

[33] A. Stips, D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. S. Liang, "On the causal structure between CO2 and global temperature," *Sci. Rep.*, vol. 6, p. 21691, Feb. 2016, doi: 10.1038/srep21691.

[34] H. Xiao, F. Zhang, L. Miao, X. S. Liang, K. Wu, and R. Liu, "Long-term trends in Arctic surface temperature and potential causality over the last 100 years," *Clim. Dyn.*, vol. 55, no. 5–6, pp. 1443–1456, Sep. 2020, doi: 10.1007/s00382-020-05330-2.

[35] C. Bai, R. Zhang, S. Bao, X. S. Liang, and W. Guo, "Forecasting the tropical cyclone genesis over the Northwest Pacific through identifying the causal factors in cyclone–climate interactions," *J. Atmospheric Ocean. Technol.*, vol. 35, no. 2, pp. 247–259, Feb. 2018, doi: 10.1175/JTECH-D-17-0109.1.

[36] G. Wang, C. Zhao, M. Zhang, Y. Zhang, M. Lin, and F. Qiao, "The causality from solar irradiation to ocean heat content detected via multi-scale Liang–Kleeman information flow," *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Oct. 2020, doi: 10.1038/s41598-020-74331-2.

[37] H. Zhang, Z. Qiu, D. Sun, S. Wang, and Y. He, "Seasonal and interannual variability of satellite-derived Chlorophyll-a (2000–2012) in the Bohai Sea, China," *Remote Sens.*, vol. 9, no. 6, p. 582, Jun. 2017, doi: 10.3390/rs9060582.

[38] D. F. Hagan, G. Wang, X. S. Liang, and H. A. J. Dolman, "A time-varying causality formalism based on the Liang–Kleeman information flow for analyzing directed interactions in nonstationary climate systems," *J. Clim.*, vol. 32, no. 21, pp. 7521–7537, Nov. 2019, doi: 10.1175/JCLI-D-18-0881.1.

[39] X. S. Liang, "Normalizing the causality between time series," *Phys. Rev. E*, vol. 92, no. 2, p. 022126, Aug. 2015, doi: 10.1103/PhysRevE.92.022126.

[40] Y. Rong and X. S. Liang, "Panel data causal inference using a rigorous information flow analysis for homogeneous, independent and identically distributed datasets," *IEEE Access*, pp. 1–1, 2021, doi: 10.1109/ACCESS.2021.3068273.

[41] D. T. Hristopulos, A. Babul, S. Babul, L. R. Brucar, and N. Virji-Babul, "Disrupted information flow in resting-state in adolescents with sports related concussion," *Front. Hum. Neurosci.*, vol. 13, 2019, doi: 10.3389/fnhum.2019.00419.

[42] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *ICML Workshop Deep Learn. Audio Speech Lang. Process.*, p. undefined-undefined, 2013.

[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," p. 30, 2014.

[44] X. Bouthillier, K. Konda, P. Vincent, and R. Memisevic, "Dropout as data augmentation," *ArXiv150608700 Cs Stat*, Jan. 2016, Accessed: Dec. 20, 2021. [Online]. Available: http://arxiv.org/abs/1506.08700

[45] D. P. Kingma and J. Ba, "A method for stochastic optimization," *ArXiv14126980 Cs*, Jan. 2017, Accessed: Nov. 01, 2020. [Online]. Available: http://arxiv.org/abs/1412.6980

[46] E. P. Chassignet *et al.*, "The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system," *J. Mar. Syst.*, vol. 65, no. 1, pp. 60–83, Mar. 2007, doi: 10.1016/j.jmarsys.2005.09.016.

[47] B. D. Beckley *et al.*, "Assessment of the Jason-2 extension to the TOPEX/Poseidon, Jason-1 sea-surface height time series for global mean sea level monitoring," *Mar. Geod.*, vol. 33, no. sup1, pp. 447–471, Aug. 2010, doi: 10.1080/01490419.2010.491029.

[48] R. E. Glazman, A. Fabrikant, and A. Greysukh, "Statistics of spatial-temporal variations of sea surface height based on Topex altimeter measurements," *Int. J.*

Remote Sens., vol. 17, no. 13, pp. 2647–2666, Sep. 1996, doi: 10.1080/01431169608949097.

[49] M. Imani, R.-J. You, and C.-Y. Kuo, “Caspian Sea level prediction using satellite altimetry by artificial neural networks,” *Int. J. Environ. Sci. Technol.*, vol. 11, no. 4, pp. 1035–1042, May 2014, doi: 10.1007/s13762-013-0287-z.

[50] F. R. S. K. Pearson, “LIII. On lines and planes of closest fit to systems of points in space,” *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901, doi: 10.1080/14786440109462720.

[51] P. Baldi and K. Hornik, “Neural networks and principal component analysis: Learning from examples without local minima,” *Neural Netw.*, vol. 2, no. 1, pp. 53–58, Jan. 1989, doi: 10.1016/0893-6080(89)90014-2.

[52] G. Fang, G. Wang, Y. Fang, and W. Fang, “A review on the South China Sea western boundary current,” *Acta Oceanol. Sin.*, vol. 31, no. 5, pp. 1–10, Sep. 2012, doi: 10.1007/s13131-012-0231-y.

[53] D. Wang *et al.*, “Progress of regional oceanography study associated with western boundary current in the South China Sea,” *Chin. Sci. Bull.*, vol. 58, no. 11, pp. 1205–1215, Apr. 2013, doi: 10.1007/s11434-012-5663-4.

[54] Q. Quan, H. Xue, H. Qin, X. Zeng, and S. Peng, “Features and variability of the South China Sea western boundary current from 1992 to 2011,” *Ocean Dyn.*, vol. 66, no. 6–7, pp. 795–810, Jul. 2016, doi: 10.1007/s10236-016-0951-1.

Yineng Rong is a Ph.D. candidate in meteorology at Nanjing University of Information Science and Technology. He is currently working on the development of causality-based machine learning algorithms with the Liang-Kleeman information flow, and their applications to problems in atmosphere-ocean science.

X. San Liang received his Ph.D. degree in Applied Mathematics from Harvard University, Massachusetts, USA. He has worked at the Second Institute of Oceanography of the State Oceanic Administration, Harvard University, Courant Institute, MIT, China Institute for Advanced Study, Central University of Finance and Economics, Nanjing Institute of Meteorology, and was a team leader of the Ninth Chinese National Antarctic Research Expedition. Presently he is a professor at Fudan University, with a joint appointment at Shanghai Qi Zhi (Andrew C. Yao) Institute, Shanghai, China. He is interested in a variety of interdisciplinary fields such as, not exclusively, information flow, causality analysis, causal AI, manifold learning, fluid dynamics, atmosphere-ocean-climate science, turbulence, complex networks, multiscale modeling and simulation, predictability, uncertainty dynamics, etc.