Article

Editor's Choice

Emulation of a Process-Based Salinity Generator for the Sacramento–San Joaquin Delta of California via Deep Learning

Minxue He, Liheng Zhong, Prabhjot Sandhu and Yu Zhou

MDPI

Emulation of a Process-Based Salinity Generator for the Sacramento–San Joaquin Delta of California via Deep Learning

Minxue He *, Liheng Zhong †, Prabhjot Sandhu and Yu Zhou

California Department of Water Resources, 1416 9th Street, Sacramento, CA 95814, USA; lihengzhong@berkeley.edu (L.Z.); Prabhjot.Sandhu@water.ca.gov (P.S.); Yu.Zhou@water.ca.gov (Y.Z.)

* Correspondence: kevin.he@water.ca.gov; Tel.: +1-916-651-9634

† Now at Descartes Labs, San Francisco, CA 94103, USA.

Abstract: Salinity management is a subject of particular interest in estuarine environments because of the underlying biological significance of salinity and its variations in time and space. The foremost step in such management practices is understanding the spatial and temporal variations of salinity and the principal drivers of these variations. This has traditionally been achieved with the assistance of empirical or process-based models, but these can be computationally expensive for complex environmental systems. Model emulation based on data-driven methods offers a viable alternative to traditional modeling in terms of computational efficiency and improving accuracy by recognizing patterns and processes that are overlooked or underrepresented (or overrepresented) by traditional models. This paper presents a case study of emulating a process-based boundary salinity generator via deep learning for the Sacramento–San Joaquin Delta (Delta), an estuarine environment with significant economic, ecological, and social value on the Pacific coast of northern California, United States. Specifically, the study proposes a range of neural network models: (a) multilayer perceptron, (b) long short-term memory network, and (c) convolutional neural network-based models in estimating the downstream boundary salinity of the Delta on a daily time-step. These neural network models are trained and validated using half of the dataset from water year 1991 to 2002. They are then evaluated for performance in the remaining record period from water year 2003 to 2014 against the process-based boundary salinity generation model across different ranges of salinity in different types of water years. The results indicate that deep learning neural networks provide competitive or superior results compared with the process-based model, particularly when the output of the latter are incorporated as an input to the former. The improvements are generally more noticeable during extreme (i.e., wet, dry, and critical) years rather than in near-normal (i.e., above-normal and below-normal) years and during low and medium ranges of salinity rather than high range salinity. Overall, this study indicates that deep learning approaches have the potential to supplement the current practices in estimating salinity at the downstream boundary and other locations across the Delta, and thus guide real-time operations and long-term planning activities in the Delta.

Keywords: salinity; deep learning; martinez boundary salinity generator; Sacramento–San Joaquin Delta

1. Introduction

Salinity has been long recognized as a critical environmental variable in estuaries which are transition zones between upstream freshwater environments and downstream saline marine environments [1]. The spatial and temporal variation pattern of salinity in estuaries plays a dominant role in the health of estuarine habitats and biota [2–7]. This pattern is typically influenced by drivers

including upstream freshwater inflow, downstream tidal forces, as well as local water diversions, precipitation, evaporation, and wind in the estuaries, among others. Understanding the variation pattern of salinity is the foremost step in predicting its future behavior and thus guiding salinity management in estuaries [8–12]. This is particularly the case for estuarine environments with paramount economic, ecological, and social significance including the Sacramento–San Joaquin Delta in California, United States.

As the fifth largest economy in the world, the state of California accommodates a population of nearly 40 million and is one of the most productive agricultural areas globally [13]. Reliable water supply is indispensable to support such a large population and sustain such a robust economy. However, the spatial and temporal distribution of precipitation, the largest water supply source for the state, largely mismatches water demands. Most of the state's population and farmlands (and thus water demand) is in the southern half, while most of the precipitation falls in the northern mountain ranges in the state. In addition, a majority of precipitation occurs in the wet winter season while the highest water demand is normally in the dry summer and fall. To balance the mismatch, a complex water storage and transfer system has been built in the state to redistribute water across different spatial and temporal scales. The most critical components of the system are the State Water Project (SWP) and the Central Valley Project (CVP) which are operated by the state and the federal governments, respectively. SWP and CVP infrastructure consist of tens of dams and reservoirs, pumping plants, hydro-power generation plants, and over 1000 km of aqueducts, tunnels, canals, and pipelines [14,15].

The hub of this statewide water redistribution system is the Sacramento–San Joaquin Delta (Delta, California, CA, USA). Physically, the delta is a patchwork of islands surrounded by about 1100 km of waterways (Figure 1). It receives freshwater from the largest two rivers in the state, namely Sacramento River on the north and San Joaquin River and its tributaries on the south-east. Freshwater inflows are either diverted to water users within and outside of the delta or serve to repel seawater intrusion from its downstream boundary at Martinez (Figure 1). Ecologically, the delta is a globally important biodiversity hot spot with the highest priority of conservation [16]. It provides habitats that support about 750 species of plants and animals including some near extinction [17]. Socioeconomically, the delta provides water to about two thirds of the state's population and over 15,000 km^2 of farmlands via SWP and CVP deliveries. Millions of people use the delta for recreation and transportation [18]. These physical, ecological, and socioeconomic features of the delta drive SWP and CVP operations with the coequal goals of a reliable water supply and an ecologically sustainable delta ecosystem [19]. The SWP and CVP pump water from southern delta (Figure 1) and transfer the water to municipal and agricultural users in the state. The pumping time and rates are dictated by state and federal regulatory requirements to ensure that: (1) flow and water quality standards at various locations in the delta are complied with, and that (2) additional regulations to protect endangered species are followed [20,21]. One critical water quality standard is that the salinity level, typically reported in units of electrical conductivity (EC) as microSiemens/cm (µs/cm), at compliance locations cannot exceed preset threshold values during certain periods in a certain type (e.g., wet, dry, critical) of water year. Traditionally empirical and process-based models have been used to simulate salinity variations at these locations to guide real-time delta operations (e.g., SWP and CVP operations) and long-term planning studies (e.g., structural changes to the delta) to ensure salinity compliance.

One of the earliest models developed for this purpose is the conceptual–empirical salinity gradient model (i.e., G-model) of [22]. The model derives salinity from antecedent delta outflow based on the assumption that there is a non-linear relationship between these variables. Hutton et al. [11] extended the G-model to simulate the low salinity zone in the delta defined as the position of a predetermined salinity isohaline. Numerical process-based models have also been developed to simulate the spatial and temporal variation of salinity in the delta. These models include, to name a few, but are not limited to, the one-dimensional Delta Simulation Model II (DSM2) [23], two-dimensional RMA10 [24] and TRIM2D [25], and three-dimensional SCHISM [26,27], UnTrim [28,29], and SUNTANS [30,31]. Although physically more rigorous and being able to provide higher spatial

resolution on flow and salinity distribution than one-dimensional models, multi-dimensional models are computationally expensive. For studies with long data temporal sequences and multiple scenarios involved, simpler one-dimensional models are still favored. This is the case for DSM2 which is widely applied in contemporary applications in the delta [32]. The DSM2 domain covers the entire delta with Martinez as its downstream boundary. For operational planning and forecasting studies, DSM2 relies on the Martinez Boundary Salinity Generator (MBSG) [33] to produce downstream salinity boundary conditions for DSM2 simulations.
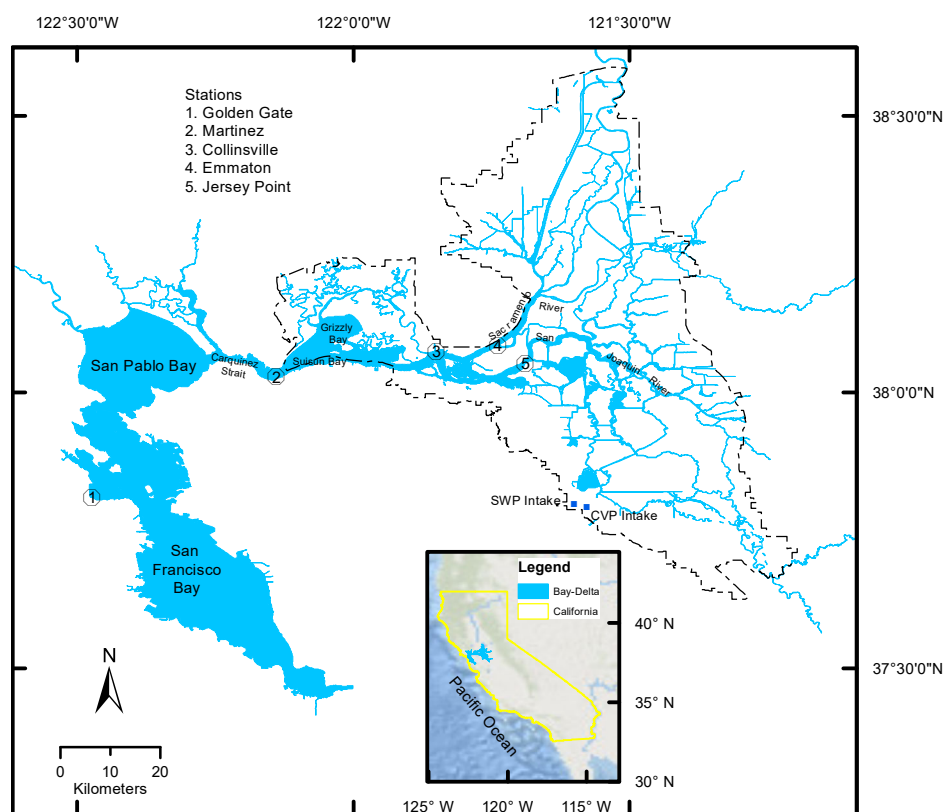

Figure 1. Location map showing the San Francisco Bay and Sacramento–San Joaquin Delta (Bay-Delta) Estuary and the study location Martinez along with several other key water quality stations including the Golden Gate, Collinsville, Emmaton, and Jersey Point. The dash line highlights the modeling domain of the Delta Simulation Model II (DSM2). The insert map illustrates the location of the Bay-Delta in California.

In addition to empirical and process-based models, data-driven models including Artificial Neural Networks (ANNs) have also been explored in deriving salinity in the delta area [34–39]. An ANN employs a mathematical network structure to implicitly identify the relationships between one or more inputs (e.g., usually measured variables such as stage or flow) and outputs (e.g., salinity at selected locations) datasets. The basic processing units in the network are called neurons. These neurons are arranged in layers and are connected to other neurons in adjacent layers. Multilayer perceptrons (MLPs) are probably the most popular ANN models applied in the field of water resources engineering [40]. The above-mentioned ANN studies generally use MLP-based models in estimating salinity in the Delta. An MLP is a feedforward ANN typically consisting of two visible layers on both ends of the network (i.e., input and output layers) and one or more hidden layers in the middle. A neuron in a specific layer takes inputs from neurons in the previous layer and outputs a linear or non-linear transformation of the combined input information to neurons in the next layer. The connections between neurons are represented by linear weights. These weights are determined in the training

process by minimizing the difference (i.e., error signal) between network predictions of the variable of interest (e.g., salinity) and the corresponding observations. The most common training method is gradient descent, which propagates the difference backward into the network and updates the weights according to the chain rule [41].

Despite their popularity, MLPs do not treat the sequential ordering of input time series as a feature during training. For non-linear systems where short- or long-term temporal dependencies exist between output and input (e.g., salinity at the current time relates to antecedent flows), MLPs may not be the most viable choice [42]. Recurrent Neural Networks (RNNs), a different category of ANN, are designed to overcome this drawback of MLPs [43]. RNNs process input in its temporal order. The output of hidden layer neurons at each time step is recurrently fed as an additional input to the next time step. This feature grants RNNs the advantage of better understanding the temporal dynamics between input and output variables. In spite of this advantage, standard RNN models are shown to have difficulties in capturing long-term dependencies [44]. This is mainly caused by two problems encountered during the training process: vanishing gradient (network weights approach zero) and exploding gradient (network weights become extremely large). These two problems occur mostly because the error signal can only be back propagated effectively for a few steps [45].

Many variants of RNNs have been proposed to avoid the vanishing and exploding gradient problems. The most well-known and successful variant is the long short-term memory (LSTM) network [45]. LSTM introduces the concept of gates, which are essentially neurons with learnable weights. The gates inform the network on what information to discard, what to retain, and for how long. This gate configuration helps the network preserve essential information over a long time and avoid rapid error signal decay. LSTM networks have only been applied recently in the field of water resources in terms of modeling rainfall-runoff process [46–49], groundwater table [50,51], water level in channels [52], water quality [53,54], and reservoir operations [55]. Given the long-term dependencies between salinity and flow/stage, LSTM should also be suitable for salinity simulation given flow and stage inputs.

Convolutional neural network (CNN) is another category of ANNs. A CNN consists of a sequence of layers that shrink in length from the input layer to the output layer. The shrinking aims to condense information learnt from previous layers to more abstract concepts in deep layers [42]. CNN is a leading network architecture in deep learning techniques and has had extremely successful applications in image pattern recognition and classification [56–59]. It has only recently received significant attention in water-related time series modeling in terms of ground water level prediction [60], precipitation estimation [61,62], and flood forecasting [63,64].

To our knowledge, few studies have explored the applicability of state-of-the-art deep learning techniques (e.g., LSTM, CNN) in salinity estimation for deltaic and/or estuarine environments, not to mention in the Sacramento–San Joaquin Delta specifically. This study explores the capacity of these techniques by presenting a case study of emulating the Martinez Boundary Salinity Generator in estimating the downstream salinity boundary (i.e., Martinez salinity) for the delta via LSTM- and CNN-based models. The salinity generator itself is applied as the benchmark model. A number of MLP-based neural network models are also proposed for comparison purpose. The rest of the paper is organized as follows: Section 2 describes the study area, the available dataset, the Martinez Salinity Boundary Generator, the proposed neural network models, and the evaluation metrics; Section 3 presents the results and findings; Section 4 discusses data stationarity, study limitations, implications, and future work; the last section concludes the paper.

2. Materials and Methods

2.1. Study Location and Dataset

The Sacramento–San Joaquin Delta (Delta) is the hub of California's vast water supply system with critical urban, agricultural, environmental, industrial, and recreational importance. The Delta is an estuary at the confluence of the largest two river systems in California, the rain-dominated Sacramento River in the north and snow-dominated San Joaquin River in the south (Figure 1). These two rivers and their tributaries contribute freshwater to the Delta. Freshwater is the main source for Delta diversions (including SWP and CVP exports) and Delta consumptive use. It is also used to repel the intrusion of seawater which enters the San Francisco Bay and San Pablo Bay via the Golden Gate (Station 1 in Figure 1). The downstream boundary of the Delta, Martinez (Station 2 in Figure 1), is connected to the San Pablo Bay via the Carquinez Strait. Martinez is under strong influence of salty tides from the San Pablo Bay. Salinity at Martinez serves as the salinity boundary for the Delta.

SWP and CVP pump water from south Delta for export to serve over 25 million people (about two thirds of the state's population) and 15,000 km^2 farmland in California. Water quality standards have been developed to ensure that the water at the intakes of SWP and CVP is appropriate for drinking water, agricultural, and other purposes [20,21]. In California, five types of water years are defined to facilitate water resources management. They are wet (W), above-normal (AN), below-normal (BN), dry (D), and critical (C) years and are defined according to the overall wetness of a specific year [20]. The water quality standards vary across different water year types. Salinity is being monitored at a range of key compliance stations including Collinsville, Emmaton, and Jersery Point (Stations 3, 4, and 5 in Figure 1) to ensure compliance with these water quality standards. For instance, in below-normal years, the salinity (represented by electrical conductivity (EC)) at Jersey Point should remain below 450 μs/cm from 1 April to 20 June and below 740 μs/cm from 21 June to 15 August. Table A1 in the Appendix A provides a more detailed list of such standards at Jersey Point and Emmaton. Martinez salinity is the major salinity source for these compliance stations. To have a clear understanding on the salinity at these locations and thus the overall compliance status in the Delta, it is imperative to have a rigorous estimate on Martinez salinity. This is particularly true in planning studies (e.g., different operation options or different structural change scenarios in the Delta) where no field measurements of salinity at the interior locations would be available.

This study utilizes the same dataset as applied in the [65] study. The dataset includes a 24 year period (water year 1991–2014) of daily observed water stage (average, minimum, and maximum) at Martinez, Martinez salinity (hereinafter "reference salinity") and the net Delta outflow (NDO) calculated based on observed or modeled inflows and water uses in the Delta [66]. The salinity exhibits a strong seasonality with the lowest value in early spring (Figure 2). It increases gradually till peaking typically near the end of fall. In the winter, when upstream reservoirs in the Sacramento and San Joaquin River system increase releases to reserve storage to manage wet season floods, the salinity at Martinez starts dropping. The NDO shows a roughly reversed pattern with its peak in the winter and its minimum in the fall. It is clear that variations in salinity mimic that of the NDO, although in reverse fashion as expected. There is a negative correlation (with a Pearson correlation coefficient, R = −0.91) between them on the annual scale. The average stage has a different cyclic pattern throughout the year, owing to a more complex relationship with both incoming freshwater as well as actual cycles in tidal elevation. Its correlation with the salinity is much weaker (R = 0.11). The variation patterns of these three variables are also evident when looking at their daily time series during the entire 24 year period (Figure A1 in the Appendix A). In this study, the first 12 year period (water year 1991–2002) is used as the training/validation period for neural network models, while the second 12 year period (water year 2003–2014) is used as the evaluation period.
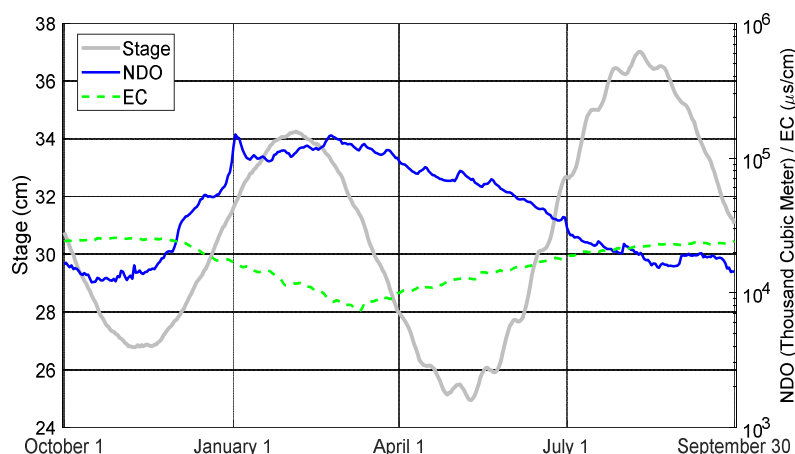
Figure 2. Long-term (water year 1991–2014) average daily Martinez water stage (Stage; left *y*-axis), net Delta outflow (NDO; right *y*-axis) and Martinez salinity represented by Electrical Conductivity (EC; right *y*-axis).

2.2. Martinez Boundary Salinty Generator

The Martinez Boundary Salinity Generator (MBSG) integrates the conceptual–empirical G-model of [22] and a linear filter proposed by [33] for planning and forecasting studies involved with DSM2 modeling. The G-model simulates antecedent flow-salinity relationship as follows:

$$S_t = (S_O - S_U) \times e^{-\alpha G_t^k} + S_U \tag{1}$$

where S_t is the salinity at time step t; S_O and S_U represent the downstream ocean and upstream river salinity, respectively; α and k are a dispersion parameter (with the effect of upstream distance consolidated) and an empirical shape parameter, respectively; and G is a function representing the antecedent flow defined as:

$$\frac{dG}{dt} = \frac{(Q - G) \times G}{\beta} \tag{2}$$

where β is an empirical constant and Q is the volumetric flow rate which is net Delta outflow for Martinez.

The linear filter models tidally varying effects from the ocean based on the assumption that "tidally-averaged salinity is the result of a uniform, harmonic advection acting on the exponential salinity profile from G-model" [33]. This study provides the mathematical forms of the filter as well as the updated salinity estimation equation for simplicity. For detailed explanation on the theory, implementation, and application of the linear filter in estimating Martinez salinity, the readers are referred to [33]. Specifically, the change made to the G-model is that a harmonic position reflecting displacement of the salinity profile, x_t, is added to Equation (1):

$$S_t = (S_O - S_U) \times e^{-\widetilde{\alpha} G_t^k x_t} + S_U \tag{3}$$

where $\widetilde{\alpha}$ is the decay parameter in Equation (1) before distance is bundled into it. x_t can be divided into a centered position (x_0) and a harmonic perturbation (x_t'): $= x_0 + x_t'$. Combining $-\widetilde{\alpha} x_0$ into a coefficient β_1 and rearranging Equation (3) yield:

$$\ln\left(\frac{S_t - S_U}{S_O - S_U}\right) = \beta_1 G_t^k + x_t' G_t^k \tag{4}$$

$$x_t' = \sum_{i=0}^{n_i} a_i \, z_{t+i_0 \Delta t - i \Delta t}$$

x_t' can be written as a convolution filter modeling displacement on lagged Martinez stage:

$$x_t' = \sum_{i=0}^{n_i} a_i z_{t+i_0 \Delta t - i \Delta t} \tag{5}$$

where z_t is the tide stage; a_i are the filter coefficients; n_i represents the length of the convolution kernel (i.e., number of lagged input values applied); and Δt stands for the spacing between lagged values; i_0 designates an offset of the filter. With Equation (5) incorporated, the governing equation of MBSG becomes:

$$ln(\frac{S_t - S_U}{S_O - S_U}) = \beta_1 G_t^k + G_t^k \sum_{i=0}^{n_i} a_i z_{t+i_0 \Delta t - i \Delta t} \tag{6}$$

The MBSG was recently recalibrated [65] using an automated parameter optimization software named Parameter Estimation (PEST) [67]. The recalibration improves model performance when compared to the original calibration [33], particularly in the high salinity range. This study uses the PEST calibrated MBSG as the baseline model to benchmark the proposed neural network models.

2.3. Artificial Nerual Networks

Multi-Layer Perceptron (MLP) is the plain form of neural networks. In MLP, each neuron in each layer is fully connected to all neurons in adjacent layers. An MLP model with one or more hidden layers is often used to evaluate the baseline performance of deep neural networks without employing any special architecture. Specialized network architectures have been developed, and the most popular ones include Recurrent Neural Network (RNN), which is naturally suitable to handle sequential data, and Convolutional Neural Network (CNN), which captures patterns in a hierarchical manner. A widely used special form of RNN is Long Short-Term Memory (LSTM), in which neurons are organized as sequential units each of which "remembers" values over arbitrary time steps, long or short. One-dimensional convolutional neural network (Conv1D) is a special form of CNN. Conv1D employs layers of one-dimensional filters to capture hierarchical patterns in series data including time series. By stacking many convolutional layers, Conv1D can effectively combine local and overall patterns to learn complex temporal features which are very hard to delineate with pre-defined mathematical models.

In this study, the output of the prediction task of neural network models is the EC of the current day. For each model used in this task, different sets of inputs were tested. Candidate input variables are daily NDO and mean, minimum, and maximum stage in the previous 60 days (including the current day) as well as the simulated EC by the PEST-calibrated MBSG model at the current day. The selection of 60 days is empirical. In the Delta, salinity is influenced by antecedent flow in the preceding several months. However, after about two months, the influence generally becomes very weak (with an absolute correlation value less than 0.5; Figure A2). We tested shorter and longer periods. The results were not as ideal as that of the case when 60 day is applied. Daily NDO is a basic indicator of the hydrologic condition in Delta, and the statistics of stage level observation provide more detailed information on daily stage dynamics. Three scenarios with different combinations of input variables were investigated. In Scenario 1, the input variables include daily NDO and average stage. Scenario 2 also employs the two daily series and adds daily minimum and maximum stage to further delineate daily stage dynamics. In Scenario 3, the input variables include all inputs in Scenario 2 as well as the simulated EC by the MBSG model.

MLP and LSTM are applied to the three scenarios of input variables. For MLP, the input size is 120 (2×60, Scenario 1), 240 (4×60, Scenario 2), or 241 ($4 \times 60 + 1$, Scenario 3). For LSTM, the main data input is expected to be time series with equal length for which the simulated EC at the current day does not fit. The input size to the neural network is 120 (2×60) for Scenario 1 and 240 (4×60) for Scenarios 2 and 3. As the simulated EC from the MBSG model is available in Scenario 3, the neural network predicts the relative difference between the simulated EC and the actual EC, using the simulated EC by the

MBSG model as an additional input to the last layer of the network. The neural network models used in Scenarios 1 and 2 predict the absolute level of EC. The special use of the simulated EC is because of the requirement of LSTM to have aligned series as inputs. In addition to looking at impacts of different input variables on network model performance, the impacts of different network hyper-parameters are also investigated. Hyper-parameters of MLP and LSTM configured with Scenario 3 input variables are fined-tuned to yield a fourth MLP and a fourth LSTM model, respectively. Table 1 lists all models investigated in this study.

Table 1. Study Models.

Model Name	Description
MBSG	PEST calibrated MBSG model in the [65] study
MLP1	MLP with NDO and average stage as input
MLP2	MLP with MLP1 inputs plus maximum and minimum stage
MLP3	MLP with MLP2 inputs plus MBSG simulated Martinez salinity
MLP4	MLP with MLP3 inputs and fine-tuned network parameters
LSTM1	LSTM with NDO and average stage as input
LSTM2	LSTM with LSTM 1 inputs plus maximum and minimum stage
LSTM3	LSTM with LSTM 2 inputs plus MBSG simulated Martinez salinity
LSTM4	LSTM with LSTM inputs and fine-tuned network parameters
Hybrid	Hybrid MBSG and Conv1D

We also test the feasibility of combining the existing process-based model and deep learning architectures as a hybrid model. One of the key steps in the MBSG model is to reduce the recent stage level series to a scalar as an indicator of current hydrodynamic conditions. The existing approach in the MBSG model is using 15 min stage data in the past 72 h at a 12 h interval, or 7 stage observations (i.e., n_i in Equation (5)) in total to quantify the relationship between short-term stage dynamics and salinity. Traditionally it is very hard to fully utilize the temporal information in the dense and noisy 15 min series. Although human experts may interpret the 15 min series to some extent, building explicit rules for model development is prohibitive. As a result, the existing approach in the MBSG model only samples the 15 min every 12 h to simplify the calculation. Conv1D is particularly capable of learning very complex patterns from one-dimensional data and the learning process requires minimal human input. We replaced the existing 12 h sampling approach with a stack of Conv1D layers which take the original 15 min series as inputs, hoping to discover and employ the patterns in the denser series data that might be neglected in the past. In this way we have a hybrid model, which includes most of the physical processes of the MBSG model as well as a deep learning-based pattern recognizer to deal with the complexity in dense stage observations.

Data from water year 1991 to 1999 are utilized to train the neural network models. To optimize hyper-parameters and select the optimal network structure of a certain type, data in water years 2000, 2001, and 2002 are used as a validation set which is not directly used in training but in the evaluation of performance by various combinations of hyper-parameters. Hyper-parameters of MLP include the number of hidden layers and the number of neurons in each layer. For LSTM, hyper-parameters include the number of LSTM units/layers, the number of filters in the recurrent units, the dropout rate between layers, and the dropout rate between time steps. These parameters are specified in Table A2 in the Appendix A. For the Conv1D component in the hybrid model, we tuned the number of Conv1D layers, the number of filters in each layer, and the dropout rate. We also tuned batch size and initial learning rate for all types of deep networks. The Adam optimizer was used in all experiments [68].

2.4. Study Metrics

This study employs five metrics which provide complementary information on the performance of the proposed models in simulating Martinez salinity. They are defined in Table 2 where S represents salinity, $\overline{S}$ means average salinity, t indicates a specific time step, n stands for the total number of time

steps, and *sim* and *ref* designate simulated and reference values, respectively. Specifically, these metrics include the commonly used percent bias and mean absolute error between the reference salinity and the corresponding model simulations. Percent bias shows whether the model over-estimates or under-estimates the reference salinity by how much on average sense. Mean absolute error indicates the average magnitude of model simulation errors. In addition, the study also includes three metrics that represent the three components of the Taylor Diagram: standard deviation (SD), correlation coefficient (R), and centered root mean square difference (RMSD). The Taylor Diagram provides a concise summary of how well different model simulations match the reference field in terms of these three components in a single diagram [69]. The correlation coefficient measures the strength of the linear relationship between model simulations and the reference. The standard deviation quantifies the amplitude of their variations. The centered root mean square difference provides the centered (with mean subtracted) model error.

Table 2. Study Metrics.

Name	Description	Formula
Bias	Percent bias	$\frac{\sum_{t=1}^{n}\left(S_{sim,t}-S_{ref,t}\right)}{\sum_{t=1}^{n}S_{ref,t}}\times 100$
MAE	Mean Absolute Error	$\frac{1}{n}\sum_{t=1}^{n}\left|S_{sim,t}-S_{ref,t}\right|$
SD	Standard Deviation	$\sqrt{\frac{\sum_{t=1}^{n}\left(S_{t}-\overline{S}\right)}{n-1}}$
R	Correlation Coefficient	$\frac{\sum_{t=1}^{n}\left[\left(S_{sim,t}-\overline{S_{sim}}\right)\times\left(S_{ref,t}-\overline{S_{ref}}\right)\right]}{n\times SD_{sim}\times SD_{ref}}$
RMSD	Centered Root Mean Square Difference	$\sqrt{\frac{\sum_{t=1}^{n}\left[\left(S_{sim,t}-\overline{S_{sim}}\right)-\left(S_{ref,t}-\overline{S_{ref}}\right)\right]^{2}}{n}}$

Model results are evaluated in three different ways. Firstly, those five metrics (Table 2) are calculated between model simulated salinity at Martinez and the reference salinity during the entire evaluation period (watery year 2003–2014) to shed light on the overall performance of these models. Secondly, model simulations within different salinity ranges (high, medium, and low) during the entire evaluation period are assessed against the corresponding reference salinity in terms of these five metrics. In practice, there are different management strategies for different salinity ranges. In general, managing high salinity (versus low–medium salinity) is more challenging. Finally, those five metrics are calculated in different water year types to illustrate whether neural network model performance varies with different categories of water years.

3. Results

The results are grouped into three sub-sections accordingly. The first sub-section presents the overall results during the entire evaluation period from water year 2003–2014. In the second sub-section, the entire evaluation period is divided into three sub-periods containing three different ranges (high, medium, and low) of salinity, respectively. Model performance in simulating different ranges of salinity is examined. In the last sub-section, the evaluation period is divided into five sub-periods representing five different water year types, respectively. Model results are evaluated in each of these five sub-periods.

3.1. Entire Evaluation Period

Standard deviation (SD) of simulated salinity at Martinez along with its correlation (R) with the reference salinity as well as its centered root mean square difference (RMSD) for each model are calculated and illustrated in Figure 3. The hybrid MBSG–CNN model slightly outperforms the process-based MBSG model (Figure 3a). The former has a smaller (by an amount of 5.7%) RMSD and a

higher (by about 0.4%) R value compared to the latter. The SD values of both models are fairly close to each other. They are both smaller than their counterpart of the reference salinity, indicating that salinity simulations of both models have relatively less variation compared to the reference salinity.
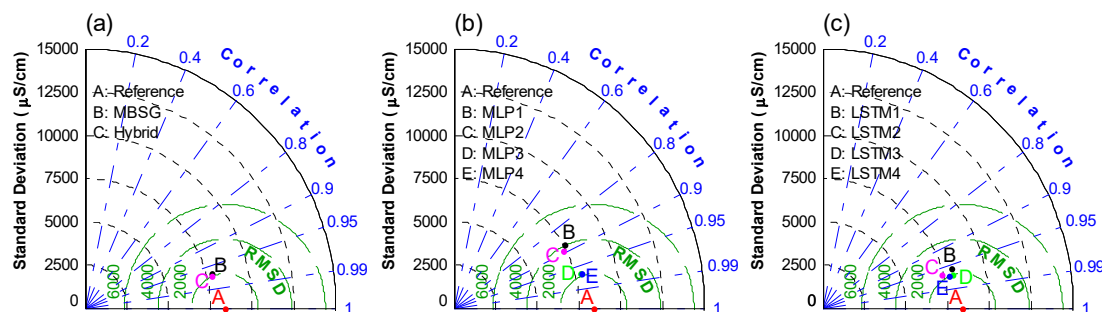

Figure 3. Taylor Diagrams illustrating the correlation (R; the azimuth position), standard deviation (SD; radial distance from the origin, shown on both vertical and horizontal axes), and centered root mean square difference (RMSD; radial distance from the reference point A which serves as the origin for RMSD) between the reference salinity at Martinez and the corresponding salinity simulations generated via (**a**) MBSG and hybrid MBSG–CNN models, (**b**) four MLP models, and (**c**) four LSTM models. Reference point A designates the statistics of the reference salinity itself (RMSD = 0; R = 1). The *X*-axis shows the values of both SD and RMSE of which the ticks are different.

For MLP models, when only using net Delta outflow (NDO) and average water stage as input (MLP1; point B in Figure 3b), the resulting salinity simulations have a smaller (by 9.5%) correlation value and a remarkably larger (by 90%) RMSD compared to MBSG simulations (point B in Figure 3a). The SD values of both MLP1 and MGSB are comparable to each other, yet both are smaller than that of the reference salinity. Adding daily maximum and minimum stage as input (MLP2; point C in Figure 3b) yields simulations with only a slightly higher R value and a marginally smaller RMSD than that of the MLP1 simulations. The SD of MLP2 differs noticeably (12% smaller) from that of the reference salinity. When further incorporating MBSG simulations as an additional input feature (MLP3; point D in Figure 3b), however, the results are improved markedly. The metrics (SD, R, and RMSD) of MLP3 become comparable that of MBSG. Fine-tuning MLP3 hyper-parameters (MLP4; point E in Figure 3b) leads to salinity simulations with even more satisfactory metrics compared to both MLP3 and MBSG.

Different from MLP1, the LSTM model using NDO and average stage information alone as input (LSTM1; point B in Figure 3c) yields comparable simulations to that of the MBSG (point B in Figure 3a). The MBSG model has slightly smaller RMSD and higher R. However, the SD value of LSTM1 is closer to that of the reference salinity compared to the SD value of MBSG. Adding daily maximum and minimum stage information as input (LSTM2; point C in Figure 3c) yields simulations with a higher R value and a lower RMSD than LSTM1. Further including MBSG simulations as input (LSTM3; point D in Figure 3c) leads to salinity simulations with smaller RMSD, higher R, and better SD (i.e., closer to the reference SD) than LSTM2 and MBSG simulations. Fine-tuning LSTM3 hyper-parameters (LSTM4; point E in Figure 3c) results in simulations with even better R and RMSD than that of LSTM3 simulations.

In addition to R, SD, and RMSD, bias, and mean absolute error (MAE) are also calculated for all models studied. Overall, the process-based MBSG model under-estimates the reference salinity (bias = −2.4%; Figure 4). Similarly, most neural network models also under-estimate the salinity except for MLP1 (14.7% bias) and LSTM3 (2.1% bias). In terms of the magnitude, MLP4 and LSTM4 are less biased than MBSG. The remaining neural network models have comparable but slightly higher bias than MBSG except for MLP1. MLP1 also has the largest mean absolute error (MAE = 3979 μs/cm). MLP2 has the second largest MAE value. The MAE values of other models are consistently smaller than 2000 μs/cm. Compared to MBSG, four neural network models, including MLP4, LSTM3, LMST4, and the hybrid model, have smaller MAE values.
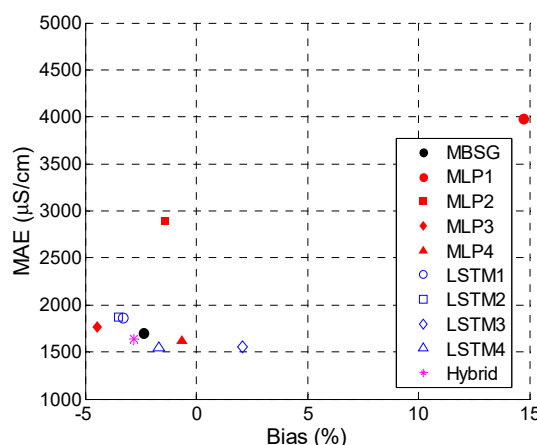
Figure 4. Percent bias (Bias; horizontal axis) and mean absolute error (MAE; vertical axis) between reference and simulated salinity at Martinez via different models during the entire evaluation period.

Looking at five metrics all together, for MLP and LSTM models, incorporating maximum and minimum stage information generally improves network performance. Adding MBSG simulations as an additional network input feature leads to further improvement. The improvement is much more significant for MLP rather than LSTM. Fine-tuning network hyper-parameters is shown to improve the general performance of both MLP and LSTM models. Put differently, among all MLP (LSTM) models, MLP4 (LSTM4) has the best performance in general during the entire evaluation period. Among all nine neural network models, LSTM4 has the smallest RMSD, highest R, and lowest MAE; MLP4 has the lowest bias; LSTM1 and LSTM3 have the closest SD to that of the reference salinity. MLP4, LSTM3, and LSTM4 are the only three models which outperform the process-based MBSG model in terms of all five metrics. In comparison, the hybrid model yields improvement over MBSG in terms of R, RMSD, and MAE. The bias and SD values of the hybrid model are comparable to that of MBSG.

3.2. Different Salinity Ranges

Martinez salinity varies seasonally, typically with low values in winter/spring and high values during summer/fall (Figure 2). Salinity management practices vary accordingly, based on the range of salinity. In addition to looking at model performance in the entire evaluation period, this section further examines its performance during different salinity ranges. Specifically, three ranges are considered, including low range (less than 25th percentile of observed Martinez salinity during the evaluation period; $<1.19 \times 10^4$ microsiemens per centimeter (μs/cm)), medium range (25th percentile to 75th percentile), and high range (over 75th percentile; $>2.53 \times 10^4$ μs/cm). The entire evaluation period is divided into three sub-periods accordingly. The length of the low salinity period is identical to that of the high salinity period, with each accounting for half of the length of the medium range salinity period.

Based on the results during the entire evaluation period presented in Section 3.1, MLP4 and LSTM4 have the best performance among all MLP and LSTM models, respectively. The hybrid model provides generally comparable or superior simulations than MBSG. The current section first evaluates the performance of these three neural network models (MLP4, LSTM4, and Hybrid) against that of the MBSG model (Figure 5). For low range salinity (Figure 5a), all three models yield higher correlation values and lower RMSD compared to MBSG. For medium range salinity (Figure 5b), all three models have higher correlation values and smaller RMSD with one exception. The RMSD of MLP4 is slightly (2%) larger than that of MBSG. For high range salinity (Figure 5c), the RMSD of MLP4 is even higher (by 9.7%) compared to its counterpart of MBSG. The correlation value of MLP4 is also smaller. Conversely, LSTM4 and the hybrid model outperform MBSG in terms of both R and RMSD. Regarding SD, for both medium and high ranges of salinity, all three neural network models and the MBSG model yield simulations with higher variations (higher SD) than the reference salinity; for low salinity, LSTM4 is the only model with a higher than reference SD value. For low, medium, and high

ranges of salinity, MLP4, LSTM4, and the hybrid model have the most satisfactory SD (closest to reference SD) values, respectively.
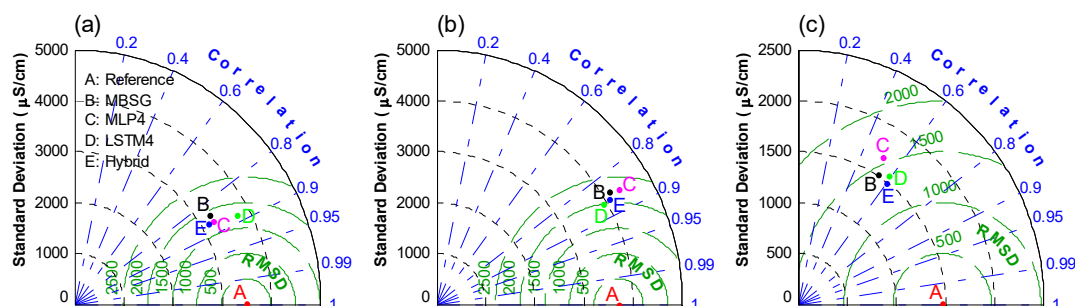

Figure 5. Taylor Diagrams displaying statistics (correlation, standard deviation, and centered root mean square difference) between the reference salinity at Martinez and the corresponding salinity simulations generated via four different models (MBSG, MLP4, LSTM4, and Hybrid MBSG–CNN) grouped in three salinity ranges including (**a**) low salinity range (less than 25% non-exceedance probability), (**b**) medium salinity range (between 25% and 75% non-exceedance probability), and (**c**) high salinity range (above 75% non-exceedance probability).

For the models not depicted in Figure 5, those three metrics (R, SD, and RMSD) are also examined (Table 3). Similar to the results presented in Section 3.1, adding additional information as network input features generally improves model performance across all three salinity ranges. Nevertheless, a noticeable difference is that fine-tuning network hyper-parameters does not necessarily lead to improved performance. The differences in these three metrics between MLP3 (LSTM3) and MLP4 (LSTM4) are minimal. MLP3 performs relatively better than MLP4 in high salinity ranges while LSTM3 generally outperforms LSTM4 in medium and high salinity ranges. Table 3 also indicates that model performance differs evidently in high salinity range versus low to -medium ranges. Specifically, SD, and RMSD values of high salinity simulations are considerably smaller than that of the low and medium salinity simulations while the R value of the former is remarkably smaller than that of the latter. This suggests that simulations on high salinity are generally less spread out (smaller SD and RMSD). However, their linear relationship with the corresponding reference salinity is remarkably weaker when compared to that of simulations on low- to medium ranges of salinity.

Table 3. Statistics of model-simulated Martinez salinity during three different salinity ranges.

Scenarios	Standard Deviation (μs/cm)			Centered Root Mean Square Difference (μs/cm)			Correlation Coefficient		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Reference	3482	3702	1467	0	0	0	1.000	1.000	1.000
MBSG	3249	4152	1509	1902	2214	1428	0.843	0.847	0.539
MLP1	4428	5392	3306	3504	3837	3306	0.631	0.703	0.222
MLP2	4210	5077	2852	3372	3488	2878	0.630	0.727	0.239
MLP3	3186	4348	1673	1774	2272	1563	0.862	0.853	0.511
MLP4	3258	4346	1681	**1758**	2259	1568	0.866	0.854	0.511
LSTM1	3915	4253	1689	2217	2558	1727	0.827	0.802	0.408
LSTM2	**3648**	3895	1439	1810	2236	1514	0.872	0.828	0.458
LSTM3	4108	**3864**	**1475**	1987	**1992**	**1318**	0.876	0.862	0.598
LSTM4	3734	3920	1565	1773	2005	1376	**0.882**	**0.863**	0.589
Hybrid	3152	4082	1493	1765	2077	1326	0.863	0.862	**0.599**

Bold numbers represent the best metrics.

In terms of bias and MAE, different models perform differently across different salinity ranges. First, all models tend to over simulate low salinity (Figure 6a). The process-based model has a bias of 7.9% and MAE of 1519 µs/cm for low salinity simulations. In comparison, only LSTM1, MLP3, and the hybrid model are less biased among all nine neural network models. Overall, MLP1 is the outlier model with significantly large bias and MAE. MLP2 shows improvement over MLP1. However, its bias and MAE values are still remarkably larger than that of the remaining models. In contrast, MLP3 and the hybrid model have the smallest bias and MAE. Second, all models except for MLP1 under-simulate high salinity (Figure 6c). The bias and MAE of MBSG are −6.4% and 1858 µs/cm, respectively, for high salinity simulations. Four neural network models including MLP4, LSTM1, LSTM3, and LSTM4 have smaller bias and MAE than MBSG. Among them, LSTM3 has the most satisfactory bias and MAE. Finally, most models also under-estimate the medium range salinity (Figure 6b). MLP1 is again the outlier model with the largest positive bias and MAE. LSTM3 is the other model with positive bias (4.2%). MBSG simulations on medium salinity have a bias of −1.2% and MAE of 1723 µs/cm, respectively. In comparison, both MLP4 and LSTM4 have smaller bias and MAE values.
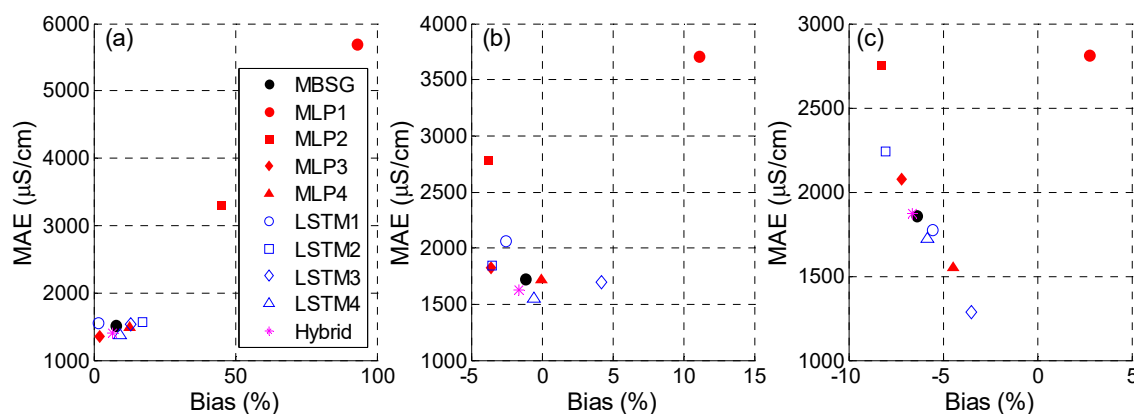

Figure 6. Percent bias (Bias; horizontal axis) and mean absolute error (MAE; vertical axis) between reference and simulated salinity at Martinez via different models grouped in three salinity ranges including (**a**) low salinity range; (**b**) medium salinity range; and (**c**) high salinity range.

All in all, no single model consistently outperforms the others in terms of all five metrics across low, medium, and high ranges of salinity. However, among all models, MLP1 and MLP2 have the worst performance measured by nearly all metrics. For high range salinity, LSTM3 has the best performance in general. It is the least bias model with the smallest RMSD and MAE and the best SD. The associated correlation coefficient (0.598) is very close to the optimal value (0.599) of the hybrid model. For medium range salinity, LSTM3 has the smallest RMSE and the best SD; LSTM4 has the highest correlation coefficient and smallest MAE, while MLP4 is the least biased. The results on low salinity are mixed. The five optimal metrics come from five different models, respectively. Nevertheless, on average, MLP4, LSTM4, and the hybrid models have relatively better performance.

3.3. Different Water Year Types

In the Delta, water quality standards vary with water year types (e.g., Table A1 in the Appendix A). Understanding model performance in different types of water years is critical to guide corresponding salinity management practices. The entire evaluation period (2003–2014) is divided into five sub-periods, with each sub-period containing the data from a specific water year type (W, AN, BN, D, C). There are two wet years, two above-normal years, three below-normal years, three dry years, and two critical years. Therefore, these five sub-periods vary (from two to three years) in length.

Following Section 3.2, this section first examines three metrics illustrated by the Taylor diagram of the process-based MBSG model and three neural network models (MLP4, LSTM4, and Hybrid).

Overall, the performance of these four models are fairly close to each other across all five types of water years (Figure 7). However, none of them consistently outperform the others. Specifically, across all types of water years, LSTM4 and the hybrid model have higher correlation values than MLP4 and MBSG. In addition, the hybrid model has smaller RMSE than MLP4 and MBSG. Regarding SD, MLP4 has the best performance in all types of water years except for above-normal years. The SD value of MBSG is the closest to the reference SD (−0.6% difference versus −2.3% of MLP4). Model performance also varies across different water year types. Highest R values of all four models occur in wet years when salinity is generally low. In contrast, R values during dry and critical years (when salinity are normally high on average) are typically the lowest. The smallest and highest RMSD values are observed in below-normal and critical years, respectively, for MBSG, MLP4, and the Hybrid model. For LSTM4, wet years have the smallest RMSD while above-normal years have the highest RMSD. In terms of SD, model performance is generally the worst in critical years, followed by dry years. On average, above-normal years have the most satisfactory SD values.
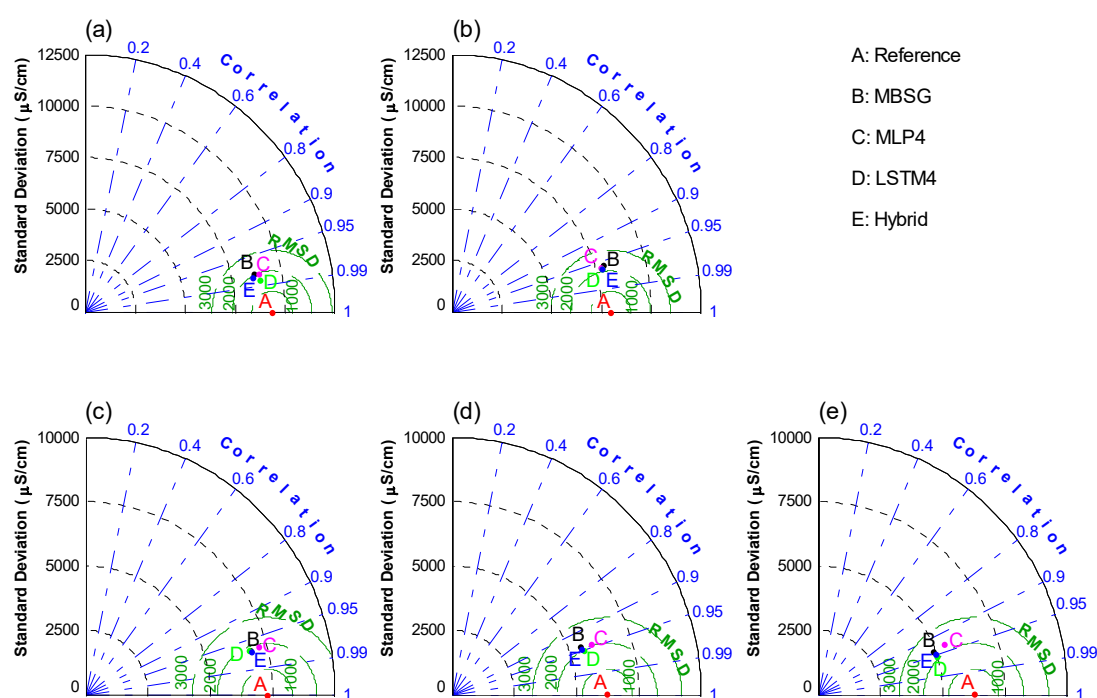

Figure 7. Taylor Diagrams displaying statistics (correlation, standard deviation, and centered root mean square difference) between the reference salinity at Martinez and the corresponding salinity simulations generated via four different models (MLP4, LSTM4, and Hybrid MBSG–CNN) grouped in five water year types including (**a**) wet year, (**b**) above-normal year, (**c**) below-normal year, (**d**) dry year, and (**e**) critical year.

The performance of those four models is also compared to that of the remaining models. Table 4 shows the RMSE of all nine neural network models along with the process-based MBSG model. For MLP models, when only NDO and stage data are considered as network input features (MLP1 and MLP2), the resulting RMSE are much larger than the process-based model across all types of water years. Adding MBSG simulations as an additional input (MLP3) largely improves model performance. Fine-tuning hyper-parameters (MLP4) leads to even smaller RMSE in all types of water years except for below-normal years. For LSTM models, when only NDO and average stage are employed (LSTM1), the resulting RMSE values are generally comparable to that of the MBSG. Adding minimum and maximum stage (LSTM2) yields smaller RMSE in general. Incorporating MBSG simulations as input (LSTM3) leads to consistently smaller RMSE values (versus RMSE values of MBSG, LSTM1, and LSTM2) in all five types of water years. Fine-tuning hyper-parameters (LSTM4) does not necessary lead to

further improvement. Similar features are also observed in other two metrics (Tables A3 and A4 in the Appendix A). Looking all models together, LSTM3 has the smallest RMSE in wet years; LSTM4 has the smallest RMSE in dry and critical years, while the hybrid model performs the best during above-normal and below-normal years.

Table 4. Centered root mean square difference between reference and simulated Martinez salinity in different types of water years.

Scenarios	Water Year Type				
	Wet (W)	Above-Normal (AN)	Below-Normal (BN)	Dry (D)	Critical (C)
MBSG	2028	2277	1836	2146	2323
MLP1	3261	3524	3648	3704	4464
MLP2	3430	3404	3343	3462	3972
MLP3	2041	2223	1870	2062	2331
MLP4	1986	2204	1872	2040	2313
LSTM1	2055	2478	2313	2368	2392
LSTM2	1939	2280	2182	2219	2411
LSTM3	**1603**	2180	1786	2022	2194
LSTM4	1644	2142	1867	**1947**	**2080**
Hybrid	1918	**2110**	**1767**	2031	2169

Bold numbers represent the best metrics.

Similar to what has been observed in the entire evaluation (Figure 4) period and in three sub-periods representing three different salinity ranges (Figure 6), MLP1 and MLP2 tend to be the outlier models with very different bias and MAE from other models (Figure 8). Their MAE values are markedly larger than that of other models. MLP1 considerably over-estimates the salinity in all types of water years except for the critical years, while MLP2 largely under-estimates Martinez salinity in dry and critical years. For the remaining models, the hybrid model performs the best in above-normal years with the smallest bias and MAE; MLP4 is the best performance model in below-normal years; LSTM3 outperforms other models in dry years. For wet years, two models (MLP4 and LSTM3) have the smallest bias and MAE, respectively. For critical years, LSTM3 and LSTM4 has the smallest bias and MAE, respectively.
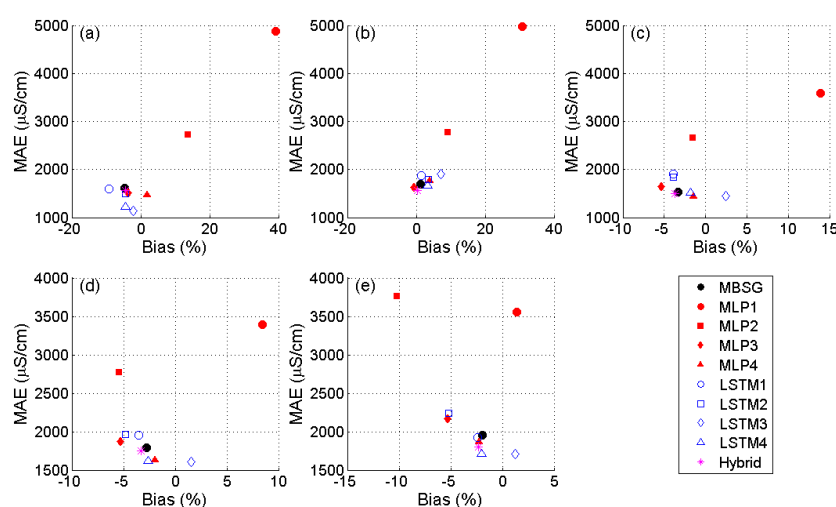

Figure 8. Percent bias (Bias; horizontal axis) and mean absolute error (MAE; vertical axis) between reference and simulated salinity at Martinez via different models grouped in five water year types including (**a**) wet year, (**b**) above-normal year, (**c**) below-normal year, (**d**) dry year, and (**e**) critical year.

Examining five metrics altogether, the neural network models can outperform the process-based MBSG model consistently across all water year types. Table 5 tabulates the improvements calculated as the percent difference between the optimal metrics (of the neural network models with the outlier models MLP1 and MLP2 excluded) and the corresponding MBSG metrics. For R, SD, RMSE, and MAE, the improvements in extreme years (wet, dry, and critical) are more noticeable than the improvements in near-normal (above-normal and below-normal) years. For R and SD (RMSE and MAE), the largest improvements occur in critical (wet) years. The optimal metrics are not associated with a single neural network model. In extreme years, the LSTM models (LSTM3 and LSTM4) tend to be the optimal models; in above-normal years, the hybrid model seems to have the best metrics (except for SD); in below-normal years, the hybrid model and MLP4 perform relatively better in terms of the number of optimal metrics associated with them.

Table 5. Improvements over MBSG metrics in different water year types.

Metrics	Water Year Type				
	W	AN	BN	D	C
R	0.7% (LSTM3/LSTM4)	0.5% (Hybrid)	0.2% (LSTM3/Hybrid)	1.2% (LSTM4)	1.5% (LSTM4)
SD	7.0% (LSTM3)	0.3% (LSTM3)	5.1% (MLP4)	7.0% (MLP4)	8.3% (MLP4)
RMSE	20.9% (LSTM3)	7.3% (Hybrid)	3.8% (Hybrid)	9.3% (LSTM4)	10.5% (LSTM4)
Bias	59.3% (MLP4)	65.3% (Hybrid)	55.1% (MLP4)	45.0% (LSTM3)	33.3% (LSTM3)
MAE	29.1% (LSTM3)	7.7% (Hybrid)	6.1% (MLP4)	10.7% (LSTM3)	12.7% (LSTM4)

4. Discussion

4.1. Data Stationarity and Availability

This study used the first half (water year 1991–2002) of the record period as training/validation period and the second half (water year 2003–2014) as evaluation period. The underlying assumption is that the relationships between salinity and NDO/stage in the first half would hold valid in the second half as well. Put differently, stationarity is assumed in the data employed. To validate this assumption, trend assessment is conducted for mean, maximum, and minimum NDO, Martinez salinity and stage variables on both annual and monthly scales. The widely used non-parametric Mann–Kendall test [70,71] is used in assessing the significance of trend in these variables with a significance level of 0.05. The slope of a significant trend is determined via the Theil–Sen approach [72,73].

Figure 9 depicts the significance level (*p*-value) of the trends in these variables. There is no statistically significant trend in salinity on annual scale or monthly scale. This is also the case for NDO with one exception; mean NDO in February has a significant decreasing trend (−4.4 million m^3/year). Similarly, there is generally no statistically trend in minimum Martinez stage with one exception in March (slope = −0.08 cm/year). For maximum stage, however, significant upward trends are identified in seven out of 12 months. The trend slopes in these months range from 0.06 cm/year (December) to 0.12 cm/year (June). On an annual scale, the trend is also significant with a slope of 0.08 cm/year. Compared to maximum stage, mean stage tends to have significant trends in most months except for January and April. The difference is that, there are downward trends in February and March (both at a rate of −0.04 cm/year). Upward trends in other months with significant trends are generally milder in slope, ranging from 0.03 cm/year to 0.06 cm/year. The trend slope is also milder on annual scale with a value of 0.01 cm/year. This upward trend in mean stage at Martinez is likely linked to the mean sea level rise recorded at the Golden Gate [74].
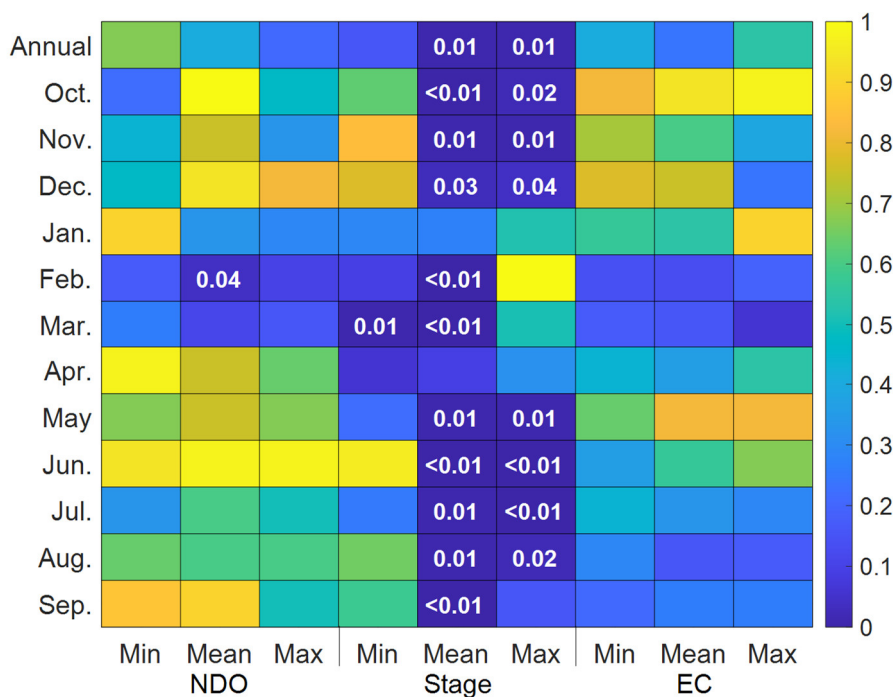
Figure 9. *p*-values of monotonic trends in minimum (Min), mean (Mean), and maximum (Max), net Delta outflow (NDO), Martinez stage (Stage), and salinity (EC) identified via the Mann–Kendall trend assessment on annual scale and monthly scale. A value less than 0.05 indicates that the trend is statistically significant. The *p*-values associated with significant trends are specified.

In this study, Martinez salinity is the predictand while Martinez stage and NDO are predictors. As previously shown in Figure 2, NDO is the primary predictor (R = −0.91) while the mean stage (R = 0.11) is the secondary predictor. Section 3 shows that adding minimum and maximum stage information as additional input features leads to marginal improvement in neural network model performance, suggesting that minimum stage and maximum stage are also minor predictors. Figure 9 illustrates that there are generally no statistically significant trends in the predictand and the primary predictor. Figure 9 also shows that the slopes of significant trend in mean and maximum Martinez stage are mostly mild. In the entire evaluation period (2003–2014), for instance, the overall increase in mean stage amounts to about 1.2 mm (at an annual rate of 0.01 cm/year). This change in the secondary predictor should have minimal impact on the predictability of the predictand.

Nevertheless, sea level rise near Golden Gate is expected to accelerate in pace in the future [75]. Consequently, Martinez stage likely increases at a higher rate. Its influence on Martinez salinity would continue to grow till becoming non-trivial. That poses challenges to process-based models in reliably simulating Martinez salinity, as those models are typically calibrated based on historical conditions which would not be representative for future conditions anymore. Under these circumstances, neural network models have the advantage of learning the trend embedded in the data and applying it into the projections.

It is worth noting that the deep learning methods proposed in the current study use only a subset of the data available in training and predicting. Specifically, the long-term (water year 1991–2014) salinity and stage data available is at 15 min time step. The deep learning methods utilize daily data (aggregated from 15 min data) which only accounts for about 1% of the original salinity and stage data. Nevertheless, the deep learning methods mostly yield superior results when compared to the process-based model. This highlights the robustness of the deep learning methods. This type of method has the potential to be applied to other riverine or estuarine environments where observations are temporally limited, given that the observations are on the key predictors of the predictand.

When observations are spatially limited, model simulations can serve as a viable option in developing and applying machine learning (including deep learning) models [38].

4.2. Estimation of High Range Salinity

This study examines five statistical metrics when evaluating the performance of proposed neural network models against that of the process-based benchmark model. The values of these metrics are solid overall, reflecting satisfactory performance of the proposed deep learning models. However, it is noticeable that the value of one metric is only fair. Specifically, the correlation coefficient between simulated high salinity and the corresponding reference high salinity is remarkably lower than its counterparts for low and medium ranges of salinity (Figure 5c; Table 3). For the benchmark model, the correlation value associated with high range salinity is 0.539, which is much smaller than that of the low range salinity (0.843) and medium range salinity (0.847). The highest correlation of high range salinity simulations is 0.599 (of the hybrid model). LSTM3 (0.598) and LSTM4 (0.589) also yield higher correlation values than the benchmark model. In comparison, the best correlation metrics for low and medium range salinity are 0.883 (LSTM4) and 0.863 (LSTM4), respectively. These observations indicate that: (1) all models are relatively poorer in simulating the variability in high (versus low or medium) range salinity; (2) even though neural network models show improvements over the benchmark model, the corresponding simulations still explain less than 36% (versus 29% of the benchmark model in terms of R^2 from simple linear regressions) of the variability in the reference high range salinity. Additionally, all models (except for the outlier model MLP1) tend to under-estimate high range salinity (Figure 6c). This negative bias is also evident when looking at the exceedance probability curves of the reference and modeled salinity together (Figure A3).

Three additional neural network models are developed to explore the possibility of better modeling the variability and reducing model bias in high range salinity. These models are based on LSTM4 which yields the most favorable R, MAE, RMSD and near-optimal bias and SD during the entire evaluation period (Figures 3 and 4). The first model (LSTM4/D120) differs from LSTM4 in that it uses data from the previous 120 (rather than 60) days to generate next day's salinity. The presumption is that a longer dataset may add new information for the model to learn and predict. The second model (LSTM4/Weight) applies a higher weight to high range salinity in the loss function, while LSTM4 utilizes equal weights to all ranges of flows in the loss function. The expectation is that the model prioritizes high range salinity over low–medium range salinity in its learning and predicting process. The third model (LSTM4/SL) incorporates mean daily sea level observations near Golden Gate as an additional model input. The hypothesis is that as a surrogate of the original salinity source of the Delta, sea level may add information which is not contained in water stage at Martinez. Results of these models are illustrated in Figure 10, along with that of the benchmark model, LSTM4, and the hybrid model.

Looking at the probability exceedance curves of the high range salinity (Figure 10a), the benchmark model, LSTM4, and the hybrid model all under-estimate the reference salinity, so do LSTM4/Weight and LSTM4/SL. This is particularly true at the lower end of the high salinity range (with an exceedance probability over 80%). For LSTM4/D120, however, the exceedance curve becomes remarkably closer to that of the reference salinity. Nevertheless, when looking at the corresponding time series (the insert figure of Figure 10a), a negative bias is still noticeable. The model (i.e., LSTM4/D120) does not capture the variability in reference high range salinity well. These are also reflected in the bias (Figure 10b) and correlation (Figure 10c) plots. The overall bias of LSTM4/D120 in high range salinity is the smallest among all models. Contrariwise, LSTM4/D120 has the largest bias in other salinity ranges, especially for low range salinity (bias = −57.1%). This suggests that LSTM4/D120 reduces bias for high range salinity at the expense of increasing bias for low–medium range salinity. Additionally, LSTM4/D120 has the smallest correlation coefficients across all ranges of salinity (Figure 10c). Except for LSTM/D120, two other proposed models (LSTM4/Weight and LSTM4/SL) have similar biases as LSTM4 (Figure 10a,b) in high range salinity. However, LSTM4/Weight has higher bias in low range salinity while LSTM4/SL has higher bias in medium range salinity compared to LSTM4 (Figure 10b). In terms of correlation,

LSTM4/SL has the most favorable value (R = 0.614) in high range salinity, while LSTM/Weight yields no improvement over LSTM4 nor the hybrid model. For low and medium ranges of salinity, however, LSTM4/SL has smaller correlation coefficients compared to LSTM4 or the hybrid model. In addition, for the entire salinity range, LSTM4 still has the highest correlation value among all models.
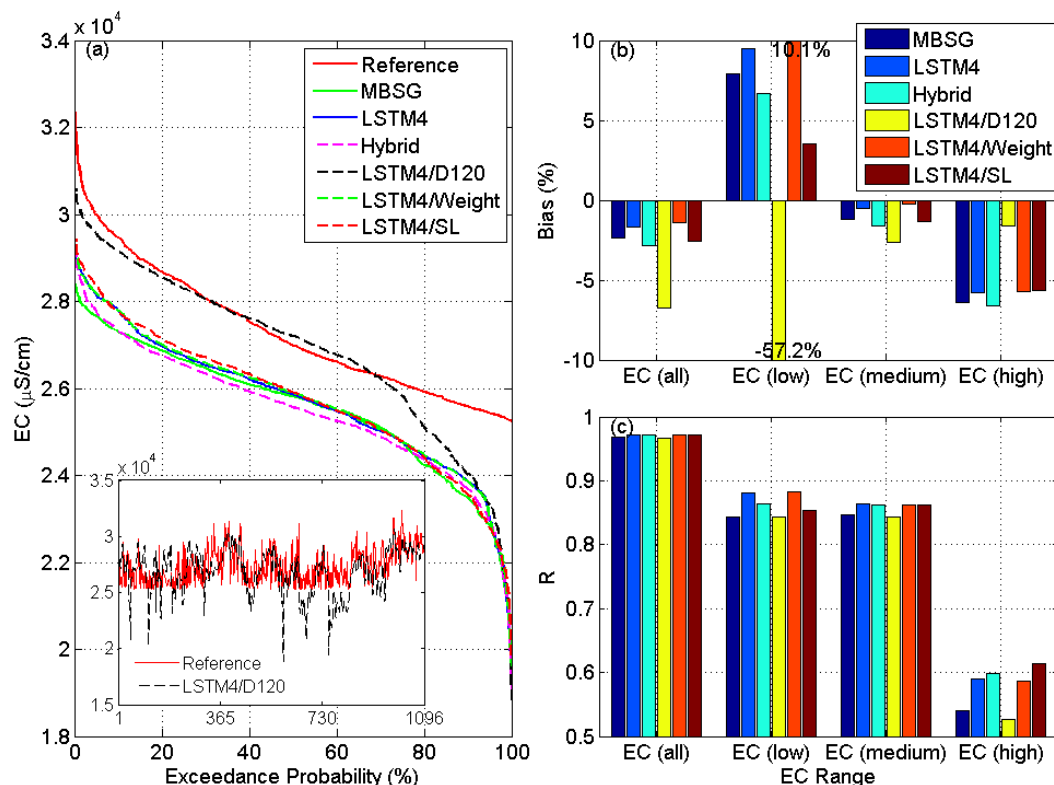

Figure 10. (**a**) Exceedance probability curves of the reference high range salinity at Martinez and corresponding simulations generated via MBSG, LSTM4, hybrid model, and three additional LSTM models (LSTM/D120, LSTM4/Weight, and LSTM4/SL). The insert figure shows time series plot of the reference high range salinity and the corresponding simulations from LSTM4/D120. (**b**) Correlation between reference salinity and simulated salinity via different models shown in (**a**). (**c**) Percent bias between reference and simulated salinity across different salinity ranges (low, medium, high, and all together).

In brief, out of three additional neural network models proposed, LSTM4/D120 yields less bias and LSTM4/SL slightly improves estimation on the variability of the high range salinity compared to the benchmark MBSG model, LSTM4, and the hybrid model. However, these improvements in high range salinity are at the cost of deteriorated model performance in low and medium ranges of salinity. Further research is warranted to better model high range salinity without compromising on model performance in low–medium range salinity.

4.3. Implications and Future Work

The findings of this study have both scientific and practical implications. From a scientific perspective, the study demonstrates the feasibility of state-of-the-art deep learning techniques in salinity estimation in the Sacramento–San Joaquin Delta (Delta) for the first time. Traditional multilayer perceptron (MLP) networks have been developed and successfully applied in estimating Delta salinity [34,37,39]. This study shows that, when driven by the same NDO and Martinez stage input features, deep learning neural networks (e.g., LSTM1 and LSTM2) distinctly outperform the classic MLP networks (e.g., MLP1 and MLP2) in estimating different ranges of salinity at Martinez and across different water year types. The study further shows that, when trained and validated using only half (versus 85% in those previous MLP studies) of the dataset in the record period, deep learning models

(e.g., LSTM3 and LSTM4) can outperform the well-calibrated process-based model (i.e., PEST-calibrated MBSG). These findings lay foundation for developing more sophisticated and carefully designed deep learning architectures to further improve salinity (particularly high range salinity as discussed in Section 4.2) estimation in the Delta. For instance, previous studies have shown that the general ability of artificial neural networks (ANNs) can be improved by combining several ANNs in an ensemble [76–79]. This study indicates that different deep learning neural network models exhibit different strengths in modeling different ranges of salinity across different water year types. Combining the strengths of different models is expected to yield better performance than using individual models alone. This can be achieved by assigning a weight to the output of each model. The weights can be determined from different methods ranging from simple averaging to more complicated Bayesian methods [80]. As another example, the hybrid model examined in this study applies the one-dimensional convolutional neural network (Conv1D) to recognize abstract patterns in dense stage observations which may be ignored by the process-based model (Equation (5)). As indicated previously (Section 2.1; Section 4.1), net Delta outflow (NDO) is the primary predictor for Martinez salinity while stage is the secondary predictor. Presumably applying Conv1D directly to NDO (versus stage) should yield even better performance. Both fronts (i.e., multi-model ensemble and Conv1D configuration for NDO) will be explored in our future work.

From a practical perspective, the PEST-calibrated Martinez Boundary Salinity Generator (MBSG) is mainly applied in generating downstream boundary salinity for the hydrodynamics and water quality model Delta Simulation Model II (DSM2). DSM2 is the operational model used in guiding real-time State Water Project (SWP) and Central Valley Project (CVP) operations and long-term Delta planning studies ranging from climate change, water system operations, to assessment of impacts of potential physical changes in the Delta (dredging, subsidence, island flooding, new water infrastructure, etc.) [81]. The deep learning models (e.g., LSTM3, LSTM4, and the hybrid model) developed in this study have the potential to supplement MBSG for this purpose. In addition, DSM2 currently simulates salinity at 90 locations including those water quality compliance locations in the Delta. The deep learning models developed in this study can be extended to emulate DSM2 in simulating salinity at all these locations. These models, once trained and validated, are expected to run much faster than DSM2. This is very meaningful and particularly appealing to time-sensitive (i.e., real-time) operations in the Delta. These models are also more flexible in terms of requiring less input data. For example, DSM2 needs channel geometry data to accurately simulate flow conditions based on which salinity is derived. The deep learning models, in comparison, does not necessarily need such input and can learn from in-situ flow observations directly. As another example, the current operational version of DSM2 uses Martinez salinity as its downstream salinity boundary as Martinez serves as the physical downstream boundary of the model, while Martinez is not a salinity source and its salinity level is dominated by salty tides from the Pacific Ocean. This study illustrates that the deep learning model can be adapted to directly include sea level as an additional model input (i.e., LSTM4/SL in Section 4.2). The results compare favorably to that of other deep learning models proposed earlier which have been shown to outperform the benchmark MBSG (Figure 10). This flexibility makes deep learning models distinctly attractive to long-term planning studies as sea level rise is projected to be a growing stress to the Delta as well as water operations in the Delta [82,83]. Emulation of DSM2 via deep learning with sea level as an additional input is ongoing and will be reported in our future work.

5. Conclusions

This study aims to explore the potential of deep learning techniques in emulating the process-based Martinez Boundary Salinity Generator (MBSG) in simulating downstream salinity boundary for the Sacramento–San Joaquin Delta of California, United States. The calibrated MBSG is used as the benchmark model. Results indicate that deep learning neural networks are able to provide Martinez salinity simulations that are competitive or superior compared with the benchmark model, particularly when the output of the latter are incorporated as an input to the former. The improvements are

generally more noticeable during extreme (wet, dry, and critical) years rather than in near-normal (above-normal and below-normal) years and during low and medium ranges of salinity rather than high range salinity. In a nutshell, this study indicates that deep learning approaches have the potential to supplement the current practices in estimating salinity at Martinez and other locations across the Delta, and thus guide real-time operations and long-term planning activities in the Delta.
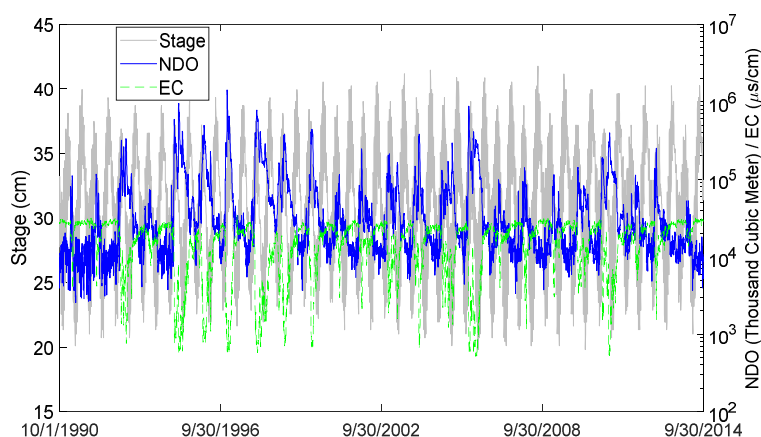
Appendix A

Figure A1. Daily Martinez water stage (Stage), net Delta outflow (NDO) and Martinez salinity represented by electrical conductivity (EC) during the study period (water year 1991–2014).
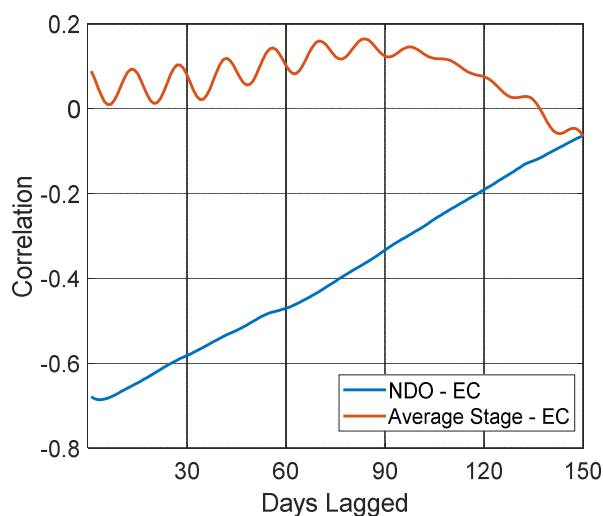

Figure A2. Correlation between daily Martinez salinity and lagged Martinez stage as well as net Delta outflow.
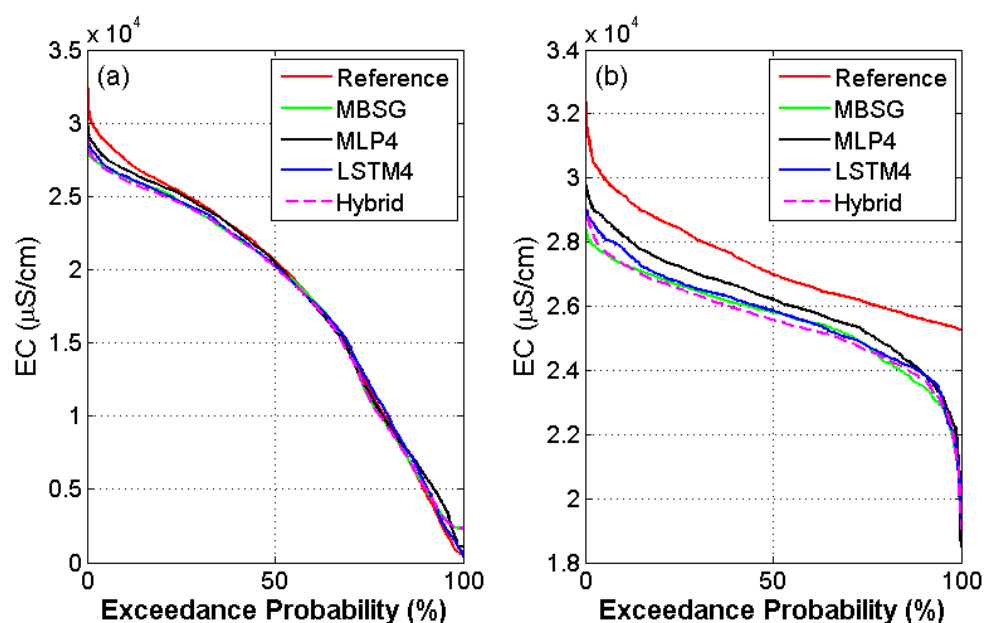
Figure A3. Exceedance probability curves of reference and simulated (via MBSG, MLP4, LSTM4, and the hybrid model) (**a**) salinity and (**b**) high range only salinity during the evaluation period from water year 2003 to 2014.

Table A1. The maximum 14 day running average of mean daily salinity (EC, µs/cm) at Emmaton and Jersey Point in different types of water year [20].

Water Year Type	Emmaton		Jersey Point	
	450 EC from 1 April to Date Shown	EC Value from Date Shown to 15 August	450 EC from 1 April to Date Shown	EC Value from Date Shown to 15 August
Wet (W)	15 August	-	15 August	-
Above-Normal (AN)	1 July	630	15 August	-
Below-Normal (BN)	20 June	1140	20 June	740
Dry (D)	15 June	1670	15 June	1350
Critical (C)	-	2780	-	2200

Table A2. Hyper-parameters of the proposed MLP and LSTM models.

Models	Number of Hidden Layers	Number of Neurons (for MLP) or Filters (for LSTM)	Batch Size	Initial Learning Rate	Dropout between Layers	Dropout between Time Steps
MLP1	2	16 and 4	128	1×10^{-3}	-	-
MLP2	2	16 and 4	64	1×10^{-3}	-	-
MLP3	2	32 and 2	128	1×10^{-3}	-	-
MLP4	2	32 and 2	128	1×10^{-6}	-	-
LSTM1	2	64 and 8	256	1×10^{-2}	0	25%
LSTM2	2	8 and 32	64	1×10^{-2}	0	25%
LSTM3	3	32, 32, and 64	128	1×10^{-2}	0	0
LSTM4	3	32, 32, and 64	128	1×10^{-6}	0	0

Table A3. Percent differences (%) between the standard deviation of reference salinity and simulated salinity at Martinez in different types of water years.

Scenarios	Water Year Type				
	W	AN	BN	D	C
MBSG	−7.5	−0.6	−6.2	−12.1	−21.5
MLP1	−1.4	−17.5	−1.2	5.6	**−1.0**
MLP2	−5.4	−21.0	−4.9	**1.1**	−2.9
MLP3	−6.2	−2.7	−1.6	−5.3	−13.2
MLP4	−5.0	−2.3	**−1.1**	−5.1	−13.2
LSTM1	−2.3	0.9	−3.6	−7.5	−17.9
LSTM2	−9.8	−8.8	−11.4	−12.9	−23.4
LSTM3	**−0.5**	**0.3**	−4.9	−11.4	−21.3
LSTM4	−5.0	−2.2	−7.0	−10.7	−19.7
Hybrid	−8.6	−1.6	−5.7	−11.7	−20.4

Bold numbers represent the best metrics.

Table A4. Correlation between reference and simulated Martinez salinity in different types of water years.

Scenarios	Water Year Type				
	W	AN	BN	D	C
MBSG	0.978	0.959	0.968	0.941	0.941
MLP1	0.938	0.900	0.873	0.834	0.741
MLP2	0.930	0.912	0.890	0.847	0.792
MLP3	0.977	0.960	0.967	0.944	0.930
MLP4	0.977	0.961	0.967	0.945	0.931
LSTM1	0.975	0.952	0.948	0.925	0.930
LSTM2	0.981	0.959	0.957	0.937	0.938
LSTM3	**0.985**	0.963	**0.970**	0.948	0.950
LSTM4	**0.985**	0.963	0.967	**0.952**	**0.955**
Hybrid	0.981	**0.964**	**0.970**	0.948	0.950

Bold numbers represent the best metrics.

References

1. Alber, M. A conceptual model of estuarine inflow policy. *Estuaries* **2002**, *25*, 1246–1261. [CrossRef]
2. Jassby, A.D.; Kimmerer, W.J.; Monismith, S.G.; Armor, C.; Cloern, J.E.; Powell, T.M.; Schubel, J.R.; Vendlinski, T.J. Isohaline position as a habitat indicator for estuarine populations. *Ecol. Appl.* **1995**, *5*, 272–289. [CrossRef]
3. Feyrer, F.; Hobbs, J.; Sommer, T. Salinity inhabited by age-0 splittail (Pogonichthys macrolepidotus) as determined by direct field observation and retrospective analyses with otolith chemistry. *San Franc. Estuary Watershed Sci.* **2010**, 10. [CrossRef]
4. Moyle, P.B.; Lund, J.R.; Bennett, W.A.; Fleenor, W.E. Habitat variability and complexity in the upper San Francisco Estuary. *San Franc. Estuary Watershed Sci.* **2010**, 24. [CrossRef]
5. Howard, R.J.; Mendelssohn, I.A. Salinity as a constraint on growth of oligohaline marsh macrophytes. I. Species variation in stress tolerance. *Am. J. Bot.* **1999**, *86*, 785–794. [CrossRef]
6. Weilhoefer, C.L.; Nelson, W.G.; Clinton, P.; Beugli, D.M. Environmental determinants of emergent macrophyte vegetation in Pacific Northwest estuarine tidal wetlands. *Estuaries Coasts* **2013**, *36*, 377–389. [CrossRef]
7. Borgnis, E.; Boyer, K.E. Salinity tolerance and competition drive distributions of native and invasive submerged aquatic vegetation in the Upper San Francisco Estuary. *Estuaries Coasts* **2016**, *39*, 707–717. [CrossRef]

8. Fernández-Delgado, C.; Baldó, F.; Vilas, C.; García-González, D.; Cuesta, J.; González-Ortegón, E.; Drake, P. Effects of the river discharge management on the nursery function of the Guadalquivir river estuary (SW Spain). *Hydrobiologia* **2007**, *587*, 125–136. [CrossRef]

9. Wang, Y.; Jiao, J.J. Origin of groundwater salinity and hydrogeochemical processes in the confined Quaternary aquifer of the Pearl River Delta, China. *J. Hydrol.* **2012**, *438*, 112–124. [CrossRef]

10. Smajgl, A.; Toan, T.Q.; Nhan, D.K.; Ward, J.; Trung, N.H.; Tri, L.; Tri, V.; Vu, P. Responding to rising sea levels in the Mekong Delta. *Nat. Clim. Chang.* **2015**, *5*, 167–174. [CrossRef]

11. Hutton, P.H.; Rath, J.S.; Chen, L.; Ungs, M.J.; Roy, S.B. Nine decades of salinity observations in the San Francisco Bay and Delta: Modeling and trend evaluations. *J. Water Resour. Plan. Manag.* **2016**, *142*, 04015069. [CrossRef]

12. Johns, N.D.; Heger, N.A. Characterizing estuarine salinity patterns with event duration and frequency of reoccurrence approaches. *Limnol. Oceanogr.-Meth.* **2018**, *16*, 180–198. [CrossRef]

13. United States Census Bureau. United States Census Bureau QuickFacts: United States. 2019. Available online: https://www.census.gov/quickfacts/fact/table/US/PST045219 (accessed on 1 May 2020).

14. Sabet, M.H.; Coe, J.Q. Models for water and power scheduling for the California State Water Project. *J. Am. Water Resours.* **1986**, *22*, 587–596. [CrossRef]

15. Becker, L.; Yeh, W.; Fults, D.; Sparks, D. Operations models for central valley project. *J. Water Resour. Plann. Manag. Div. Am. Soc. Civ. Eng.* **1976**, *102*, 101–115.

16. Myers, N.; Mittermeier, R.A.; Mittermeier, C.G.; Da Fonseca, G.A.; Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **2000**, *403*, 853. [CrossRef]

17. Moyle, P.B.; Brown, L.R.; Durand, J.R.; Hobbs, J.A. Delta smelt: Life history and decline of a once-abundant species in the San Francisco Estuary. *San Franc. Estuary Watershed Sci.* **2016**, 30. [CrossRef]

18. Healey, M.; Dettinger, M.; Norgaard, R. Perspectives on Bay–Delta Science and Policy. *San Franc. Estuary Watershed Sci.* **2016**, 25.

19. DSC. *The Delta Plan*; Delta Stewardship Counci: Sacramento, CA, USA, 2013.

20. CSWRCB. *Water right Decision 1641*; CSWRCB: Sacramento, CA, USA, 1999; p. 225.

21. USFWS. *Formal Endangered Species Act Consultation on the Proposed Coordinated Operations of the Central Valley Project (CVP) and State Water Project (SWP)*; USFWS: Sacramento, CA, USA, 2008; p. 410.

22. Denton, R.A. Accounting for antecedent conditions in seawater intrusion modeling—Applications for the San Francisco Bay-Delta. In *Hydraulic Engineering*; CSWRCB: Sacramento, CA, USA, 1993; pp. 448–453.

23. CDWR. Calibration and verification of DWRDSM. In *Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh: 12th Annual Progress Report*; CDWR: Sacramento, CA, USA, 1991.

24. DeGeorge, J.F. *A Multi-Dimensional Finite Element Transport Model Utilizing a Characteristic-Galerkin Algorithm*; University of California: Davis, CA, USA, 1996.

25. Cheng, R.T.; Casulli, V.; Gartner, J.W. Tidal, residual, intertidal mudflat (TRIM) model and its applications to San Francisco Bay, California. *Estuar. Coast. Shelf Sci.* **1993**, *36*, 235–280. [CrossRef]

26. CDWR. Bay-Delta SCHISM Model Developments and Applications. In *Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh: 36th Annual Progress Report*; CDWR: Sacramento, CA, USA, 2015.

27. Chao, Y.; Farrara, J.D.; Zhang, H.; Zhang, Y.J.; Ateljevich, E.; Chai, F.; Davis, C.O.; Dugdale, R.; Wilkerson, F. Development, implementation, and validation of a modeling system for the San Francisco Bay and Estuary. *Estuar. Coast. Shelf Sci.* **2017**, *194*, 40–56. [CrossRef]

28. Casulli, V.; Zanolli, P. Semi-implicit numerical modeling of nonhydrostatic free-surface flows for environmental problems. *Math. Comput. Model* **2002**, *36*, 1131–1149. [CrossRef]

29. MacWilliams, M.; Bever, A.J.; Foresman, E. 3-D simulations of the San Francisco Estuary with subgrid bathymetry to explore long-term trends in salinity distribution and fish abundance. *San Franc. Estuary Watershed Sci.* **2016**, 24. [CrossRef]

30. Fringer, O.; Gerritsen, M.; Street, R. An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. *Ocean Model.* **2006**, *14*, 139–173. [CrossRef]

31. Hsu, K.; Stacey, M.T.; Holleman, R.C. Exchange between an estuary and an intertidal marsh and slough. *Estuaries Coasts* **2013**, *36*, 1137–1149. [CrossRef]

32. MacWilliams, M.L.; Ateljevich, E.S.; Monismith, S.G.; Enright, C. An Overview of Multi-Dimensional Models of the Sacramento–San Joaquin Delta. *San Franc. Estuary Watershed Sci.* **2016**, 35.

33. Ateljevich, E. Improving Estimates of Salinity at the Martinez Boundary. In *Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh: 22nd Annual Progress Report*; CDWR: Sacramento, CA, USA, 2001.

34. Finch, R.; Sandhu, N. *Artificial Neural Networks with Application to the Sacramento–San Joaquin Delta*; California Department of Water Resources Delta Modeling Section, Division of Planning: Sacramento, CA, USA, 1995.

35. CDWR. Integration of CALSIM and Artificial Neural Networks Models for Sacramento-San Joaquin Delta Flow-Salinity Relationships. In *Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh: 22nd Annual Progress Report*; CDWR: Sacramento, CA, USA, 2001.

36. Chung, F.I.; Seneviratne, S.A. Developing artificial neural networks to represent salinity intrusions in the Delta. In Proceedings of the World Environmental and Water Resources Congress 2009: Great Rivers, Kansas City, MO, USA, 17–21 May 2009; pp. 1–10.

37. Rath, J.S.; Hutton, P.H.; Chen, L.; Roy, S.B. A hybrid empirical-Bayesian artificial neural network model of salinity in the San Francisco Bay-Delta estuary. *Environ. Model. Softw.* **2017**, *93*, 193–208. [CrossRef]

38. Chen, L.; Roy, S.B.; Hutton, P.H. Emulation of a process-based estuarine hydrodynamic model. *Hydrol. Sci. J.* **2018**, *63*, 783–802. [CrossRef]

39. Jayasundara, N.C.; Seneviratne, S.A.; Reyes, E.; Chung, F.I. Artificial Neural Network for Sacramento–San Joaquin Delta Flow–Salinity Relationship for CalSim 3.0. *J. Water Resour. Plan. Manag.* **2020**, *146*, 04020015. [CrossRef]

40. Maier, H.R.; Jain, A.; Dandy, G.C.; Sudheer, K.P. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* **2010**, *25*, 891–909. [CrossRef]

41. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]

42. Shen, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* **2018**, *54*, 8558–8593. [CrossRef]

43. Elman, J.L. Finding structure in time. *Cogn. Sci.* **1990**, *14*, 179–211. [CrossRef]

44. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef]

45. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

46. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [CrossRef]

47. Gu, H.; Xu, Y.-P.; Ma, D.; Xie, J.; Liu, L.; Bai, Z. A surrogate model for the Variable Infiltration Capacity model using deep learning artificial neural network. *J. Hydrol.* **2020**, 125019. [CrossRef]

48. Tian, Y.; Xu, Y.-P.; Yang, Z.; Wang, G.; Zhu, Q. Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. *Water* **2018**, *10*, 1655. [CrossRef]

49. Fan, H.; Jiang, M.; Xu, L.; Zhu, H.; Cheng, J.; Jiang, J. Comparison of Long Short Term Memory Networks and the Hydrological Model in Runoff Simulation. *Water* **2020**, *12*, 175. [CrossRef]

50. Zhang, J.; Zhu, Y.; Zhang, X.; Ye, M.; Yang, J. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* **2018**, *561*, 918–929. [CrossRef]

51. Bowes, B.D.; Sadler, J.M.; Morsy, M.M.; Behl, M.; Goodall, J.L. Forecasting Groundwater Table in a Flood Prone Coastal City with Long Short-term Memory and Recurrent Neural Networks. *Water* **2019**, *11*, 1098. [CrossRef]

52. Liang, C.; Li, H.; Lei, M.; Du, Q. Dongting lake water level forecast and its relationship with the three gorges dam based on a long short-term memory network. *Water* **2018**, *10*, 1389. [CrossRef]

53. Yan, J.; Chen, X.; Yu, Y.; Zhang, X. Application of a Parallel Particle Swarm Optimization-Long Short Term Memory Model to Improve Water Quality Data. *Water* **2019**, *11*, 1317. [CrossRef]

54. Zhou, J.; Wang, Y.; Xiao, F.; Wang, Y.; Sun, L. Water Quality Prediction Method Based on IGRA and LSTM. *Water* **2018**, *10*, 1148. [CrossRef]

55. Zhang, D.; Peng, Q.; Lin, J.; Wang, D.; Liu, X.; Zhuang, J. Simulating Reservoir Operation Using a Recurrent Neural Network Algorithm. *Water* **2019**, *11*, 865. [CrossRef]

56. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; NIPS: Lake Tahoe, CA, USA, 2012; pp. 1097–1105.

57. Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; Van Gool, L. Domain adaptive faster r-cnn for object detection in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3339–3348.

58. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.* **2019**, *38*, 1–12. [CrossRef]

59. Zhong, L.; Hu, L.; Zhou, H. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [CrossRef]

60. Afzaal, H.; Farooque, A.A.; Abbas, F.; Acharya, B.; Esau, T. Groundwater estimation from major physical hydrology components using artificial neural networks and deep learning. *Water* **2020**, *12*, 5. [CrossRef]

61. Pan, B.; Hsu, K.; AghaKouchak, A.; Sorooshian, S. Improving precipitation estimation using convolutional neural network. *Water Resour. Res.* **2019**, *55*, 2301–2321. [CrossRef]

62. Miao, Q.; Pan, B.; Wang, H.; Hsu, K.; Sorooshian, S. Improving monsoon precipitation prediction using combined convolutional and long short term memory neural network. *Water* **2019**, *11*, 977. [CrossRef]

63. Kimura, N.; Yoshinaga, I.; Sekijima, K.; Azechi, I.; Baba, D. Convolutional Neural Network Coupled with a Transfer-Learning Approach for Time-Series Flood Predictions. *Water* **2020**, *12*, 96. [CrossRef]

64. Wang, Y.; Fang, Z.; Hong, H.; Peng, L. Flood susceptibility mapping using convolutional neural network frameworks. *J. Hydrol.* **2020**, *582*, 124482. [CrossRef]

65. Sandhu, P.; Zhou, Y. Calibrating the Martinez Boundary Salinity Generator Using PEST. In *Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh: 36th Annual Progress Report*; CDWR: Sacramento, CA, USA, 2015.

66. CDWR. *On Estimating Net Delta Outflow (NDO): Approaches to Estimating NDO in the Sacramento-San Joaquin Delta*; CDWR: Sacramento, CA, USA, 2016.

67. Doherty, J.; Johnston, J.M. Methodologies for calibration and predictive analysis of a watershed model. *J. Am. Water Resour. Assoc.* **2003**, *39*, 251. [CrossRef]

68. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

69. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [CrossRef]

70. Mann, H.B. Nonparametric tests against trend. *Econometrica* **1945**, *13*, 245–259. [CrossRef]

71. Kendall, M. *Rank Correlation Methods*; Charles Griffin: London, UK, 1975.

72. Thiel, H. A rank-invariant method of linear and polynomial regression analysis, Part 3. In Proceedings of the Koninalijke Nederlandse Akademie van Weinenschatpen A, Amsterdam, The Netherlands, 25 February 1950; volume 53, pp. 1397–1412.

73. Sen, P.K. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **1968**, *63*, 1379–1389. [CrossRef]

74. Cloern, J.E.; Knowles, N.; Brown, L.R.; Cayan, D.; Dettinger, M.D.; Morgan, T.L.; Schoellhamer, D.H.; Stacey, M.T.; Van der Wegen, M.; Wagner, R.W. Projected evolution of California's San Francisco Bay-Delta-River system in a century of climate change. *PLoS ONE* **2011**, *6*, e24465. [CrossRef]

75. Pierce, D.W.; Kalansky, J.F.; Cayan, D.R. *Climate, Drought, and Sea Level Rise Scenarios for California's Fourth Climate Change Assessment*; Technical Report CCCA4-CEC-2018-006; California Energy Commission: Sacramento, CA, USA, 2018.

76. Shu, C.; Burn, D.H. Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resour. Res.* **2004**, *40*. [CrossRef]

77. Tiwari, M.K.; Adamowski, J. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resour. Res.* **2013**, *49*, 6486–6507. [CrossRef]

78. Berkhahn, S.; Fuchs, L.; Neuweiler, I. An ensemble neural network model for real-time prediction of urban floods. *J. Hydrol.* **2019**, *575*, 743–754. [CrossRef]

79. Elkiran, G.; Nourani, V.; Abba, S. Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *J. Hydrol.* **2019**, *577*, 123962. [CrossRef]

80. Hoeting, J.A.; Madigan, D.; Raftery, A.E.; Volinsky, C.T. Bayesian model averaging: A tutorial. *Stat. Sci.* **1999**, 382–401.

81. Anderson, J.; Chung, F.; Anderson, M.; Brekke, L.; Easton, D.; Ejeta, M.; Peterson, R.; Snyder, R. Progress on incorporating climate change into management of California's water resources. *Clim. Chang.* **2008**, *87*, 91–108. [CrossRef]

82. Dettinger, M.; Anderson, J.; Anderson, M.; Brown, L.R.; Cayan, D.; Maurer, E. Climate change and the Delta. *San Franc. Estuary Watershed Sci.* **2016**, 26. [CrossRef]

83. Ruckert, K.L.; Oddo, P.C.; Keller, K. Impacts of representing sea-level rise uncertainty on future flood risks: An example from San Francisco Bay. *PLoS ONE* **2017**, *12*, e0174666. [CrossRef] [PubMed]