Environmental Science & Technology

pubs.acs.org/est

Article

Predicting Dynamic Riverine Nitrogen Export in Unmonitored Watersheds: Leveraging Insights of AI from Data-Rich Regions

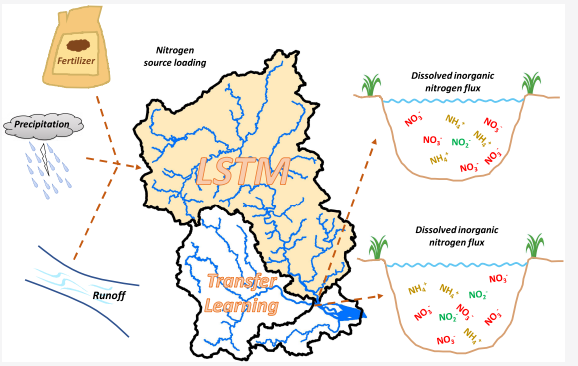
Rui Xiong, Yi Zheng,* Nengwang Chen, Qing Tian, Wei Liu, Feng Han, Shijie Jiang, Mengqian Lu, and Yan Zheng

Cite This: https://doi.org/10.1021/acs.est.2c02232

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Terrestrial export of nitrogen is a critical Earth system process, but its global dynamics remain difficult to predict at a high spatiotemporal resolution. Here, we use deep learning (DL) to model daily riverine nitrogen export in response to hydrometeorological and anthropogenic drivers. Long short-term memory (LSTM) models for the daily concentration and flux of dissolved inorganic nitrogen (DIN) were built in a coastal watershed in southeastern China with a typical subtropical monsoon climate. The DL models exhibited excellent accuracy for both DIN concentration and flux, with Nash-Sutcliffe efficiency coefficients (NSEs) up to 0.67 and 0.92, respectively, a performance unlikely to be achieved by generic process-based models with comparable data quality. The flux model ensemble, without retraining, performed well (mean NSE = 0.32−0.84) in seven distinct watersheds in Asia, Europe, and North America, and retraining with multi-watershed data further improved the lowest NSE from 0.32 to 0.68. DL interpretation confirmed that interbasin consistency of riverine nitrogen export exists across different continents, which stems from the similarities in rainfall−runoff relationships. The multi-watershed flux model projects 0.60−12.4% increases in the nitrogen export to oceans from the studied watersheds under a 20% increase in fertilizer consumption, which rises to 6.7−20.1% with a 10% increase in runoff, indicating the synergistic effect of human activities and climate change. The DL-based method represents a successful case of explainable artificial intelligence in environmental science, providing a potential shortcut to a consistent understanding of the global daily-resolution dynamics of riverine nitrogen export under the currently limited data conditions.



KEYWORDS: *nitrogen, riverine export, nonpoint sources, deep learning, LSTM, transfer learning, artificial intelligence*

1. INTRODUCTION

Biogeochemical flows pose huge threats to terrestrial and marine ecosystems and damage the Earth system's resilience and habitability.[1,2] Quantifying the riverine export of nitrogen has long been the key to understanding the significance of global biogeochemical flows.[3−5] Given the current state of technology, the long-term and high-frequency monitoring of riverine nitrogen remains costly and applicable only to a small number of rivers worldwide. Thus, modeling approaches have advantages in estimating historical exports and projecting future exports under climate change. Process-based watershed models such as SWAT,[6] HSPF,[7] and INCA[8] can accurately simulate riverine nitrogen export with a high temporal resolution. However, it is time-consuming to develop such models, and fidelity usually requires a complete dataset and skillful modelers. In contrast, classic empirical approaches such as the export coefficient method[9] and Load Estimator (LOADEST) model[10] require fewer data and less adept modeling skills but can hardly generate models with adequate spatial transferability (i.e., the ability to apply a model in settings beyond where it has been calibrated and validated). As a compromise, hybrid approaches such as the spatially referenced regression on watershed attributes (SPARROW) technique[5,11] that integrate empirical modeling with the consideration of watershed processes have also been developed. Nevertheless, although to some extent such diversified modeling approaches are suitable at different spatial and temporal scales, they require diverse types of datasets and exhibit different levels of simulation accuracy, making it a great challenge to achieve a coherent understanding of the global riverine export of nitrogen.

Machine learning (ML), the most representative technology in artificial intelligence (AI), has been employed for water and environmental studies for more than two decades.[12−14] However, ML was largely deemed a supplement to process-based approaches and a branch of empirical approaches until

Received: March 31, 2022
Revised: June 21, 2022
Accepted: June 22, 2022

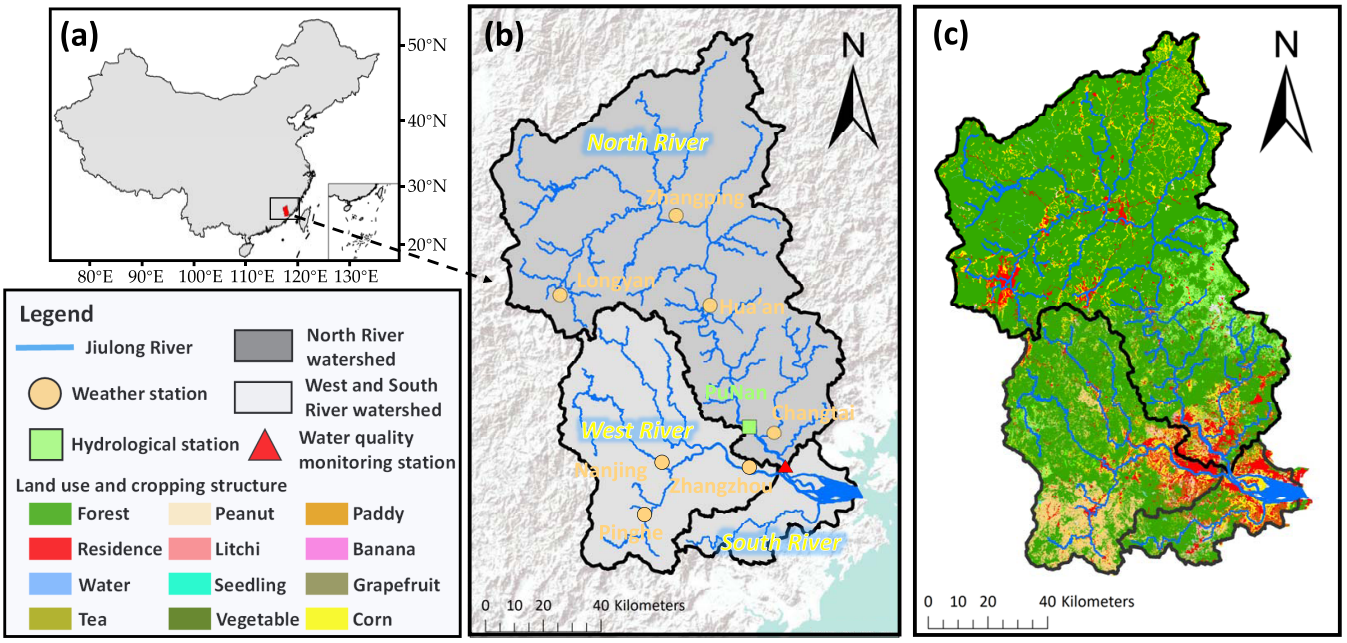


Figure 1. Study area. (a) Location of the Jiulong River watershed in China; (b) the river network and hydrological and weather stations in the watershed; and (c) land uses and cropping structure in the watershed.

recent breakthroughs in big data and deep learning (DL) redefined its role in water sciences.[15] In particular, Kratzert et al. demonstrated the great potential of long short-term memory (LSTM),[16,17] a special type of recurrent neural network (RNN), in the simulation of runoff; at present, LSTM is the most popular DL approach in hydrological modeling.[18−20] More recently, Jiang et al. developed a novel framework to implant process understanding into RNN architectures that further enhanced the simulation/prediction accuracy and intelligence in DL deductions on hydrological modeling tasks.[21] These studies achieved DL-based hydrologic models with excellent spatial transferability.[22−24] The fact that DL excels at transferring prelearned hydrological knowledge to different basins indicates the interbasin consistency of hydrological processes. Two further questions of greater importance are whether this interbasin consistency in hydrological processes manifests in biogeochemical flows, e.g., the riverine export of nitrogen, and whether DL can effectively utilize such consistency, if it exists, in transfer learning at regional, continental, and even global scales.

Recent investigations have reported a burst of LSTM applications in water quality modeling,[25−29] most of which were focused on the conditions of the target waterbody (e.g., water level, temperature, and chemical concentrations) without considering basin-scale driving forces (e.g., precipitation, irrigation, and fertilizer application). For example, Liang et al. forecasted the concentration of chlorophyll a with a 1- to 30-day lead time using six water quality parameters (including the chlorophyll a concentration itself) as the input features,[26] and Song et al. predicted the dissolved oxygen (DO) concentration 1 week ahead based on the DO concentrations over the preceding 10 weeks.[28] Partially, the good performance of LSTM is due to the collinearity between different water quality parameters and/or the strong autocorrelation of the parameters. Zhi et al. built an LSTM model to predict DO concentrations in unmonitored rivers with hydrometeorology data and streamflow data.[25] The model was trained using a large set of data collected from 236 basins across the United States, and the results indicated the interbasin consistency of water quality processes; however, the underlying mechanisms were not explored. Compared with streamflow and DO, nitrogen is involved in more complex watershed processes, and much less observational data are available. Accordingly, how to fully exploit the predictive power of DL in nitrogen modeling under such conditions and how to achieve prediction accuracies comparable to those of hydrologic modeling remain unaddressed. In addition, the black-box nature of DL models has raised many arguments about whether such models are sufficiently comprehensible.[30] Consequently, interpretative DL is now an emerging research front in many fields, including hydrology,[31−33] as it can provide insights into black-box model processes regarding how DL models produce outputs for given inputs and thereby enhance users' confidence in the prediction. A wide variety of post hoc interpretation methods, e.g., layerwise relevance propagation,[34] integrated gradients,[35] and expected gradients,[36] have been developed. However, watershed biogeochemical flows have not yet been examined with interpretative DL.

This study developed a DL-based approach to model the dynamic riverine export of nitrogen that explicitly considers nonpoint sources of nitrogen. The dissolved inorganic nitrogen (DIN, including nitrite, nitrate, and ammonia) flux of a coastal watershed in southeastern China with a typical subtropical monsoon climate was modeled using LSTM, and post hoc interpretation was performed on the DL models. The transferability of the trained models was examined in seven distinct watersheds distributed on three continents (i.e., Asia, Europe, and North America). This study aims to address three major research questions: (1) From the perspective of DL-based AI, what are the critical basin-scale driving forces of the daily riverine export of nitrogen? (2) Does the riverine export of nitrogen exhibit interbasin consistency, and if so, what is the mechanism? and (3) Can DL-based AI leverage interbasin consistency for transfer learning from data-rich watersheds to unmonitored watersheds? Overall, this study demonstrates that the deep learning method can substantially improve the

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

predictability of riverine nitrogen export, in both time and space dimensions, while respecting the existing knowledge, which represents a successful case of explainable AI[37,38] in environmental science.

2. MATERIALS AND METHODS

2.1. Study Area. The Jiulong River watershed (JRW) in southeastern China (Figure 1a) is a good testbed for studying the biogeochemical processes of nutrients under a subtropical monsoon climate.[39−41] Long-term studies in this area have accumulated an adequate amount of nutrient data for practicing DL-based modeling. The JRW has a total area of 14 740 km^2, with an average annual temperature of 20.9 °C and an average annual precipitation of 1600 mm. Three river branches, the North River, West River, and South River, drain water from the JRW into the ocean (Figure 1b). The riverine nitrogen export from the JRW was estimated to have increased nearly 5 times from 6×10^3 tN/a in 1980 to 28×10^3 tN/a in 2010.[42] The North River watershed (NRW) is the largest subbasin, with a drainage area of 9570 km^2 and a mean annual runoff of 8.33 billion m^3 (approximately 2/3 of the total runoff from the JRW); thus, the NRW was selected to train the DL model in this study.

The NRW covers five administrative areas, including the entire city of Longyan, Zhangping city, Hua'an County, Changtai County, and a part of Zhangzhou city. Seventy-eight percent of the area of the NRW is forest, while 16% comprises arable land, 3% is urban and residential land, 2% is water, and 1% is covered with bare land and grassland (Figure 1c). Agriculture is the dominant industry in the NRW, and fertilizer accounts for more than half of the nitrogen source input over this subbasin, while livestock breeding makes up 20% of the nitrogen source input.[43] Paddy rice is the main crop in the NRW, with a cropping area of 617.3 km^2. In addition, many commercial crops are planted, including tea, litchee, banana, grapefruit, seedlings, peanuts, corn, and vegetables. Peanuts, litchi, paddy rice, banana, and corn are fertilized mainly in spring and summer, while the other crops are fertilized throughout the year. The plant structure in each of the five administrative areas and the fertilization schedules for different crops are summarized in the Supporting Information (SI) in Tables S1 and S2, respectively.

2.2. Data Sources and Hypotheses of Fertilizer Application. Six years of daily streamflow data between 2014 and 2019 were collected from the Punan station on the North River, and daily meteorological data (including temperature, precipitation, relative humidity, wind speed, surface pressure, and solar radiation) during the same time period were obtained from four weather stations (Figure 1b) run by the local governments. The concentration of DIN, denoted C_{DIN}, was monitored at the outlet of the NRW by the Marine Monitoring and Information Service Center, Xiamen University (MMIS). We obtained the daily C_{DIN} data from 2014−2019 from MMIS. Additional information on the water sample processing and measurement procedures is provided in Text S1 in the SI.

One distinctive feature of the DL-based modeling scheme in this study is that a nonpoint source of nitrogen (i.e., fertilizer application and livestock breeding) is considered a basin-scale driving force, which is critical if the model is used to project future scenarios and/or to support watershed management practices. To drive daily C_{DIN} simulations, data on daily nitrogen fertilizer application rates (FARs, kg N/day) and livestock breeding are needed. However, such high-resolution information is largely unavailable in real-world situations. In this study, only yearly estimates of fertilizer application and livestock breeding at the city/county level are available from the statistical yearbooks of the cities/counties within the NRW. Therefore, to match the modeling time scale, we first derived the monthly FARs (kg N/month) of each administrative area from its yearly FARs (see Text S2 in the SI for details regarding the calculation) based on the cropping structure and fertilization schedule (see Tables S1 and S2, respectively). As no reliable information is available to further downscale the monthly FARs to daily values, an inverse approach was adopted in this study.[44] We proposed and compared the following four hypotheses.

Hypothesis 1: The monthly FAR is randomly distributed among each day in the month. With this hypothesis, the daily FAR (denoted DayFAR_i) can be expressed as

$$\mathrm{DayFAR}_i = \frac{\mathrm{RandNum}_i}{\sum_{j=1}^{n} \mathrm{RandNum}_j} \times \mathrm{MonFAR}_m,\ (i = 1, 2, \ldots, n;\ m = 1, 2, \ldots, 12) \quad (1)$$

where n is the number of days in month m, MonFAR_m is the total nitrogen FAR in month m, and RandNum is a random number that has a uniform distribution between 0 and 1.

Hypothesis 2: The monthly FAR is equally distributed among each day in the month. The daily FAR can then be expressed as

$$\mathrm{DayFAR}_i = \frac{1}{n} \times \mathrm{MonFAR}_m,\ (i = 1, 2, \ldots, n;\ m = 1, 2, \ldots, 12) \quad (2)$$

Hypothesis 3: The monthly FAR is equally distributed on each rainy day. The daily FAR can be expressed as

$$\mathrm{DayFAR}_i = \frac{\varepsilon(p_i)}{\sum_{j=1}^{n} \varepsilon(p_j)} \times \mathrm{MonFAR}_m,\ (i = 1, 2, \ldots, n;\ m = 1, 2, \ldots, 12) \quad (3)$$

where p_i is the precipitation on day i. The value of $\varepsilon(p_i)$ is 0 if $_{pi}$ = 0 and 1 if $_{pi} > 0$.

Hypothesis 4: The monthly FAR is distributed among rainy days in proportion to the daily precipitation. The daily FAR can be expressed as

$$\mathrm{DayLoad}_i = \frac{p_i}{\sum_{j=1}^{n} p_j} \times \mathrm{MonLoad}_m,\ (i = 1, 2, \ldots, n;\ m = 1, 2, \ldots, 12) \quad (4)$$

Hypotheses 1 and 2 represent two common strategies for temporal downscaling of pollutant source loading data when no auxiliary information is available, while Hypotheses 3 and 4 reflect our speculation on the potential relationship between weather conditions and fertilization behavior. All of these hypotheses were examined during the establishment of the DL models for the DIN concentration and flux, and the most plausible hypothesis was identified inversely based on model performance testing. Considering that there are no significant variations in livestock breeding scale within a year, we assumed that the pollution from livestock and poultry is equally distributed on each day.

2.3. DL Models. This study built DL models using LSTM, a network with a special RNN architecture proposed in 1997.[45]

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

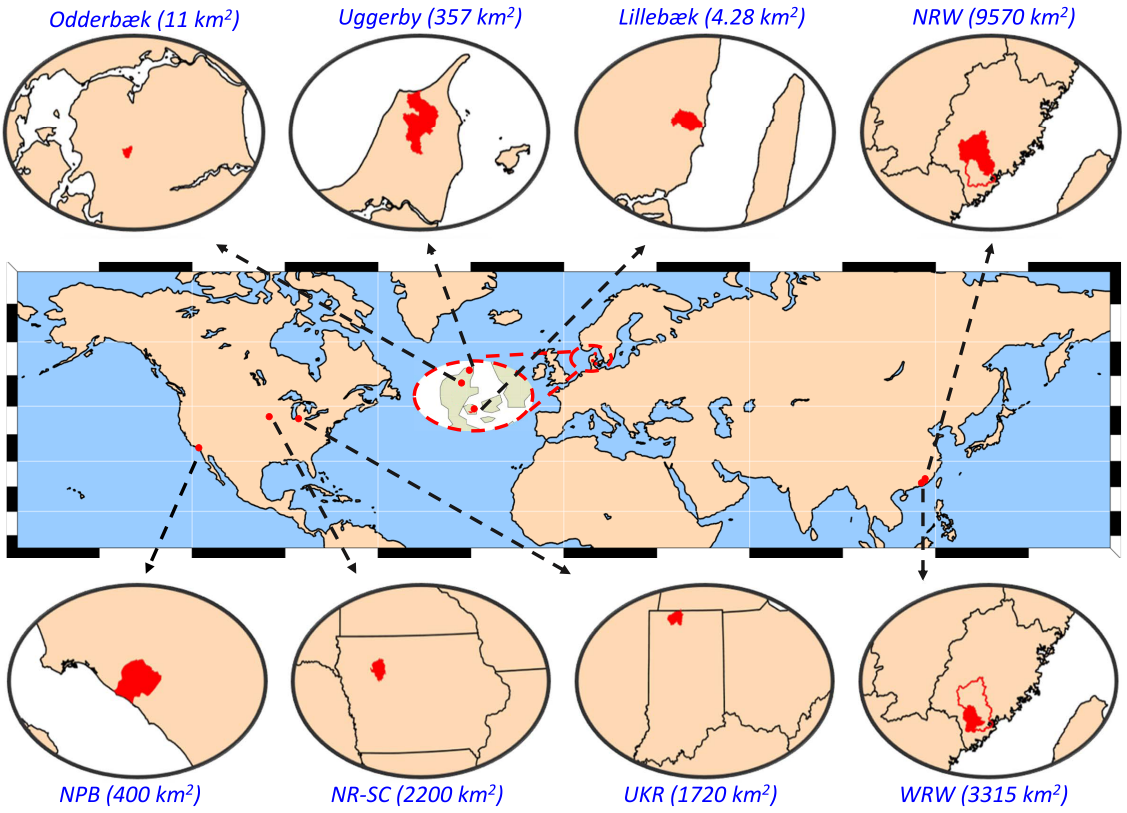


Figure 2. Locations of the North River watershed (NRW) for building the nitrogen flux models and the seven watersheds for transfer learning, including the West River watershed (WRW) in China; the Lillebæk, Odderbæk, and Uggerby watersheds in Denmark; and the Newport Bay (NPB), North Raccoon River-Sac City (NR-SC), and Upper Kankakee River (UKR) watersheds in the United States.

Compared with traditional RNNs that can remember only sequences within 10 time steps,[46] LSTM contains a memory cell that can store information over a long time period. LSTM was originally designed to solve the vanishing gradient problem encountered during the training of traditional RNNs on time series tasks. Text S3 and Figure S1 in the SI provide additional information on LSTM, and more technical details can be found in the literature.[17,47,48]

In building the LSTM models for the NRW, the data from 2014 to 2017 were used for training, while the data from 2018 to 2019 were used for testing. The runoff (Q), nonpoint sources of nitrogen (NSN), that is, nitrogen FARs and livestock breeding of different administrative areas, and meteorological variables on the present and preceding 20 days (preliminary tests showed that features beyond 20 days would have a negligible impact) were candidate model inputs (i.e., input features). The meteorological variables included temperature, precipitation (P), relative humidity, wind speed, surface pressure, and solar radiation, and basin-wide daily mean values were considered. Using basin-wide mean values of meteorological inputs is a common practice in LSTM-based hydrological modeling,[16,22,33] although recent studies show that high-resolution data fields of meteorological inputs may further enhance the prediction accuracy of neural networks.[49,50] This study used the average values derived from the four meteorological stations in the NRW (Figure 1) for the following main reasons: (1) reanalyzed meteorological data of high spatial resolution are not available for the two Chinese watersheds, and the publicly accessible reanalysis datasets are of coarse resolution (e.g., the ERA-5[51] and ERA-interim[52] datasets have grid sizes of approximately 31 and 80 km, respectively); (2) the four stations are national stations whose locations were carefully determined to ensure representativeness, and the quality of data collected at these stations is high. All input features were normalized through min-max normalization. The mean square error was adopted as the loss function, and the Nash-Sutcliffe efficiency coefficient (NSE) was used to evaluate the model performance. We performed two numerical experiments (see Table S3 in the SI for a summary) to build a variety of LSTM models for the NRW.

Experiment 1: Examining the four hypotheses of daily nitrogen FAR. In this experiment, we built four groups of LSTM models corresponding to the four hypotheses, each with 250 random replicates. The LSTM models corresponding to Hypothesis 1 were assigned fixed initial weights but a randomly selected daily FAR, which is the dominant source of uncertainty in this experiment, while the LSTM models in the other three groups were run 250 times with random initial weights. Meanwhile, for each group of LSTM models, the features that contributed the most to the C_{DIN} and F_{DIN} predictions were identified using a DL interpretation algorithm (see Section 2.4), and other features with minor contributions were excluded from the models. The hypothesis leading to the best model performance was adopted in the following analyses, and the models with the best performance in predicting C_{DIN} and F_{DIN} under this hypothesis (denoted as LSTM-C and LSTM-F, respectively) were further interpreted (see Section 2.4).

Experiment 2: Examining model transferability. In this experiment, we chose two ensembles of models under the most plausible hypothesis in Experiment 1, one for C_{DIN} (denoted as LSTM-C-TL) and the other for F_{DIN} (denoted as LSTM-F-TL), to perform transfer learning across continents

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

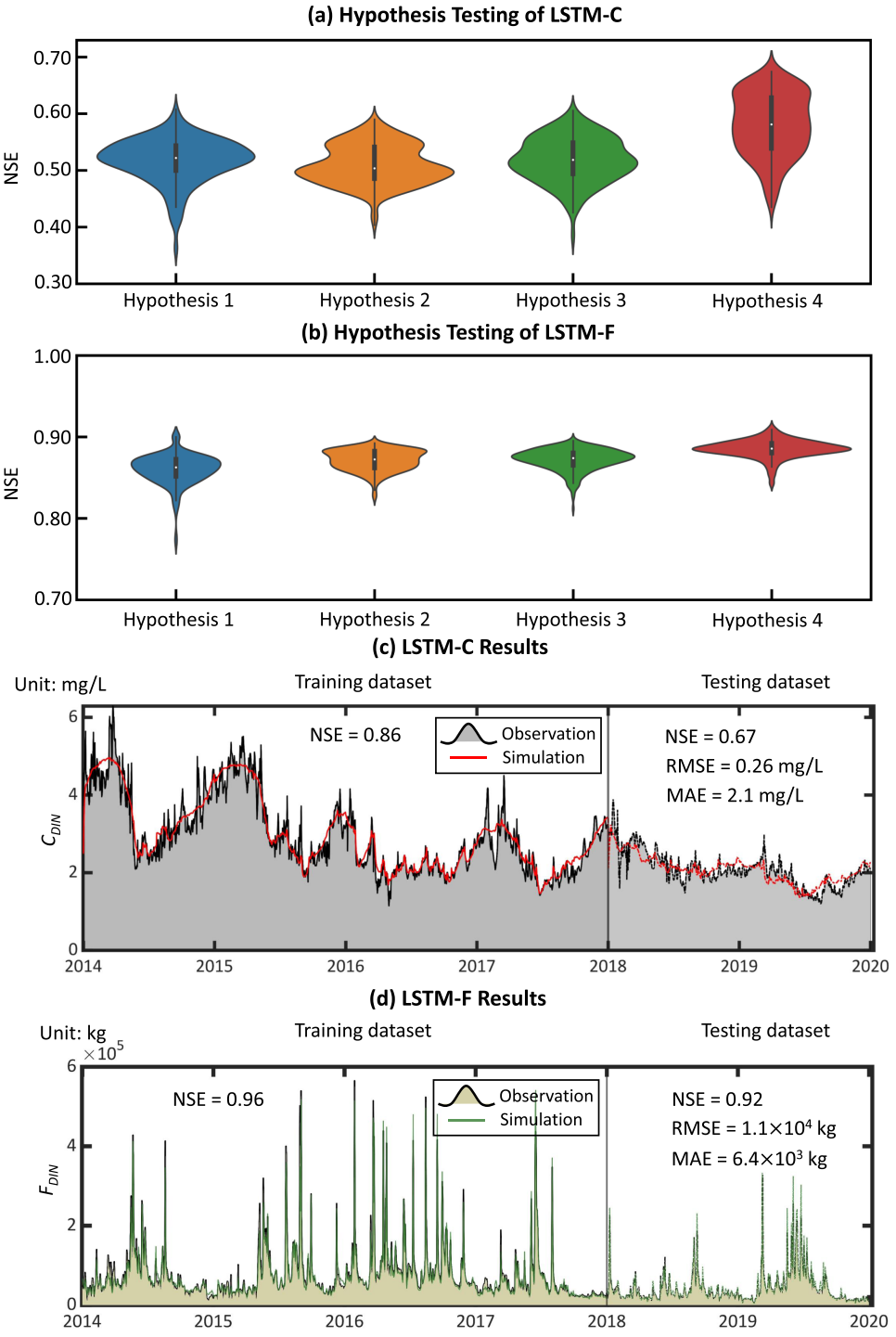


Figure 3. Performance of the deep learning models built under different hypotheses of daily fertilizer application rates (FARs). (a, b) Violin plots of NSE (testing period) under the four hypotheses for the prediction of C_{DIN} and F_{DIN}, respectively. (c, d) Simulation results of the best concentration model (LSTM-C) and flux model (LSTM-F), respectively, where colored (red and green) lines represent the simulations and black lines with shading represent the observations. The training period is 2014−2017, and the testing period is 2018−2019.

(Section 2.5). The selected models' NSE should be (1) in the range of 5−95 quantiles and (2) above 0.5.

2.4. Method of DL Interpretation. We adopted the post hoc approach known as Shapley additive explanations (SHAP)[53] to further interpret the behaviors of LSTM models. The SHAP approach is based on the Shapley value in cooperative game theory. In a cooperative game, the Shapley value assigns a unique and fair distribution of the total generated surplus among all players of a coalition.[54] The Shapley value of a feature (i.e., predictor) is its contribution to the prediction: the larger the absolute Shapley value is, the higher the contribution of the feature. Given a characteristic function val in a coalition game, the Shapley value is defined as

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

$$\phi_j(\mathrm{val}) = \sum_{S \subseteq \{x_1,\ldots,x_p\}\setminus\{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} (\mathrm{val}(S \cup \{x_j\}) - \mathrm{val}(S)) \quad (5)$$

where S is a subset of the features used in the model; x is the vector of feature values of the instance to be explained; p is the feature dimensions; and $\mathrm{val}(S\cup\{x_j\} - \mathrm{val}(S))$ refers to the marginal contribution of x_j to the prediction.

The contribution of each feature to the prediction in the SHAP approach is manifested by the SHAP value, which is an estimation of the Shapley value. Given a prediction model f, the SHAP value is calculated as

$$\phi_j = \sum_{S \subseteq \{x_1,\ldots,x_p\}\setminus\{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} (f_x(S \cup \{x_j\}) - f_x(S)) \quad (6)$$

where $f_x(S) = E[f(x)|x_s]$. By averaging the absolute SHAP values per feature across the data, we obtain the feature's global importance:

$$I_j = \sum_{i=1}^{n} |\phi_j^{(i)}| \quad (7)$$

where n is the sample size of feature j in the training dataset.

2.5. Transfer Learning across Continents. To investigate the potential interbasin consistency of riverine nitrogen export, Experiment 2 was conducted, in which the selected models (i.e., LSTM-C-TL and LSTM-F-TL) trained and tested in the NRW were "naively" (i.e., without retraining) used in seven distinct watersheds located on different continents (Figure 2). These seven watersheds included the West River Watershed (WRW), another subbasin of the JRW (Figure 1b); the Lillebæk,[55] Odderbæk,[56] and Uggerby[57] watersheds in Denmark; and the Newport Bay (NPB),[58] North Raccoon River-Sac City (NR-SC),[59] and Upper Kankakee River (UKR)[60] watersheds in the United States. These seven watersheds are significantly different in size, climate, and geographic and socioeconomic conditions (Figure 2). More detailed descriptions of these watersheds can be found in the SI (Text S5, Figure S2, and Table S4), indicating their distinctive watershed conditions. In these watersheds, we have adequate input data to drive the flux models, and the data were prepared in the same format as that in the NRW case (see Text S4 and Tables S3 and S5). The model's transferability was then assessed by comparing the simulation results against the respective observations.

For each of the seven watersheds, Pearson correlation analysis and wavelet coherence analysis[61,62] were performed to characterize the dynamic response of streamflow to rainfall. The relationships between the response characteristics and the model transferability in different watersheds were further examined, based on which insights into the interbasin consistency of riverine nitrogen export were achieved. A technical overview of the wavelet coherence analysis and results is included in Text S6 and Figure S3 in the SI.

3. RESULTS

3.1. LSTM Performance. LSTM models of DIN concentration and flux were built for the NRW (Figure 1). Post hoc interpretation using the SHAP approach quantified the global importance of each feature (i.e., the absolute SHAP value), which indicates that Q, NSN, and P are the most significant features for the prediction of C_{DIN} and F_{DIN}, while the remaining features have negligible contributions (see Figure S4 in the SI). While the importance ranking of the three features is the same for C_{DIN} and F_{DIN} (Q > NSN > P), NSN appears to have a more substantial contribution to the prediction of C_{DIN} than that to the prediction of F_{DIN}. The normalized importance score of NSN for C_{DIN} is approximately triple that for F_{DIN}. This difference may imply that the dynamics of nitrogen concentration are more dependent on biogeochemical processes than the dynamics of nitrogen flux. In this study, Q, NSN, and P over the present and preceding 20 days were used as the input features of the LSTM models.

Figure 3a illustrates the performance of the LSTM models for C_{DIN} in the testing stage under the four hypotheses of intramonthly variation in FARs (Experiment 1). The violin plot shows the distribution of the NSE as the performance measure under each hypothesis based on the 250 model replicates. The daily FARs of each administrative area under Hypothesis 1 were generated randomly in each of the 250 simulations, creating a wider NSE range (approximately 0.24). The variations in the NSE values under the other three hypotheses resulted from the uncertainties associated with model training (i.e., the random initialization of weights 250 times). The LSTM-C models under Hypothesis 4 (i.e., FARs occur on rainy days in proportion to the daily precipitation) generally exhibit significantly better performance than the models under the other three hypotheses and the same for the LSTM-F models (Figure 3b). These results indicate that Hypothesis 4 may be the most plausible assumption regarding the daily nitrogen FARs in the NRW. The plausibility of Hypothesis 4 was further confirmed in a field survey conducted in August 2020. Because overfertilization is very common in the study region, if fertilizer is applied in dry periods, farmers would have to irrigate extra water to dilute the fertilizer to ensure that the fertilizer could be effectively used by crops and to avoid burning the seedlings. Naturally, farmers would prefer applying fertilizer immediately before or during rainfall events, particularly significant events, to reduce labor and water usage. Therefore, only Hypothesis 4 was considered in transfer learning (Experiment 2).

Figure 3c,d shows the simulation results by the LSTM-C and LSTM-F models with the best performance in the testing period, respectively. For the LSTM-C model, the values of NSE on the training dataset and testing dataset are 0.86 and 0.67, respectively, and the mean predicted C_{DIN} (=2.08 mg/L) is very close to the observation (=2.07 mg/L) on the testing dataset, with the root mean square error (RMSE) and mean absolute error (MAE) being 0.26 and 2.1 mg/L, respectively. Although some extreme values are not well matched, the decreasing trend across the years and seasonal variation in C_{DIN} are nicely captured by the model. The performance of LSTM-C is closely related to the FAR data quality because daily FAR data are a rough approximation of the real situation. The mismatches of some extreme values may be attributed to the inaccuracy of the daily FAR data, which were downscaled from yearly and monthly estimates based on Hypothesis 4 (see Section 2.3). For F_{DIN}, the performance of the LSTM-F model is superior on both the training dataset and the testing dataset (NSE equaling 0.96 and 0.92, respectively). Such good performance is unlikely to be achieved by generic process-based models (e.g., SWAT) for daily nutrient modeling with comparable data quality. For benchmarking, we also built multiple linear regression models for C_{DIN} and F_{DIN} for comparison, and the respective NSE values

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

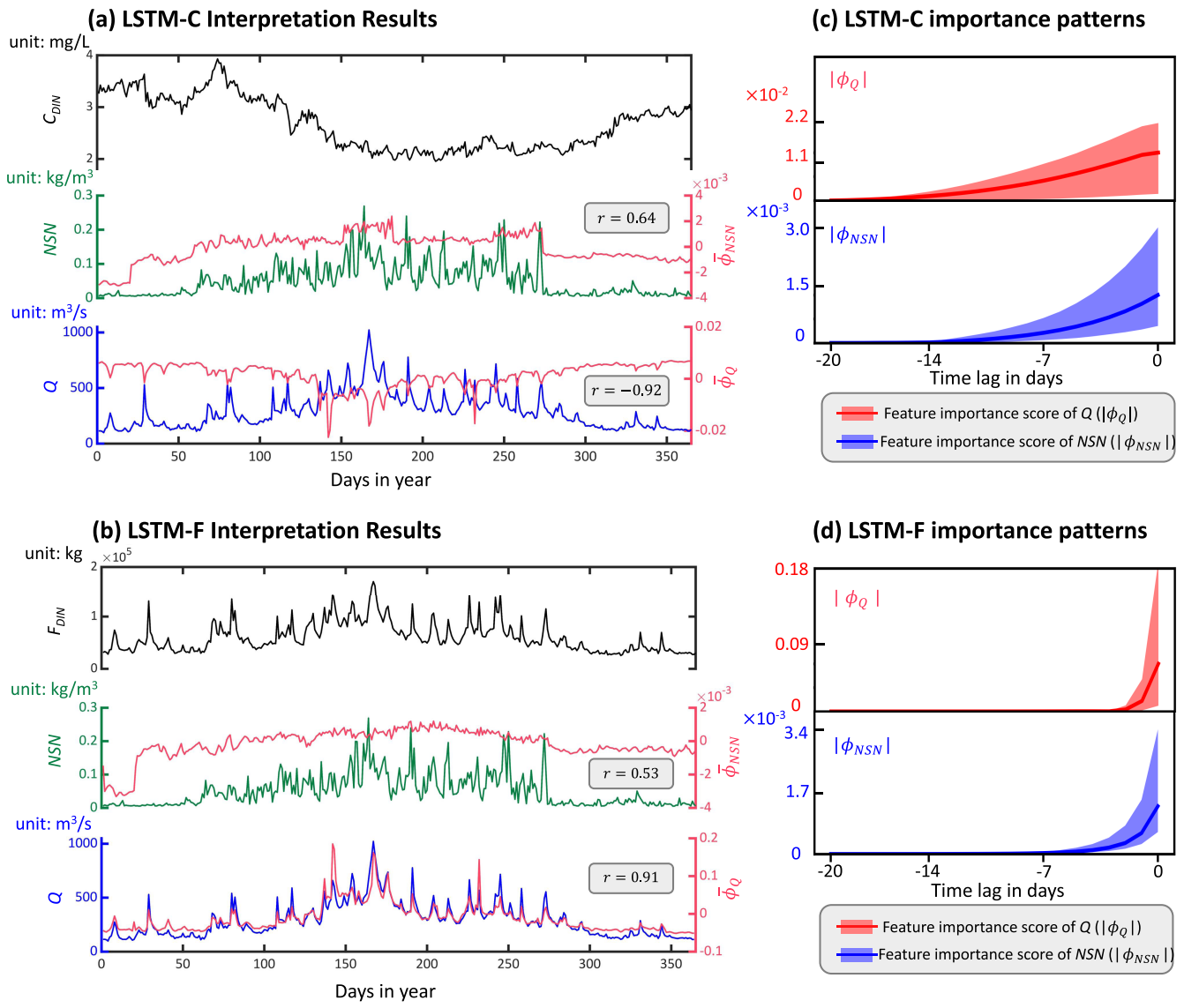


Figure 4. LSTM model interpretation results. (a, b) Mean SHAP values $\overline{\phi}_{NSN}$ and $\overline{\phi}_Q$ (averaged over the 4 years of the training period) for C_{DIN} and F_{DIN}, respectively. The time series of $\overline{\phi}_{NSN}$ and $\overline{\phi}_Q$ (red lines) reflect the accumulative (considering a time lag of up to 20 days) contributions of NSN and Q, respectively, on each calendar day. The values in the gray boxes indicate the Pearson correlation coefficients of the respective time series. The NSN (green lines) and Q (blue lines) features, as well as the observed concentration and flux (black lines), are plotted along with $\overline{\phi}_{NSN}$ and $\overline{\phi}_Q$. (c, d) Absolute SHAP values of Q ($|\phi_Q|$) and NSN ($|\phi_{NSN}|$) for C_{DIN} and F_{DIN}, respectively. $|\phi_Q|$ and $|\phi_{NSN}|$ were computed for different time lags, ranging from 0 to 20 days, across the training dataset. The upper and lower boundaries of the color bands represent the 95th and 5th percentiles, respectively, and color lines indicate the mean values.

on the testing dataset are -2.51 and 0.70, much lower than those of LSTM-C and LSTM-F.

3.2. Interpreted Model Behaviors. Figure 4a,b presents the time series of the mean SHAP values, $\overline{\phi}_{NSN}$ and $\overline{\phi}_Q$, which represent the accumulative (considering a time lag of up to 20 days) contributions of NSN and Q, respectively, to the predictions of C_{DIN} and F_{DIN} on each calendar day averaged over the 4 years of the training period. While Q is more influential than NSN on the predictions of both C_{DIN} and F_{DIN}, the magnitude of $\overline{\phi}_Q$ is 1 order of magnitude higher than $\overline{\phi}_{NSN}$ for the DIN concentration (Figure 4a) and 2 orders of magnitude higher for the flux (Figure 4b). Q is largely determined by rapid watershed processes such as surface flow formation and concentration, while nitrogen in fertilizer undergoes both fast (e.g., surface flow) and slow (e.g., groundwater flow and plant uptake) processes before it reaches the watershed outlet. Thus, Figure 4a,b implies that F_{DIN} is dominated largely by rainfall−runoff processes, while C_{DIN} involves more complex phenomena, which may explain the better performance of LSTM-F demonstrated before. Figure 4c,d illustrates the absolute SHAP values of Q ($|\phi_Q|$) and NSN ($|\phi_{NSN}|$) for different time lags; these plots further confirm that C_{DIN} involves more complex and slow processes and therefore has a longer "memory" of historical inputs than F_{DIN}. For LSTM-F (Figure 4d), the responses of F_{DIN} to runoff generation (reflected by Q), livestock breeding, and fertilizer application (reflected by NSN) diminish exponentially within 1 week, while for LSTM-C (Figure 4c), the influence of Q and NSN on C_{DIN} lasts approximately 2 weeks.

In addition, $\overline{\phi}_{NSN}$ demonstrates a significant positive correlation with NSN for both LSTM-C (middle plot in Figure 4a) and LSTM-F (middle plot in Figure 4b) and is generally

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

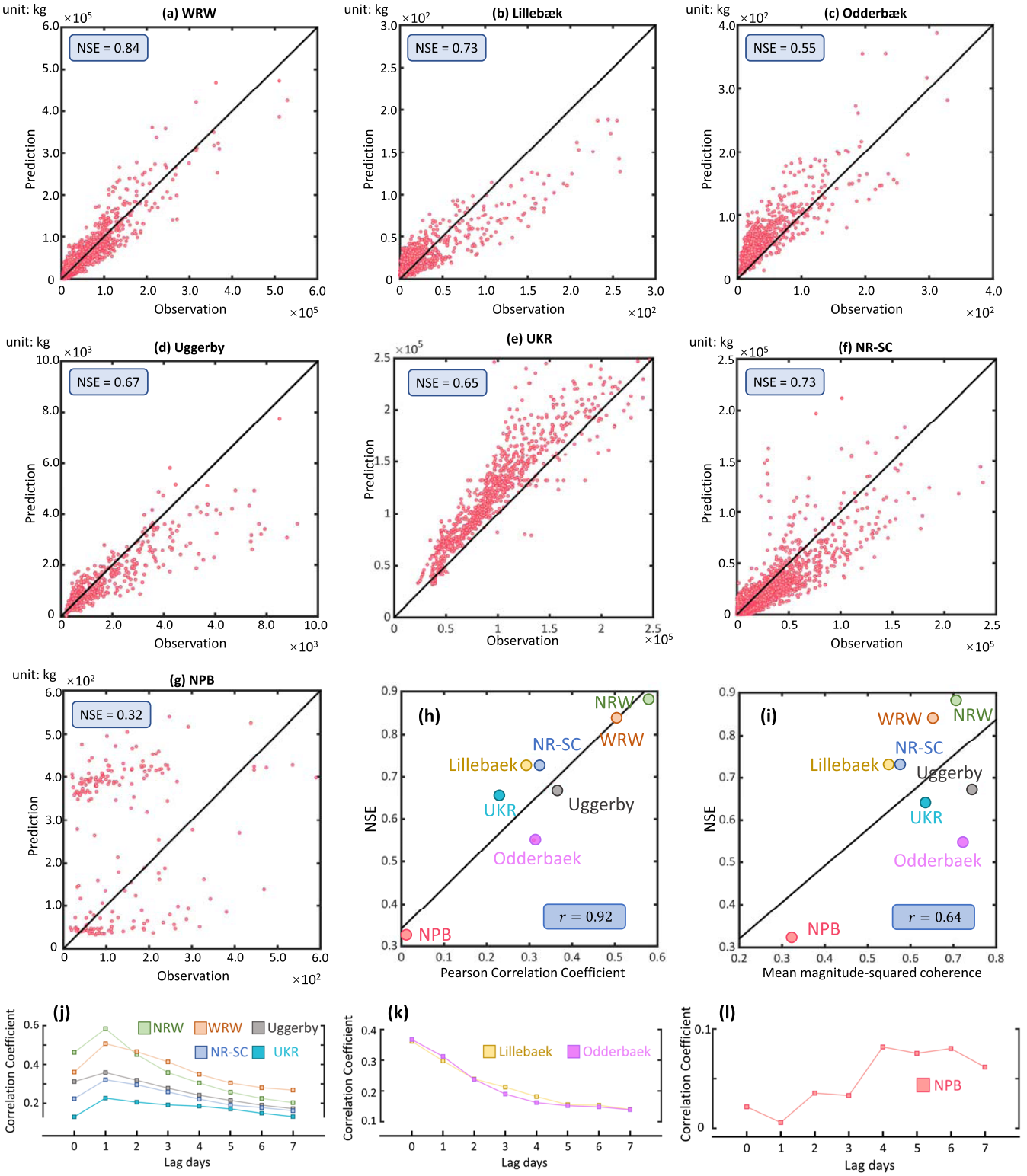


Figure 5. Performance of LSTM-F-TL in the seven watersheds for transfer learning. The goodness of fit of the model ensemble in the (a) WRW, (b) Lillebæk, (c) Odderbæk, (d) Uggerby, (e) UKR, (f) NR-SC, and (g) NPB watersheds. The *y*-axis represents the ensemble mean predictions. The mean NSE (i.e., the observations against the ensemble mean predictions) is indicated in the top left corner. Covariations of the mean NSE and (h) 1-day-lagged Pearson correlation coefficients between rainfall and runoff and (i) mean magnitude-squared coherence between rainfall and runoff. (j, k, l) Time-lagged cross-correlation between rainfall and runoff in the eight watersheds.

small in fall and winter and high in spring and summer. This pattern coincides with the timing of fertilizer application in the study area. During fall and winter, fertilizer consumption in the NRW is small, as most crops are harvested in the fall. In spring and summer, multifarious agricultural activities dramatically increase fertilizer consumption, and a large amount of reactive N is introduced into the basin. In contrast, $\overline{\phi}_Q$ is negatively correlated with Q for LSTM-C (bottom plot in Figure 4a) but

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

positively correlated with Q for LSTM-F (bottom plot in Figure 4b). In the NRW, the wet season coincides with the growing season of crops, in which fertilization application occurs. Thus, the contribution of Q to the flux is more significant in the wet season (i.e., a higher Q produces a greater nonpoint source loading of nitrogen), which explains the positive correlation between ϕ_Q and Q for LSTM-F. In the dry season (low Q), when fertilization application is minimal, an increase in Q has a significant dilution effect (i.e., the concentration is sensitive to Q), while in the wet season, the nonpoint source input and dilution effect cancel each other out to a certain degree; therefore, the concentration is less sensitive to Q, which explains the negative correlation for LSTM-C.

Overall, the temporal patterns recognized by the interpretive DL are consistent with the physical knowledge, which indicates that our LSTM models have learned the correct knowledge from the data.

3.3. Model Transferability. The LSTM-C-TL and LSTM-F-TL model ensembles built in the NRW (selected in Experiment 2) were further tested without retraining at seven distinct watersheds across three continents (Figure 2). LSTM-C-TL showed poor transferability (Figure S5 in the SI), while LSTM-F-TL achieved surprisingly good agreement between the predicted and observed F_{DIN} (Figures 5a–g and S6). This difference in transferability is consistent with outcomes of observational studies at various scales; that is, discharge varies by over several orders of magnitude, while nitrogen concentrations vary only a few factors.[63−66] Therefore, the nitrogen flux will be strongly governed by discharge. If the model accurately simulates discharge (Figure S7 in the SI shows the excellent performance of LSTM in simulating the discharge from the NRW), the nitrogen flux prediction will be reliable. In contrast, nitrogen concentrations in streams vary within only a few factors and are controlled by a more complex interplay of hydrological, geochemical, and microbiological processes; as a result, it is more difficult to predict nitrogen concentrations. In addition, inaccurate data on nonpoint sources of nitrogen may also contribute to the difference in transferability. In this study, the fertilization data of different watersheds are inconsistent (i.e., they were collected from different sources, are in different data formats and of different spatial and temporal resolutions, and have different levels of accuracies). However, the transferability of LSTM-F-TL is not significantly impacted by this data inconsistency because the importance of nonpoint sources of nitrogen to the flux simulation is much lower than their importance to the concentration simulation (Figure S4 in the SI).

LSTM-F-TL performs the best in the WRW (mean NSE = 0.84), followed by the NR-SC (mean NSE = 0.73) and Lillebæk watersheds (mean NSE = 0.73). The performance degrades slightly in the two Denmark watersheds and one U.S. watershed but is still fair. In the Uggerby (mean NSE = 0.67) watershed, LSTM-F-TL tends to underestimate large F_{DIN} values, while in the Odderbæk watershed (mean NSE = 0.55) and the UKR watershed (mean NSE = 0.65), LSTM-F-FL tends to overestimate small F_{DIN} values. The mean NSE in the NPB watershed equals 0.32, which is the lowest among the seven watersheds.

We speculate that the transferability of LSTM-F-TL is related to the rainfall−runoff relationship because F_{DIN} is strongly governed by discharge, as discussed above. Figure 5j,k,l shows the time-lagged cross-correlation between rainfall and runoff for the NRW and the seven watersheds for transfer learning, which provides a way to compare the rainfall−runoff relationships among the seven watersheds. As Figure 5j shows, the WRW, Uggerby, UKR, and NR-SC watersheds have the same relationships as the NRW, featuring a relatively high cross-correlation with the peak at a 1-day lag. The Lillebæk and Odderbæk watersheds also have relatively high cross-correlations, but the rainfall and runoff therein are almost synchronous (i.e., the correlation peaks at lag = 0), which may be due to their small drainage areas (Figure 2). The LSTM-F-TL model ensemble exhibits satisfactory transferability in these six watersheds. In contrast, the relationship in the NPB watershed is distinctive from those in the other watersheds, featuring a low, multimodal cross-correlation pattern. This unique rainfall−runoff relationship may explain the unsatisfactory transferability in the NPB watershed.

4. DISCUSSION

4.1. Interbasin Consistency. Although the seven watersheds employed for transfer learning are significantly different from the NRW in size, climate, and geographic and socioeconomic conditions, the LSTM-F-TL model built in the NRW exhibits good performance in six of them, which offers strong support for the existence of interbasin consistency. However, how does the DL model recognize and utilize this consistency? The answer may be rooted in the special architecture of LSTM that can mimic watershed nitrogen processes. The architecture of the LSTM cell has four key components, namely, the cell state, forget gate, input gate, and output gate (see Figure S1 in the SI), which may correspond to specific processes when modeling physical variables (although the correspondence is untransparent). In modeling the nitrogen flux, the cell state can analogize the watershed storage of nitrogen (e.g., soil and groundwater storage). Moreover, the update of the cell state depends on the three gates, and each gate can analogize certain processes that influence the export of nitrogen at the watershed outlet. In each time step, the input gate maps the input features (Q, NSN, and P in our case) into inputs (e.g., fertilization, livestock breeding, atmospheric deposition, and biological nitrogen fixation) that are significant enough to influence the cell state. The forget gate determines what information of the cell state in the previous time step can be ignored and can analogize processes that remove nitrogen from the watershed system, such as plant uptake, volatilization, and denitrification. Finally, the output gate determines the response of the output to the change in the cell state in each time step; intuitively, it determines how much nitrogen in the storage tank can be flushed out and delivered to the watershed outlet. Essentially, the interbasin consistency is reflected by the generalizability of the trainable parameters of the neural network across different watersheds.

The similarity between the rainfall−runoff relationship in the NRW and those in six of the seven tested watersheds (see Section 3.3) implies that the network parameters mapping the runoff generation process (possibly associated with the output gate), after being trained in the NRW, are highly generalizable in the six watersheds. Moreover, the DL interpretation in Section 3.2 indicates that the DIN export flux is routed mainly by fast flows such as surface runoff; therefore, the network parameters mapping the runoff generation process may largely determine the prediction of DIN export. Together, these pieces of evidence explain why LSTM-F-TL exhibits good performance in all of the tested watersheds except the NPB watershed and suggest that

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

similarity among hydrological responses is an important dimension of the interbasin consistency of nitrogen export.

Another dimension of the interbasin consistency may be the similarity in nitrogen input, which is possibly related to the input gate of LSTM. Agriculture is the dominant industry in all of the basins we selected for transfer learning. The high interbasin consistency of the dominant nitrogen source further endorses the transferability of LSTM-F-TL. However, the degradation in the performance of LSTM-F-TL in the three Denmark basins is possibly induced by their different patterns of nitrogen input. Farms in the NRW are small in size, and the fertilization schedule is highly decentralized and irregular, whereas Denmark has a flat terrain with relatively large farms and a much higher level of agricultural mechanization than the NRW, which results in a relatively fixed schedule of fertilization.[67]

In contrast, LSTM-C-TL demonstrates poor transferability (Figure S5), which implies inconsistency across the watersheds. As the DL interpretation results show, C_{DIN} is influenced by higher-complexity processes in addition to runoff formation, such as plant uptake and denitrification. Such processes may significantly vary among the tested watersheds. For example, the main crop in the NRW is paddy rice, which is different from the other tested watersheds except the WRW. Climate and geographic conditions may also cause dissimilarities in the water environment, for example, by varying the denitrifying bacterial population and the conditions favoring denitrification (e.g., oxygen concentration and pH). Such inconsistencies may lower the generalizability of the network parameters that are possibly associated with the forget gate. In addition, the inconsistency in the data of nonpoint source of nitrogen may also degrade the transferability of the parameters, as discussed in Section 3.3.

4.2. Toward a Global Model of Dynamic Nitrogen Export into Oceans. It remains greatly challenging to reach a consistent understanding of the global dynamics of riverine nitrogen export. This study holistically demonstrated that interbasin consistency exists for the daily DIN flux and that this consistency can be effectively utilized by DL-based AI, thereby providing a promising pathway to a global model of dynamic nitrogen export into oceans worldwide. In runoff modeling, it has been demonstrated that the tandem training of LSTM models in a number of watersheds to achieve a general model is an effective way to increase the model's transferability.[17,21,22] In this study, we also conducted a tandem training experiment for the trial. We trained a uniform LSTM model using data from seven watersheds (i.e., the NRW, WRW, Lillebæk, Odderbæk, Uggerby, UKR, and NR-SC watersheds) and tested the trained model in the NPB (featuring the least similarity with the other watersheds). Q, P, and NSN (under Hypothesis 4) on the present and preceding 20 days were used as the model inputs. The LSTM models were run 250 times with random initial weights, and the mean NSE increased from 0.32 (Figure 5g) to 0.68, which is evidence that sufficient data may enable LSTM to comprehensively identify this interbasin consistency (i.e., highly general patterns in hydrological processes, biochemical processes, and human activities). While the above co-training procedure with only seven watersheds is preliminary, it does shed light on the ability to construct a regional or even global DIN export model with broad transferability using DL.

In addition to the compilation of a global dataset for training a general model, two critical issues must be addressed before such a global model can eventually be developed. The first issue is the accurate simulation of global runoff. With the emergence of global hydrology,[68] many global hydrological models have been developed in recent decades, such as WBM,[69] WaterGAP,[70] and PCR-GLOBWB.[71] However, these process-based global models are difficult to implement and calibrate, and the simplification of the hydrological processes in these models further leads to low simulation accuracies.[72] Recently, LSTM has achieved remarkable successes in predicting streamflow, indicating that it is feasible to build an LSTM-based global streamflow model. We expect some promising results to emerge in the near future.

The second issue is the treatment of different types of interbasin consistency. As indicated by the tandem training experiment, although including more watersheds in the training set may enhance the model's transferability, the improvement may be limited if the training and testing groups of watersheds have significantly inconsistent nitrogen export characteristics. We speculate that LSTM-F-TL in this study would not exhibit good performance in watersheds where point sources (e.g., wastewater discharge) dominate the flux. One possible way to address this problem is to build submodels for different categories of watersheds. An appropriate watershed categorization may consider two aspects: (1) the rainfall−runoff relationship (the Pearson correlation coefficient and mean magnitude-squared coherence used in this study may be considered metrics) and (2) nitrogen sources. A more systematic solution is to incorporate static watershed properties (e.g., land use and soil attributes) as additional input features such that different types of interbasin consistency are internally considered by the DL model. Innovative DL architectures have been proposed in the field of hydrological modeling,[21] and further studies for nutrient modeling are needed.

4.3. Nitrogen Export into Oceans under Rising Fertilizer Consumption and Climate Change. Global fertilizer consumption is forecasted to continue to rise in the future.[73] To examine the sensitivity of nitrogen export to global oceans to rising consumption, the co-trained LSTM models (Section 4.2) were run to emulate a hypothetical scenario with a 20% increase in nitrogen fertilizer consumption. This degree of rise is consistent with the existing predictions.[74] The hypothetical scenario assumes that daily FARs are evenly increased and that the pollution from livestock and poultry remains unchanged. The increase in the mean annual DIN export in the seven watersheds ranges from 0.60 to 12.4% (0.60% for Uggerby, 0.66% for the UKR, 0.73% for Odderbæk, 1.1% for WRW, 3.3% for Lillebæk, 5.1% for NRW, and 12.4% for NPB), which reflects the significant heterogeneity across different regions. We also found that the spatially differentiated responses would be highly dependent on climate conditions. If the future climate leads to a 10% increase in runoff in the studied watersheds, the predicted increase in mean annual DIN export will range from 6.7 to 20.1% (6.7% for the UKR, 7.7% for Uggerby, 8.3% for NRW, 9.1% for Odderbæk, 10.2% for WRW, 16.6% for Lillebæk, and 20.1% for NPB). While the above analysis is preliminary, the heterogeneous response and the synergistic effect between fertilizer consumption and climate change shed light on the importance and complexity of nitrogen management in a rapidly changing global environment. Our study demonstrates that deep learning is a novel and promising approach to tackle complexity and therefore deserves more attention.

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.2c02232.

Technical details regarding the used methods; information on planting structure and fertilization schedule; detailed descriptions of the watersheds; and supplementary results (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yi Zheng − *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China; Shenzhen Municipal Engineering Lab of Environmental IoT Technologies, Southern University of Science and Technology, Shenzhen 518055 Guangdong Province, China;* orcid.org/0000-0001-8442-182X; Email: zhengy@sustech.edu.cn

Authors

Rui Xiong − *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China; Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR 999077, China*

Nengwang Chen − *Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, College of the Environment and Ecology, Xiamen University, Xiamen 361102, China;* orcid.org/0000-0002-5200-1035

Qing Tian − *Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, College of the Environment and Ecology, Xiamen University, Xiamen 361102, China*

Wei Liu − *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*

Feng Han − *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*

Shijie Jiang − *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China; Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research, Leipzig 04318, Germany*

Mengqian Lu − *Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR 999077, China*

Yan Zheng − *School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.2c02232

Notes

The authors declare no competing financial interest.

The yearly estimates of fertilizer application at the city/county level were retrieved from the statistical yearbooks issued by local governments within the NRW (http://www.longyan.gov.cn/sj/tjnj/; http://tjj.zhangzhou.gov.cn/cms/html/zzstjj/tjnj/index.html). The meteorological data were provided by the China Meteorological Administration (https://data.cma.cn/). The streamflow and nitrogen monitoring data were provided by the Marine Monitoring and Information Service Center, Xiamen University (MMIS) (http://marinecloud.xmu.edu.cn/arcgis/geoplat/viewer/functionGallery.html).

■ ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (nos. 51961125203 and 92047302) and the Shenzhen Science and Technology Innovation Commission (no. KCXFZ202002011006491).

■ REFERENCES

(1) Rockström, J.; Steffen, W.; Noone, K.; Persson, Å.; Chapin, F. S., III; Lambin, E.; Lenton, T. M.; Scheffer, M.; Folke, C.; Schellnhuber, H. J.; et al. Planetary boundaries: exploring the safe operating space for humanity. *Ecol. Soc.* **2009**, *14*, No. 32.

(2) Steffen, W.; Richardson, K.; Rockström, J.; Cornell, S. E.; Fetzer, I.; Bennett, E. M.; Biggs, R.; Carpenter, S. R.; De Vries, W.; De Wit, C. A.; et al. Planetary boundaries: Guiding human development on a changing planet. *Science* **2015**, *347*, No. 1259855.

(3) Tada, A.; Tanakamaru, H. Unbiased estimates and confidence intervals for riverine loads. *Water Resour. Res.* **2021**, *57*, No. e2020WR028170.

(4) Cohn, T. A.; Delong, L. L.; Gilroy, E. J.; Hirsch, R. M.; Wells, D. K. Estimating constituent loads. *Water Resour. Res.* **1989**, *25*, 937−942.

(5) Alexander, R. B.; Johnes, P. J.; Boyer, E. W.; Smith, R. A. A comparison of models for estimating the riverine export of nitrogen from large watersheds. *Biogeochemistry* **2002**, *57*, 295−339.

(6) Akhavan, S.; Abedi-Koupai, J.; Mousavi, S.-F.; Afyuni, M.; Eslamian, S.-S.; Abbaspour, K. C. Application of SWAT model to investigate nitrate leaching in Hamadan−Bahar Watershed, Iran. *Agric., Ecosyst. Environ.* **2010**, *139*, 675−688.

(7) Filoso, S.; Vallino, J.; Hopkinson, C.; Rastetter, E.; Claessens, L. Modeling nitrogen transport in the Ipswich River Basin, Massachusetts, using a hydrological simulation program in FORTRAN (HSPF) 1. *J. Am. Water Resour. Assoc.* **2004**, *40*, 1365−1384.

(8) Wade, A. J.; Durand, P.; Beaujouan, V.; Wessel, W. W.; Raat, K. J.; Whitehead, P. G.; Butterfield, D.; Rankinen, K.; Lepisto, A. A nitrogen model for European catchments: INCA, new model structure and equations. *Hydrol. Earth Syst. Sci.* **2002**, *6*, 559−582.

(9) Ding, X.; Shen, Z.; Hong, Q.; Yang, Z.; Wu, X.; Liu, R. Development and test of the export coefficient model in the upper reach of the Yangtze River. *J. Hydrol.* **2010**, *383*, 233−244.

(10) Runkel, R. L.; Crawford, C. G.; Cohn, T. A. *Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers*; US Geological Survey, 2004; Vol. 69.

(11) Smith, R. A.; Schwarz, G. E.; Alexander, R. B. Regional interpretation of water-quality monitoring data. *Water Resour. Res.* **1997**, *33*, 2781−2798.

(12) Bowden, G. J.; Maier, H. R.; Dandy, G. C. Optimal division of data for neural network models in water resources applications. *Water Resour. Res.* **2002**, *38*, 2-1−2-11.

(13) Maier, H. R.; Dandy, G. C. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Modell. Soft.* **2000**, *15*, 101−124.

(14) Bowden, G. J.; Dandy, G. C.; Maier, H. R. Input determination for neural network models in water resources applications. Part 1—background and methodology. *J. Hydrol.* **2005**, *301*, 75−92.

(15) Shen, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* **2018**, *54*, 8558−8593.

(16) 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.

(17) Chen, Z.; Zhu, Z.; Jiang, H.; Sun, S. Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. *J. Hydrol.* **2020**, *591*, No. 125286.

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(18) Fang, K.; Pan, M.; Shen, C. The value of SMAP for long-term soil moisture estimation with the help of deep learning. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2221−2233.
(19) Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **2020**, *56*, No. e2019WR025326.
(20) Ma, Y.; Montzka, C.; Bayat, B.; Kollet, S. Using Long Short-Term Memory networks to connect water table depth anomalies to precipitation anomalies over Europe. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 3555−3575.
(21) Jiang, S.; Zheng, Y.; Solomatine, D. Improving AI system awareness of geoscience knowledge: symbiotic integration of physical approaches and deep learning. *Geophys. Res. Lett.* **2020**, *47*, No. e2020GL088229.
(22) Feng, D.; Fang, K.; Shen, C. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resour. Res.* **2020**, *56*, No. e2019WR026793.
(23) Vandaele, R.; Dance, S. L.; Ojha, V. Deep learning for automated river-level monitoring through river-camera images: an approach based on water segmentation and transfer learning. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 4435−4453.
(24) Willard, J. D.; Read, J. S.; Appling, A. P.; Oliver, S. K.; Jia, X.; Kumar, V. Predicting Water Temperature Dynamics of Unmonitored Lakes With Meta-Transfer Learning. *Water Resour. Res.* **2021**, *57*, No. e2021WR029579.
(25) Zhi, W.; Feng, D.; Tsai, W.-P.; Sterle, G.; Harpold, A.; Shen, C.; Li, L. From hydrometeorology to river water quality: Can a deep learning model predict dissolved oxygen at the continental scale? *Environ. Sci. Technol.* **2021**, *55*, 2357−2368.
(26) Liang, Z.; Zou, R.; Chen, X.; Ren, T.; Su, H.; Liu, Y. Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. *J. Hydrol.* **2020**, *581*, No. 124432.
(27) Zhou, Y. Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. *J. Hydrol.* **2020**, *589*, No. 125164.
(28) Song, C.; Yao, L.; Hua, C.; Ni, Q. A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. *J. Hydrol.* **2021**, *603*, No. 126879.
(29) Than, N. H.; Ly, C. D.; Van Tat, P. The performance of classification and forecasting Dong Nai River water quality for sustainable water resources management using neural network techniques. *J. Hydrol.* **2021**, *596*, No. 126099.
(30) Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206−215.
(31) Yang, Y.; Chui, T. F. M. Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 5839−5858.
(32) Fleming, S. W.; Vesselinov, V. V.; Goodbody, A. G. Augmenting geophysical interpretation of data-driven operational water supply forecast modeling for a western US river using a hybrid machine learning approach. *J. Hydrol.* **2021**, *597*, No. 126327.
(33) Kratzert, F.; Herrnegger, M.; Klotz, D.; Hochreiter, S.; Klambauer, G. NeuralHydrology−Interpreting LSTMs in Hydrology. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer, 2019; pp 347−362.
(34) Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; Samek, W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* **2015**, *10*, No. e0130140.
(35) Sundararajan, M.; Taly, A.; Yan, Q. In *Axiomatic Attribution for Deep Networks*, International Conference on Machine Learning, PMLR, 2017; pp 3319−3328.
(36) Erion, G.; Janizek, J. D.; Sturmfels, P.; Lundberg, S. M.; Lee, S. I. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. *Nat. Mach. Intell.* **2021**, *3*, 620−631.
(37) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56−67.
(38) Samek, W.; Montavon, G.; Vedaldi, A.; Hansen, L. K.; Müller, K.-R. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer Nature, 2019; Vol. *11700*.
(39) Huang, Y.; Huang, J.; Ervinia, A.; Duan, S.; Kaushal, S. S. Land use and climate variability amplifies watershed nitrogen exports in coastal China. *Ocean Coastal Manage.* **2021**, *207*, No. 104428.
(40) Chen, N.; Wu, J.; Hong, H. Effect of storm events on riverine nitrogen dynamics in a subtropical watershed, southeastern China. *Sci. Total Environ.* **2012**, *431*, 357−365.
(41) Cao, W.; Hong, H.; Yue, S. Modelling agricultural nitrogen contributions to the Jiulong River estuary and coastal water. *Global Planet. Change* **2005**, *47*, 111−121.
(42) Yu, D.; Yan, W.; Chen, N.; Peng, B.; Hong, H.; Zhuo, G. Modeling increased riverine nitrogen export: source tracking and integrated watershed-coast management. *Mar. Pollut. Bull.* **2015**, *101*, 642−652.
(43) Chen, N.; Hong, H.; Zhang, L.; Cao, W. Nitrogen sources and exports in an agricultural watershed in Southeast China. *Biogeochemistry* **2008**, *87*, 169−179.
(44) Han, F.; Zheng, Y. Joint analysis of input and parametric uncertainties in watershed water quality modeling: A formal Bayesian approach. *Adv. Water Resour.* **2018**, *116*, 77−94.
(45) Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735−1780.
(46) Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks* **1994**, *5*, 157−166.
(47) Graves, A. Long Short-Term Memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer, 2012; pp 37−45.
(48) Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D* **2020**, *404*, No. 132306.
(49) Jiang, S.; Zheng, Y.; Babovic, V.; Tian, Y.; Han, F. A computer vision-based approach to fusing spatiotemporal data for hydrological modeling. *J. Hydrol.* **2018**, *567*, 25−40.
(50) Barzegar, R.; Aalami, M. T.; Adamowski, J. Coupling a hybrid CNN-LSTM deep learning model with a Boundary Corrected Maximal Overlap Discrete Wavelet Transform for multiscale Lake water level forecasting. *J. Hydrol.* **2021**, *598*, No. 126196.
(51) Copernicus Climate Change Service Climate Data Store (CDS). *ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate*, 2017; Vol. 15.
(52) Dee, D. P.; Uppala, S. M.; Simmons, A. J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.; Balsamo, G.; Bauer, dP.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **2011**, *137*, 553−597.
(53) Lundberg, S.; Lee, S. In *A Unified Approach to Interpreting Model Predictions*, Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017; Vol. 30.
(54) Roth, A. E. *The Shapley Value: Essays in Honor of Lloyd S. Shapley*; Cambridge University Press, 1988.
(55) Lu, S.; Andersen, H. E.; Thodsen, H.; Rubæk, G. H.; Trolle, D. Extended SWAT model for dissolved reactive phosphorus transport in tile-drained fields and catchments. *Agric. Water Manage.* **2016**, *175*, 78−90.
(56) Kim, H.; Sandersen, P. B.; Jakobsen, R.; Kallesøe, A. J.; Claes, N.; Blicher-Mathiesen, G.; Foged, N.; Aamand, J.; Hansen, B. A 3D hydrogeochemistry model of nitrate transport and fate in a glacial sediment catchment: A first step toward a numerical model. *Sci. Total Environ.* **2021**, *776*, No. 146041.
(57) Liu, W.; Park, S.; Bailey, R. T.; Molina-Navarro, E.; Andersen, H. E.; Thodsen, H.; Nielsen, A.; Jeppesen, E.; Jensen, J. S.; Jensen, J. B.;

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Trolle, D. Quantifying the streamflow response to groundwater abstractions for irrigation or drinking water at catchment scale using SWAT and SWAT−MODFLOW. *Environ. Sci. Eur.* **2020**, *32*, 1−25.
(58) Zheng, Y.; Han, F. Markov Chain Monte Carlo (MCMC) uncertainty analysis for watershed water quality modeling and management. *Stochastic Environ. Res. Risk Assess.* **2016**, *30*, 293−308.
(59) Schilling, K.; Wolter, C.; Christiansen, D.; Schnoebelen, D.; Jha, M. *Water quality improvement plan for Raccoon River, Iowa*, TMDL Report; Watershed Improvement Section, Iowa Department of Natural Resources: Iowa, 2008.
(60) Sidle, W.; Roose, D.; Shanklin, D. *Isotopic Evidence for Naturally Occurring Sulfate Pollution of Ponds in the Kankakee River Basin, Illinois-Indiana*; 0047-2425; Wiley Online Library, 2000.
(61) Torrence, C.; Compo, G. P. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 61−78.
(62) Grinsted, A.; Moore, J. C.; Jevrejeva, S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes Geophys.* **2004**, *11*, 561−566.
(63) Godsey, S. E.; Kirchner, J. W.; Clow, D. W. Concentration−discharge relationships reflect chemostatic characteristics of US catchments. *Hydrol. Process.* **2009**, *23*, 1844−1864.
(64) Kim, H.; Høyer, A.-S.; Jakobsen, R.; Thorling, L.; Aamand, J.; Maurya, P. K.; Christiansen, A. V.; Hansen, B. 3D characterization of the subsurface redox architecture in complex geological settings. *Sci. Total Environ.* **2019**, *693*, No. 133583.
(65) Moatar, F.; Abbott, B. W.; Minaudo, C.; Curie, F.; Pinay, G. Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions. *Water Resour. Res.* **2017**, *53*, 1270−1287.
(66) Musolff, A.; Fleckenstein, J.; Rao, P.; Jawitz, J. Emergent archetype patterns of coupled hydrologic and biogeochemical responses in catchments. *Geophys. Res. Lett.* **2017**, *44*, 4143−4151.
(67) Rydberg, T.; Haden, A. C. Emergy evaluations of Denmark and Danish agriculture: Assessing the influence of changing resource availability on the organization of agriculture and society. *Agric., Ecosyst. Environ.* **2006**, *117*, 145−158.
(68) Eagleson, P. S. The emergence of global-scale hydrology. *Water Resour. Res.* **1986**, *22*, 6S−14S.
(69) Vörösmarty, C. J.; Moore, B., III; Grace, A. L.; Gildea, M. P.; Melillo, J. M.; Peterson, B. J.; Rastetter, E. B.; Steudler, P. A. Continental scale models of water balance and fluvial transport: An application to South America. *Global Biogeochem. Cycles* **1989**, *3*, 241−265.
(70) Alcamo, J.; Döll, P.; Kaspar, F.; Siebert, S. Global Change and Global Scenarios of Water Use and Availability: An Application of WaterGAP 1.0. In *Center for Environmental Systems Research*; University of Kassel, Kassel: Germany, 1997.
(71) Van Beek, L.; Bierkens, M. *The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification*; Utrecht University: Utrecht, The Netherlands, 2009; pp 25−26.
(72) Beck, H. E.; van Dijk, A. I.; De Roo, A.; Miralles, D. G.; McVicar, T. R.; Schellekens, J.; Bruijnzeel, L. A. Global-scale regionalization of hydrologic model parameters. *Water Resour. Res.* **2016**, *52*, 3599−3622.
(73) Tilman, D.; Balzer, C.; Hill, J.; Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 20260−20264.
(74) Zhang, W.; Zhang, X. A forecast analysis on fertilizers consumption worldwide. *Environ. Monit. Assess.* **2007**, *133*, 427−434.

https://doi.org/10.1021/acs.est.2c02232
Environ. Sci. Technol. XXXX, XXX, XXX−XXX