



Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Analysis of coastal wind speed retrieval from CYGNSS mission using artificial neural network

Xiaohui Li^{a,b}, Dongkai Yang^a, Jingsong Yang^{b,c}, Gang Zheng^{b,c}, Guoqi Han^d, Yang Nan^e, Weiqiang Li^{f,g,*}

^a School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

^b State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

^c Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

^d Fisheries and Oceans Canada, Institute of Ocean Sciences, Sidney, BC V8L 4B2, Canada

^e GNSS Research Center, Wuhan University, Wuhan 430079, China

^f Institute of Space Sciences (ICE, CSIC), Barcelona 08193, Spain

^g Institut d'Estudis Espacials de Catalunya (IEEC), Barcelona 08034, Spain

ARTICLE INFO

Keywords:

Global navigation satellite system reflectometry (GNSS-R)
Cyclone GNSS (CYGNSS)
Sea surface wind speed
Coastal
Artificial neural network (ANN)

ABSTRACT

This paper demonstrates the capability and performance of sea surface wind speed retrieval in coastal regions (within 200 km away from the coastline) using spaceborne Global Navigation Satellite System Reflectometry (GNSS-R) data from NASA's Cyclone GNSS (CYGNSS) mission. The wind speed retrieval is based on the Artificial Neural Network (ANN). A feedforward neural network is trained with the collocated CYGNSS Level 1B (version 2.1) observables and the wind speed from European Centre for Medium-range Weather Forecast Reanalysis 5th Generation (ECMWF ERA5) data in coastal regions. An ANN model with five hidden layers and 200 neurons in each layer has been constructed and applied to the validation set for wind speed retrieval. The proposed ANN model achieves good wind speed retrieval performance in coastal regions with a bias of -0.03 m/s and a RMSE of 1.58 m/s, corresponding to an improvement of 24.4% compared to the CYGNSS Level 2 (version 2.1) wind speed product. The ANN based retrievals are also compared to the ground truth measurements from the National Data Buoy Center (NDBC) buoys, which shows a bias of -0.44 m/s and a RMSE of 1.86 m/s. Moreover, the sensitivities of the wind speed retrieval performance to different input parameters have been analyzed. Among others, the geolocation of the specular point and the swell height can provide significant contribution to the wind speed retrieval, which can provide useful reference for more generic GNSS-R wind speed retrieval algorithms in coastal regions.

1. Introduction

Sea surface wind is an essential variable in both marine environment monitoring and climate change study (Huang et al., 2003; Peng and Jin, 2019). The stability of the wind field plays an important role in local and mesoscale atmospheric circulation, and the frequency of different stability conditions in coastal areas is very important information (Barthelmie, 1999). Sea surface wind is the direct power of ocean wave formation and the key power of regional and global ocean circulation (Kirincich, 2016). In the coastal area, the sea surface wind field has a strong influence on the circulation, fog formation, coastal upwelling, and tidal mixing, which also influence stability and hence turbulent

mixing and momentum transfer (Kaimal and Finnigan, 1994). In addition, the global offshore wind farms also bring us renewable wind power resources. Wind power generation can reduce environmental pollution and save coal, oil and other conventional energy (Kirincich, 2016). Traditional observing techniques (e.g. weather stations, buoys and ships) have been playing important roles in marine forecasting, disaster prevention and mitigation. However, these traditional techniques are limited by small coverage and high cost (Tang et al., 2014).

Satellite remote sensing, both active (radar, scatterometer and altimeter) and passive (radiometer), has been providing significant understanding of ocean by providing global observations (Tang et al., 2014; Li et al., 2020). During the last decades, Global navigation satellite

* Corresponding author at: Institute of Space Sciences (ICE, CSIC), Barcelona 08193, Spain.

E-mail addresses: xh_lee@buaa.edu.cn (X. Li), dkyang@buaa.edu.cn (D. Yang), jsyang@sio.org.cn (J. Yang), zhenggang@sio.org.cn (G. Zheng), Guoqi.Han@dfmpo.gc.ca (G. Han), nanyang@whu.edu.cn (Y. Nan), weiqiang@ice.csic.es (W. Li).

<https://doi.org/10.1016/j.rse.2021.112454>

Received 22 November 2020; Received in revised form 8 March 2021; Accepted 11 April 2021

Available online 21 April 2021

0034-4257/© 2021 Elsevier Inc. All rights reserved.

system reflectometry (GNSS-R) has become a valid option for ocean remote sensing (Garrison and Katzberg, 2000). The feasibilities of the GNSS-R technique for sensing different geophysical parameters, such as ocean wind speed (Clarizia and Ruf, 2016; Ruf and Balasubramaniam, 2018), significant wave height (SWH; Soulat et al., 2004; Roggenbuck et al., 2019), sea ice detection (Alonso-Arroyo et al., 2017; Yan and Huang, 2018), ocean altimetry (Li et al., 2017; Hu et al., 2020; Tabibi et al., 2020), and soil moisture (Alonso Arroyo et al., 2014; Chew and Small, 2018; Yan and Huang, 2020), have been demonstrated with theoretical analyses and massive ground-based, airborne, and spaceborne experiments.

Among other applications, GNSS-R ocean scatterometry has great potential to become an operational one. The reflected GPS signal was successfully received from space by the GPS receiver onboard UK's disaster monitoring constellation (DMC) satellite in 2003 (Gleason et al., 2005), which demonstrated the feasibility of spaceborne GNSS-R ocean scatterometry (Zuffada et al., 2003). Following the successful detection of reflected signals by UK-DMC, the UK TechDemoSat-1 (TDS-1) was launched in 2014 (Unwin et al., 2017), which delivered a unique dataset of globally distributed spaceborne GNSS-R data, spanning a period of four years. With the data collected by the TDS-1 satellite, the first global ocean wind and wave data service derived from spaceborne GNSS-R has been publicly available (Foti et al., 2015; Hammond et al., 2020). On December 15, 2016, NASA's Cyclone GNSS (CYGNSS) constellation, consisting of 8 micro-satellites, were launched with the specific aim of observing the inner core of cyclones with high temporal resolution in precipitating conditions (Ruf et al., 2013; Ruf et al., 2016b). In addition, China has also launched the Bufeng-1 A/B satellites in 2019, which is part of the first Chinese GNSS-R satellite constellation for measuring global sea surface wind field and typhoon monitoring (Jing et al., 2019).

With the accumulation of the spaceborne GNSS-R data from these spaceborne missions, different wind speed retrieval algorithms have been developed to improve the wind speed retrieval performance (Clarizia et al., 2014; Clarizia and Ruf, 2016; Clarizia et al., 2017; Ruf et al., 2018; Ruf and Balasubramaniam, 2018). The performance of the CYGNSS Level 2 (L2) baseline wind speed retrieval has been improved gradually. The root mean square error (RMSE) of the wind speed retrieval has been improved to 1.4 m/s for low-to-moderate wind speed (<20 m/s), and the RMSE for high wind speed (>20 m/s) retrieval has been improved from 17% of the wind speed (Ruf et al., 2018) to 11.3% (Ruf et al., 2019). Note that the CYGNSS wind speed error at high wind speed is normally presented as the relative RMSE in percentage from the mean values and RMSE of the retrieved wind speeds (Ruf et al., 2018, Eq. (4)). However, there are few studies on GNSS-R wind speed retrieval around the coastal regions. Different from GNSS-R measurements over the open ocean, the reflected signals from the coastal regions are strongly affected by the land contamination, land terrain properties (Carreno-Luengo et al., 2019), and the near-shore shallow water. Table 1 shows the performance of CYGNSS L2 version 2.1 (v2.1) wind speed product, which clearly shows a performance degradation around the

coastal regions (within 200 km from the coastline) compared to the open ocean (i.e. ~2.1 vs. 1.88 m/s). Note that the data within 25 km from the land has been removed in CYGNSS L2 v2.1 retrieval, which is not included in the computation of the performance statistics in Table 1. Indeed, the near-shore shallow water can significantly reduce the sea surface roughness response to the wind speed, which could make a transition from a strongly diffuse scattering regime to a dominantly or partially coherent reflection regime in coastal regions (Cardellach et al., 2020; Al-Khaldi et al., 2020). As a result, in essence all of the underlying assumptions associated with the conversion of CYGNSS's observables to ocean surface wind speed estimates are no longer valid thereby leading to this degradation.

Most of the spaceborne GNSS-R wind retrievals are based on empirical or semi-empirical geophysical model function (GMF) methods, i.e. by fitting the observables [such as σ_0 , leading edge slope (LES), or signal-to-noise ratio (SNR)] to collocated wind speed (Foti et al., 2015; Hammond et al., 2020). However, the relationship between wind speed and GNSS-R observables can be also affected by other factors, such as the wave age and fetch length. For example, the CYGNSS L2 wind speed retrieval is based on two separate GMFs, corresponding to fully developed and young sea/limited fetch regimes. For coastal regions, the response of the ocean wave to sea surface wind is affected by even more factors, such as the topography, orientation of the coastline, prevailing monsoon, coastal ocean currents and limited fetch length (Barthelmie, 1999). These factors make the relationship between the sea surface wind speed and the GNSS-R observations more complex, which is difficult to be characterized with a simple GMF.

Machine learning, especially the artificial neural network (ANN), can make full use of the self-learning and self-adaptive ability of neural cells to deal with complex nonlinear problems, and thus can be used for GNSS-R wind speed retrieval by considering the impacts of different geophysical and geographical parameters. ANN is a nonlinear and adaptive information processing system, consisting of a number of interconnected neural node(s), which can solve problems by providing a minimum variance (Kwok and Yeung, 1997). ANNs have been widely applied in global wind speed retrieval from scatterometer and synthetic-aperture radar (SAR; Hornik, 1991; Stiles and Dunbar, 2010; Stiles et al., 2014). More recently, several studies have shown that ANNs can also improve the accuracy of GNSS-R wind speed retrieval using ground-based and spaceborne data (Kasantikul et al., 2018; Liu et al., 2019; Gao et al., 2019a; Asgarimehr et al., 2019), which have shown promising performance by using data collected by the TDS-1 (Wang et al., 2018; Asgarimehr et al., 2019) and CYGNSS missions (Liu et al., 2019; Reynolds et al., 2020). Moreover, this approach has been also attempted in some other GNSS-R applications, such as sea ice detection (Yan and Huang, 2018), soil moisture (Feng et al., 2018; Eroglu et al., 2019), hurricane tracking (Alshaye et al., 2020), and inland water detection (Ghasemigoudarzi et al., 2020).

As the ANN method has the advantages in establishing multi-parameter model, it can provide powerful alternatives for GNSS-R wind speed retrieval in coastal regions. This work aims to investigate the feasibility and performance of coastal wind retrieval based on ANN using the data collected by the CYGNSS mission. The rest of the paper is organized as follows. In Section 2, we describe the datasets and pre-processing methods. Section 3 explains the setup and configuration of the ANN model, together with the sensitivity analyses of the wind speed retrieval to different input parameters. The performances of the proposed ANN model are evaluated with the European Centre for Medium-range Weather Forecast Reanalysis 5th Generation (ERA5) wind speed data and the National Data Buoy Center (NDBC) buoys wind speed measurements, which are presented in Section 4. And Section 5 draws the conclusions.

Table 1

The statistics of CYGNSS Level 2 version 2.1 wind speed product compared to the European Centre for Medium-range Weather Forecast Reanalysis 5th Generation (ECMWF ERA5) values in January 2018. Note, the Bias, Root-mean-square error (RMSE) and the correlation coefficient (R) are computed following Eqs. (1), (2) and (3), respectively.

	Global	25–50 km	50–100 km	100–200 km	200–300 km	300–500 km
Bias (m/s)	0.21	0.01	0.27	0.31	0.20	0.32
RMSE (m/s)	1.91	2.08	2.20	2.17	1.87	1.89
R	0.80	0.80	0.80	0.80	0.80	0.80

2. Materials and data filtering

2.1. Datasets

CYGNSS is a constellation of eight small satellites, which can receive both direct and reflected signals from Global Positioning System (GPS) satellites. The direct signal can accurately locate the positions of the CYGNSS observatories. The reflected signal carries the information of ocean surface roughness, from which wind speed can be retrieved. CYGNSS measurements can provide rapid coverage over the tropical regions with mean and median revisit times of 7.2 and 2.8 hours over $0.25^\circ \times 0.25^\circ$ latitude-longitude grids (Morris and Ruf, 2017). CYGNSS Level 1B (L1B) product includes delay-Doppler map (DDM) power of the reflected signal, the calibrated Normalized Bistatic Radar Cross Section (NBRCS), together with other ancillary information and metadata (Gleason et al., 2019). These L1B observables are then inverted to the Level 2 wind speed products using an empirical GMF (Ruf and Balasubramaniam, 2018).

In this work, we focus on the CYGNSS L1B v2.1 product, which is available at the Physical Oceanography Distributed Active Archive Center (PO.DAAC; Ruf et al., 2016a). Table 2 lists the brief information about the L1B variables used in our analyses, which include NBRCS, LES, SNR, Range Corrected Gain (RCG), incidence angle, azimuth angle, and the longitude and latitude of the specular point. In addition to the L1B data, the CYGNSS L2 v2.1 product is also collected for comparison, which includes the sea surface wind speed measurements with a spatial resolution of 25 km (Ruf et al., 2019).

As the proposed ANN method is based on training the retrieval model with the GNSS-R measurements and the ground true wind speed, ECMWF/Copernicus Climate Change Service (C3S) ERA5 wind fields (including the u and v components of the sea surface wind speed), at 12.5 km and hourly spatio-temporal resolutions, have been also collected as the reference.

2.2. Data filtering

The study covers the period between Jan. 1 and Dec. 31, 2018. By using the global coastline derived from global self-consistent, hierarchical, high-resolution geography database (GSHHG; Wessel and Smith, 1996), the distance between each CYGNSS L1 measurement and the coastline has been computed. Only the CYGNSS observables within 200 km away from the coastline have been selected.

To ensure the quality of the CYGNSS data, the selected coastal L1B measurements are quality controlled and filtered before the ANN training. The initial quality control is mainly based on the following criteria: (Ruf and Balasubramaniam, 2018; Reynolds et al., 2020; Clarizia and Ruf, 2020).

- 1) All NaN values of observables are discarded.
- 2) All negative values of observables are discarded.
- 3) The RCG values should be higher than 3.

In addition, the standard quality control (QC) flags in the CYGNSS L1B data are also used in the data filtering. Instead of using the overall

Table 2
List of input variables used in CYGNSS coastal wind speed retrieval.

Input variables	Description	Type
NBRCS	Normalized bistatic radar cross section	Surface attribute
LES	Leading edge slope	Surface attribute
SNR	DDM signal to noise ratio	Surface attribute
sp_lat	Specular point latitude	Geometry attribute
sp_lon	Specular point longitude	Geometry attribute
sp_inc_angle	Specular point incidence angle	Geometry attribute
sp_az_body	Specular point azimuth angle	Geometry attribute
RCG	Range Corrected Gain	Geometry attribute

QC flag bit (“*poor_overall_quality*”, the least significant bit or bit 0), the data filtering is applied by using the individual QC flag bits. This discards cases in which the spacecraft has large attitude error or attitude anomalies (bit 2, 3 and 28), the transmitter power or antenna gain has high uncertainties (bit 16 and 27), the DDM is invalid or has anomalies (e.g. black body load and test pattern DDMs indicated by bit 4, 7, 8, 9, 10, 15, 17–20), or there are some other instrumental, data transmission and calibration issues (bit 1, 5, 6, 13, 14, 21–26). It is noted that the L1B observables from GPS IIF satellites are removed due to the QC flag “*low_quality_gps_ant_knowledge*” (bit 27). Moreover, as the QC flags “*sp_very_near_land*” (bit 11) and “*sp_near_land*” (bit 12) are not used in the data filtering, there remain the L1B observables with the specular point within 50 km of land.

After the quality control and data filtering, the remaining CYGNSS L1B measurements are then collocated to the ERA5 sea surface wind speed and ocean wave data, which is performed with the constraint of the maximum spatial and temporal differences of 12.5 km and 15 min, respectively. There are ~9.2 million collocated CYGNSS L1B measurements around the coastal region, which are shown in Fig. 1. It can be seen that there are more measurements in the Gulf of Mexico and the other low latitude area (near $30^\circ\text{N}/^\circ\text{S}$), and no measurement can be obtained in the Arctic Ocean or the sea around Antarctica, which is due to the orbit inclination of the CYGNSS satellites ($\sim 35^\circ$).

3. Construction and analyses of the ANN based wind speed retrieval model

In this study, the feed-forward backpropagation (BP) network is implemented based on the Keras framework, which generally consists of input layer, hidden layer, and output layer. Neural network can learn the relationship between input and output parameters without providing any analytical equations (Dayhoff and DeLeo, 2001). Its learning rule is to use the gradient steepest descent method to continuously adjust the weight and threshold of the network through error back propagation and minimize the square error of the network. Overfitting or underfitting is a common potential problem in the application of neural network (Reynolds et al., 2020). Increasing the amount of data is one of the most effective methods. Moreover, on the basis of the early stopping rule, validation can be powerful for overfitting after each training the network, training stops when the early stopping rule is reached (Kwok and Yeung, 1997). Three metrics are chosen to evaluate the performance of the ANN model, including the mean deviation (or bias), the root mean square error (RMSE) and the correlation coefficient (R), which are defined as

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3)$$

where X is CYGNSS retrieved wind speed, Y is ERA5 wind speed, n is the number of total matchups, \bar{X} is the mean of the CYGNSS retrieved wind speed, and \bar{Y} is the mean of the ERA5 wind speed.

3.1. Determining the ANN configuration

3.1.1. Basic setting

- 1) Activation Function: The activation function can be any differentiable activation function such as Tanh, sigmoid, Rectified Linear

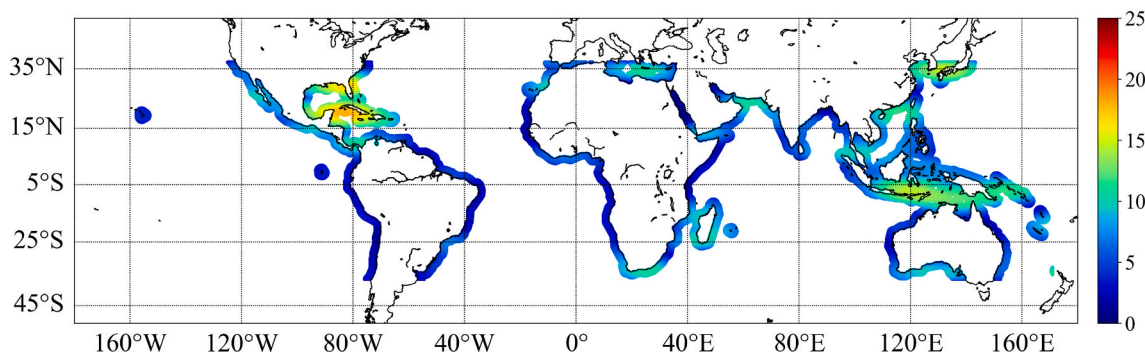


Fig. 1. Distribution of CYGNSS-ERA5 matchups in coastal regions in 2018. The color scale is 1/10000 of the density of points.

Unit. We chose the Tanh function as the activation function between the input layers and the hidden layers. This setting is also used in other studies, e.g. in (Gao et al., 2019b).

- 2) Loss Function: The loss function is an important parameter for compiling the model. In this paper, the mean squared error function was used as the loss function.
- 3) Optimization Method: Optimizer is another important configuration of the training model. After comparison, we choose the Adam optimization algorithm. Adam is an update of the RMSProp optimizer, which is like RMSprop with momentum. When training the neural networks, it is helpful to set a suitable learning rate. In this paper, we implemented the adaptive learning algorithm by using the Adam optimizer. The neural network training times is set a constant of 100. The batch size for the neural network training is 1000.

3.1.2. Implementation of ANN

The selection of the CYGNSS and ERA5 matchups for ANN training and validation is shown in Fig. 2. To train the ANN model, 15% of the total matchups obtained in Section 2 are selected randomly. As the performance of the neural network training is affected the quality and spatial/temporal difference of the training data sets, we further refine the matchups with the time difference between the CYGNSS measurements and the ERA5 data smaller than 5 min. The training dataset is further split randomly into a training subset and a test subset, composed of 15% and 85% of the training data, respectively. The first subset is used for training network, and the test subset serves to assess the performance of each network configuration. In the model training stage, the number of hidden layers and the number of neurons in each layer are specified. The configuration with the best performance is selected for the final implementation.

Here we test different ANN configurations with the number of hidden

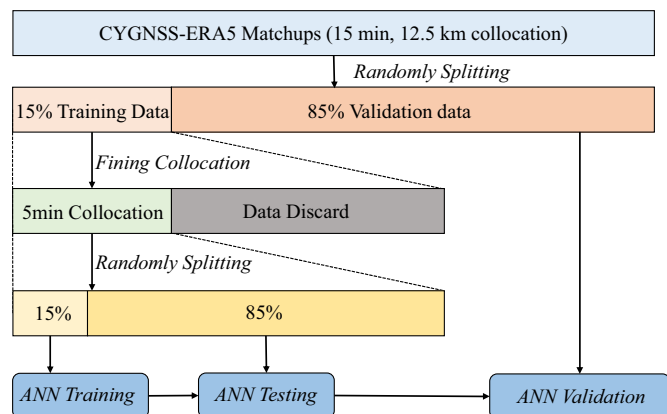


Fig. 2. Subsets selection of the CYGNSS-ERA5 matchups for ANN training and validation.

layers from 1 to 6, and with 10 to 200 neurons in each hidden layer. For each ANN configuration, the network structure is optimized with the above-mentioned training data following the parameters in Section 3.1.1. These ANN configurations are then applied to the CYGNSS observables in the testing data to retrieve the wind speed. The resulting wind speed retrievals are then compared to the collocated ERA5 values, from which the RMSE for each ANN configuration is computed. Fig. 3 gives the RMSEs of each ANN configuration, which allows us to make a detailed evaluation of different networks configurations. It shows that the performance of ANN model with multi-layer is better than that with only one hidden layer. The RMSE of one hidden layer is close to 1.64 m/s, which basically remains unchanged even by increasing number of neurons up to 200. Overall, the trained ANN structure with five hidden layers and 170 neurons in each layer can reach the optimal performance with a RMSE of ~1.49 m/s.

Sea state conditions, including the swell and the degree of wave development, complicate the ocean surface wave spectra, and increase the uncertainty of the GNSS-R wind speed retrieval. For example, the CYGNSS standard L2 wind speed retrieval is based on two separate GMFs (fully developed and young sea) by considering the development of the ocean wave (Ruf and Balasubramaniam, 2018); and NOAA’s GMF for CYGNSS wind speed retrieval (Said et al., 2019) includes the SWH as the input parameter. Similarly, the information of wave height has been also used as the prior knowledge for sea surface wind speeds retrieval in radar altimeter (Chelton and McCabe, 1985; Glazman and Greysukh, 1993). In the same way, the contribution of swell information in neural network-based wind speed retrieval is also evaluated in this study. To decouple the extra wind speed information from the wave height data, the ERA5 variable “significant wave height of total swell” (SWHTS) is added as the input parameter.

To characterize the impact of the wave height on the ANN based retrieval, an input parameter combination, including NBRCS, LES, SNR, RCG, incidence angle, azimuth angle, longitude and latitude of the specular point and SWHTS, is attempted. By using different ANN configurations, the networks are trained by using the same training dataset. As shown in Fig. 3, the RMSE decrease with increasing the numbers of layers and neurons, especially the network of five hidden layers with 200 neurons in each layer can reach better performances with good correlation (0.88) and low RMSE (1.41 m/s) between the retrieved wind speeds and validation ones. Compared to the case without the SWHTS information, the new network including the swell wave height can provide better retrieval performance with the RMSE decreased from 1.49 m/s to 1.41 m/s (~6%). The comparison also highlights that the ANN approach can capture the nonlinear relationship between multi-observables and ERA5 winds.

3.2. Sensitivity analyses

To further assess the contribution of different input parameters in ANN-based wind speed retrieval, different combinations of the input

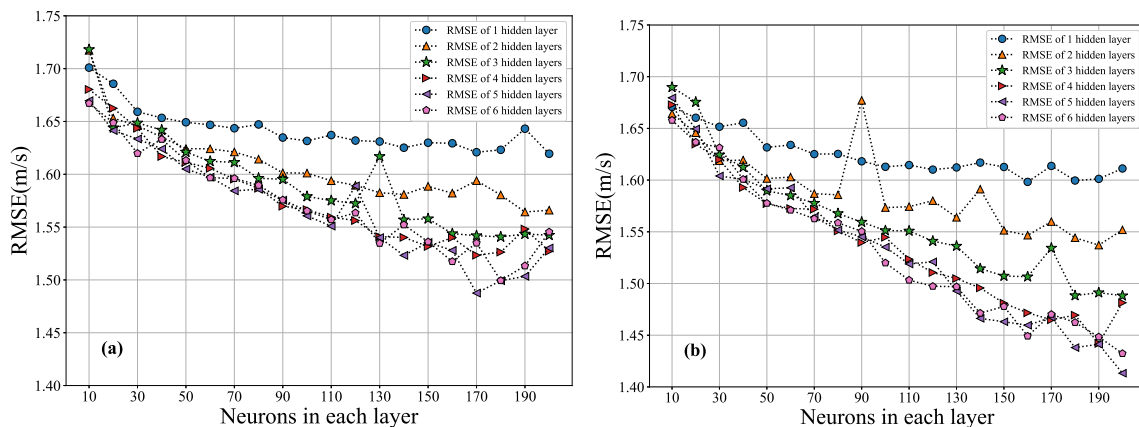


Fig. 3. RMSEs of the wind speed retrieval by using ANN models with different number of layers and number of neurons in each layer. (a) Without the wave height information as the input parameter. (b) With the wave height information as the input parameter.

parameters were designed as follow:

- Case 1: NBRCS, LES, SNR (Baseline configuration).
- Case 2: NBRCS.
- Case 3: LES.
- Case 4: NBRCS, LES, SNR, longitude and latitude.
- Case 5: NBRCS, LES, SNR, incidence angle.
- Case 6: NBRCS, LES, SNR, RCG.
- Case 7: NBRCS, LES, SNR, azimuth angle.
- Case 8: NBRCS, LES, SNR, SWHTS.
- Case 9: NBRCS, LES, SNR, RCG, incidence angle, azimuth angle, longitude and latitude.
- Case 10: NBRCS, LES, SNR, RCG, incidence angle, azimuth angle, longitude and latitude, SWHTS.
- Case 11: NBRCS, SNR, RCG, incidence angle, azimuth angle, longitude and latitude, SWHTS.

The ANN model is applied to the testing data with different input parameters combinations, and the retrieved wind speed are then compared to the ERA5 values, from which the performance metrics are computed. Table 3 compiles the wind speed retrieval performance for each input parameters combination. The ANN configuration with the input parameters of NBRCS, LES and SNR is considered as the baseline configuration, as these parameters are normally used in other wind speed retrieval methods. From the comparison, the following aspects are remarkable.

First of all, the CYGNSS NBRCS observable provides the main contribution to the wind speed retrieval, which is clearly justified by the fact that the RMSE is only slightly increased by adding the SNR and LES observables (Case 2 vs. Case 1, and Case 11 vs. Case 10).

Secondly, the comparisons also highlight that the RMSE of the ANN-based wind speed retrieval with the baseline combination is better than that obtained with the official CYGNSS L2 v2.1 baseline retrieval in the coastal regions (1.74 m/s in Case 1 vs. 2.09 m/s in L2 v2.1). Such improvement implies that ANN has the advantages in nonlinear regression problems relative to traditional methods (Kwok and Yeung,

1997).

Thirdly, it shows that the geographical information of the specular point can provide positive contribution with the correlation coefficient increased by 3.7% and RMSE reduced by 7.4% (Case 4 vs. Case 2), which is consistent with the results obtained in (Reynolds et al., 2020). Such improvement can be explained by the fact that geographic location of the specular point can provide extra information, such as the coastal topography and the orientation of the coastline, which may also affect the relationship between the wind speed and the CYGNSS observables.

In addition, specular point incidence angle can also provide contribution to the wind speed retrieval with RMSE reduced by 3.4% (Case 5 vs. Case 1). While RCG, azimuth angle and swell height can only provide slightly improvements, if each of them is added separately to the baseline case (Case 6, 7 and 8). However, by adding all parameters as the inputs of the ANN (Case 10), the RMSE of the retrieved wind speed can improved to 1.41 m/s, corresponding to an improvement of 19% comparing with baseline one. It is remarked that the wind speed retrieval performance can be improved significantly (5.2%) by adding the information of swell height into the input parameters (Case 10 vs. Case 9), which is consistent with the analysis in Section 3.1.2.

4. Wind speed retrieval and performance analyses

In this section, we evaluate the wind speed retrieval performance of the proposed ANN model with five hidden layers and 200 neurons in each layer. To achieve an optimal wind speed retrieval performance, all the parameters listed in Case 10 in Section 3.2 are used as the inputs of the ANN model. The proposed ANN model is applied to the rest of CYGNSS data (85%) for validation. The retrieved wind speeds are then compared to the collocated ERA5 wind speed data and the wind speed measured from the National Data Buoy Center (NDBC) buoys, from which the main performance metrics are computed.

Table 3

The performance metrics of each input parameter combination for the ANN based wind speed retrieval. The performance metrics of the official CYGNSS Level 2 version 2.1 wind retrieval is computed by using the data collected in year 2018. Note that the differences and improvements are relative to the case 1.

	L2	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Case 11
Bias(m/s)	0.395	0.255	0.166	0.102	0.040	-0.129	-0.055	0.176	-0.083	0.059	0.010	0.011
R	0.793	0.81	0.80	0.73	0.83	0.82	0.81	0.81	0.82	0.86	0.88	0.88
RMSE(m/s)	2.094	1.74	1.75	2.02	1.62	1.68	1.72	1.71	1.71	1.50	1.41	1.44
Difference	-	-	-	-	0.12	0.06	0.02	0.03	0.03	0.24	0.33	0.30
Improvement	-	-	-	-	6.9%	3.4%	1.1%	1.7%	1.7%	13.8%	19.0%	17.2%

4.1. Validation with the ERA5 data

4.1.1. Overall performance

The wind speed retrieved from the validation data is compared to the collocated ERA5 data in a density scatter plot as shown in Fig. 4, in which the CYGNSS L2 v2.1 wind speed is also compared to the ERA5 data as the reference. It can be seen that the ANN model retrieved wind speed shows an overall better performance than the CYGNSS L2 v2.1 wind speed product with more data point symmetrically centered along the $y = x$ line and less data spreading around this line.

In Table 4, the performance metrics of both retrievals are presented for each month in 2018, from which no significant temporal dependence can be found for both retrievals. However, ANN generally shows better performance than the CYGNSS L2 baseline retrieval for each month. By using all the matchups through the year, the ANN-based wind speed retrieval can achieve a bias of -0.03 m/s and a RMSE of 1.58 m/s, while the CYGNSS L2 wind speed product is with a bias of 0.37 m/s and a RMSE of 2.09 m/s. This comparison shows that the RMSE can be reduced by 24.4% with the ANN-based retrieval. Note that the CYGNSS baseline wind speed retrieval removes the measurements within 25 km away from the land, which are not included in this comparison. While the mean bias and the RMSE of ANN retrieved wind speed is 0.012 m/s and 1.68 m/s in this region.

Fig. 5 (left) presents the probability density functions (PDF) of the GNSS-R wind speeds and the ERA5 data in the range of 0 to 20 m/s, which shows that the PDF of ANN based coastal wind is in general consistent with the ERA5 one. While the PDF peak of the CYGNSS L2 v2.1 wind speed is lower than that of ERA5 winds. On the contrary, PDF peak of the ANN-based wind speed is slightly higher than the ERA5 one. In addition, the comparison also highlights that the probability of ANN wind speed is significantly lower than the ERA5 one between 0 and 2 m/s, which means that the ANN model exhibits an underestimation in lower wind speed conditions. This is mainly due to the lack of low wind speed samples in the training dataset so that ANN is not able to learn sufficiently from them. The ANN can be further optimized by improving the number of wind speeds at different ranks. Moreover, our ANN model is trained only with ERA5 data, while the GFM of CYGNSS L2 v2.1 winds is fitted by both ERA5 and the Global Data Assimilation System (GDAS) matchups, which could be another reason for the deviation of the CYGNSS L2 v2.1 products distribution. The consistency between the PDF of the ANN retrieval and the ERA5 wind speed can be further improved by applying a cumulative distribution function (CDF) matching algorithm to the ANN outputs. Such algorithm has been applied to both

Table 4

The performance metrics of CYGNSS ANN based coastal wind speed retrieval and the CYGNSS Level 2 version 2.1 wind speed product for each month in 2018. The ECMWF/C3S ERA5 wind speed data is used as the reference for the computation of these performance metrics.

	ANN retrievals			CYGNSS L2 v2.1		
	Bias (m/s)	RMSE (m/s)	R	Bias(m/s)	RMSE(m/s)	R
Jan.	-0.28	1.74	0.83	0.16	2.37	0.80
Feb.	-0.18	1.60	0.83	0.15	1.99	0.80
Mar.	-0.13	1.57	0.85	0.20	2.00	0.80
Apr.	0.04	1.55	0.83	0.46	2.32	0.76
May	-0.11	1.57	0.85	0.18	1.93	0.81
Jun.	-0.18	1.59	0.84	0.17	1.92	0.80
Jul.	-0.16	1.61	0.84	0.17	1.98	0.81
Aug.	-0.12	1.61	0.84	0.30	1.98	0.80
Sept.	0.09	1.54	0.85	0.48	2.02	0.81
Oct.	0.20	1.53	0.84	0.60	2.06	0.81
Nov.	0.27	1.55	0.84	0.77	2.18	0.80
Dec.	0.18	1.62	0.85	0.84	2.30	0.81
Average	-0.03	1.58	0.84	0.37	2.09	0.80

GMF-based and ANN-based CYGNSS wind speed retrieval (Clarizia and Ruf, 2020; Reynolds et al., 2020) with improved performances. Fig. 5 (right) presents the distribution histogram of the deviation between the CYGNSS ANN retrievals relative to the ERA5 values, which shows that the difference is concentrated between -3 m/s and 3 m/s (about double standard deviation of entire difference set).

4.1.2. Geographical statistics

Fig. 6 presents the wind speed retrieval performance (the RMSE of the wind speed) in the regions with different distance to the coastline. In general, the ANN based wind speed retrieval shows better performance away from the land, which achieves the best performance in the range of 75–150 km away from the coastline. The evolution of the RMSE also reveals that the GNSS-R observables within 10 km are strongly affected by the land contamination and near-shore shallow water, so that the retrieval performance is significantly degraded with a RMSE of ~ 1.76 m/s. Fig. 6 also highlights that the wind speed retrieval performance also degrades in the region 175–200 km away from the land (closer to the open ocean), which could be due to the different wind stress and ocean wave generation mechanism in the open ocean and coastal regions.

Fig. 7 shows that the global distribution of the wind speed bias

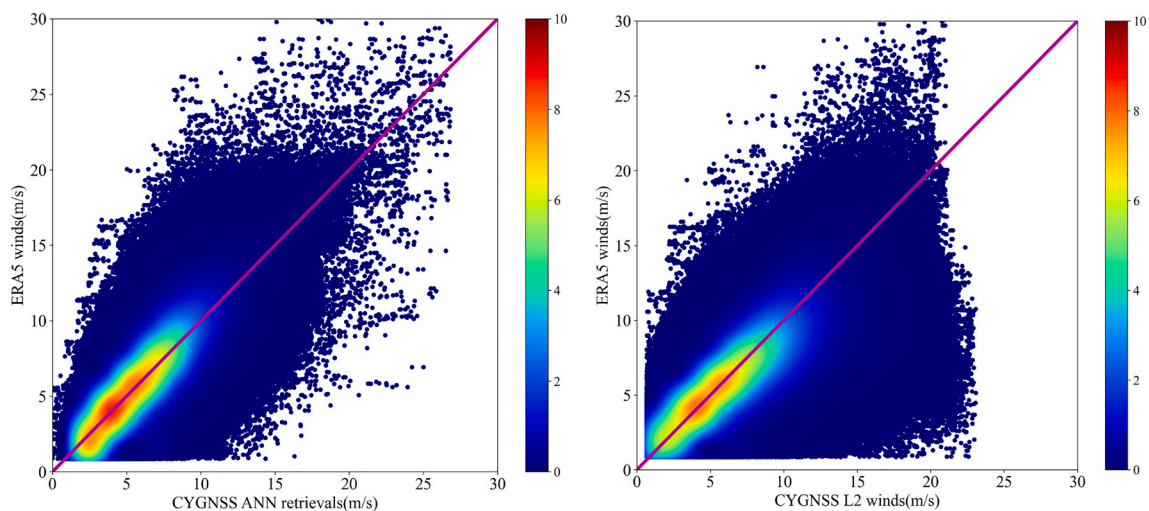


Fig. 4. Comparison between the CYGNSS ANN retrieved wind speed and the ECMWF/C3S ERA5 wind speed values (left). Comparison between the CYGNSS baseline L2 v2.1 wind speed and the ECMWF/C3S ERA5 wind speed values (right). The color scale is $1/100000$ of the density of points. The purple line shows the 1:1 diagonal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

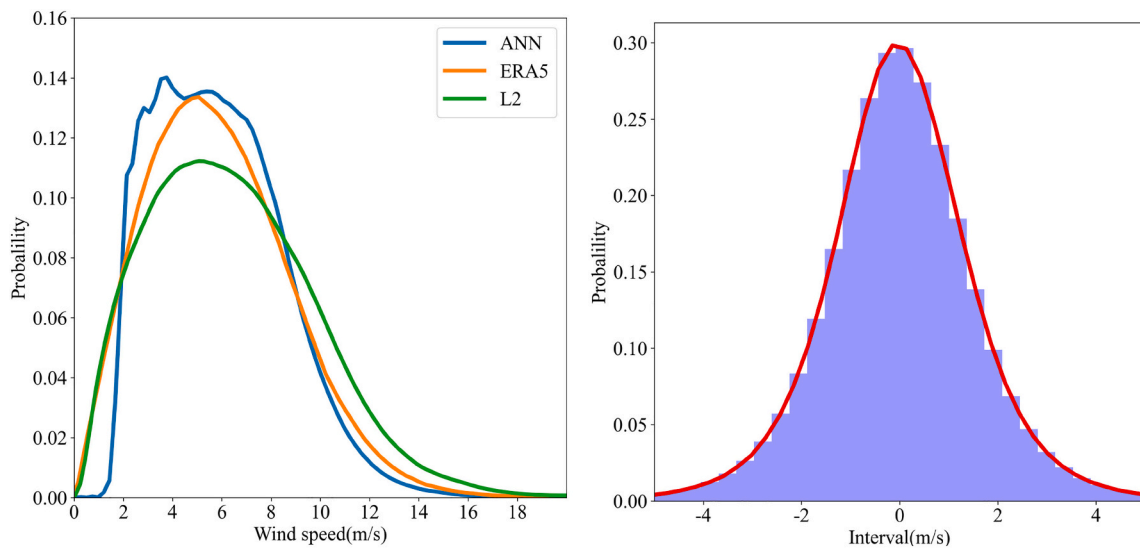


Fig. 5. Probability density functions of CYGNSS ANN retrieved coastal wind speed, CYGNSS Level 2 wind speed and ERA5 wind speed (left). The distribution of the wind speed deviation between the CYGNSS ANN retrieved coastal wind speed and ERA5 data (right).

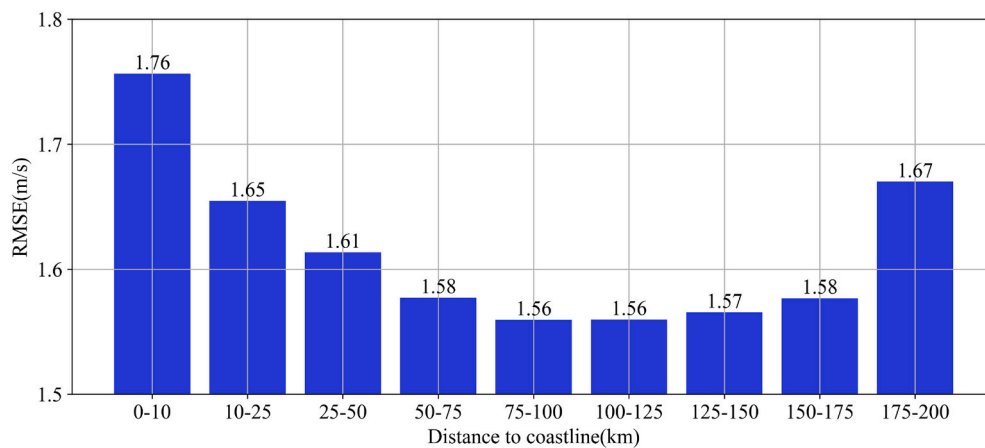


Fig. 6. The root-mean-square error of the CYGNSS ANN-based wind speed retrievals in different regions with different distance to the coastline.

during the period between January 2018 and December 2018, in which a global map of gridded wind speed biases (with the grid size being $0.25^\circ \times 0.25^\circ$ in latitude and longitude) is generated. It is clearly shown that positive biases appear along the coast line of the Asia-Pacific region in the CYGNSS L2 v2.1 product (Fig. 7 - top), which are removed in the ANN retrieval (Fig. 7 - bottom). Such improvement is expected as the CYGNSS L2 v2.1 product is generated with a global GMF while the ANN wind speed retrieval model function is trained by using the geographical information (latitude and longitude of the specular point) as the additional parameters.

4.1.3. Systematic effects

As a non-cooperative bistatic radar system, there is a large diversity of geometries and complexity of elements involved in the overall CYGNSS system: eight CYGNSS spacecraft (SC) with two reflectometry antennas (port and starboard) for each of them, receiving reflected signals from more than 30 GPS satellites with different designs (Block types) and the capability to dynamically change the transmitted power levels. Hence, the CYGNSS L1 observables depend on the individual CYGNSS SC, the reflectometry antenna, as well as the targeting GPS satellite. In addition to the geophysical related errors, these diversity and complexity of the system are also important sources of error in wind speed retrieval.

To characterize the inter-CYGNSS-SC and inter-antenna biases, the wind speed retrieval performance metrics are computed and evaluated per CYGNSS SC, as well as per reflectometry antenna. The wind speed biases and RMSEs for each CYGNSS SC and each reflectometry antenna are presented in Table 5 for both the CYGNSS L2 v2.1 product and the ANN retrieval. Fig. 8 presents the variations of the inter-CYGNSS-SC biases $[(Bias_{port} + Bias_{starboard})/2]$ and the inter-antenna biases $(Bias_{port} - Bias_{starboard})$ of both the ANN retrieval and the CYGNSS L2 v2.1 products for different CYGNSS SCs, from which it is clearly seen that both biases still appear in the ANN retrievals. In addition, it is also shown that both the inter-CYGNSS-SC and inter-antenna biases show similar variations among different CYGNSS SC for the ANN retrieval and the CYGNSS L2 products.

In addition to the inter-CYGNSS-SC and inter-antenna biases, the errors due to different GPS Block types are also assessed for the ANN retrieval. As all the GPS IIF satellites are removed in previous analyses, a new ANN training and retrieval procedure has been performed by including all the GPS satellites [i.e. to include CYGNSS L1 measurements with the quality control flag of “low_quality_gps_ant_knowledge” (QC bit 27)]. The performances of the ANN retrieval for GPS Block IIR, IIR-M and IIF satellites are computed and shown in Table 6 for two different cases: the IIR-model case means to train the ANN model only with the CYGNSS L1 observables from the GPS Block IIR and IIR-M satellites,

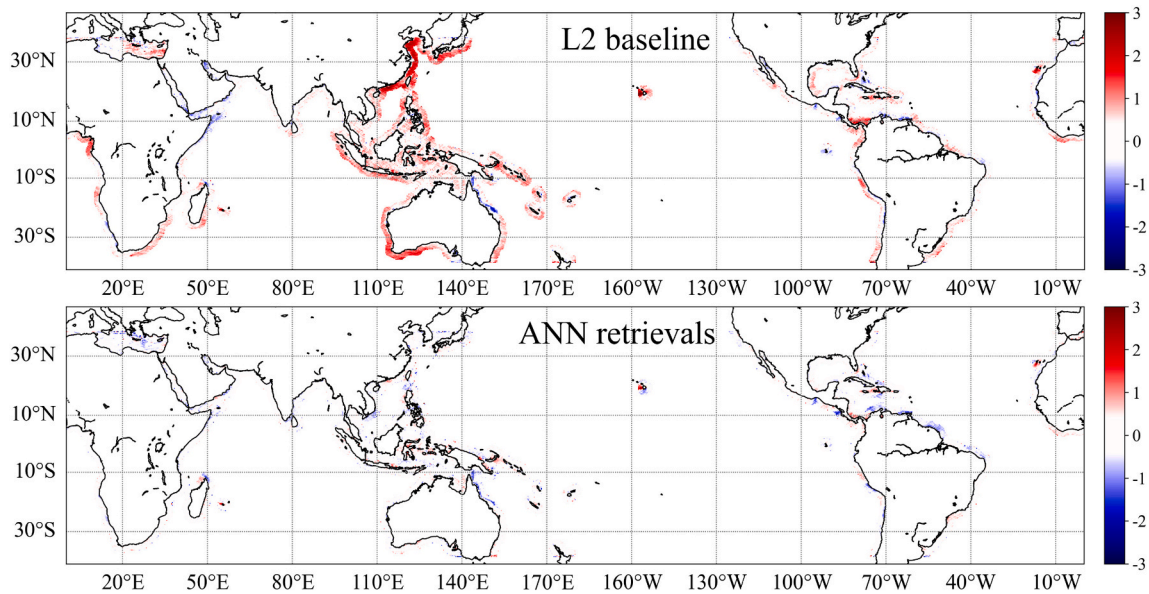


Fig. 7. Geographical map of the wind speed biases (CYGNSS wind speed - ERA5 wind speed) along the coastlines. Top: Wind speed bias of the CYGNSS Level 2 v2.1 products. Bottom: Wind speed bias of the ANN based retrieval.

Table 5
Comparisons of CYGNSS ANN based coastal wind speed retrieval and the CYGNSS L2 v2.1 product for each CYGNSS spacecraft and each reflectometry antenna.

Retrieval	CYGNSS SC	Antenna	Bias (m/s)	RMSE (m/s)	R
ANN	cyg01	starboard	-0.29	1.59	0.85
		port	0.06	1.53	0.85
	cyg02	starboard	0.03	1.48	0.86
		port	0.00	1.53	0.84
	cyg03	starboard	0.01	1.51	0.86
		port	-0.23	1.59	0.85
	cyg04	starboard	-0.22	1.53	0.86
		port	0.04	1.52	0.85
	cyg05	starboard	0.14	1.59	0.84
		port	-0.23	1.59	0.85
	cyg06	starboard	-0.21	1.56	0.85
		port	0.05	1.58	0.84
	cyg07	starboard	-0.12	1.51	0.86
		port	0.02	1.57	0.84
	cyg08	starboard	-0.02	1.51	0.86
		port	-0.06	1.57	0.84
CYGNSS L2 v2.1	cyg01	starboard	0.09	1.92	0.80
		port	0.63	2.22	0.80
	cyg02	starboard	0.54	2.02	0.82
		port	0.32	1.98	0.80
	cyg03	starboard	0.63	2.18	0.81
		port	0.09	1.95	0.79
	cyg04	starboard	0.19	1.91	0.81
		port	0.49	2.07	0.80
	cyg05	starboard	0.83	2.49	0.77
		port	0.06	1.91	0.80
	cyg06	starboard	0.22	1.92	0.81
		port	0.58	2.18	0.80
	cyg07	starboard	0.34	1.94	0.82
		port	0.50	2.21	0.78
	cyg08	starboard	0.55	2.08	0.81
		port	0.38	2.05	0.80

while the full-block-model case means to train the ANN model with the CYGNSS observables from all the GPS satellites. Table 6 shows that the performances of CYGNSS ANN based coastal wind speed retrieval from the Block IIR and IIR-M satellites are more stable with smaller biases (-0.08 to -0.06 m/s) and RMSE (1.55 to 1.6 m/s), while the wind speeds retrievable using GPS Block II-F satellite show significant performance degradation with larger biases (-0.25 to -0.18 m/s) and

RMSE (1.76 to 1.97 m/s).

Overall, the current ANN configurations cannot significantly reduce these systematic effects, e.g. inter-CYGNSS-SC, inter-antenna or inter-GNSS-Block biases, in CYGNSS wind speed retrieval. It is expected as the current ANN does not include these factors (e.g. CYGNSS SC ID, reflectometry antenna and GNSS PRN) as the input parameters in ANN training and retrieval.

4.2. Validation with the NDBC Buoy observations

As the ANN model is trained and validated with the same “ground truth” data, i.e. the ECMWF/C3S ERA5 wind speed, it is necessary to assess the wind speed retrieval performance with an independent source of wind speed measurements. For this purpose, the wind speed measurements collected with the National Data Buoy Center (NDBC) buoys have been also collected as the reference. Wind speed measurements from a total of 15 NDBC buoys are selected from January 1 to December 31, 2018. The selected NDBC buoys are located along the coastal line as the red pentacle depicted in Fig. 9, and Table 5 presents the brief information of these buoys, including the buoy ID, their geographic coordinates and the distances to coastal line (L).

The NDBC buoys normally provide one sea surface wind observation every ten minutes to one hour. As the anemometer heights of the NDBC buoys are different, it is necessary to adjust the wind speed measurements at different heights to the standard height of 10 m above the sea surface. The wind speed adjustment is implemented by following the power-law wind profile (Hsu et al., 1994).

All ANN-derived wind speed measurements are considered within 25 km and ± 15 min of buoy locations and measurement times. There are ~11 k matchups in total with the numbers of the matchups for each NDBC buoys listed in Table 7. The RMSEs between the CYGNSS ANN-based wind speeds and the adjusted anemometer measurements are computed and presented in Table 7 for different buoys. It can be seen that the wind speed retrieval performance shows significant dispersion among the analyzed buoys. The CYGNSS-NDBC matchups for each buoy station are then compiled and presented as a density scatter plot in Fig. 10, which shows that both measurements have good correlations in general with most of the data point centered along the 1:1 diagonal. However, for the wind speed larger than 5 m/s, the scatterplot presents a small but noticeable offset off the 1:1 line towards higher NDBC wind speed. By applying (1)–(3), the CYGNSS ANN retrievals have a small

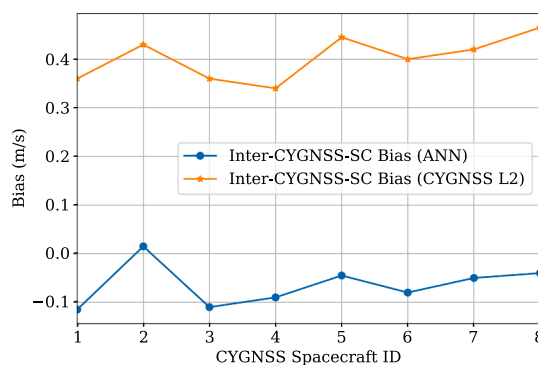
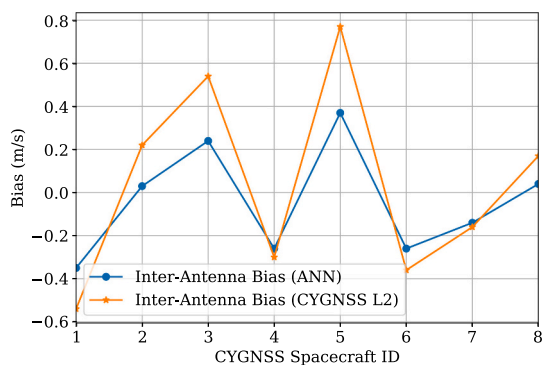


Fig. 8. Comparison between inter CYGNSS spacecraft biases and inter reflectometry antenna biases for the ANN retrieval and the CYGNSS L2 products. Left: Inter reflectometry antenna biases. Right: Inter CYGNSS spacecraft biases.

Table 6

The performance metrics of the ANN based CYGNSS coastal wind speed retrieval for different GPS satellites Block types (IIR, IIR-M and IIF). The IIR-model case: the ANN model is trained only with the GPS Block IIR and IIR-M satellites; the Full-block-model case: the ANN model is trained with all the GPS satellites.

	IIR-model retrieval			Full-block-model retrieval		
	Bias (m/s)	RMSE (m/s)	R	Bias(m/s)	RMSE(m/s)	R
IIR	-0.08	1.55	0.85	-0.06	1.60	0.82
IIR-M	-0.06	1.55	0.85	-0.06	1.59	0.82
IIF	-0.25	1.97	0.81	-0.18	1.76	0.82
Average	-0.13	1.69	0.84	-0.10	1.65	0.82

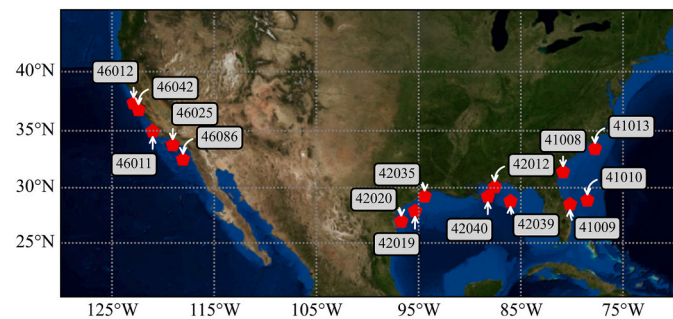


Fig. 9. Distribution of the National Data Buoy Center's buoy stations along the coastal line, which are selected for the validation of the wind speed retrieval in this study.

Table 7

Brief information about the National Data Buoy Center's buoy stations used for the validation of the wind speed retrievals in this study.

Buoy ID	Lat (°N)	Lon (°W)	L (km)	Number of matchups	RMSE (m/s)
42,012	30.06	-87.55	20	464	1.78
42,035	29.23	-94.41	30	503	2.15
46,042	36.78	-122.39	30	534	1.90
46,011	34.95	-121.02	32	447	2.04
41,008	31.40	-80.86	32	698	1.70
46,025	33.76	-119.05	33	516	1.65
41,009	28.50	-80.18	34	975	2.08
46,012	37.36	-122.88	35	464	1.78
41,013	33.43	-77.74	50	1512	2.05
42,020	26.96	-96.69	68	880	2.02
42,040	29.20	-88.22	75	1567	1.75
46,086	32.49	-118.05	75	738	1.76
42,019	27.90	-95.35	100	1004	1.77
42,039	28.78	-86.00	115	652	1.81
41,010	28.87	-78.48	200	303	1.63

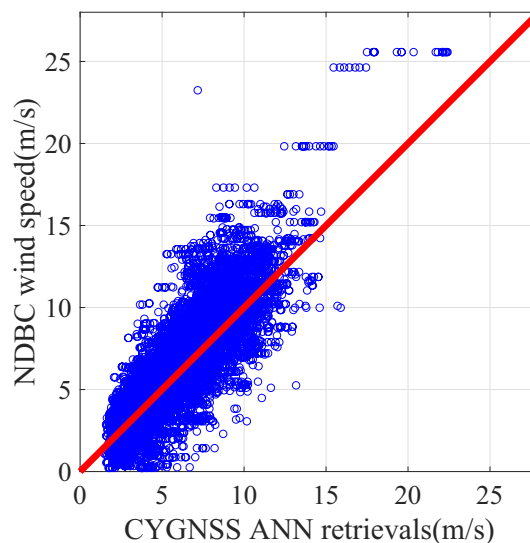


Fig. 10. A 2D density plot of collocated CYGNSS and the National Data Buoy Center's buoy wind speeds. The purple line shows the 1:1 diagonal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

deviation from buoys measurements with a bias of -0.44 m/s, a correlation coefficient of 0.83 and a RMSE is 1.86 m/s, which demonstrate a good correlation between the CYGNSS ANN wind retrievals and the buoy measurements. The negative bias (CYGNSS minus buoy) indicates that the CYGNSS ANN-derived wind speed is in general underestimated with respected the NDBC values.

In addition to assessing the overall performance metrics, it is important to understand the ANN-based wind speed retrieval's capability for capturing wind speed temporal variations. Fig. 11 shows the per-site comparisons for selected buoys between the CYGNSS ANN-based wind speeds and NDBC measurements. These figures show both time-series to visualize the annual trends to illustrate the consistence between the CYGNSS ANN-based measurements and the in-situ values from NDBC. It is clearly shown that the CYGNSS wind speed estimates closely follow the temporal trend of the NDBC buoys wind speed observations and correctly capture the high wind events during Tropical Storm Alberto, Hurricane Florence and Michael (e.g. in Fig. 11 a and b), although the peak winds are still underestimated. It can be explained by the low occurrence of the high wind samples in the ANN training, so that the ANN wind speed retrieval model is not able to represent the relationship between the GNSS-R observables and the wind speed at high wind conditions. In addition, the sensitivity of the CYGNSS observables to the change in wind speed is reduced as the winds increase, the ANN

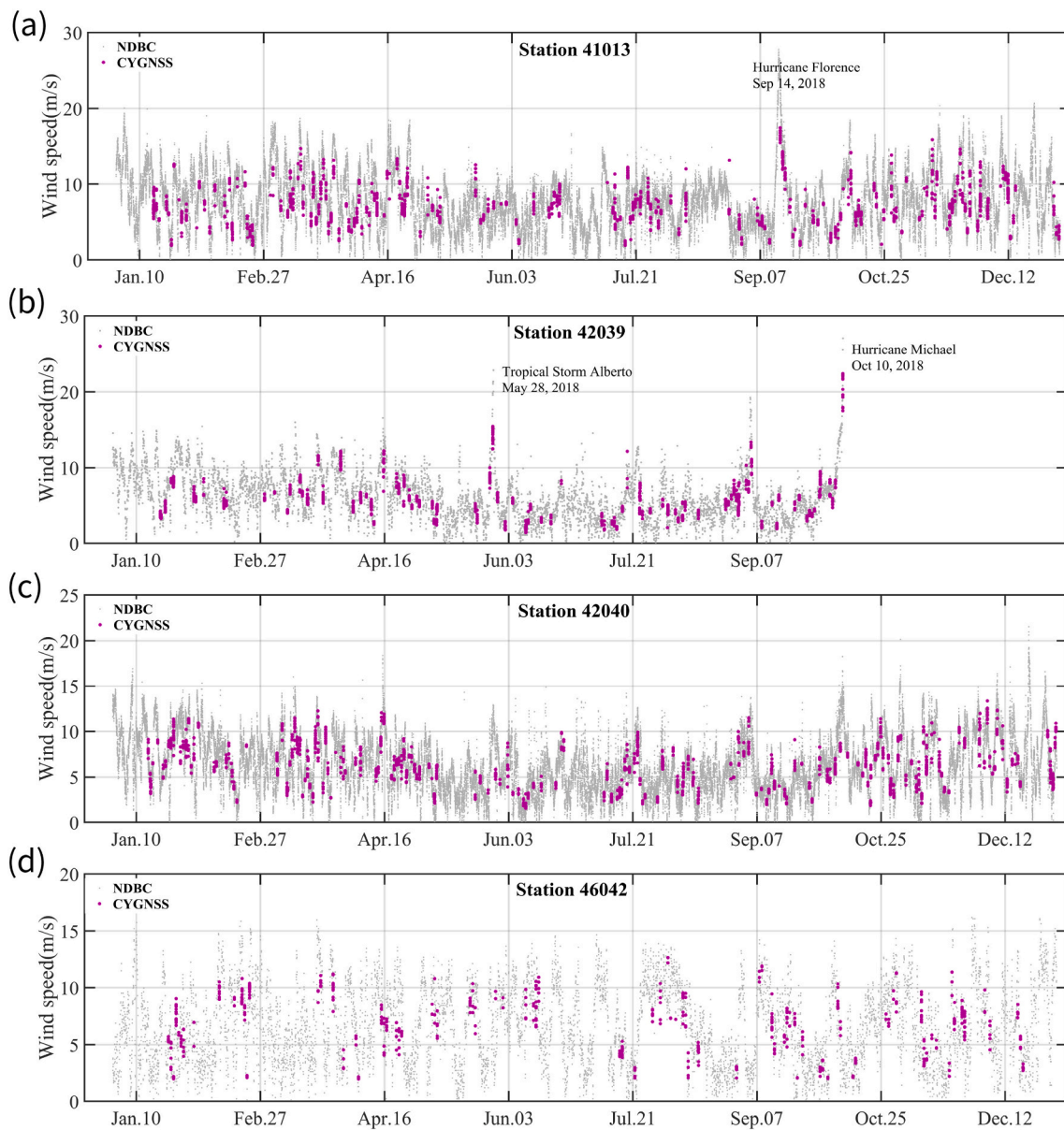


Fig. 11. Time series of the wind speeds measured by the NDBC buoy and retrieved based on ANN-based model.

model may underestimate the wind speed by following the same relationship between the CYGNSS observables and u_{10} at low-to-moderate wind speed.

5. Conclusion and discussion

In this paper, the capability of spaceborne GNSS-R on measuring near sea surface wind speed around the coastal regions has been demonstrated. The CYGNSS standard Level 2 wind speed product shows a significant performance degradation in coastal region (25–200 km away from the coastline) with respect to in open ocean (RMSE 2.10 m/s vs. 1.88 m/s according to the data in 2018). In addition, the CYGNSS L2 baseline wind speed retrieval has removed the data within 25 km away from the coastline, so that no wind speed data is provided in this region.

This work is based on a learning-based wind speed retrieval methodology, in which a feedforward neural network is trained with the collocated CYGNSS Level 1B observables and the wind speed from ECMWF ERA5 data in coastal regions (<200 km away from the

coastline). In addition to the CYGNSS L1 baseline observables (NBRCS and LES), some other geophysical and geographical parameters, such as the incidence and azimuth angles, the geolocation of the specular point and the swell wave height, are also used as the input parameters for the ANN training. By testing different ANN configurations, the network with five hidden layers and 200 neurons in each layer is selected to be with the best wind speed retrieval performance. Moreover, the contributions of different input parameters to wind speed retrieval has been also characterized. It is found that the position of the specular point and the swell wave height, among others, can provide significant contribution to wind speed retrieval, which provides useful reference for more generic GNSS-R wind speed retrieval algorithms for the coastal regions.

The wind speed retrieval performance of the proposed approach has been evaluated by applying the trained ANN model to the testing datasets (i.e. 85% of the randomly selected CYGNSS Level 1B data in 2018). The retrieved wind speeds are compared to the collocated ERA5 values, from which the performance metrics have been obtained with a bias of 0.002 m/s and a RMSE of 1.58 m/s. These performance metrics show significant improvements (i.e. 24.4% in wind speed RMSE)

compared to the CYGNSS Level 2 wind speed products around the coastal region. In addition, it is also found that the ANN based wind speed retrieval shows better performance away from the land, which achieves the best performance (1.53 m/s RMSE) in the range of 75–125 km away from the coastline. Even affected by the land contamination and shallow water, our wind speed retrieval still shows an acceptable performance (1.76 m/s RMSE) in near-shore region within 10 km away from the coastline.

The ANN-based retrieval model is also assessed with an independent sources of wind speed measurements from the NDBC buoys. The CYGNSS ANN-based wind speeds show good correlation with the collocated NDBC measurements ($R = 0.83$) with a mean bias of -0.44 m/s and a RMSE is 1.86 m/s. However, the CYGNSS ANN-based model significantly underestimates the high wind speed during extreme wind events, which could be due to the low occurrence of the high wind samples in the ANN training and the low sensitivity of the CYGNSS observables to the change in wind speed in such cases.

The proposed learning-based wind speed retrieval methodology generates promising overall performance, demonstrating the capability of GNSS-R measurements on sea surface wind speed sensing in near-shore regions. The performance of the proposed model can be improved potentially, which will be conducted in future works.

- 1) The ANN structure and configurations can be further optimized by training over a much larger and more representative data sets. Current wind speed data for ANN training is from the ECMWF ERA-5 data, which itself has limitations in coastal regions and at high wind speed. A more comprehensive training data by merging wind speed data from different models/sensors (Ruf and Balasubramaniam, 2018), especially with reliable high wind speed measurements, would benefit such processing. In addition, the current training data is randomly selected from the overall CYGNSS-ERA5 matchups, which makes a low occurrence of the low wind speed and high wind speed samples. The sampling strategy of the training set selection should be also investigated in detail in further studies. Alternatively, the ANN retrieved wind speed can be post-processed by applying the CDF matching algorithm to the network outputs, which can also improve the consistency between the PDFs of the retrieved and reference wind speeds.
- 2) As the ocean wave information can significantly contribute to our ANN-base wind speed retrieval, it could be also helpful to use more ocean wave related parameters as the inputs of the ANN training, such as the water level, wave age, wave length, and wave direction. However, the complexity of the network and the computational complexity for the network training would be significantly increased, which can affect the convergence of the network training. In addition, the input ocean wave parameters should be carefully checked to avoid adding extra wind speed information to the retrieval.
- 3) CYGNSS observations and data products encounter a number of uncertainties even after applying the full quality control flag. The ANN itself can also provide an effective way to detect the outliers (Balasubramaniam and Ruf, 2020) and to characterize the wind speed uncertainties for different observing configurations (e.g. geometry, antenna gain) and observable parameters (such as SNR), which will be also attempted in future works.
- 4) With the evolution of the CYGNSS data products and wind speed retrieval algorithms, the qualities of the CYGNSS L1B observables and the performance of the CYGNSS L2 wind speed retrieval have been significantly improved. For example, the CYGNSS L1B version 3.0 product is generated by using the real time transmit power monitoring and correction algorithm, which allows GPS Block IIF satellite to be used in wind speed retrieval and improves the accuracy and precision of the L1B observables. By using the latest version of the CYGNSS L1B data product, the performance of the ANN based wind speed retrieval can be further improved. In addition, the CYGNSS L2 Climate Data Record (CDR) products v1.0 (CYGNSS,

2020) and v1.1 (CYGNSS, 2021) and NOAA CYGNSS L2 Wind Speed product v1.1 (NOAA/NESDIS/STAR/SOCD, 2021) use the trackwise correction algorithms (Ruf and Twigg, 2020; Said et al., 2019) to compensate the systematic effects due to uncertainties in the GPS transmitting powers and antenna gain patterns as well as the inter CYGNSS spacecraft and inter reflectometry antenna biases. As these systematic effects also have significant effects on our ANN retrieval results (as shown in Section 4.1.3), further improvements in the performance of the ANN retrieval can be expected by using the pre-processed CYGNSS L1B observables with a similar trackwise correction.

CRedit authorship contribution statement

Xiaohui Li: Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing, Visualization. **Dongkai Yang:** Writing - original draft, Supervision, Funding acquisition. **Jingsong Yang:** Software, Validation, Funding acquisition. **Gang Zheng:** Validation. **Guoqi Han:** Validation. **Yang Nan:** Software. **Wei-qiang Li:** Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests, or personal relationships that could have appeared to influence the work, reported in this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China under Grant No. 41774028 and Grant No. 41676167, the National Key Research and Development Plan under Grant No. 2017YFB050802, Zhejiang Provincial Natural Science Foundation of China under Grant No. LR21D060002, the Key R&D Project of Shandong Province under Grant No. 2019JZZY010102, the Project of State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography under Grant No. SOEDZZ2003, and the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDA19090103 and Grant No. XDB42000000. This work of W. Li has been supported in part by the Spanish Ministry of Economy and Competitiveness and EU/FEDER (ESP2015-70014-C2-2-R) and the Ramón y Cajal Program (RYC2019-027000-I). The authors are grateful to the CYGNSS team for making the Level 1 and Level 2 data publicly available through the NASA EOSDIS Physical Oceanography Distributed Active Archive Center (PO.DAAC; doi:<https://doi.org/10.5067/CYGNSS-L1X21> and <https://doi.org/10.5067/CYGNSS-L2X21>).

References

- Al-Khaldi, M.M., Johnson, J.T., Gleason, S., Loria, E., O'Brien, A.J., Yi, Y., 2020. An algorithm for detecting coherence in cyclone global navigation satellite system mission level-1 delay-Doppler maps. *IEEE Trans. Geosci. Remote Sens. Early Access* 1–10. <https://doi.org/10.1109/TGRS.2020.3009784>.
- Alonso Arroyo, A., Camps, A., Aguiasca, A., Forte, G.F., Monerri, A., Rudiger, C., Walker, J.P., Park, H., Pascual, D., Onrubia, R., 2014. Dual-polarization GNSS-R interference pattern technique for soil moisture mapping. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 7, 1533–1544. <https://doi.org/10.1109/JSTARS.2014.2320792>.
- Alonso-Arroyo, A., Zavorotny, V.U., Camps, A., 2017. Sea ice detection using U.K. TDS-1 GNSS-R data. *IEEE Trans. Geosci. Remote Sens.* 55, 4989–5001. <https://doi.org/10.1109/TGRS.2017.2699122>.
- Alshaye, M., Alawwad, F., Elshafiey, I., 2020. Hurricane tracking using Multi-GNSS-R and deep learning. In: 2020 3rd International Conference on Computer Applications and Information Security (ICCAIS), Riyadh, Saudi Arabia, pp. 1–4. <https://doi.org/10.1109/ICCAIS48893.2020.9096717>.
- Asgarimehr, M., Zhelavskaya, I., Foti, G., Reich, S., Wickert, J., 2019. A GNSS-R geophysical model function: machine learning for wind speed retrievals. *IEEE*

- Geosci. Remote Sens. Lett. 17, 1333–1337. <https://doi.org/10.1109/LGRS.2019.2948566>.
- Balasubramaniam, R., Ruf, C., 2020. Neural network based quality control of CYGNSS wind retrieval. *Remote Sens.* 12, 2859. <https://doi.org/10.3390/rs12172859>.
- Barthelmie, R.J., 1999. The effects of atmospheric stability on coastal wind climates. *Meteorol. Appl.* 6, 39–47. <https://doi.org/10.1017/S1350482799000961>.
- Cardellach, E., Li, W., Rius, A., Semmling, M., Wickert, J., Zus, F., Ruf, C.S., Buontempo, C., 2020. First precise spaceborne sea surface altimetry with GNSS reflected signals. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 13, 102–112. <https://doi.org/10.1109/JSTARS.2019.2952694>.
- Carreno-Luengo, H., Luzzi, G., Crosetto, M., 2019. First evaluation of topography on GNSS-R: an empirical study based on a digital elevation model. *Remote Sens.* 11, 2556. <https://doi.org/10.3390/rs11122556>.
- Chelton, D.B., Mccabe, P.J., 1985. A review of satellite altimeter measurement of sea surface wind speed: with a proposed new algorithm. *J. Geophys. Res.-Oceans* 90, 4707–4720. <https://doi.org/10.1029/JC090i03p04707>.
- Chew, C.C., Small, E.E., 2018. Soil moisture sensing using spaceborne GNSS reflections: comparison of CYGNSS reflectivity to SMAP soil moisture. *Geophys. Res. Lett.* 45, 4049–4057. <https://doi.org/10.1029/2018GL077905>.
- Clarizia, M.P., Ruf, C.S., 2016. Wind speed retrieval algorithm for the cyclone global navigation satellite system (CYGNSS) mission. *IEEE Trans. Geosci. Remote Sens.* 54, 4419–4432. <https://doi.org/10.1109/TGRS.2016.2541343>.
- Clarizia, M.P., Ruf, C.S., 2020. Statistical derivation of wind speeds from CYGNSS data. *IEEE Trans. Geosci. Remote Sens.* 6, 3955–3964.
- Clarizia, M.P., Ruf, C.S., Jales, P., Gommenginger, C., 2014. Spaceborne GNSS-R minimum variance wind speed estimator. *IEEE Trans. Geosci. Remote Sens.* 52, 6829–6843. <https://doi.org/10.1109/TGRS.2014.2303831>.
- Clarizia, M.P., Ruf, C.S., Gleason, S., Balasubramaniam, R., Mckague, D., 2017. Generation of CYGNSS level 2 wind speed data products. *IEEE Int. Geosci. Remote Sens. Symp.* <https://doi.org/10.1109/IGARSS.2017.8127539>.
- CYGNSS, 2020. CYGNSS Level 2 Climate Data Record Version 1.0. Ver. 1.0. PO.DAAC. <https://doi.org/10.5067/CYGNSS-L2C10> (accessed: 2021-03-08).
- CYGNSS, 2021. CYGNSS Level 2 Climate Data Record Version 1.1. Ver. 1.1. PO.DAAC. <https://doi.org/10.5067/CYGNSS-L2C11> (accessed: 2021-03-08).
- Dayhoff, J.E., DeLeo, J.M., 2001. Artificial neural networks. *Cancer* 91, 1615–1635. [https://doi.org/10.1002/1097-0142\(20010415\)91:8+<1615::AID-CNCR1175>3.0.CO;2-L](https://doi.org/10.1002/1097-0142(20010415)91:8+<1615::AID-CNCR1175>3.0.CO;2-L).
- Eroglu, O., Kurum, M., Boyd, D., Gurbuz, A.C., 2019. High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks. *Remote Sens.* 11, 2272. <https://doi.org/10.3390/rs11192272>.
- Feng, Q., Zheng, N., Liu, C., Zhou, X., 2018. BP neural network assisted soil moisture retrieval using GNSS satellite reflected signal. In: *Ence of Surveying and Mapping*, vol. 43, pp. 157–162.
- Foti, G., Gommenginger, C., Jales, P., Unwin, M., Rosello, J., 2015. Spaceborne GNSS reflectometry for ocean winds: first results from the UK TechDemoSat-1 mission. *Geophys. Res. Lett.* 42, 5435–5441. <https://doi.org/10.1002/2015GL064204>.
- Gao, H., Bai, Z., Fan, D., Co, D.S., Ltd, 2019a. GNSS-R Sea surface wind speed inversion based on BP neural network. *Acta Aeronaut. Astronaut. Sin.* 40, 323261.
- Gao, H., Yang, D., Wang, F., Wang, Q., Li, X., 2019b. Retrieval of ocean wind speed using airborne reflected GNSS signals. *IEEE Access* 7, 71986–71998. <https://doi.org/10.1109/ACCESS.2019.2915193>.
- Garrison, J.L., Katzberg, S.J., 2000. The application of reflected GPS signals to ocean remote sensing. *Remote Sens. Environ.* 73, 175–187. [https://doi.org/10.1016/S0034-4257\(00\)00092-4](https://doi.org/10.1016/S0034-4257(00)00092-4).
- Ghasemigoudarzi, P., Huang, W., De Silva, O., Yan, Q., Power, D., 2020. A machine learning method for inland water detection using CYGNSS data. *IEEE Geosci. Remote Sens. Lett.* Early Access 1–5. <https://doi.org/10.1109/LGRS.2020.3020223>.
- Glazman, R.E., Greysukh, A., 1993. Satellite altimeter measurements of surface wind. *J. Geophys. Res.-Oceans* 98, 2475. <https://doi.org/10.1029/92JC02659>.
- Gleason, S., Hodgart, S., Sun, Y., Gommenginger, C., Mackin, S., Adjrak, M., Unwin, M., 2005. Detection and processing of bistatically reflected GPS signals from low earth orbit for the purpose of ocean remote sensing. *IEEE Trans. Geosci. Remote Sens.* 43, 1229–1241. <https://doi.org/10.1109/TGRS.2005.845643>.
- Gleason, S., Ruf, C.S., O'Brien, A.J., McKague, D.S., 2019. The CYGNSS level 1 calibration algorithm and error analysis based on on-orbit measurements. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 12, 37–49. <https://doi.org/10.1109/JSTARS.2018.2832981>.
- Hammond, M.L., Foti, G., Gommenginger, C., Srokosz, M., 2020. Temporal variability of GNSS-Reflectometry Ocean wind speed retrieval performance during the UK TechDemoSat-1 mission. *Remote Sens. Environ.* 242, 111744. <https://doi.org/10.1016/j.rse.2020.111744>.
- Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* 4, 251–257. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T).
- Hsu, S.A., Meindl, E.A., Gilhousen, D.B., 1994. Determining the power-law wind-profile exponent under near-neutral stability conditions at sea. *J. Appl. Meteorol.* 33, 757–765. [https://doi.org/10.1175/1520-0450\(1994\)03302.CO;2](https://doi.org/10.1175/1520-0450(1994)03302.CO;2).
- Hu, C., Benson, C.R., Qiao, L., Rizos, C., 2020. The validation of the weight function in the leading-edge-derivative path delay estimator for space-based GNSS-R altimetry. *IEEE Trans. Geosci. Remote Sens.* 58, 6243–6254. <https://doi.org/10.1109/TGRS.2020.2975817>.
- Huang, B., Stone, P.H., Sokolov, A.P., Kamenkovich, I.V., 2003. Ocean heat uptake in transient climate change: mechanisms and uncertainty due to subgrid-scale eddy mixing. *J. Clim.* 16, 3344–3356. [https://doi.org/10.1175/1520-0442\(2003\)0162.0.CO;2](https://doi.org/10.1175/1520-0442(2003)0162.0.CO;2).
- Jing, C., Niu, X., Duan, C., Lu, F., Yang, X., 2019. Sea surface wind speed retrieval from the first chinese GNSS-R mission: technique and preliminary results. *Remote Sens.* 11, 3013. <https://doi.org/10.3390/rs11243013>.
- Kaimal, J.C., Finnigan, J.J., 1994. *Atmospheric Boundary Layer Flows: Their Structure and Measurement*. Oxford University Press, New York.
- Kasantikul, K., Yang, D., Wang, Q., Lwin, A., 2018. A novel wind speed estimation based on the integration of an artificial neural network and a particle filter using BeiDou GEO reflectometry. *Sensors* 18, 3350. <https://doi.org/10.3390/s18103350>.
- Kirincich, A., 2016. Remote sensing of the surface wind field over the coastal ocean via direct calibration of HF radar backscatter power. *J. Atmos. Ocean. Technol.* 33, 1377–1392. <https://doi.org/10.1175/JTECH-D-15-0242.1>.
- Kwok, T.Y., Yeung, D.Y., 1997. Constructive algorithms for structure learning in feedforward neural networks for regression problems. *IEEE Trans. Neural Netw.* 8, 630. <https://doi.org/10.1109/72.572102>.
- Li, W., Cardellach, E., Fabra, F., Rius, A., Ribó, S., Martín-Neira, M., 2017. First spaceborne phase altimetry over sea ice using TechDemoSat-1 GNSS-R signals. *Geophys. Res. Lett.* 44, 8369–8376. <https://doi.org/10.1002/2017GL074513>.
- Li, X., Yang, D., Han, G., Yang, L., Wang, J., Yang, J., Chen, D., Zheng, G., 2020. Exploiting the potential of coastal GNSS-R for improving storm surge modelling. *IEEE Geosci. Remote Sens. Lett.* 1–5. <https://doi.org/10.1109/LGRS.2020.2996415>.
- Liu, Y., Collett, I., Morton, Y.J., 2019. Application of neural network to GNSS-R wind speed retrieval. *IEEE Trans. Geosci. Remote Sens.* 57, 9756–9766. <https://doi.org/10.1109/TGRS.2019.2929002>.
- Morris, M., Ruf, C.S., 2017. Determining tropical cyclone surface wind speed structure and intensity with the CYGNSS satellite constellation. *J. Appl. Meteorol. Climatol.* 56, 1847–1865. <https://doi.org/10.1175/JAMC-D-16-0375.1>.
- NOAA/NESDIS/STAR/SOCD, 2021. NOAA CYGNSS Level 2 Science Wind Speed 25-km Product Version 1.1. Ver. 1.1. PO.DAAC. <https://doi.org/10.5067/CYGNSS-22511> (accessed: 2021-03-08).
- Peng, Q., Jin, S., 2019. Significant wave height estimation from space-borne cyclone-GNSS reflectometry. *Remote Sens.* 11, 584. <https://doi.org/10.3390/rs11050584>.
- Reynolds, J., Clarizia, M.P., Santi, E., 2020. Wind speed estimation from CYGNSS using artificial neural networks. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 13, 708–716. <https://doi.org/10.1109/JSTARS.2020.2968156>.
- Roggenbuck, O., Reinking, J., Lambertus, T., 2019. Determination of significant wave heights using damping coefficients of attenuated GNSS SNR data from static and kinematic observations. *Remote Sens.* 11, 409. <https://doi.org/10.3390/rs11040409>.
- Ruf, C., Twigg, D., 2020. Algorithm theoretical basis document. In: *Level 1 & 2 Trackwise Corrected Climate Data Record*. Technical Report. CYGNSS Project Document (Accessed: 2021-03-08).
- Ruf, C., Unwin, M., Dickinson, J., Rose, R., Rose, D., Vincent, M., Lyons, A., 2013. CYGNSS: enabling the future of hurricane prediction. *IEEE Geosci. Remote Sens. Mag.* 1, 52–67. <https://doi.org/10.1109/MGRS.2013.2260911>.
- Ruf, C., Posselt, D., Majumdar, S., Gleason, S., Morris, M., 2016a. *CYGNSS Handbook*. Michigan Pub, Ann Arbor, MI (Accessed: 2021-03-08).
- Ruf, C., Asharaf, S., Balasubramaniam, R., Gleason, S., Waliser, D., 2019. In-orbit performance of the constellation of CYGNSS hurricane satellites. *Bull. Am. Meteorol. Soc.* 100, 2009–2023. <https://doi.org/10.1175/BAMS-D-18-0337.1>.
- Ruf, C.S., Balasubramaniam, R., 2018. Development of the CYGNSS geophysical model function for wind speed. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 12, 66–77. <https://doi.org/10.1109/JSTARS.2018.2833075>.
- Ruf, C.S., Atlas, R., Chang, P.S., Clarizia, M.P., Garrison, J.L., Gleason, S., Katzberg, S.J., Jelenak, Z., Johnson, J.T., Majumdar, S.J., et al., 2016b. New Ocean winds satellite mission to probe hurricanes and tropical convection. *Bull. Am. Meteorol. Soc.* 97, 385–395. <https://doi.org/10.1175/BAMS-D-14-00218.1>.
- Ruf, C.S., Gleason, S., Mckague, D.S., 2018. Assessment of CYGNSS wind speed retrieval uncertainty. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 12, 87–97. <https://doi.org/10.1109/JSTARS.2018.2825948>.
- Said, F., Jelenak, Z., Park, J., Soisuvarn, S., Chang, P.S., 2019. A 'track-wise' wind retrieval algorithm for the CYGNSS mission. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Yokohama, Japan, pp. 8711–8714. <https://doi.org/10.1109/IGARSS.2019.8898099>.
- Soulat, F., Caparrini, M., Germain, O., Lopez-Dekker, P., Taani, M., Ruffini, G., 2004. Sea state monitoring using coastal GNSS-R. *Geophys. Res. Lett.* 31, 133–147. <https://doi.org/10.1029/2004GL020680>.
- Stiles, B.W., Dunbar, R.S., 2010. A neural network technique for improving the accuracy of scatterometer winds in rainy conditions. *IEEE Trans. Geosci. Remote Sens.* 48, 3114–3122. <https://doi.org/10.1109/TGRS.2010.2049362>.
- Stiles, B.W., Danielson, R.E., Poulsen, W.L., Brennan, M.J., Hristova-Velleva, S., Shen, T., Fore, A.G., 2014. Optimized tropical cyclone winds from QuikSCAT: a neural network approach. *IEEE Trans. Geosci. Remote Sens.* 52, 7418–7434. <https://doi.org/10.1109/TGRS.2014.2312333>.
- Tabibi, S., Geremia-Nievinski, F., Francis, O., van Dam, T., 2020. Tidal analysis of GNSS reflectometry applied for coastal sea level sensing in Antarctica and Greenland. *Remote Sens. Environ.* 248, 111959. <https://doi.org/10.1016/j.rse.2020.111959>.
- Tang, R., Liu, D., Han, G., Ma, Z., de Young, B., 2014. Reconstructed wind fields from multi-satellite observations. *Remote Sens.* 6, 2898–2911. <https://doi.org/10.3390/rs6042898>.
- Unwin, M., Jales, P., Tye, J., Gommenginger, C., Foti, G., Rosello, J., 2017. Spaceborne GNSS-reflectometry on TechDemoSat-1: early mission operations and exploitation. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 9, 4525–4539. <https://doi.org/10.1109/JSTARS.2016.2603846>.
- Wang, F., Yang, D., Zhang, B., Li, W., 2018. Waveform-based spaceborne GNSS-R wind speed observation: demonstration and analysis using UK TechDemoSat-1 data. *Adv. Space Res.* 61, 1573–1587. <https://doi.org/10.1016/j.asr.2018.01.013>.

- Wessel, P., Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. *J. Geophys. Res.* 101, 8741–8743. <https://doi.org/10.1029/96JB00104>.
- Yan, Q., Huang, W., 2018. Sea ice sensing from GNSS-R data using convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* 15, 1510–1514. <https://doi.org/10.1109/LGRS.2018.2852143>.
- Yan, Q., Huang, W., 2020. Sea ice thickness measurement using spaceborne GNSS-R: first results with TechDemoSat-1 data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 13, 577–587. <https://doi.org/10.1109/JSTARS.2020.2966880>.
- Zuffada, C., Elfouhaily, T., Lowe, S., 2003. Sensitivity analysis of wind vector measurements from ocean reflected GPS signals. *Remote Sens. Environ.* 88, 341–350. [https://doi.org/10.1016/S0034-4257\(03\)00175-5](https://doi.org/10.1016/S0034-4257(03)00175-5).