Deep Learning to Near-Surface Humidity Retrieval from Multi-sensor Remote Sensing Data over the China Seas

Rongwang Zhang[1,2,3], Weihao Guo[1,4], Xiaofeng Li[5,6], Xin Wang[1,2,3*]

[1]State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China; rwzhang@scsio.ac.cn (R.Z.); guoweihao@scsio.ac.cn (W.G.)

[2]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

[3]Innovation Academy of South China Sea Ecology and Environmental Engineering, Chinese Academy of Sciences, Guangzhou 511458, China

[4]University of Chinese Academy of Sciences, Beijing 100049, China

[5]Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; lixf@qdio.ac.cn (X.L.)

[6]Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China

*Correspondence: wangxin@scsio.ac.cn

Abstract

Near-surface humidity (Q_a) is a key parameter that modulates oceanic evaporation and influences the global water cycle. Remote sensing observations act as feasible sources for long-term and large-scale Q_a monitoring. However, existing satellite Q_a retrieval models are subject to apparent uncertainties due to model errors and insufficient training data. Based on in situ observations collected over the China Seas over the last two decades, a deep learning approach named Ensemble Mean of Target deep neural networks (EMTnet) was proposed to improve the satellite Q_a retrieval over the China Seas for the first time. The EMTnet model outperforms five representative existing models by nearly eliminating the mean bias and significantly reducing the root-mean-square error in satellite Q_a retrieval. According to its target deep neural networks selection process, the EMTnet model can obtain more objective learning results when the observational data are divergent. The EMTnet model was subsequently applied to produce a 30-year monthly gridded Q_a data over the China Seas. It indicates that the climbing rate of Q_a over the China Seas under the background of global warming are probably underestimated by current products.

Key words: near-surface humidity, remote sensing, deep learning, China Seas

1. Introduction

As the primary source of global evaporation and precipitation, the ocean plays an important role in the transportation and redistribution of water resources on Earth [1–3]. On this basis, the near-surface humidity (Q_a) over global oceans is crucial as it modulates oceanic evaporation and influences the global water cycle [4–6]. Nevertheless, there are non-negligible uncertainties in Q_a estimates in satellite-derived products [7–8] and reanalysis products [9–10]. The imperfection of Q_a data quality has been reported as one of the leading error sources of uncertainties in freshwater exchanges across the air-sea interface and in global water budgets [11–14]. Even in coupled general circulation models, the performance of Q_a is highly related to simulations of oceanic evaporation [15]. Accurate estimates of Q_a are thus necessary for studies on the global water cycle, air-sea interactions, and climate change [16].

The measurements of Q_a can be generally divided into two approaches: in situ observations and remote sensing observations. The former are direct measurements of Q_a and have relatively high credibility, but these observations are subject to poor continuity in time and space. The latter have the advantage of long-term and large-scale Q_a monitoring. Still, remote sensing is an indirect approach that requires a relevant retrieval model to convert satellite measurements into Q_a. With the development of space-borne technology and microwave radiometers, the last several decades have experienced the prosperity of investigations in model development for satellite Q_a retrieval. Considering that a large portion of the total column precipitable water (TPW) is confined in the atmospheric surface layer,

pioneering work by [17] and [18] (hereafter L86) linked the TPW to Q_a in light of the Q_a-W relation. It was reported that the Q_a-W relationship has excellent performance with training data on a monthly scale [18] and can also work well with synoptic-scale data [19]. Considering the decoupling of the atmospheric boundary layer from the higher atmosphere, [20] proposed replacing the TPW data in the Q_a-W relation with the precipitable water constrained in the lowest 500 m, which can be derived from brightness temperature (TB) measurements. To reduce the propagation of uncertainties within input data, [21] established a direct linear regression between Q_a and TB. Under the scheme of the Q_a-W relation, [22] developed an empirical orthogonal function method for satellite Q_a retrieval. A neutral network combining TPW and sea surface temperature (SST) was first developed to estimate Q_a by [23]. Subsequently, estimates of Q_a with multichannel TBs as input data by multivariate linear regression [24–27] or nonlinear regression [16,28,29](Yu & Jin, 2018; Tomita et al., 2018; Gao et al., 2019) prevailed in the last two decades.

The models above were primarily designed for global oceans. However, [28] reported that satellite Q_a retrieval differed in regions of the tropics and high latitudes and a high-latitude enhancement was considered to their model. It indicates that attention should be given to different regional features of satellite Q_a retrieval. The China Seas, consisting of the Bohai Sea, Yellow Sea (YS), East China Sea (ECS), and South China Sea (SCS), is the largest marginal group of seas in the northwestern Pacific and is strongly influenced by complex continental environments. Previous investigations have pointed out that the Q_a data

in this region suffer from significant uncertainties and are the leading error source of air-sea heat fluxes [30–31].

In recent years, machine learning, especially deep learning, has been widely used to provide new insights into traditional and/or emerging research in earth science [32–36]. With the accumulation of high-quality in situ observations of Q_a over the China Seas in recent two decades, the main objective of this study is to develop a deep learning-based model to improve the satellite Q_a retrieval over the China Seas. The data and methods are introduced in Section 2. The main results are presented in Section 3. Section 4 discusses the interpretability of the deep learning-based model. Finally, Section 5 draws the main conclusion of this study.

2. Data and methods

2.1. In situ observations

In situ observations information in this study are listed in Table 1. There are 20 observational stations in the coastal and open ocean (Figure 1a), including 18 buoys, one offshore platform, and one flux tower on an island. Compared to ship observations, these fixed-point observations usually have more stable data quality performance. The data collected in coastal areas are valuable touchstones to validate the performance of remote sensing observations, including Q_a, surface wind, and sea surface temperature. The data span from 1998 to 2018, and the sampling intervals vary from 1 minute to 30 minutes. All the raw data were processed with quality control procedures as suggested by [31,37,38]. For all stations, the Q_a and surface wind data were adjusted to standard heights of 2 m and 10 m, respectively, according to the COARE 3.0 algorithm [39].

The basic characteristics of in situ observations of Q_a are examined to check their representativeness, considering many missing data and the uneven sampling in time and space. As shown in Figs. 1b to 1d, Q_a varies from 5.5 to 24.0, 1.5 to 24.5, and 0.8 to 24.8 g/kg in the SCS, ECS, and YS. Q_a's mean values plus/minus one standard deviation (STD) are 16.6±4.1, 10.1±5.5, and 8.5±5.7 g/kg in the SCS, ECS, and YS. For the probability density distributions (PDDs) of Q_a, a left-skewed distribution in the SCS and right-skewed distributions in the ECS and YS can be observed. These lower limits, mean values, and PDDs of Q_a in the three seas are coincide well with the latitudes they locate in, as Q_a ususally decreases from low to high latitudes. With the data in the three seas considered as a whole, Q_a presents a fairly even distribution in the range of 2~22 g/kg, which varies basically around a steady density across each bin (Figure 1e). The highly uniform PDD of Q_a shows an acceptable representativeness of in situ observations collected here. Therefore, it is expected that the analyses based on those data could be relatively objective and with high significance.

2.2. Remote sensing data

Remote sensing observations of TPW, wind speed (U), cloud liquid water (CLW), and SST from various satellite microwave radiometers are utilized in this study. The sensors include the Special Sensor Microwave Imager (SSM/I), the Special Sensor Microwave Imager Sounder (SSMIS), the Advanced Microwave Scanning Radiometer series (AMSR-E and AMSR-2), the WindSat Polarimetric Radiometer, the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), and the Global Precipitation Measurement Microwave Imager (GMI). For SSM/I and SSMIS, the instruments are referred to by satellite number

starting with F08. Here F15 from SSMI/I and F16 to F18 from SSMIS are employed because

of their relatively long time coverage. Detailed descriptions for each sensor can be found at

the Remote Sensing Systems (RSS; www.remss.com).Due to the satellite swath width and

orbit seam, banded gaps exist in the ascending and descending daily maps of

satellited-derived variables. To facilitate the matching of satellite data and in situ data, the

products incorporating both the ascending and descending data by using 3-day running

average are utilized, which can achieve more homogeneous spatial distributions for these

variables. Note that there exist differences in data from various sensors (Figure 2) and

uncertainties in results may be caused by single data source. Consequently, multi-sensor

inputs from SSM/I F15, SSMIS F16, SSMIS F17, SSMIS F18, AMSR-E, AMSR-2, WindSat,

TMI, and GMI are utilized in the following study.

TB data from six channels are used for the models listed in Table 2 that directly retrieve

Q_a from multichannel TBs. These channels are 19H, 19V, 22V, and 37V GHz from SSMIS

F17 and 52V GHz from Advanced Microwave Sounding Unit-A (AMSU-A), where H and V

denote the horizontal and vertical polarizations, respectively. As underscored by RSS, TBs

from the SSMIS are produced using uniform processing techniques. They are intercalibrated

by considering the differences in sensor frequencies, channel resolutions, instrument

operation, and other radiometer characteristics [40]. The AMSU-A is a multichannel

microwave radiometer that performs atmospheric sounding of temperature and moisture by

passively recording atmospheric microwave radiation at multiple wavelengths. Detailed

descriptions of how TBs from the AMSU-A are processed can be found in [41–42]. The

collocating strategy between satellite data and in situ observations for each station is as follows:

(i) For variables TPW, U, and CLW, which are already in daily values, first, average the high-frequency in situ observations to daily values with the local standard time adjusted to Coordinated Universal Time and then apply a 3-day running average. Second, locate the $0.25°\times0.25°$ box in the satellite data where the corresponding observational station lies. Subsequently, the mean value of the four corners of that box is used as a proxy for satellite data. Attempts such as extending the search area to $1°\times1°$ box and/or applying an inverse-distance-weighted average present similar results.

(ii) For TBs with ascending and descending measurements per day, temporal and spatial windows of 90 minutes and 50 km following Yu and Jin (2018) are used. If multiple points of satellite data meet the criterion, the average of those points is taken. If no satellite data match the in situ observations, a missing value is set.

2.3. Reanalysis data

Two reanalysis products are employed to make comparisons with the Q_a data derived from the model proposed in this study. They are the European Centre for Medium Range Weather Forecast (ECWMF) fifth generation (ERA5) reanalysis product [43] and the National Centers for Environmental Prediction/Department of Energy Global Reanalysis 2 (NCEP2) product [44]. Both products are the latest versions of their corresponding series and improvements have been made in their data assimilations and model physics. Monthly Q_a data

from 1990 to 2019 are extracted from the two reanalyses and are interpolated onto 1°×1° grid

maps.

2.4. Existing satellite Qa retrieval models

Five representative satellite Q_a retrieval models are employed to intercompare with the

deep learning-based model proposed in this study. Table 2 summarizes the basic information

of these five models. The L86 model uses a fifth-order polynomial regression approach to

estimate Q_a with TPW data. The model reported in [23] (hereafter J99) uses a nonlinear

neural network approach to estimate Q_a with TPW and SST. The models reported in [24]

(hereafter B03) and [25] (hereafter J06) use multivariate linear regression to estimate Q_a with

multichannel TBs. Recently, the model reported in [28] (hereafter Y18) uses multivariate

nonlinear regression to estimate Q_a with multichannel TBs and considers enhancement in high

latitudes.

2.4. Ensemble Mean of Target deep neural networks development

A model named Ensemble Mean of Target deep neural networks (EMTnet) was

proposed to improve the satellite Q_a retrieval over the China Seas (Figure 3). The EMTnet

model is generated from a large number of deep neural networks (DNNs). Each DNN is based

on the error backpropagation (BP) algorithm [45] and consists of an input layer, three hidden

layers, and an output layer. The BP algorithm takes advantage of the gradient descent and

error backpropagation methods to adjust the connection weights of corresponding neurons to

achieve its nonlinear learning ability. The EMTnet model works as the following steps:

(i) Four satellite-derived variables, TPW, CLW, U, and SST, are put in the input layer of each DNN. Note that different combinations of input variables can lead to different learning abilities of the EMTnet model, which will be discussed in Section 4. Attempts using pure multi-channel TBs or the mix of level-3 variables and multichannel TBs show similar or even slightly worse results.

(ii) Normalize these four input variables to the range between 0 and 1 according to their maximum and minimum values. It is known that the DNNs are quite sensitive to the magnitude difference of various input variables. Therefore, normalizing the input variables in advance can lead to better computational efficiency and results.

(iii) Determine the specific configuration of each DNN. Critical parameters are eventually determined through a number of testings to make the DNNs suitable for the learning task here. For instance, three hidden layers, each with ten neurons, are used. The number of iterations is set to 5000, and the learning rate is set to 0.005. In addition, the activation function is essential in forming nonlinear learning abilities for the DNNs. Three widely used activation functions, sigmoid, rectified linear units (ReLU), and hyperbolic tangent function (tanh), have been tested and compared. These activation functions have some defects, for example, the neuronal death for the ReLU activation function and a vanishing gradient for the sigmoid activation function. Activation function tanh is free of neuronal death, and the vanishing gradient problem has been alleviated to some extent. In addition, tanh has a faster convergence speed and a lower number of iterations. Preliminary tests show that the sigmoid and tanh activation functions perform better than the ReLU activation

function in this study. Further inspections find that the accuracy of the result using the tanh

activation function is 2~4% higher than that using the sigmoid activation function. Therefore,

the tanh activation function is employed in this study.

(iv) 75% of Q_a observations are randomly sampled as training data, while the rest 25%

are used as independent testing data. As for traditional DNN training and testing, this

operation is usually conducted once. However, it is found that samplings of different training

and testing data can result in different uncertainty levels for the DNNs. To objectively

measure the uncertainties, the EMTnet model trains n different DNNs with randomly sampled

n sets of training and testing data to produce an ensemble of DNNs. In this study, the n is set

to 1000. For each DNN, the testing data are used to compute the mean bias and

root-mean-square error (RMSE). The sum of absolute values of mean bias and RMSE, that is,

the absolute error, is taken as the uncertainty of each DNN.

According to the PDD of uncertainties constructed by the DNN ensemble, the top 10% DNNs

with uncertainties falling into the highest density intervals are selected as target DNNs. The

ensemble mean of target DNNs is then used to produce the EMTnet model outputs.

3. Results

3.1. EMTnet model validation

The Q_a predictions from the EMTnet model and five existing models are intercompared

with respect to Q_a observations. Three representative stations from the three China seas are

selected according to their data quality, continuity and integrity to facilitate the

intercomparison. They are the Xisha Tower station in the SCS, the DH11 station in the ECS,

and the HH09 station in the YS. A whole year of data from 2016 is used for each station to

reduce the possible seasonal dependence of the results.

Figure 4 shows the scatter diagram between the Q_a predictions and observations and the

corresponding correlation coefficient (CC), mean bias, and RMSE for each model in each sea.

The CCs all exceed the 99% confidence level, varying from 0.59-0.91, 0.89 to 0.98, and 0.92

to 0.98 in the SCS, ECS, and YS, respectively. Among them, the EMTnet model has the

highest CCs at each station. The mean biases and RMSEs present a large spread in different

models and stations. In the SCS, the EMTnet model slightly overestimates Q_a by 0.06 g/kg,

while the rest models underestimate Q_a from 1.07 (L86 model) to 7.33 (Y18 model) g/kg. In

the ECS, except for the EMTnet model and the L86 model which overestimate Q_a by 0.13 and

0.78 g/kg, all the models underestimate Q_a by 0.41 (B03 model) to 4.12 (J06 model) g/kg. In

the YS, except for the J99 and B03 models which underestimate Q_a by 3.22 and 1.37 g/kg, all

the models overestimate Q_a by 0.06 (EMTnet model) to 4.32 (J06 model) g/kg. The RMSEs

of these model in the SCS, ECS, and YS are varying from 1.10 (EMTnet model) to 2.72 (Y18

model) g/kg, 1.17 (EMTnet model) to 3.36 (Y18 model) g/kg, and 1.22 (EMTnet model) to

3.08 (J99 model) g/kg.

The absolute values of the mean bias and RMSEs from the three seas are averaged for

each model to compare their overall uncertainty level. The mean biases plus/minus RMSEs

are 0.08±1.16, 0.72±1.77, 4.35±2.56, 1.64±2.09, 5.26±2.49, and 3.39±2.61 for the EMTnet,

L86, J99, B03, J06, and Y18 models. Quantitatively, the EMTnet model has the lowest mean

bias and RMSE on average.

3.2. EMTnet model application

All the Q_a observations collected here are subsequently used to fully train the EMTnet model. Q_a predictions of the L86 model, which has the best performance among these five existing models, are used as a reference here. It is noted that both the EMTnet model and the L86 model take TPW as an input variable, which confirms the good relationship between TPW and Q_a over the China Seas. Figure 5a compares the Q_a predictions of the EMTnet model and the L86 model in the form of the Q_a-W relation. The dots determined by Q_a and TPW data cluster around the classical curve of the L86 model. The data density distribution shows that the majority of the data coincide well with the L86 model. Compared to the medians of Q_a observations, however, biases of the L86 model occur primarily under moderate Q_a values. For example, in the range of 10~20 g/kg, the L86 model overestimates Q_a from 0.44 to 1.98 g/kg, while the biases of the EMTnet model are almost negligible.

Figures 5b and 5c depict the PDDs of the mean biases and RMSEs of the EMTnet model and the L86 model according to the 1000 samples of testing data used in the EMTnet model. The mean biases of the L86 model is concentrated from 0.60 to 0.80 (87% data). The mean biases of the EMTnet model have two peaks, which are around 0.10 to 0.30 g/kg and -0.10 to -0.30 g/kg. On average, the mean biases of the L86 model and the EMTnet model are 0.72±0.06 and -0.02±0.19 g/kg. The RMSEs of the L86 and EMTnet models are concentrated from 2.45 to 2.65 (94% data) and 1.55 to 1.70 (95% data) g/kg, which are on average 2.56±0.05 g/kg and 1.64±0.04 g/kg. Thus, the EMTnet model reduces the mean bias and RMSE of the L86 model by approximately 0.70 and 0.90 g/kg, respectively. The mean bias

for the EMTnet model in satellite Q_a retrieval is almost zero, reducing the RMSE of the L86 model by 36%.

A monthly gridded Q_a data over the China Seas was produced with satellite multi-sensor inputs by applying the fully trained EMTnet model. Both the input and output data are on 0.25°×25° gridded maps and span from 1990 to 2019. Figures 6a to 6d show the climatologies of Q_a from two satellite Q_a retrieval models (EMTnet and L86) and two reanalyses (ERA5 and NCEP2). Except for some differences in detail, apparent gradients from south to north in the mean state and seasonal variation of Q_a can be observed in all the four data sources, which is higher in the south and lower in the north. Here the intensity of seasonal variations is defined by the standard deviation from January to December.

The atmosphere's capacity to hold water vapor will increase in a warming climate according to the Clausius-Clapeyron relation [46]. The long-term trend of Q_a over the China Seas is depicted in Figures 6e to 6g. With the global warming in recent decades, the four data sources show consistent upward trends of Q_a (Figure 6e). However, the L86 model, ERA5, and NCEP2 have relatively lower climbing rates of Q_a, which are 0.08, 0.11, and 0.10 g/kg per decade, compared to the results of the EMTnet model (0.23 g/kg per decade). It is found that the long-term trends of Q_a are probably underestimated in both the southern (the SCS) and northern (the ECS and YS) sections , especially for the latter. In the southern, the long-term trends of Q_a are 0.22, 0.20 0.18, and 0.13 g/kg per decade in the EMTnet model, L86 model, ERA5, and NCEP2 (Figure 6f). Except for NCEP2, all the data show quite similar climbing rates of Q_a. In contrast, the long-term trends of Q_a show larger spread in the northern,

which are 0.21, -0.04, 0.04, and 0.07 g/kg per decade in the EMTnet model, L86 model,

ERA5, and NCEP2 (Figure 6g). Therefore, the possible underestimation of upward trends of

Q_a in the L86 model, ERA5, and NCEP2 can be mainly attributed to their too weak trends of

Q_a variations in the northern section.

4. Discussions

The interpretability of deep learning is of great significance for its development and

application. The EMTnet model and the L86 model, which take TPW as an input variable are

the top two best performing models investigated here. The possible reasons why the EMTnet

model can further improve the satellite Q_a retrieval compared to the L86 model is discussed.

Taking the result of the L86 model as a reference, eight sensitivity experiments (Exp1 to

Exp8) are designed to examine whether the improvement of the EMTnet model is due to the

model itself or the additional training data such as CLW, U, and SST compared to the L86

model. All the sensitivity experiments employ TPW as a fixed variable and adopt the eight

combinations of CLW, U, and SST to construct their training data. Note that Exp1, including

the full CLW, U, and SST information, is the result shown in Figure 5. The statistical results

for Exp1 to Exp8 are shown in Table 3. As revealed in Table 3, all the experiments show

improvements in satellite Q_a retrieval compared to the L86 model. They reduce the mean

biases and RMSEs to varying extents. If only TPW data are used as training data as in the L86

model, the absolute error is reduced by 23% (Exp8). Taking into account CLW, U, and SST,

the absolute errors are reduced by 35% (Exp5), 24% (Exp6), and 42% (Exp7). The three

pairwise combinations of CLW, U, and SST are considered in Exp2 to Exp4. The reductions in absolute error in Exp2 to Exp4 are 42%, 47% and 48%.

The results of Exp2 to Exp7 suggests that factors CLW, U, and SST are helpful to improve the deep learning for satellite Q_a retrieval. If these three factors are superimposed together, a most significant improvement of 49% (Exp1) can be archived. The abilities of CLW, U, and SST in improving satellite Q_a retrieval are probably due to their roles in reflecting the environmental information. In the following, examples of the Q_a-W relation under different CLW, U, and SST conditions are shown in Figures 7 to 9, respectively. The CLW is a measure of the total liquid water contained in a cloud in a vertical column of the atmosphere. As a component of TPW, the content of CLW will undoubtedly have an impact on the determination of the Q_a-W relation. However, none of the existing models for satellite Q_a retrieval incorporate cloud information. Figure 7 shows the Q_a-W relation under two conditions, one under CLW less than 50 μm (46% data) and the other greater than 100 μm (54% data). Note that the criterion of 50 μm here is only determined by the PDD of CLW, which ensures that the data balance in both cases. Under a relatively low CLW (Figure 7a), the reference curve of the L86 model passes through most of the medians of Q_a observations, presenting high consistency with observations. Under a relatively high CLW (Figure 7b), however, the reference curve of the L86 model is nearly above all the medians of the Q_a observations. This result indicates that a relatively high CLW condition can interfere with the determination of the Q_a-W relation and lead to evident overestimations in satellite Q_a retrieval if no CLW information is considered.

The surface wind plays a vital role in reflecting the weather conditions near the sea surface and influencing the Q_a variations. Consequently, the surface wind is expected to be a potential factor that may improve the skill of satellite Q_a retrieval, which has not been considered in existing models. Figure 8 shows the Q_a-W relation with U less than 10 m/s (90% data) and greater than 10 m/s (10% data). It can be observed that the reference curve of the L86 model fits the Q_a observations well under lower to moderate U (Figure 8a). In contrast, the L86 model overestimates Q_a in almost all ranges of Q_a under relatively high U (Figure 8b). The different performances of the L86 model here imply that the Q_a-W relation is sensitive to surface wind conditions. One possible reason is that the water vapor distributions in the vertical column of the atmosphere are relatively stable under relatively weak U, which is conducive to the estimation of Q_a from TPW. As a portion of water vapor can be carried away by horizontal advection under relatively high U, the observed Q_a will be smaller than the model-predicted Q_a.

SST is an important variable that reflects information on the marine environment and underlying atmospheric surface. For example, under a relatively warm SST (Figure 9b), the predictions of the L86 model are more consistent with the observations. In contrast, Q_a is overestimated in the range of Q_a from 10 to 20 g/kg under rather cold SSTs (Figure 9a). A colder SST means weaker sea surface evaporation, which might lead to less actual moisture than predicted values. Therefore, it is suggested that attention should be given to the Q_a-W relation under different SST conditions.

5. Summary and conclusions

Previous satellite Q_a retrieval models suffer significant uncertainties due to factors such as model errors, scarce in situ observations, environmental interference, and so on. In this study, a deep learning approach, the EMTnet model, is proposed to improve the satellite Q_a retrieval over the China Seas. The EMTnet model is based on multiple DNNs and the ensemble mean of target DNNs is used to produce output predictions, which can obtain more objective learning results when the observational data are quite divergent. The Q_a predictions from the EMTnet model outperforms five existing models by nearly eliminating the mean bias and significantly reducing the RMSE. Compared to the L86 model which has the best performance among five existing models, the outperformance of the EMTnet model can be attributed to two aspects. Firstly, if only TPW data are used as training data as in the L86 model, the EMTnet model reduces the absolute error by 23% (Table 3). This level of improvement can be attributed to the EMTnet model itself. Secondly, if CLW, U, and SST are added, the EMTnet model reduces the absolute error by 49% (Table 3). The approximately doubled increase of absolute error reduction benefits from the good interpretability of CLW, U, and SST on the determination of the Q_a-W relation.

The fully trained EMTnet model has been applied to learn from remote sensing data to produce a 30-year monthly gridded Q_a data over the China Seas. It is found that current products perform well in depicting the mean state and seasonal variations of Q_a. However, they show much weaker upward trends of Q_a in the context of global warming, which are less than half of the EMTnet model result. As a locally well-trained and well-validated model, the different perspectives on the long-term variations of Q_a suggested by the EMTnet model may

help to provide new understandings for humidity-related multi-disciplinary researches over the China Seas. In addtion, the EMTnet model is capable of merging Q_a observations from other regions as training data, which is to be applied to more oceans globally.

https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 and

https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html, respectively. The in situ

observational data can be acquired via the link https://pan.cstcloud.cn/s/4hLRaDMoRfM.

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

References

1. Baumgartner, A.; Reichel, E. The World Water Balance. Elsevier, New York, **1975**; pp. 179.

2. Chahine, M.T. The hydrological cycle and its influence on climate. *Nature* **1992**, *359*, 373–380. https://doi.org/10.1038/359373a0.

3. Trenberth, K.E.; Smith, L.; Qian, T.; Dai, A.; Fasullo, J. Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydrometeorol.* **2007**, *8*, 758–769. https://doi.org/10.1175/JHM600.1.

4. Yu, L. Global variations in oceanic evaporation (1958–2005): The role of the changing wind Speed. *J. Clim.* **2007**, *20*, 5376–5390. https://doi.org/10.1175/2007JCLI1714.1.

5. Lorenz, D.J.; DeWeaver, E.T.; Vimont, D.J. Evaporation change and global warming: the role of net radiation and relative humidity. *J. Geophys. Res.* **2010**, *115*, D20118. http://dx.doi.org/10.1029/2010JD013949.

6. Jin, X.; Yu, L.; Jackson, D.L.; Wick, G.A. An improved near-surface specific humidity and air temperature climatology for the SSM/I satellite period. *J. Atmos. Ocean. Technol.* **2015**, *32*, 412–433. https://doi.org/10.1175/JTECH-D-14-00080.1.

7. Tomita, H.; Kubota, M.; Cronin, M.F.; Iwasaki, S.; Konda, M.; Ichikawa, H. An assessment of surface heat fluxes from J-OFURO2 at the KEO and JKEO sites. *J. Geophys. Res.* **2010**, *115*, C03018. https://doi.org/10.1029/2009JC005545.

8. Kinzel, J.; Fennig, K.; Schröder, M.; Andersson, A.; Bumke, K.; Hollmann, R. Decomposition of random errors inherent to HOAPS-3.2 near-surface humidity estimates

using multiple triple collocation analysis. *J. Atmos. Ocean. Tech.* **2016**, *33*, 1455–1471.

https://doi.org/10.1175/JTECH-D-15-0122.1.

9.   Brunke, M.A.; Wang, Z.; Zeng, X.; Bosilovich, M.; Shie, C.-L. An assessment of the

uncertainties in ocean surface turbulent fluxes in 11 reanalysis, satellite-derived, and

combined global datasets. *J. Clim.* **2011**, *24*, 5469–5493.

https://doi.org/10.1175/2011JCLI4223.1.

10. Kent, E.C.; Berry, D.I.; Prytherch, J.; Roberts, J.B. A comparison of global marine

surface-specific humidity datasets from in situ observations and atmospheric reanalysis.

Int. J. Climatol. **2014**, *34*, 355–376. https://doi.org/10.1002/joc.3691.

11. Robertson, F.R.; Bosilovich, M.G.; Roberts, J.B.; Reichle, R.H.; Adler, R.; Ricciardulli,

L.; Berg, W.; Huffman, G.J. Consistency of estimated global water cycle variations over

the satellite era. *J. Clim.* **2014**, *27*, 6135–6154. https://doi.org/10.1175/JCLI-D-13-

00384.1.

12. Rodell, M.; Beaudoing, H.K.; L'Ecuyer, T.S.; Olson, W.S.; Famiglietti, J.S.; Houser, P.R.;

Adler, R.; Bosilovich, M.G.; Clayson, C.A.; Chambers, D.; Clark, E.; Fetzer, E.J.; Gao,

X.; Gu, G.; Hilburn, K.; Huffman, G.J.; Lettenmaier, D.P.; Liu, W.T.; Robertson, F.R.;

Schlosser, C.A.; Sheffield, J.; Wood, E.F. The observed state of the water cycle in the

early twenty-first century. *J. Clim.* **2015**, *28*, 8289–8318. https://doi.org/10.1175/JCLI-D-

14-00555.1.

13. Liman, J.; Schröder, M.; Fenning, K.; Andersson, A.; Hollmann, R. Uncertainty characterization of HOAPS 3.3 latent heat-flux-related parameters. *Atmos. Meas. Tech.* **2018**, *11*, 1793–1815. https://doi.org/10.5194/amt-11-1793-2018.

14. Roberts, J.B.; Clayson, C.A.; Robertson, F.R. Improving near-surface retrievals of surface humidity over the global open oceans from passive microwave observations. *Earth Space Sci.* **2019**, *6*, 1220–1233. https://doi.org/10.1029/2018EA000436.

15. Zhang, R.; Wang, X.; Wang, C. On the simulations of global oceanic latent heat flux in the CMIP5 multimodel ensemble. *J. Clim.* **2018**, *31*, 7111–7128. https://doi.org/10.1175/JCLI-D-17-0713.1.

16. Tomita, H.; Hihara, T.; Kubota, M. Improved satellite estimation of near-surface humidity using vertical water vapor profile information. *Geophys. Res Lett.* **2018**, *45*, 899–906. https://doi.org/10.1002/2017GL076384.

17. Liu, W.T.; Niiler, P.P. Determination of monthly mean humidity in the atmospheric surface layer over oceans from satellite data. *J. Phys. Oceanogr.* **1984**, *14*, 1451–1457. https://doi.org/10.1175/1520-0485(1984)014<1451:DOMMHI>2.0.CO;2.

18. Liu, W.T. Statistical relation between monthly precipitable water and surface-level humidity over global oceans. *Mon. Weather Rev.* **1986**, *114*, 1591–1602. https://doi.org/10.1175/1520-0493(1986)114<1591:SRBMMP>2.0.CO;2.

19. Hsu, S.A.; Blanchard, B.W. The relationship between total precipitable water and surface-level humidity over the sea surface: a further evaluation. *J. Geophys. Res.* **1989**, *94*, 14,539–14,545. https://doi.org/10.1029/JC094iC10p14539.

20. Schulz, J.; Schluessel, P.; Grassl, H. Water vapour in the atmospheric boundary layer over oceans from SSM/I measurements. *Int. J. Remote Sens* **1993**, *14*, 2773–2789. https://doi.org/10.1080/01431169308904308.

21. Schlüssel, P.; Schanz, L.; Englisch, G. Retrieval of latent heat flux and longwave irradiance at the sea surface from SSM/I and AVHRR measurements. *Adv. Space Res.* **1995**, *16*, 107–116. https://doi.org/10.1016/0273-1177(95)00389-V.

22. Chou, S.-H.; Atlas, R.M.; Shie, C.-L.; Ardizzone, J. Estimates of surface humidity and latent heat fluxes over oceans from SSM/I Data. *Mon. Weather Rev.* **1995**, *123*, 2405–2425. https://doi.org/10.1175/1520-0493(1995)123<2405:EOSHAL>2.0.CO;2.

23. Jones, C.; Peterson, P.; Gautier, C. A new method for deriving ocean surface specific humidity and air temperature: an artificial neural network approach. *J. Appl. Meteorol.* **1999**, *38*, 1229–1245. https://doi.org/10.1175/1520-0450(1999)038<1229:ANMFDO>2.0.CO;2.

24. Bentamy, A.; Katsaros, K.B.; Mestas-Nunoz, A.M.; Drennan, W.M.; Forde, E.B.; Roquet, H. Satellite estimates of wind speed and latent heat flux over the global oceans. *J. Clim.* **2003**, *16*, 637–656. https://doi.org/10.1175/1520-0442(2003)016<0637:SEOWSA>2.0.CO;2.

25. Jackson, D.L.; Wick, G.A.; Bates, J.J. Near-surface retrieval of air temperature and specific humidity using multi-sensor microwave satellite observations. *J. Geophys. Res.* **2006**, *111*, D10306. http://dx.doi.org/10.1029/2005JD006431.

26. Kubota, M.; Hihara, T. Retrieval of surface air spesific humidity over the ocean using AMSR-E measurements. *Sensors* **2008**, *8*, 8016–8026. https://doi.org/10.3390/s8128016.

27. Jackson, D.L.; Wick, G.A.; Robertson, F.R. Improved multisensor approach to satellite-retrieved near-surface specific humidity observations. *J. Geophys. Res.* **2009**, *114*, D16303. http://dx.doi.org/10.1029/2008JD011341.

28. Yu, L.; Jin, X. A regime-dependent retrieval algorithm for near-surface air temperature and specific humidity from multi-microwave sensors. *Remote Sens.Environ.* **2018**, *215*, 199–216. https://doi.org/10.1016/j.rse.2018.06.001.

29. Gao, Q.; Wang, S.; Yang, X. Estimation of surface air specific humidity and air–sea latent heat flux using FY-3C microwave observations. *Remote Sens.* **2019**, *11*, 466. https://doi.org/10.3390/rs11040466.

30. Wang, D.; Zeng, L.; Li., X.; Shi, P. Validation of satellite-derived daily latent heat flux over the South China Sea, compared with observations and five products. *J. Atmos. Ocean. Technol.* **2013**, *30*, 1820–1832, https://doi.org/10.1175/JTECH-D-12-00153.1.

31. Wang, X.; Zhang, R.; Huang, J.; Zeng, L.; Huang, F. Biases of five latent heat flux products and their impacts on mixed-layer temperature estimates in the South China Sea. *J. Geophys. Res. Oceans* **2017**, *122*, 5088–5104. https://doi.org/10.1002/2016JC012332.

32. Roberts, J.B.; Clayson, C.A.; Robertson, F.R.; Jackson, D.L. Predicting near-surface atmospheric variables from Special Sensor Microwave/Imager using neural networks with a first-guess approach. *J. Geophys. Res.* **2010**, *115*, D19113. https://doi.org/10.1029/2009JD013099.

33. Reichstein, M.; Camps-Valls G.; Stevens, B.; Juang, M.; Denzier, J.; Carvalhais, N.; Prabhat Deep learning and process understanding for data-driven earth system science. *Nature* **2019**, *566*, 195–204. http://dx.doi.org/10.1038/s41586-019-0912-1.

34. Liu, B.; Li X.; Zheng, G. Coastal inundation mapping from bitemporal and dual-polarization SAR imagery based on deep convolutional neural networks. *J. Geophys. Res. Oceans* **2019**, *124*, 9101–9113. https://doi.org/10.1029/2019JC015577.

35. Li, X.; Liu, B.; Zheng, G.; Zhang, S.; Liu, Y.; Gao, L.; Liu, Y.; Zhang, B.; Wang, F. Deep-learning-based information mining from ocean remote-sensing imagery. *Nat. Sci. Rev.* **2020**, *7*, 1585–1606. https://doi.org/10.1093/nsr/nwaa047.

36. Wang, Y.; Li, X.; Song, J.; Li, X.; Zhong, G.; Zhang, B. Carbon sinks and variations of pCO2 in the Southern Ocean from 1998 to 2018 based on a deep learning approach. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 3495–3503, https://doi.org/10.1109/JSTARS.2021.3066552.

37. Zhang, R.; Huang, J.; Wang, X.; Zhang, J. A.; Huang, F. Effects of precipitation on sonic anemometer measurements of turbulent Fluxes in the atmospheric surface layer. *J. Ocean Univ. China* **2016**, *15*, 389–398. https://doi.org/10.1007/s11802-016-2804-4.

38. Zhou, F.; Zhang, R.; Shi, R.; Chen, J.; He, Y.; Wang, D.; Xie, Q. Evaluation of OAFlux datasets based on in situ air–sea flux tower observations over the Yongxing Islands in 2016. *Atmos. Meas. Tech.* **2018**, *11*, 6091–6106. https://doi.org/10.5194/amt-11-6091-2018.

39. Fairall, C.W.; Bradley, E.F.; Hare, J.E.; Grachev, A.A.; Edson, J.B. Bulk parameterization of air-sea fluxes: updates and verification for the COARE algorithm. *J. Clim.* **2003**, *16*, 571–591. https://doi.org/10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

40. Wentz, F.J. SSM/I Version-7 Calibration Report (Report number 011012). *Remote Sensing Systems*, Santa Rosa, CA, **2013**; pp. 46 pp.

41. Zou, C.-Z.; Wang, W. Climate Algorithm Theoretical Basis Document (C-ATBD) - AMSU Radiance Fundamental Climate Data Record Derived From Integrated Microwave Inter-calibration Approach. Tech. Rep., NOAA, Asheville, NC, **2013**.

42. Zou, C.-Z.; Hao, X. AMSU-A Brightness Temperature FCDR - Climate Algorithm Theoretical Basis Document. NOAA Climate Data Record Program CDRP-ATBD-0345, Rev. 2.0. **2016**; Available at http://www.ncdc.noaa.gov/cdr/operationalcdrs.html.

43. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, P.; Horanyi, S.; Munoz-Sabater, J.; Nicolas, J., Peubey, C.; Radu, R.; Schepers, D.; Simmons, A.; Soci, C.; Abdalla. S.; Abellan, X.; Balsamo, G.; Bechtold. P.; Biavati, G.; Bidlot, J.; Bonavita, M.; De Chiara. G.; Dahlgren, P.; Dee, D.; Diamantakis, M.; Dragani. R.; Flemming, J.; Forbes, R.; Fuentes, M.; Geer, A.; Haimberger, L.; Healy, S.; Hogan, R.J.; Holm, E.; Janiskova, M.; Keeley, S.; Laloyaux, P.; Lopez, P.; Lupu C.; Radnoti, G.; de Rosnay, P.; Rozum, I.; Vamborg, F.; Villaume. S.; Thepaut, J. The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.* **2020**, *146*, 1999–2049. https://doi.org/10.1002/qj.3803.

44. Kanamitsu, M.; Ebisuzaki, W.; Woollen, J.; Yang, S.-K.; Hnilo, J. J.; Fiorino, M.; Potter, G.L. NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteorol. Soc.* **2002**, *83*, 1631–1644. https://doi.org/10.1175/BAMS-83-11-1631.

45. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning internal representations by error propagation. *Readings in Cognitive Science* **1988**, *323*, 399–421. https://doi.org/10.1016/B978-1-4832-1446-7.50035-2.

46. Held, I. M.; Soden B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* **2006**, *19*, 5686–5699. https://doi.org/10.1175/JCLI3990.1.

Figure captions

Figure 1. (a) Geographical distribution of observational stations. Shading denotes the ocean depth. (b)-(d) are the PDDs of in situ observations of Q_a over the SCS, ECS, and YS. (e) is the mean results for all data. The range, mean value and STD of the data in each panel are shown in blue text with unit of g/kg.

Figure 2. Comparison of 3-day running averaged TPW data on June 1, 2014, from various types of sensors. (a)-(f) are sensors SSM/I F15, SSMIS F18, WindSat, AMSR, TMI and GMI, respectively. Red dots are stations located in coastal areas. Note that different sensors have different scopes of data coverage in coastal areas.

Figure 3. Architecture of the EMTnet. Each DNN_n (n=1, 2, 3, ...) uses 75% randomly sampled data from all in situ observations as training data, while the rest 25% are used as testing data.

Figure 4. Comparisons between Q_a predictions (ordinate) and observations (abscissa). (a) to (f) are the results for the Xisha Buoy station in the SCS using the EMTnet model and five existing models summarized in Table 2. (g) to (l) and (m) to (r) are the same as (a) to (f) but for the DH11 station in the ECS and the HH09 station in the YS, respectively.. The blue texts show the CC, mean bias, and RMSE in each panel. Bars on the rightmost side show the mean results of the three stations for each model. The units of mean bias and RMSE are g/kg.

Figure 5. Comparisons of the EMTnet model and the L86 model. (a) is the scatter diagram of Q_a (ordinate) and TPW (abscissa), while the bars show the model biases. In (a), the red dot denotes the data density in each 0.5 cm bin of TPW and 0.5 g/kg bin of Q_a with units of ‰. The blue square and error bar are the median and one STD of Q_a observations in each 0.5 cm

bin of TPW. (b) and (c) are the PDDs of the mean bias and RMSE in 1000 sets of DNN computations. Black and green lines and bars are the results of the L86 model and the EMTnet model. The units of mean bias and RMSE are g/kg.

Figure 6. (a) to (d) are climatology of Q_a distributions (shading) over the China Seas from the EMTnet, L86, ERA5, and NCEP2. Contours denote the intensity of seasonal variation, which is defined by one standard deviation from January to December on each grid. (e) is the ime series of Q_a anomalies over the China Seas. (f) and (g) are the same as (e) but for the southern (SCS) and northern (ECS and YS) sections of the China Seas. The values in parentheses denote the long-term trend of Q_a during the period from 1990 to 2019 with unit of g/kg per decade. The units are g/kg.

Figure 7. Scatter diagram of Q_a (ordinate) and TPW (abscissa) for data under conditions of CLW (a) below and (b) above 50 μm. The black line denotes the L86 model. The red dot denotes the data density in each 0.5 cm bin of TPW and 0.5 g/kg bin of Q_a with units of ‰. The blue square and error bar are the median and one STD of Q_a in each bin of 0.5 cm of TPW, respectively. The bar plot is the probability density distribution of CLW, with red bars representing the data range used in the corresponding panel.

Figure 8. The same as Fig. 7 but for two conditions of U.

Figure 9. The same as Fig. 7 but for two conditions of SST.

Table 1. Information on in situ observations collected in this study. The station names with prefixes "DH" and "HH" are located in the ECS and the YS, respectively. The remaining 13 stations are located in the SCS. The data span from 1998 to 2018 and the sampling intervals vary from 1 minute to 30 minutes.

Name	Location	Ocean depth	Type	Sampling interval	Period
Maoming	111.66°E, 20.75°N	~100 m	buoy	1 min	2010.05-2011.09
Shantou	117.34°E, 22.33°N	~100 m	buoy	1 min	2010.10-2011.05
Bohe	111.32°E, 21.46°N	~15 m	offshore platform	10 min	2009.11.26-2010.05.15 2011.01.04-2011.04.28 2012.03.13-2012.06.03
Xisha flux tower	112.33°E, 16.83°N	island	tower	1~10 min	2008.04-2008.10 2013.07-2017.01
Xisha buoy	112.33°E, 16.86°N	~1000 m	buoy	10 min	2009.09-2013.05 2018.05-2018.06
Kexue 1	110.26°E, 6.41°N	~1300 m	buoy	15 min	1998.05.07-1998.06.20
Shiyan 3	117.40°E, 20.60°N	~1000 m	buoy	15 min	1998.05.06-1998.06.23
SCS1	115.60°E, 8.10°N	~3000 m	buoy	15 min	1998.04.19-1998.04.29
SCS3	114.41°E, 12.98°N	~4500 m	buoy	15 min	1998.06.08-1998.06.16
SCS3[+]	114.00°E, 13.00°N	~4000 m	buoy	15 min	1998.04.13-1998.05.29
QF301	115.59°E, 22.28°N	~100 m	buoy	30 min	2011.03.01-2011.05.31
QF302	114.00°E, 21.50°N	~100 m	buoy	30 min	2011.03.01-2011.05.31
QF303	112.83°E, 21.12°N	~100 m	buoy	30 min	2011.03.01-2011.05.31
DH06	123.13°E, 30.72°N	<100 m	buoy	30 min	2012.03.29-2013.12.30
DH10	122.00°E, 31.37°N	<100 m	buoy	30 min	2013.09.01-2015.12.02
DH11	122.82°E, 31.00°N	<100 m	buoy	30 min	2014.01.01-2016.12.30
DH20	122.75°E, 29.75°N	<100 m	buoy	30 min	2014.11.06-2016.11.01
HH07	122.58°E, 37.01°N	<100 m	buoy	30 min	2012.03.29-2013.12.30
HH09	120.27°E, 35.90°N	<100 m	buoy	30 min	2014.01.01-2016.12.30
HH19	119.60°E, 35.42°N	<100 m	buoy	30 min	2014.11.06-2016.12.30

Table 2. Summary of the methods of surface humidity retrieval validated in this study.

Algorithm	Equation	RMSE (g/kg)
Liu et al. (1986)	$Q_a = C_1 * W + C_2 * W^2 + C_3 * W^3 + C_4 * W^4 + C_5 * W^5$, where C_1=0.006088244, C_2=0.1897219, C_3=0.1891893, C_4=-0.07549036, and C_5=0.006088244.	0.40 in tropics and 0.80 in globe
Jones et al. (1999)	$Q_a = C_0 + C_1 * SST + C_2 * SST^2 + C_3 * W^1 + C_4 * W^2$, where C_0=2.1052, C_1=-0.0551, C_2=0.0138, C_3=0.2435, and C_4=-0.0019.	0.77±0.39
Bentamy et al. (2003)	$Q_a = C_0 + C_1 T_{19V} + C_2 T_{19H} + C_3 T_{22V} + C_4 T_{37V}$, where C_0=-55.9227, C_1=0.4035, C_2=-0.2944, C_3=0.3511, and C_4=-0.2395.	1.40
Jackson et al. (2006)	$Q_a = C_0 + C_1 T_{52V} + C_2 T_{19V} + C_3 T_{19H} + C_4 T_{37V}$, where C_0=-105.117, C_1=0.31743, C_2=0.62754, C_3=-0.12056, and C_4=-0.33940.	0.83
Yu and Jin (2018)	$Q_a = a_0 + a_1 T_{19v} + a_2 T_{22v} + a_3 T_{37v} + a_4 T_{52v} + b_1 T_{19v}^2 + b_2 T_{22v}^2 + b_3 T_{37v}^2 + b_4 T_{52v}^2$, where a_0=1423.34, a_1=0.46967, a_2=0.43401, a_3=-0.92292, a_4=-11.494, b_1=-0.00071, b_2=-0.00072, b_3=0.00155, and b_4=0.02336 for the global model, a_0=-127.10, a_1=-0.21113, a_2=0.71712, a_3=-0.78268, a_4=1.1918, b_1=0.00062, b_2=-0.00139, b_3=0.00153, and b_4=-0.00222 for the high-latitude model.	0.82

1   **Table 3.** The mean bias and RMSE of each sensitivity experiment with EMTnet model. "Reference" refers to the result of the L86 model. In the

2   nomenclature of Exp1 to Exp7, postfixes C, U, and S denote parameters CLW, U, and SST considered in the corresponding experiment,

3   respectively. In Exp8, the postfix "none" means no additional information is considered. The percent change means the ratio of changes in

4   absolute error compared to the reference value. The units of mean bias, RMSE and absolute error are g/kg.

5

	Reference	Exp1_CUS	Exp2_CU	Exp3_CS	Exp4_US	Exp5_C	Exp6_U	Exp7_S	Exp8_none
Bias	0.72	-0.02	0.08	0.13	-0.05	-0.31	-0.22	-0.08	-0.18
RMSE	2.56	1.64	1.81	1.62	1.64	1.81	2.28	1.83	2.36
Absolute error	3.28	1.66	1.89	1.75	1.69	2.12	2.50	1.91	2.54
Percent change	—	-49%	-42%	-47%	-48%	-35%	-24%	-42%	-23%

Figure 1. (a) Geographical distribution of observational stations. Shading denotes the ocean depth. (b)-(d) are the PDDs of in situ observations of Q_a over the SCS, ECS, and YS. (e) is the mean results for all data. The range, mean value and STD of the data in each panel are shown in blue text with unit of g/kg.
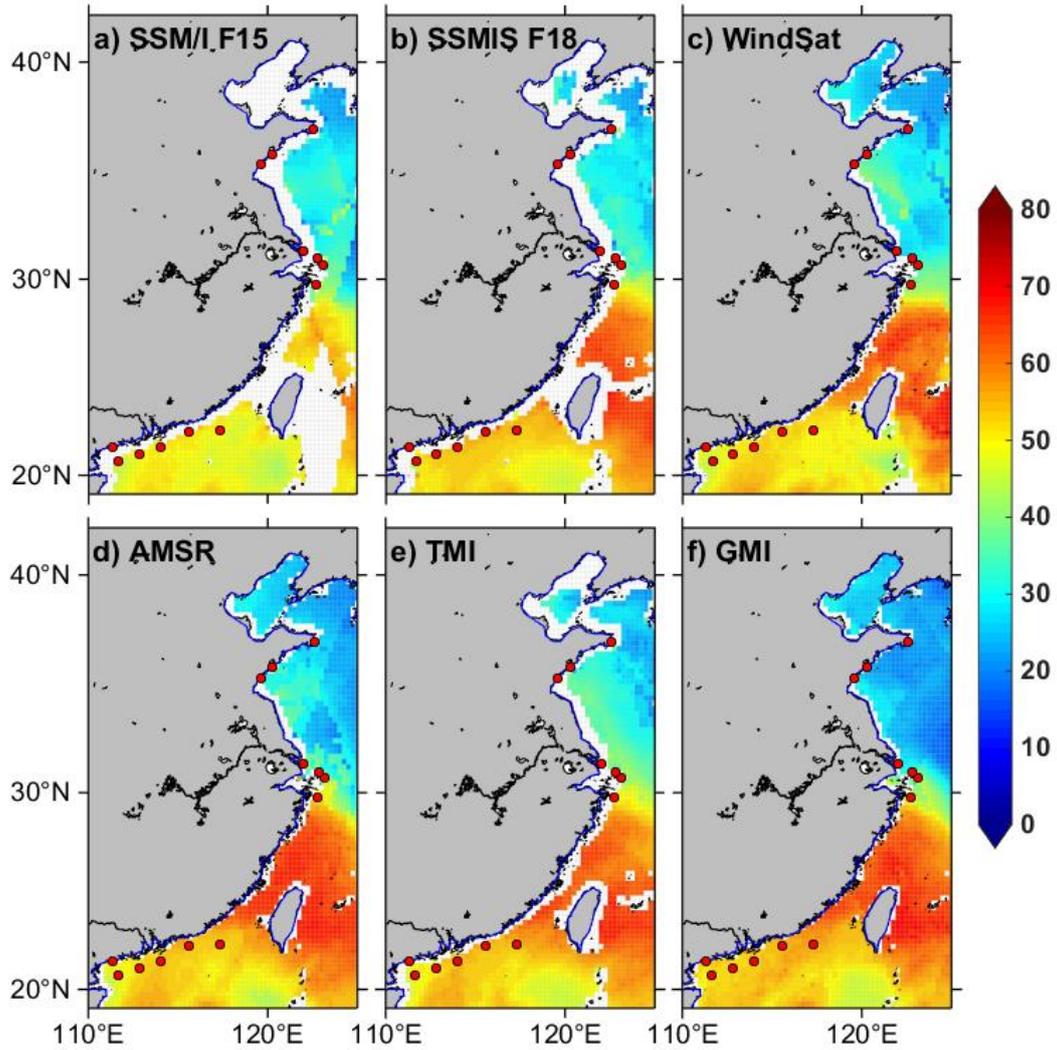
Figure 2. Comparison of 3-day running averaged TPW data on June 1, 2014, from various types of sensors. (a)-(f) are sensors SSM/I F15, SSMIS F18, WindSat, AMSR, TMI and GMI, respectively. Red dots are stations located in coastal areas. Note that different sensors have different scopes of data coverage in coastal areas.

Figure 3. Architecture of the EMTnet. Each DNN_n (n=1, 2, 3, ...) uses 75% randomly

sampled data from all in situ observations as training data, while the rest 25% are used as

testing data.

Figure 4. Comparisons between Q_a predictions (ordinate) and observations (abscissa). (a) to (f) are the results for the Xisha Buoy station in the SCS using the EMTnet model and five existing models summarized in Table 2. (g) to (l) and (m) to (r) are the same as (a) to (f) but for the DH11 station in the ECS and the HH09 station in the YS, respectively.. The blue texts show the CC, mean bias, and RMSE in each panel. Bars on the rightmost side show the mean results of the three stations for each model. The units of mean bias and RMSE are g/kg.

Figure 5. Comparisons of the EMTnet model and the L86 model. (a) is the scatter diagram of

Q_a (ordinate) and TPW (abscissa), while the bars show the model biases. In (a), the red dot

denotes the data density in each 0.5 cm bin of TPW and 0.5 g/kg bin of Q_a with units of ‰.

The blue square and error bar are the median and one STD of Q_a observations in each 0.5 cm

bin of TPW. (b) and (c) are the PDDs of the mean bias and RMSE in 1000 sets of DNN

computations. Black and green lines and bars are the results of the L86 model and the

EMTnet model. The units of mean bias and RMSE are g/kg.

Figure 6. (a) to (d) are climatology of Q_a distributions (shading) over the China Seas from the EMTnet, L86, ERA5, and NCEP2. Contours denote the intensity of seasonal variation, which is defined by one standard deviation from January to December on each grid. (e) is the ime series of Q_a anomalies over the China Seas. (f) and (g) are the same as (e) but for the southern (SCS) and northern (ECS and YS) sections of the China Seas. The values in parentheses denote the long-term trend of Q_a during the period from 1990 to 2019 with unit of g/kg per decade. The units are g/kg.
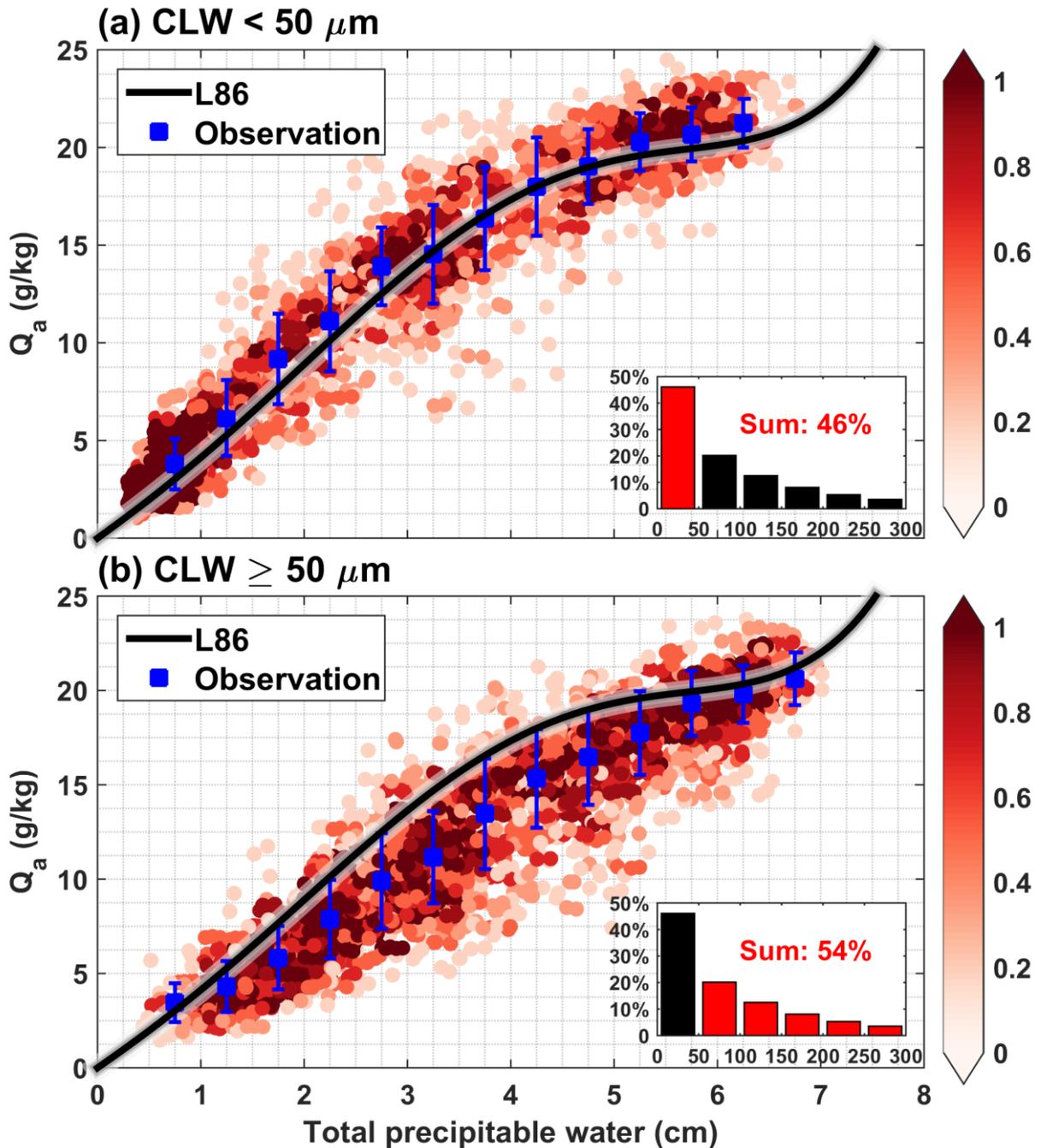
Figure 7. Scatter diagram of Q_a (ordinate) and TPW (abscissa) for data under conditions of

CLW (a) below and (b) above 50 μm. The black line denotes the L86 model. The red dot

denotes the data density in each 0.5 cm bin of TPW and 0.5 g/kg bin of Q_a with units of ‰.

The blue square and error bar are the median and one STD of Q_a in each bin of 0.5 cm of

TPW, respectively. The bar plot is the probability density distribution of CLW, with red bars

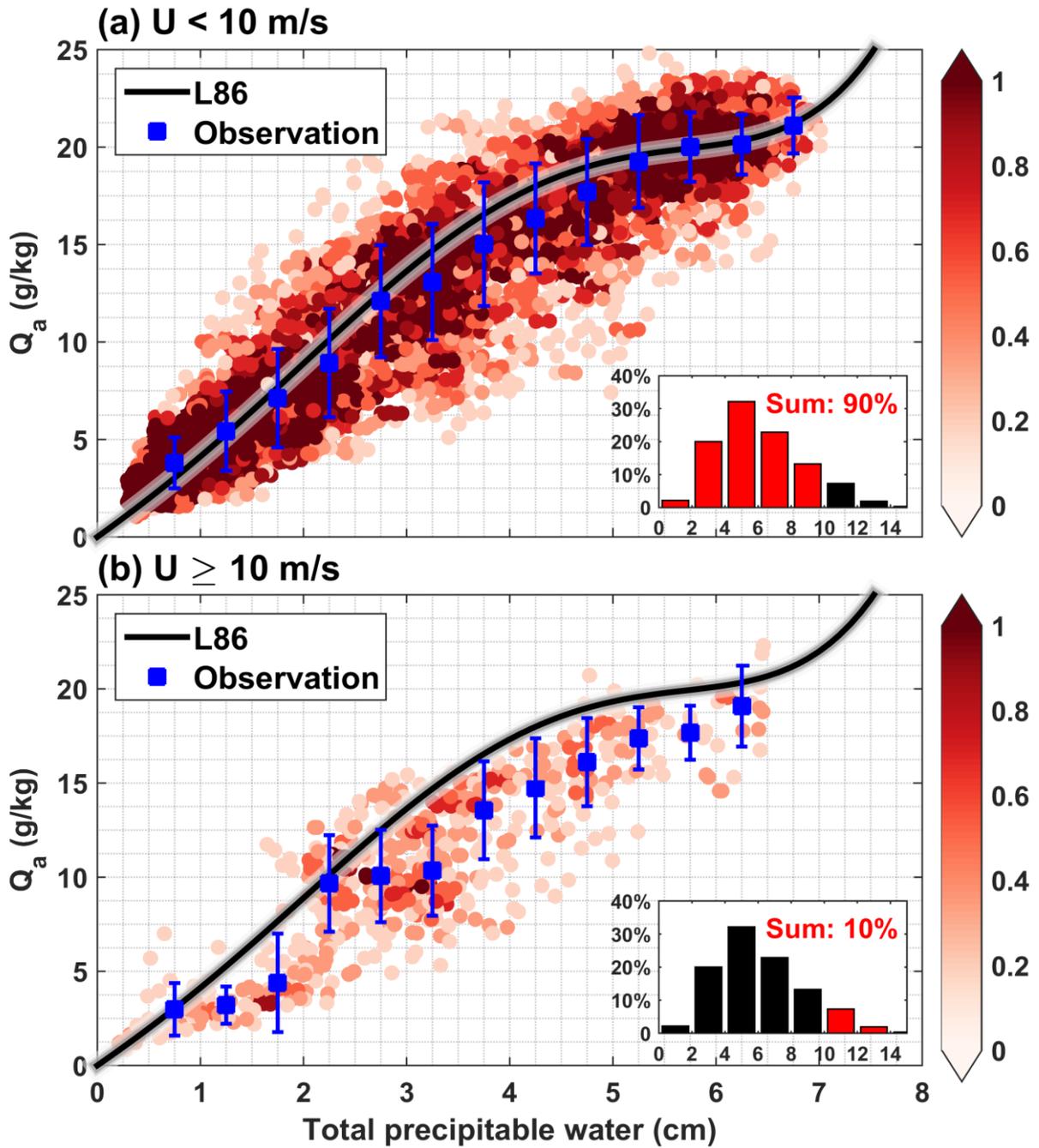representing the data range used in the corresponding panel.

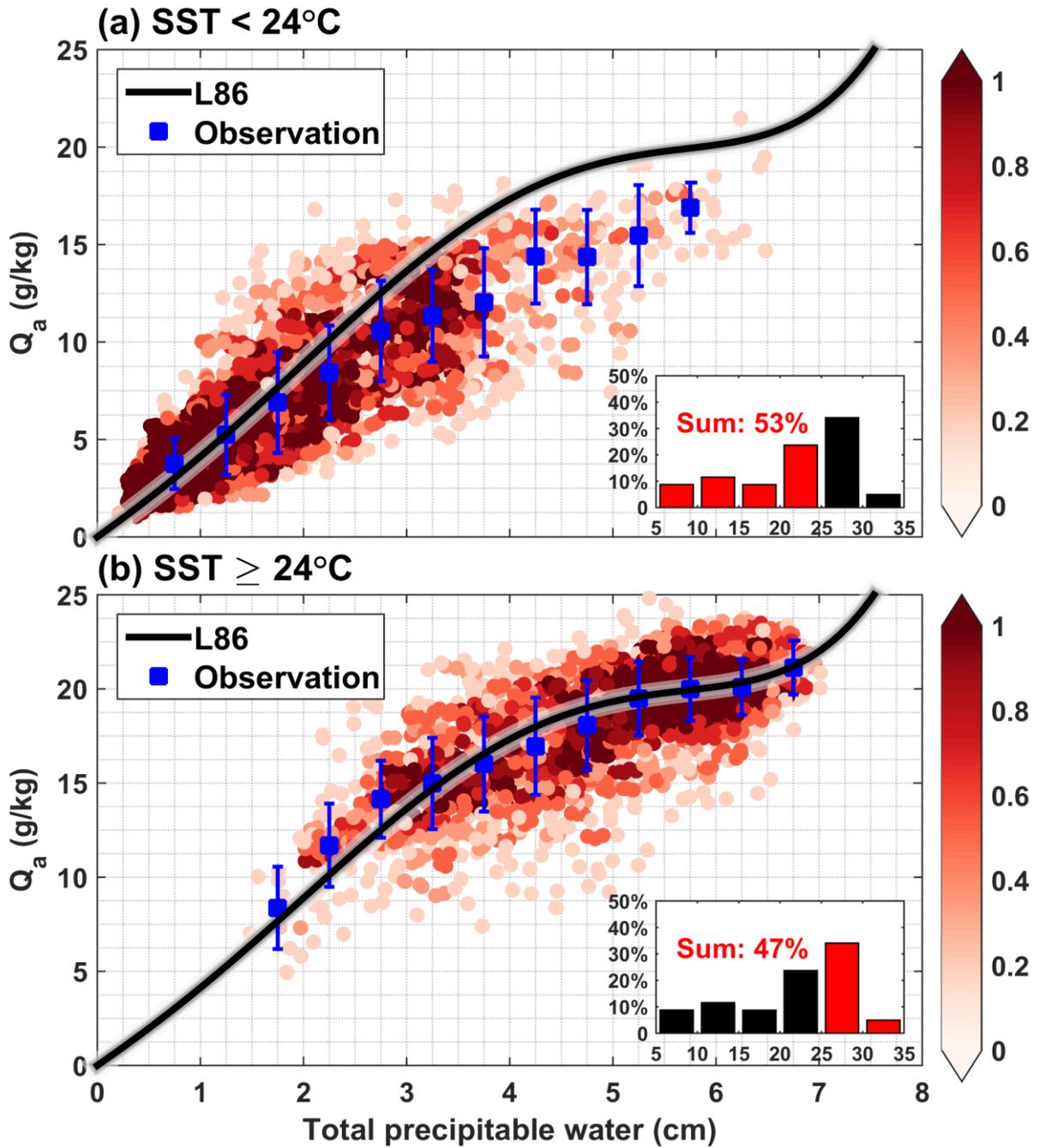**Figure 8.** The same as Fig. 7 but for two conditions of *U*.

Figure 9. The same as Fig. 7 but for two conditions of SST.