Atmospheric correction for HY-1C CZI images using neural network in western Pacific region

Jilin Men, Jianqiang Liu, Guangping Xia, Tong Yue, Ruqing Tong, Liqiao Tian, Kohei Arai & Linyu Wang

Published online: 08 Feb 2022.

Submit your article to this journal ⬚

Article views: 944

View related articles ⬚

View Crossmark data ⬚

Citing articles: 2 View citing articles ⬚

Taylor & Francis
Taylor & Francis Group

Atmospheric correction for HY-1C CZI images using neural network in western Pacific region

Jilin Men [a], Jianqiang Liu[b], Guangping Xia[a], Tong Yue[c], Ruqing Tong[a], Liqiao Tian[a], Kohei Arai[d] and Linyu Wang[e]

[a]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; [b]Ministry of Natural Resource of the People's Republic of China, National Satellite Ocean Application Service, Beijing, China; [c]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China; [d]Department of Information Science, Saga University, Saga, Japan; [e]Department of Urban Development, Wuhan Land Containment and Reserve Center, Wuhan, China

ABSTRACT

With a spatial resolution of 50 m, a revisit time of three days, and a swath of 950 km, the coastal zone imager (CZI) offers great potential in monitoring coastal zone dynamics. Accurate atmospheric correction (AC) is needed to exploit the potential of quantitative ocean color inversion. However, due to the band setting of CZI, the AC over coastal waters in the western Pacific region with complex optical properties cannot be realized easily. This research introduces a novel neural network (NN) AC algorithm for CZI data over coastal waters. Total 100,000 match-ups of HY-1 C CZI-observed reflectance at the top-of-atmosphere and Operational Land Imager (OLI)-retrieved high-quality remote sensing reflectance (R_{rs}) at the CZI bands are built to train the NN model. These reflectance data are obtained from the standard AC algorithm in the SeaDAS. Results indicate that the distributions of the CZI retrieved R_{rs} were consistent with the quasi-synchronous OLI data, but the spatial information from the CZI is more detailed. Then, the accuracy of the CZI data for AC is evaluated using the multi-source in-situ data. Results further show that the NN-AC can successfully retrieve R_{rs} for CZI and the coefficients of determination in the blue, green, red, and near-infrared bands were 0.70, 0.77, 0.76, and 0.67, respectively. The NN algorithm does not depend on shortwave-infrared bands and runs very fast once properly trained.

1. Introduction

As a new generation ocean color sensor entailing four spectral bands with central wavelengths of 460, 560, 650, and 825 nm, the Coastal Zone Imager (CZI) onboard the HY-1C satellite launched in 2018 has provided continuous observations of China's coastal waters. The corresponding details of the spectral response function (SRF) are shown in Figure 1. The CZI camera is a new wide-band remote sensor with a swath width of approximately 950 km and a spatial resolution of 50 m. The revisit period is 3 d, and the overpass time is approximately 10:30 local time. This sensor has been designed to support the research of coastal zone areas, which are the regions closely related to human activities and where strong land-sea interaction occurs.

The traditional Atmospheric Correction (AC) methods are not suitable for CZI because of its band setting. The standard AC algorithms (Gordon 1997; Gordon and Wang 1994) in Sea-viewing Wide Field-of-view Sensor Data Analysis System (SeaDAS) use two near-infrared (NIR) bands to estimate the contribution of aerosols at visible bands. This method is sufficient in the

open ocean areas but is challenging to operate in coastal areas where the black ocean assumption often fails (Gordon and Wang 1994; Zhang et al. 2007). Consequently, an alternative AC algorithm was constructed for AC in coastal waters with non-zero NIR water-leaving radiances (L_w) by using Shortwave Infrared (SWIR) bands. In this scheme, the black ocean assumption may still be tenable owing to the strong absorption of water in SWIR, and the L_w in the visible bands can be obtained by extrapolation (Wang and Shi 2007). Several modified algorithms based on this principle were subsequently developed, and improvements in accuracy, at least for selected validation areas, were achieved (Jiang and Wang 2014; Wang, Shi, and Jiang 2012; Wang and Shi 2007). However, as CZI only comprises the four bands mentioned above, the lack of SWIR renders the AC for CZI a daunting task.

The Neural Network (NN) method is a powerful tool for prediction, recognition, function approximation, and pattern classification (Haykin and Network 2004). The feasibility of applying the NN method to AC has been investigated in several previous studies
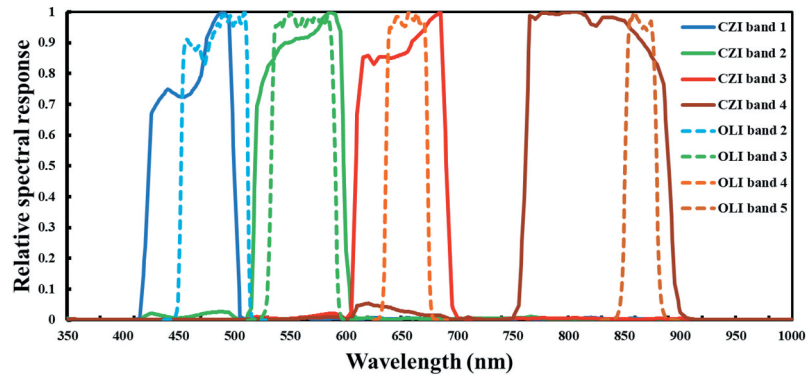

Figure 1. Relative response functions for CZI bands (solid) and OLI bands (dashed).

(Brajard et al. 2012; Brajard, Moulin, and Thiria 2008; Chen et al. 2014; Fan et al. 2017; Hamre et al. 2013; Jin and Stamnes 1994; Li et al. 2020; Schroeder et al. 2007). Brajard, Moulin, and Thiria (2008) used a trained NN to replace the radiative transfer model in the forward simulation to obtain the Top-Of-Atmosphere (TOA) radiances. Fan et al. (2017) employed a coupled atmosphere-ocean radiative transfer model to simulate the radiance at the TOA (L_{TOA}) and the remote sensing reflectance (R_{rs}) just above the surface. They also trained two NNs to directly derive the aerosol optical depth and R_{rs} from the L_{TOA}, respectively (Fan et al. 2017). Although the studies demonstrated that the NN method could be successfully used for AC, it could hardly be applied to CZI images. For example, only a few *in-situ* data could be synchronized with CZI, hence the lack of representativeness of the training samples. An alternative solution is to use an auxiliary sensor, but its setting needs to be like that of the CZI in the visible bands. The auxiliary sensor should also possess NIR and SWIR bands to be able to produce R_{rs} with the standard AC methods. The Operational Land Imager (OLI)

onboard Landsat-8 generates high-quality aquatic science products (Franz et al. 2015; Pahlevan et al. 2017), that can be applied as an auxiliary sensor for CZI owing to the following reasons: (1) OLI has a wider spectrum coverage than CZI, including NIR and SWIR bands. (2) The CZI camera visible bands (central wavelength: 460, 560, 650, and 825 nm; bandwidth: 420–500, 520–600, 610–690, and 760–890 nm) roughly correspond to those of Landsat-8 OLI bands 2, 3, 4, and 5 (central wavelength: 482, 561, 655 and 865 nm; bandwidth: 450–510, 530–590, 640–670, 850–880 nm) after band conversion, as shown in Figure 2. (3) The relatively high Signal-To-Noise Ratio (SNR) and the similar spatial resolution from OLI will be beneficial for the AC of CZI. (4) The training strategy, as to be described in the following paragraph, can help the model to further consider the differences (e.g. bandwidths) between two sensors.

This study proposes a new AC algorithm for CZI based on the NN, a scheme that differs from those in previous studies in terms of NN structure and training process (Figure 2). First, we calculate the TOA reflectance (ρ_{TOA}) from the CZI data and R_{rs} from the OLI
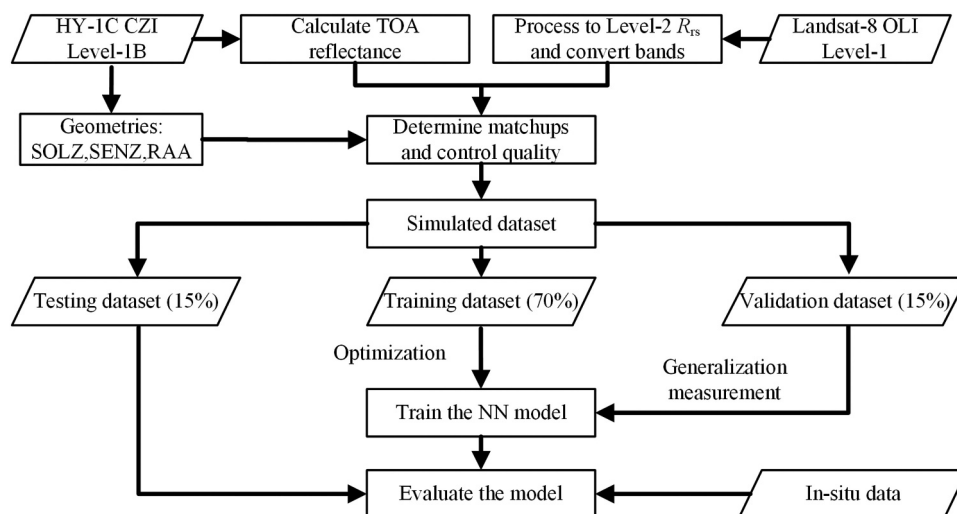

Figure 2. Flowchart of atmospheric correction process by using the NN approach. (NN: neural network model, SOLZ: solar zenith angle, SENZ: sensor zenith angle, RAA: relative azimuth angle).

data by using the standard AC method (hereafter called as SWIR-AC) embedded in SeaDAS, which is one of the best AC methods for OLI (Xu et al. 2020). Then, as shown in Figure 1, we apply the band conversion of R_{rs} from OLI to CZI based on the band SRF to build the simulated dataset for model training (Pinto et al. 2016). Based on the dataset, we take the ρ_{TOA} of the four bands and observation geometries (Solar Zenith Angle (SOLZ), Sensor Zenith Angle (SENZ), and Relative Azimuth Angle (RAA)) as the inputs of NN and R_{rs} as the outputs to train the network. Finally, we validate the NN method on a testing dataset and *in-situ* data. The detailed validation results are detailed in the succeeding sections.

2. Data and method

2.1. Simulated dataset

For the model training and validation, a simulated dataset was built by selecting the synchronous remote sensing images of OLI Level 1 and CZI Level 1B for 2019 from https://earthexplorer.usgs.gov/ and https://osdds.nsoas.org.cn/#/, respectively. The corresponding locations are marked in red boxes in Figure 3(a). A total of three match-up pairs of the images were obtained with the cloud coverage < 3%, and the time difference < 3 hours.

The top-of-atmosphere reflectance (ρ_{TOA}) from the top-of-atmosphere radiance (L_{TOA}) of CZI is calculated using

$$\rho_{TOA}(\lambda) = \pi L_{TOA}(\lambda) / \cos(\theta_0) F_0(\lambda) \quad (1)$$

where θ_0 and F_0 are the SOLZ and extraterrestrial solar irradiance, respectively. The observation geometries (cosine of the SOLZ, cosine of the SENZ, and cosine of the RAA) and ρ_{TOA} at four bands were used as the inputs of NN.

As for the outputs of NN, R_{rs} for four bands were obtained from the corresponding OLI images via the SWIR-AC embedded in SeaDAS (ver. 7.5). The spectral response difference (Figure 1) between the two sensors was compensated by the spectral band adjustment factor $\beta(\lambda)$, as calculated by Equation (2), using the relative response functions and field measurements R_{rs}^{field} ($N= 83$) obtained for the South China Sea in September, 2018.

$$\beta(\lambda) = \frac{R_{rs}^{CZI}(\lambda)}{R_{rs}^{OLI}(\lambda)}$$
$$= \frac{R_{rs}^{field}(\lambda) \times \mathrm{SRF}_{CZI}(\lambda)d\lambda / \ \mathrm{SRF}_{CZI}(\lambda)d\lambda}{R_{rs}^{field}(\lambda) \times \mathrm{SRF}_{OLI}(\lambda)d\lambda / \ \mathrm{SRF}_{OLI}(\lambda)d\lambda} \quad (2)$$

$$R_{rs}^{CZI}(\lambda) = \beta(\lambda) R_{rs}^{OLI}(\lambda) \quad (3)$$

where SRF is the spectral response function, and the β values at the 460, 560, 650, and 825 nm bands are 0.886211, 0.9995, 1.020512, and 1.240233, respectively. The spectral band conversion of R_{rs} from OLI to CZI was accomplished using Equations (2) and (3). The OLI images were resampled to the spatial resolution of 50 m to ensure pixel homogenization.

Quality control methods for Level-3 global products were used to ensure the quality of the simulated data. All pixels were screened using the 12-flags bit
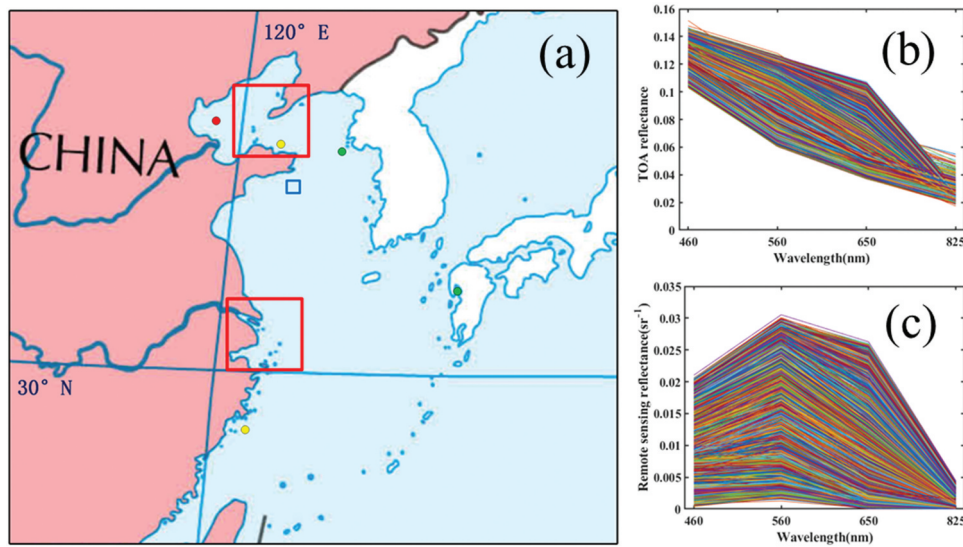

Figure 3. (a) shows locations of training data and *in-situ* data. The red boxes show the boundaries for each area where the training images were extracted; the blue box is the testing area used for algorithm evaluation in Section 3.3; the red, green and yellow dots show the locations of the *in-situ* observations (red dots: ship-borne measurements, green dots: Socheongcho and ARIAKE_TOWER stations, yellow dots: MuPing and DongTou stations) described in Table 2. (b) and (c) are the spectra of top-of-atmosphere reflectance and remote sensing reflectance of the training dataset, respectively, obtained from the red boxes in (a).

operations. Pixels involving any of the following quality-control flags were discarded: ATMFAIL (00), LAND (01), HIGLINT (03), HILT (04), HISATZEN (05), STRAYLIGHT (08), CLDICE (09), COCCOLITH (10), HISOLZEN (12), LOWLW (14), CHLFAIL (15), NAVWARN (16), MAXAERITER (19), MODGLINT (20), CHLWARN (21), ATMWARN (22), and NAVFAIL (25). In the list, the numbers in the parentheses represent the bit position of the 32-bit flag value. Then, the following criteria were used to extract the high-quality R_{rs} samples:

(1) The percentages of pixels with valid R_{rs} (460 nm) values in the 3×3 box are checked. If the valid percentage is larger than 50%, then the data in the box are further checked by the following quality assessment, otherwise it is discarded.

(2) The means and standard deviations (SDs) of the valid R_{rs} (460 nm) values within the box are calculated. Pixels with R_{rs} (460 nm) values beyond the range of mean±1.5SD in the box are discarded.

(3) The means and SDs of valid R_{rs} (460 nm) values are recalculated for the remaining valid pixels, and the coefficients of variation (CVs) are determined (i.e. SD divided by mean) to check for spatial heterogeneity. If the CV was is than 0.15, then the box is adopted in the next step; otherwise, it is discarded.

Standards (1) to (3) can help to ensure the uniformity of the spatial range of data and avoided the noise caused by instruments or stray clouds.

In total, we achieved 2,268,327 match-up pairs. Among them, 100,000 pairs were randomly selected as the simulated dataset. Then, the selected pairs were divided into three independent groups: a training subset (70,000 data points), a testing subset (15,000 data points), and a validation subset (15,000 data points). The spectral shapes of the simulated dataset are shown in Figures 3(b,c), and the maximum, mean, and minimum values are shown in Table 1. In view of ensuring the representativeness of the training dataset, a simple empirical criterion was applied to classify the different types of water: R_{rs} (650) < 0.0005 [sr^{-1}] for clear water, R_{rs} (650) > 0.012 [sr^{-1}] for turbid water, and in-between values of these two ranges for moderately turbid water. The rates of the three types of water in the training dataset were 79%, 17%, and 4% for clear water, turbid water, and moderately turbid water, respectively. The percentages indicate that the training dataset comprised various water types.

2.2. Neural network training

Previous studies have demonstrated that NN can approximate nonlinear functions (Chen et al. 2014; Liu et al. 2021; D'Alimonte, Zibordi, and Berthon 2004; D'Alimonte and Zibordi 2003; Géron 2019; Cybenko 1989; Hornik, Stinchcombe, and White 1989; Pinkus 1999). Therefore, the NN is suitable for solving our inverse problem of deriving R_{rs} from ρ_{TOA}. This section describes the NN model designed for the CZI sensor.

In the construction of our NN, an important issue was to find the optimum number of hidden layers and neurons. The choice generally depends on many variables, but this approach hinders the identification of the best solution in many cases. We used a practical approach to determine the number of hidden layers and neurons by utilizing many NN configurations with different numbers of hidden neurons (In particular, 10, 11, 13, and 15 neurons were selected, and we did not add more hidden layers due to the high-accuracy objective for the testing dataset). We recorded the process and outcomes of these NNs as a means of determining the

Table 1. Statistics of top-of-atmosphere reflectance and remote sensing reflectance used for training and evaluation of NN. CV is coefficient of variation computed by the ratio of standard deviation to the mean.

	ρ_{TOA}				$R_{rs}(sr^{-1})$			
	460 nm	560 nm	650 nm	825 nm	460 nm	560 nm	650 nm	825 nm
Maximum	0.1515	0.1278	0.1072	0.0549	0.0210	0.0305	0.0263	0.0044
Mean	0.1174	0.0772	0.0488	0.0264	0.0057	0.0072	0.0026	0.0003
Minimum	0.1029	0.0604	0.0369	0.0173	0.0004	0.0013	0.0001	0.0001
CV	0.0750	0.1701	0.2276	0.1451	0.7187	0.8259	1.7893	2.2331

Table 2. Details of the four stations: region, location, distance from land and the number of measurements within the period.

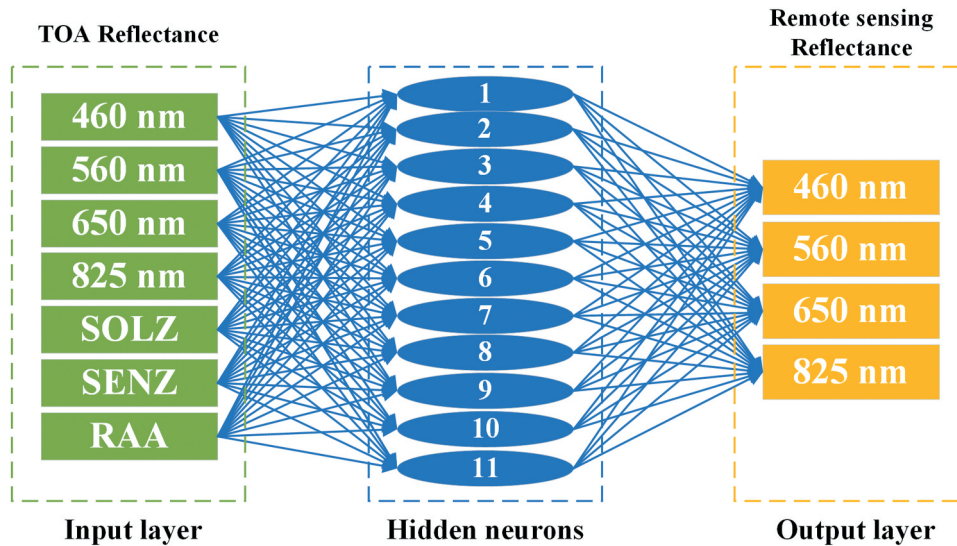| Station	Region	Latitude	Longitude	Distance from land	Time period	Number of points
Socheongcho	Yellow Sea	37.423°N	124.738°E	34 km	2019.4–2020.12	29
ARIAKE_TOWER	Sea of Japan	33.104°N	130.272°E	5 km	2019.4–2020.12	9
MuPing	Yellow Sea	37.681°N	121.700°E	21 km	2020.1–2020.11	11
DongTou	East China Sea	27.675°N	121.355°E	24 km	2019.9–2020.12	7

Figure 4. Architecture of the neural network that models the atmospheric correction for the HY-1 C CZI data. The SOLZ, SENZ and RAA is cosine of the solar zenith angle, sensor zenith angle, relative azimuth angle, respectively.

optimal performance and computing time configuration for the training process. Finally, we used a single-hidden layer configuration with 11 neurons, in which 0.99 was obtained for the coefficient of determination (R^2).

Our NN architecture is shown in Figure 4. The input layer entails seven elements, including the cosines of SOLZ, SENZ, and RAA, and the ρ_{TOA} at the four wavelengths of 460, 560, 650 and 825 nm. The output layer entails four elements corresponding to the R_{rs} values of the four wavelengths mentioned above. The input and output parameters were normalized using the mapminmax function embedded in the MATLAB NN toolbox, as follows:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (4)$$

where x and y are the values of the training data before and after normalization, and x_{min} and x_{max} are the minimum and maximum values of x. A hyperbolic tangent sigmoid function was utilized as the activation function of the neurons, and a linear function was employed to transfer the hidden layer neurons to the output. Moreover, we adopted the Levenberg-Marquardt back-propagation algorithm for the training process of the NN.

In the training process, the simulated dataset with 100,000 match-up pairs, was divided into model training, validation, and testing datasets based on the approximated random selection of 70%, 15%, and 15%, respectively. The training dataset was used to optimize the weights and biases of the NN according to the error. The validation dataset was used to measure the network generalization and to halt the training when the generalization shows no apparent improvement. The testing dataset was not involved in the training process, because it would be used for

the unbiased evaluation of the performance of the trained network. The initial weights and biases of the network were randomly obtained with values from 0 to 1 (Nguyen and Widrow 1990). Then, the training algorithm was used to iteratively update the weights and biases in accordance with the training error between the simulated R_{rs} values and the NN outputs. In each iteration, the validation dataset was used to monitor the performance of the current NN by computing the mean square error (MSE). If the MSE increases by more than ten iterations, then the training process is completed; otherwise, the training stops after 1,000 iterations. The relatively uncomplicated model can also help to prevent over-fitting.

2.3. In-situ data

To test the performance of the proposed AC method, we collected *in-situ* R_{rs} data from different sources, including (1) station observation data in western Pacific region, such as the Socheongcho, ARIAKE_TOWER, MuPing, and DongTou stations, and (2) ship-borne measurements data in the Bohai Sea. The red, green, and yellow dots in Figure 3 represent the specific locations of the different observation sources. The following aspects were considered in the match-up determination: time difference of < 3 hours, valid pixels of >50% in the 3 × 3 box after quality control, and CV of < 0.15 (Section 2.1). After removing some outliers and data without CZI matches, there are 65 match-up pairs left. The mean R_{rs} at the 460, 560, 650, and 825 nm were 0.0133, 0.0172, 0.0065, and 0.0009 sr^{-1}, and the corresponding CVs were 0.8744, 0.6718, 0.9697, and 1.080. The detailed *in-situ* data information is discussed in the succeeding sections.

2.3.1. Station observation data

Aerosol Robotic Network-Ocean Color (AERONET-OC), equipped with additional deployments of autonomous radiometer systems on offshore platforms, provides consistent and accurate measurements to support satellite ocean color validations (Zibordi et al. 2009). Two AERONET-OC stations (Socheongcho and ARIAKE_TOWER stations) were selected to obtain Level 2.0 data from NASA (https://aeronet.gsfc.nasa.gov/new_web/data.html). The two stations represent different aquatic optical properties. The Socheongcho station in the Yellow Sea is located far away from land with clear surrounding water, while the ARIAKE_TOWER station in the Sea of Japan is located close to land with surrounding turbid water. As these stations have long-term observations, they offer a superiority alternative approach of validating the time consistency of the NN algorithm's performance. All stations use the same instrument, including calibration and post-processing procedure, indicating that discrepancies introduced by the differences in the measurement methods can be avoided. The normalized water-leaving radiance ($L_{wn-}f/Q$), which was corrected for non-isotropic effects, was used in this research. The $L_{wn-}f/Q$ data at the wavelengths corresponding to the CZI were subsequently converted into R_{rs} by using the same method employed in the literature (Pahlevan et al. 2017). Due to the lack of corresponding bands for the CZI, a linear interpolation method was used to realize the spectral band matching between satellite data and measured data (Li et al. 2020). 443, 490, 551, and 667 nm were used for interpolation to obtain R_{rs} (460), R_{rs} (560), and R_{rs} (650). Due to the lack of the AERONET-OC band corresponding to 825 nm, only three visible bands were compared.

MuPing and DongTou stations are in the Yellow Sea and the East China Sea and they were designed by the National Satellite Ocean Application Service (NSOAS) to obtain long-term *in-situ* values of apparent optical properties, inherent optical properties, atmospheric optics, hydrology, meteorology, and main components concentration of ocean color. Both stations can provide high-frequency *in-situ*R_{rs} (i.e. nearly half an hour per observation) as measured by a CE-318 spectroradiometer. The data processing scheme follows the NASA ocean optics protocol (Mueller et al. 2003). 442, 490, 620, 667, 779, and 865 nm were used for linear interpolation to obtain R_{rs}(460), R_{rs}(650), and R_{rs}(825).

2.3.2 Ship-borne measurements data

The ship-borne measurements data used in this study were measured by an autonomous ship-borne above-water hyperspectral radiometer (350–2500 nm with ~1 nm increments) during a cruise on 29 October 2018 in Bohai Sea. The data processing scheme followed the

NASA ocean optics protocol (Mueller et al. 2003). The total distance of the ship was 5.4 km and 9 sample points were measured on the way within four hours starting at 10:00 am local time. The values of the sampling points were collected in a short time, indicating the unique superiority for validation of AC methods.

2.4. Accuracy assessment

The accuracy of the R_{rs} prediction was assessed by comparing the retrieved R_{rs} values with the true values obtained from the simulated dataset and *in-situ* data. The comparison was quantified using the mean Absolute Percentage Deviation (APD) and the Root-Mean-Square Error (RMSE) between the retrieved R_{rs} and true R_{rs}:

$$APD(\%) = 100\% \times \frac{1}{n} \sum_{i=1}^{n} \left| \frac{R_{r,i} - R_{t,i}}{R_{t,i}} \right| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(R_{r,i} - R_{t,i} \right)^2} \quad (6)$$

where $R_{r,i}$, $R_{t,i}$, and n are the retrieved values, true values, and sample number, respectively.

3. Results and discussion

3.1. Cross-sensor agreement

We extracted the concurrent OLI and CZI TOA reflectance values within three hours to analyze the correlations for all four bands of CZI via linear regression. As shown in Figure 1, differences exist between the spectral coverage and center wavelengths of OLI and CZI in all four bands. The agreement and difference over CZI and OLI images, were visualized and quantified by randomly selecting and comparing 100,000 ρ_{TOA} pairs of CZI and OLI As shown in Figure 5, although the data have some discretization, most of them were centered around the 1:1 line for all four bands, indicating a strong agreement between OLI and CZI. The positive linear trends can be used to describe the relationships between the CZI and OLI data. The slope values were extremely close to 1, and the R^2 values are higher than 0.79. The results depict a generally strong agreement between the two sensors.

3.2. Performance of the new NN algorithm with simulated data

The performance of our NN algorithm was initially evaluated with the simulated dataset. The training and testing datasets (70% and 15% of the simulated dataset) were used to evaluate the trained NN. The performance assessment was based on statistical parameters, including R^2, APD, and RMSE.
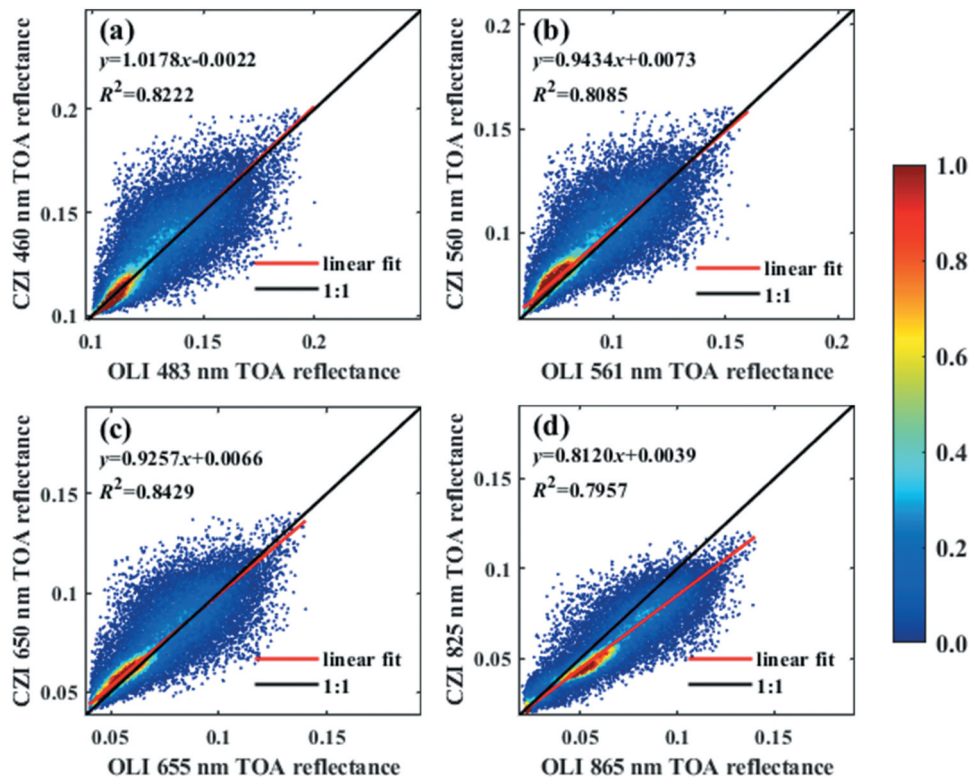
Figure 5. Scatterplots showing 100,000 matchups of Landsat-8 OLI and HY-1 C CZI TOA reflectance for (a) blue, (b) green, (c) red and (d) NIR bands, respectively. Color indicates density of matchups. Black line is the 1:1 reference line while the red line is the linear fitting line.

Figure 6 shows a comparison of the scatter plots of the R_{rs} values retrieved by the NN-AC and the simulated R_{rs} values obtained from the training and testing datasets. Figure 6(a) shows the performance of the NN on the training dataset. The retrieved R_{rs} (λ) values from the OLI images are consistent with the simulated values. The R^2 values are larger than 0.98 at the blue, green, and red bands, and larger than 0.93 at the NIR band. The *APD* values of the training datasets are lower than 15% at the visible bands, and the RMSE values do not exceed 0.0006 sr^{-1} for all bands. More importantly, the results obtained from statistical parameters on the testing datasets are extremely close to those of the training datasets, as shown in Figure 6(b).

The comparison of the retrieved R_{rs} values and the true values on the training dataset indicate that the newly built NN-AC method can be used to simulate the process of radiative transfer. The homologous results on the testing dataset indicate that the NN-AC can accurately learn the training dataset.

To examine the impact of uncertainty, we added a 3% random error to the ρ_{TOA} of the testing dataset. The results are shown in Figure 7. An increase in APD from 8.65%, 4.14%, 14.95%, and 13.59% to 18.45%, 31.53%, 35.66%, and 45.25% can be observed for the four bands, respectively. Regardless of the high scattering, the scatter points are almost within the range of 1.5 SDs at the visible bands. In the NIR band, the APD increases by 30% due to the random error; this trend may be caused by the relatively low signal and SNR values. Overall, the NN-AC can

resist the influence of random error to a certain extent. However, when the signal values are extremely low, the instability of the method is apparent.

3.3. *Evaluation of R_{rs} retrieval with CZI data*

The performance of the NN-AC was also evaluated by applying the algorithm to an HY-1 C CZI image in an optically complex coastal area. Taking the Yellow Sea as an example, as shown by the blue box in Figure 3. This region has relatively clean water and turbid water. Figures 8(a,b) show the contrast between the NN-AC and SWIR-AC for the turbid water and clear water areas, respectively. The retrieved R_{rs} values by NN-AC were extremely close to the results of SWIR-AC for two types of water with R^2 higher than 0.97 and RMSE lower than 0.0085. The results indicate that the NN-AC method is suitable for water bodies of varying degrees of turbidity.

Figures 9(a-h) show the highly similar spatial distributions of the R_{rs} values retrieved from NN-AC and SWIR-AC. Many black points can be observed in Figures 9(e-h) presenting the AC failure of SWIR-AC; this problem does not exist in the NN-AC results. The details are magnified in Figures 9(a,b,e,f) for easy viewing. The comparison indicates that the NN-AC method can achieve the AC of CZI, and its spatial distribution of R_{rs} is smoother than that of SWIR-AC. Figures 9(i-l) show the APD spatial distributions at the four bands for the NN-AC and SWIR-AC, while Figure 9(i) shows the mean APD
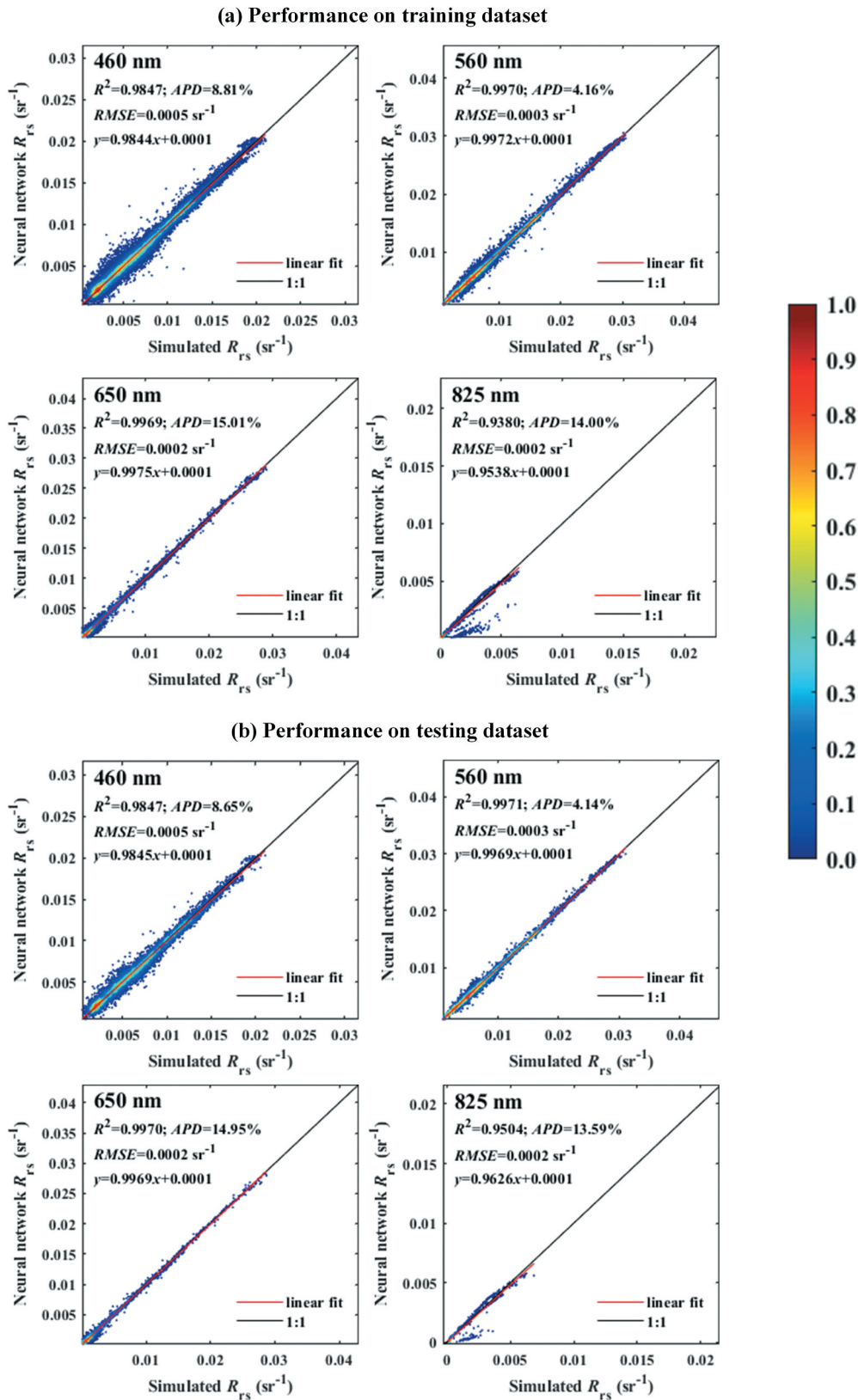
Figure 6. (a): Scatterplots showing comparison between the retrieved R_{rs} values by NN-AC algorithm and simulated R_{rs} values from the training dataset. (b): Same as (a) but obtained from testing dataset. Color indicates density of matchups.

histogram. The APDs at the first three bands are stable at approximately 20%. The relatively small error for the high-turbidity water indicates the NN-AC method can adapt to coastal waters. Notably, the APD in the NIR band is close to 40%, which may be caused by the low signal in this band; this finding indicates that the NN-AC method is unstable in clean water with low signal. The uncertainty caused by the band conversion may also lead to a relatively high APD, as depicted by spectral distance of
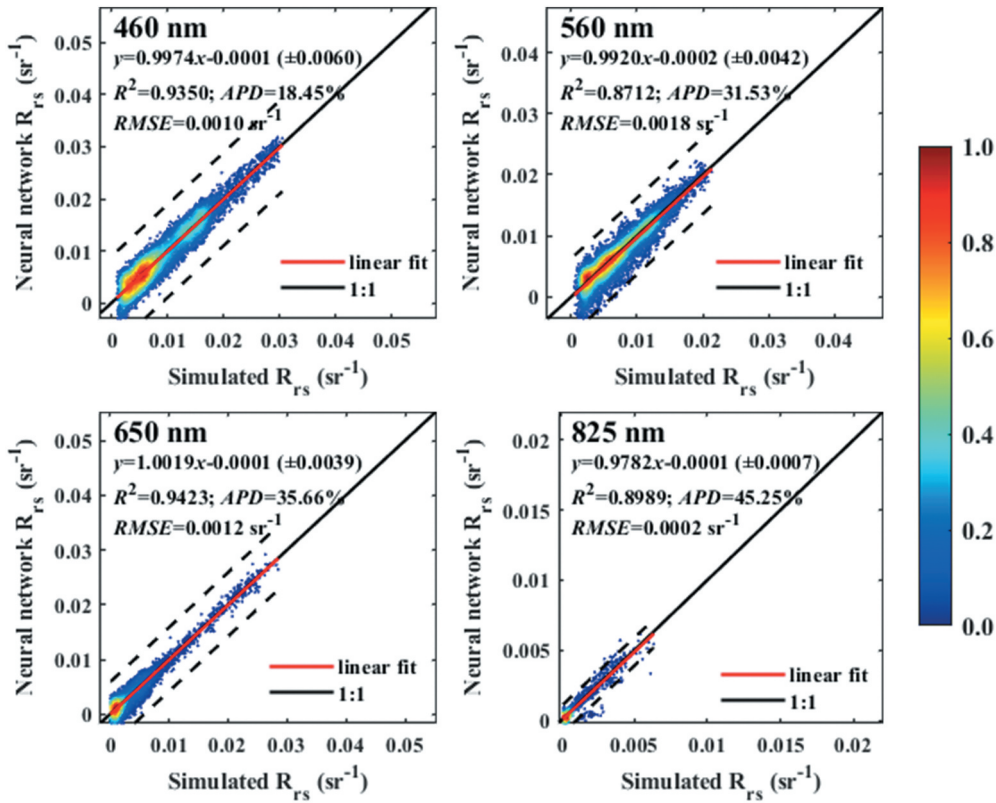
Figure 7. (a): Scatterplots showing matchups of the retrieved R_{rs} values of the NN-AC algorithm after adding 3% random error and simulated values on the testing dataset. Dotted lines show 1.5 standard deviations of the marginal distributions. Color indicates density of matchups.
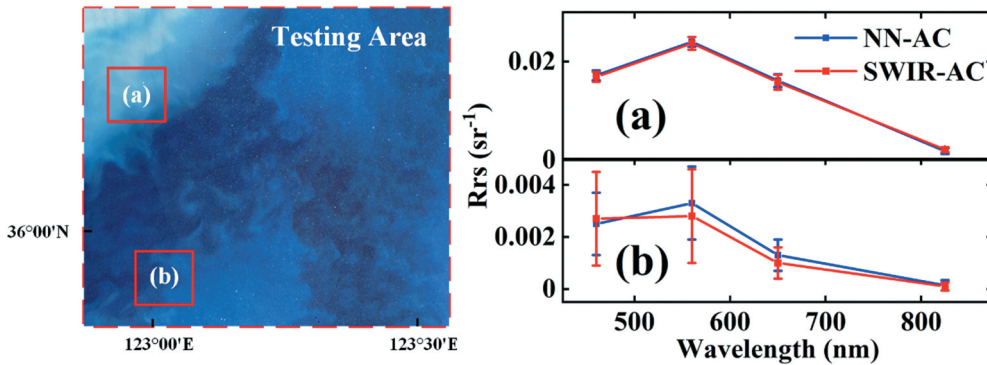

Figure 8. Comparison of R_{rs} spectra retrieved by NN-AC (CZI) and SWIR-AC (OLI) in water with different turbidity on 03/07/2019. (a) Turbid water and (b) clear water. The error bars represent the standard deviation of the red box.

approximately 40 nm in the central wavelength of the NIR band between the OLI and CZI data. The CZI's bandwidth is nearly twice than that of OLI.

3.4. Validation of the NN algorithm with in-situ data

The accuracy of the established algorithm was further validated using the multi-source *in-situ* data. The detailed information about the match-up selection has been previously presented in Section 2.3. Figure 10 shows a scatterplot of the R_{rs} values at 460, 560, 650, and 825 nm retrieved by the NN-AC algorithm and the

comparison with the *in-situ* data. Socheongcho and ARIAKE_TOWER data were not used for comparison at 825 nm since the absence of the corresponding band. The results are as follows: the RMSE values at 460, 560, 650 and 825 nm are 0.0071, 0.0107, 0.0064, and 0.0007, respectively; the R^2 values are 0.70, 0.77, 0.76, and 0.67, respectively; and the APD values are 25%, 22%, 33%, and 53%, respectively. Most of the data are centered around the 1:1 line for all four bands with no negative value obtained, suggesting that the NN-AC algorithm performs well in these bands. Moreover, the average R^2 values at the four bands in clear water (DongTou, MuPing, and Socheongcho data) are 0.73, 0.81, 0.78,
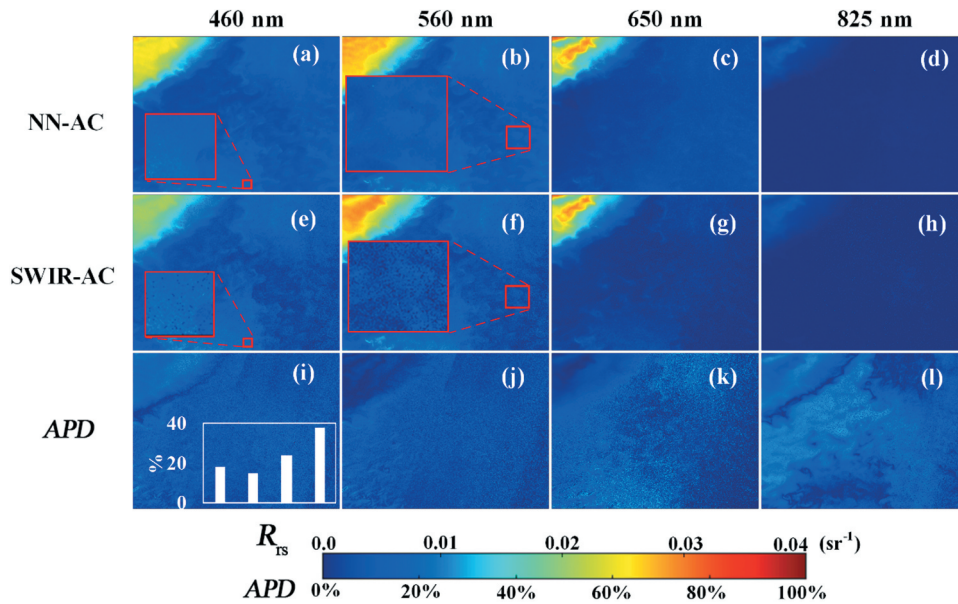
Figure 9. Comparison of retrieved R_{rs} values by NN-AC and SWIR-AC at the four CZI bands on 03/07/2019. (a)-(d) are NN-AC's results for CZI data; (e)-(h) are same as (a)-(d) but for SeaDAS SWIR-AC algorithm; (i)-(l) are APD of these two algorithms with mean APD histogram of the four bands in (i). The red boxes in (a)-(e) and (b)-(f) show AC failure with black points. Noted that the blue color in the images means the lower values and the red color means the higher values.
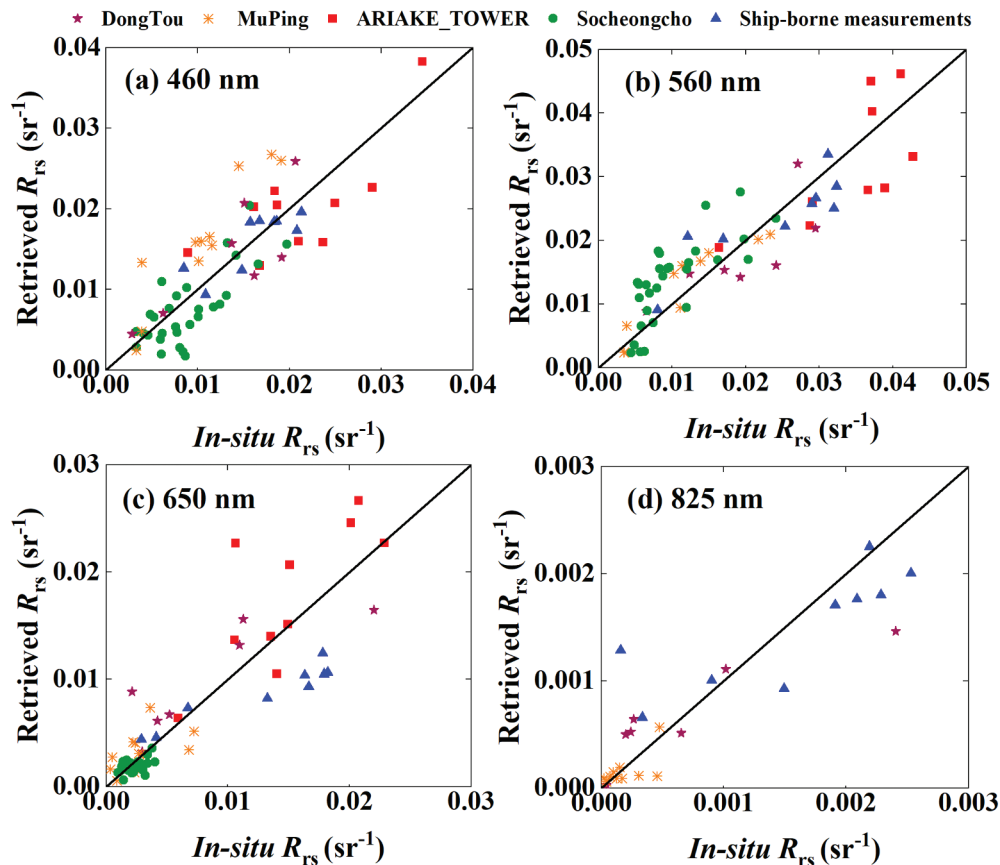

Figure 10. Comparison of CZI-retrieved R_{rs} by NN-AC algorithms with *in-situ* data at four CZI bands.

and 0.69, respectively, whereas those in turbid water (ARIAKE_TOWER and Ship-borne measurements) are 0.69, 0.70, 0.59, and 0.45, respectively. The results indicate that the method proposed in this study can

perform better in clear water and achieve competitive performance in turbid water because of the good consistency with the *in-situ* data, especially at the visible bands. More importantly, the results show that the

NN-AC algorithm can accomplish the AC for CZI in coastal waters regardless of the spatial or temporal difference.

The accuracy of the NN algorithm is limited by the original algorithm. This limitation suggests that the issue may be solved by improving the AC algorithm used for OLI images. However, a much higher precision training dataset is difficult to obtain, as complex algorithms are required to handle the AC of water with complex optical properties. As a probable solution, instead of training a single gigantic NN for handling different ocean conditions, the NN-AC can utilize a targeted dataset representing different types of atmospheric and marine conditions for model fine-tuning. Spatial location and data statistical analytic technique may be introduced when selecting the NN that can best match the marine conditions based on the geographical coordinates and the satellite measurements of ρ_{TOA} to yield the optimum retrieval results.

4. Conclusions

The implementation of AC of CZI images over coastal waters is challenging due to the lack of NIR and SWIR bands. In this study, we proposed a new AC algorithm based on the NN for CZI data to solve the problem. Different from the approach of traditional AC algorithms that use a small volume of *in-situ* data, we generated numerous high-quality simulated matchups from CZI and OLI images for the NN training.

The new NN algorithm was validated using the simulated dataset and the results showed that it could accurately and effectively learn the training dataset. Furthermore, the algorithm attained a similar spatial distribution pattern and even a higher spatial continuity of R_{rs} when evaluated with the retrieved R_{rs} from the OLI image by using SeaDAS. Then, we further validated the algorithm by multi-source *in-situ* data, and the results showed that the NN algorithm could accomplish the goal of AC for the CZI sensor in coastal waters regardless of the spatial or temporal difference. The NN algorithm runs very fast once the network has been properly trained, and is therefore, suitable for operational use.

Meanwhile, there are some deficiencies in this study. Our method can sufficiently simulate the results of SWIR-AC, but some errors still exist between retrieved R_{rs} and *in-situ* data. If the quality of the training data can be further improved in the future, then the performance of this algorithm can also be enhanced.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Notes on contributors

Jilin Men is currently pursuing his PhD degree in Photogrammetry and Remote Sensing at LIESMARS (State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing), Wuhan University. His research interests are ocean color remote sensing and atmospheric correction.

Jianqiang Liu received the MS degree in marine meteorology from the National Marine Environmental Forecasting Center, State Oceanic Administration, Beijing, China, in 1989. He is the Chief Designer of the ground application system for the Chinese-France ocean satellite and new generational ocean color satellites and was a Deputy Chief Designer of the ground application system for the HY-1 and HY-2 satellites. He is one of the founders of satellite ocean remote sensing in China and plays an important role in the development of Chinese Ocean Satellite and manned space flight. His research interests include ocean color and sea wave satellite data processing, applications of ocean remote sensing data and Antarctic research.

Guangping Xia is currently pursuing her MS degree in Photogrammetry and Remote Sensing at LIESMARS, Wuhan University. She is interested in ocean color remote sensing.

Tong Yue is currently pursuing her BS degree in spatial information and digital technology at Wuhan University. She is interested in ocean color remote sensing.

Ruqing Tong is currently pursuing her MS degree in Photogrammetry and Remote Sensing at LIESMARS, Wuhan University. She is interested in ocean color remote sensing.

Liqiao Tian is a professor in LIESMARS, Wuhan University. He received the MS and PhD degrees in Cartography and GIS from Wuhan University. His current research interest is the application of remote sensing.

Kohei Arai received the BS, MS, and PhD degrees from Nihon University in 1972, 1974, and 1978, respectively. He moved to Saga University as a Professor in the Department of Information Science in April 1990. Since 1998, he has been appointed as an Adjunct Professor of the University of Arizona, Tucson. He is Science Calibration Team Leader for the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) project. His major concerns are

atmospheric optics, vicarious calibration, nonlinear optimization theory, inverse problem solving, and pattern recognition/understanding.

Linyu Wang received the MS and BS degrees in Land Resource Management from China University of Geosciences (Wuhan), in 2014 and 2017. He is interested in the application of remote sensing.

ORCID

Jilin Men http://orcid.org/0000-0002-6973-2520

References

Brajard, J., C. Moulin, and S. Thiria. 2008. "Atmospheric Correction of SeaWiFS Ocean Color Imagery in the Presence of Absorbing Aerosols off the Indian Coast Using a Neuro-variational Method." *Geophysical Research Letters* 35 (20): L20604. doi:10.1029/2008GL035179.

Brajard, J., R. Santer, M. Crépon, and S. Thiria. 2012. "Atmospheric Correction of MERIS Data for Case-2 Waters Using a Neuro-variational Inversion." *Remote Sensing of Environment* 126: 51–61. doi:10.1016/j.rse.2012.07.004.

Chen, J., T. Cui, J. Ishizaka, and C. Lin. 2014. "A Neural Network Model for Remote Sensing of Diffuse Attenuation Coefficient in Global Oceanic and Coastal Waters: Exemplifying the Applicability of the Model to the Coastal Regions in Eastern China Seas." *Remote Sensing of Environment* 148: 168–177. doi:10.1016/j.rse.2014.02.019.

Cybenko, G. 1989. "Approximation by Superpositions of a Sigmoidal Function." *Mathematics of Control, Signals and Systems* 2 (4): 303–314. doi:10.1007/bf02551274.

D'Alimonte, D., and G. Zibordi. 2003. "Phytoplankton Determination in an Optically Complex Coastal Region Using a Multilayer Perceptron Neural Network." *IEEE Transactions on Geoscience and Remote Sensing* 41 (12): 2861–2868. doi:10.1109/TGRS.2003.817682.

D'Alimonte, D., G. Zibordi, and J.-F. Berthon. 2004. "Determination of CDOM and NPPM Absorption Coefficient Spectra from Coastal Water Remote Sensing Reflectance." *IEEE Transactions on Geoscience and Remote Sensing* 42 (8): 1770–1777. doi:10.1109/TGRS.2004.831444.

Fan, Y., W. Li, C.K. Gatebe, C. Jamet, G. Zibordi, T. Schroeder, and K. Stamnes. 2017. "Atmospheric Correction over Coastal Waters Using Multilayer Neural Networks." *Remote Sensing of Environment* 199: 218–240. doi:10.1016/j.rse.2017.07.016.

Franz, B.A., S.W. Bailey, N. Kuring, and P.J. Werdell. 2015. "Ocean Color Measurements with the Operational Land Imager on Landsat-8: Implementation and Evaluation in SeaDAS." *Journal of Applied Remote Sensing* 9 (1): 096070. doi:10.1117/1.JRS.9.096070.

Géron, A. 2019. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA, USA: O'Reilly Media.

Gordon, H.R. 1997. "Atmospheric Correction of Ocean Color Imagery in the Earth Observing System Era." *Journal of Geophysical Research: Atmospheres* 102 (D14): 17081–17106. doi:10.1029/96JD02443.

Gordon, H.R., and M. Wang. 1994. "Retrieval of Water-leaving Radiance and Aerosol Optical Thickness over the Oceans with SeaWiFS: A Preliminary Algorithm." *Applied Optics* 33 (3): 443–452. doi:10.1364/ao.33.000443.

Hamre, B., S. Stamnes, K. Stamnes, and J. Stamnes. 2013. "C-DISORT: A Versatile Tool for Radiative Transfer in Coupled Media like the Atmosphere-ocean System." Paper presented at the AIP Conference Proceedings, Kuala Lumpur, Malaysia, July 24-26.

Haykin, S., and N. Network. 2004. "A Comprehensive Foundation." *Neural Networks* 2 (2004): 41.

Hornik, K., M. Stinchcombe, and H. White. 1989. "Multilayer Feedforward Networks are Universal Approximators." *Neural Networks* 2 (5): 359–366. doi:10.1016/0893-6080(89)90020-8.

Jiang, L., and M. Wang. 2014. "Improved Near-infrared Ocean Reflectance Correction Algorithm for Satellite Ocean Color Data Processing." *Optics Express* 22 (18): 21657. doi:10.1364/OE.22.021657.

Jin, Z., and K. Stamnes. 1994. "Radiative Transfer in Nonuniformly Refracting Layered Media: Atmosphere–ocean System." *Applied Optics* 33 (3): 431. doi:10.1364/AO.33.000431.

Li, H., X. He, Y. Bai, P. Shanmugam, Y.-J. Park, J. Liu, Q. Zhu, F. Gong, D. Wang, and H. Huang. 2020. "Atmospheric Correction of Geostationary Satellite Ocean Color Data under High Solar Zenith Angles in Open Oceans." *Remote Sensing of Environment* 249: 112022. doi:10.1016/j.rse.2020.112022.

Liu, H., X. He, Q. Li, S. Kratzer, J. Wang, T. Shi, Z. Hu, et al. 2021. "Estimating Ultraviolet Reflectance from Visible Bands in Ocean Colour Remote Sensing." *Remote Sensing of Environment* 258:112404. doi:10.1016/j.rse.2021.112404.

Mueller, J.L., A. Morel, R. Frouin, C. Davis, R. Arnone, K. Carder, Z. Lee, et al. 2003. *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4*. Vol. III. Greenbelt, MD, USA: Radiometric Measurements and Data Analysis Protocols.

Nguyen, D., and B. Widrow. 1990. "Improving the Learning Speed of 2-layer Neural Networks by Choosing Initial Values of the Adaptive Weights." Paper presented at the 1990 IJCNN International Joint Conference on Neural Networks, San Diego, USA, June 17-21.

Pahlevan, N., J.R. Schott, B.A. Franz, G. Zibordi, B. Markham, S. Bailey, C.B. Schaaf, M. Ondrusek, S. Greb, and C. M. Strait. 2017. "Landsat 8 Remote Sensing Reflectance (Rrs) Products: Evaluations, Intercomparisons, and Enhancements." *Remote Sensing of Environment* 190: 289–301. doi:10.1016/j.rse.2016.12.030.

Pinkus, A. 1999. "Approximation Theory of the MLP Model in Neural Networks." *Acta Numerica* 8: 143–195. doi:10.1017/s0962492900002919.

Pinto, C., F. Ponzoni, R. Castro, L. Leigh, M. Kaewmanee, D. Aaron, and D. Helder. 2016. "Evaluation of the Uncertainty in the Spectral Band Adjustment Factor (SBAF) for Cross-calibration Using Monte Carlo Simulation." *Remote Sensing Letters* 7: 837–846. doi:10.1080/2150704X.2016.1190474.

Schroeder, T., I. Behnert, M. Schaale, J. Fischer, and R. Doerffer. 2007. "Atmospheric Correction Algorithm for MERIS above Case-2 Waters." *International Journal of Remote Sensing* 28 (7/8): 1469–1486. doi:10.1080/01431160600962574.

Wang, M., and W. Shi. 2007. "The NIR-SWIR Combined Atmospheric Correction Approach for MODIS Ocean Color Data Processing." *Optics Express* 15 (24): 15722–15733. doi:10.1364/oe.15.015722.

Wang, M., W. Shi, and L. Jiang. 2012. "Atmospheric Correction Using Near-infrared Bands for Satellite Ocean Color Data Processing in the Turbid Western Pacific Region." *Optics Express* 20 (2): 741–753. doi:10.1364/oe.20.000741.

Xu, Y., L. Feng, D. Zhao, and J. Lu. 2020. "Assessment of Landsat Atmospheric Correction Methods for Water Color Applications Using Global AERONET-OC Data." *International Journal of Applied Earth Observation and Geoinformation* 93: 102192. doi:10.1016/j.jag.2020.102192.

Zhang, K., W. Li, K. Stamnes, H. Eide, R. Spurr, and S.-C. Tsay. 2007. "Assessment of the Moderate-Resolution Imaging Spectroradiometer Algorithm for Retrieval of Aerosol Parameters over the Ocean." *Applied Optics* 46 (9): 1525–1534. doi:10.1364/ao.46.001525.

Zibordi, G., F. Mélin, J.-F. Berthon, B. Holben, I. Slutsker, D. Giles, D. D'Alimonte, D. Vandemark, H. Feng, and G. Schuster. 2009. "AERONET-OC: A Network for the Validation of Ocean Color Primary Products." *Journal of Atmospheric and Oceanic Technology* 26 (8): 1634–1651. doi:10.1175/2009JTECHO654.1.