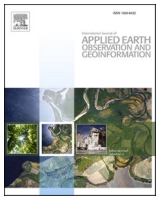
International Journal of Applied Earth Observations and Geoinformation 114 (2022) 103050

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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



Enhanced all-weather precipitable water vapor retrieval from MODIS near-infrared bands using machine learning

Jiafei Xu [a,b], Zhizhao Liu [a,b,*]

[a] Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong
[b] Research Institute for Sustainable Urban Development (RISUD), Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Global Positioning System
Machine learning
Moderate Resolution Imaging Spectroradiometer
Near-infrared
Precipitable water vapor retrieval

ABSTRACT

Four novel PWV retrieval approaches based on machine learning methods are for the first time developed to estimate all-weather precipitable water vapor (PWV) from near-infrared (NIR) measurements of the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument. The four retrieval approaches are Back Propagation Neural Network (BPNN), Gradient Boosting Decision Tree (GBDT), Generalized Regression Neural Network (GRNN), and eXtreme Gradient Boosting (XGBoost). The transmittance, latitude, longitude, elevation, cloud, season, and solar zenith angle information, in association with the MODIS NIR PWV's performance, are utilized. The in-situ one-year PWV data collected in 2017 from 453 Global Positioning System (GPS) sites in Australia and 214 GPS sites in China are utilized as target water vapor estimates for model training. Independent of the 2017 training data, two-year data observed in 2018–2019 in Australia and China are utilized to validate the four models' performance. The results indicate that the retrieval algorithms can greatly improve the PWV retrieval accuracy from MODIS NIR observations under all-weather conditions, reducing the impact of clouds on NIR PWV retrieval. The new all-weather PWV estimates obtain R^2 in the range of 0.83 ~ 0.86, root-mean-square-error (RMSE) in the range of 4.71 mm ~ 5.28 mm, and mean bias (MB) in the range of 0.18 mm ~ 0.51 mm, significantly outperforming the official MODIS NIR PWV product (R^2 = 0.31, RMSE = 12.03 mm, and MB = -3.04 mm). The reduction in RMSE is 60.85 % for BPNN, 59.68 % for GBDT, 56.69 % for GRNN, and 57.27 % for XGBoost. The new all-weather PWV results show a superior retrieval accuracy compared to the official MODIS NIR confident-clear PWV product, illustrating the effectiveness of the models. This could be because the retrieval models have considered multiple dependence parameters that affect the performance of MODIS-observed NIR PWV. The retrieval algorithms exhibit little spatial or temporal dependence and they can be applied to other regions and periods. This work provides a more accurate way to retrieve all-weather PWV estimates from satellite NIR measurements considering multiple dependence parameters – location, cloud, season, and solar zenith angle information.

1. Introduction

Water vapor is considered the most abundant and the most important atmospheric greenhouse component on the planet (Li et al., 2003; Stocker et al., 2013). It exhibits a fundamental role in atmospheric circulation (Schneider et al., 2010), energy transport (Myhre et al., 2013), hydrological cycle (Muller et al., 2009; Wang et al., 2021; Ye et al., 2014) and it has a strong association with climate change locally or globally (Held and Soden, 2000; Sherwood et al., 2010; Soden et al., 2002). Atmospheric water vapor is also considered one of the most essential parameters in the atmospheric correction process of optical satellite data (Chavez, 1988; Chen et al., 2019). The atmospheric water vapor distribution on the planet presents a high spatial–temporal variability (Ccoica-Lopez et al., 2019; Li et al., 2003; Trenberth et al., 2005). Precipitable water vapor (PWV) is equal to the total atmospheric water vapor amount measured in a vertical column and in a cross-section unit (Ichoku et al., 2002; King et al., 1992). To advance the monitoring of atmospheric-water–vapor-related processes, it is critical to obtain high-quality atmospheric PWV distribution measurements at a reasonable spatial–temporal resolution.

The instruments for in-situ PWV observations consist of sun photometer (Fragkos et al., 2019; Holben et al., 1998; Liu et al., 2013),

* Corresponding author.
E-mail address: lszzliu@polyu.edu.hk (Z. Liu).

https://doi.org/10.1016/j.jag.2022.103050
Received 3 August 2022; Received in revised form 5 October 2022; Accepted 7 October 2022
Available online 15 October 2022

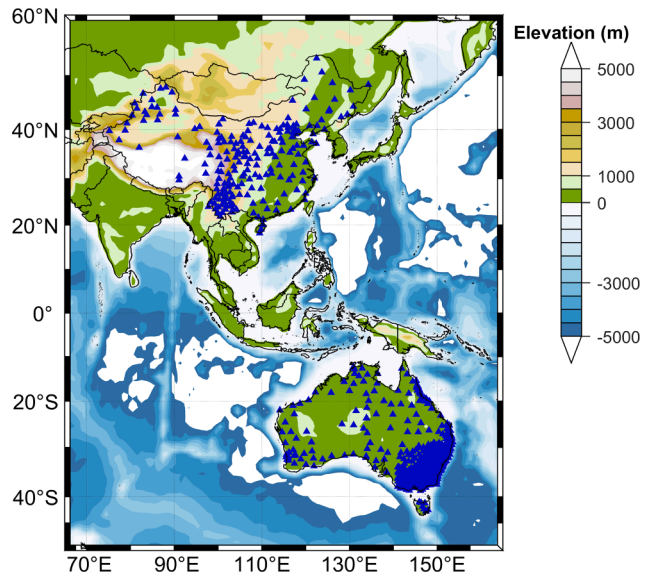


Fig. 1. Geographical locations of 453 GPS sites in Australia and 214 GPS sites in China. The color bar shows the altitude of the GPS sites.\.

Table 1
Summary of the characteristics of the GPS and MODIS datasets collected between January 1, 2017 and December 31, 2019 over both Australia and China.

Data	Time period	Time resolution	Space resolution	Region	Function
GPS	January 1, 2017 to December	Every hour	667 stations, point data	Australia and China	Model training
MOD021KM	31, 2017	Every day	1000 *m*		
MOD03		Every day	1000 *m*		
MOD35		Every day	1000 *m*		
GPS	January 1, 2018 to December	Every hour	667 stations, point data	Australia and China	Model validation
MOD021KM	31, 2019	Every day	1000 *m*		
MOD03		Every day	1000 *m*		
MOD05		Every day	1000 *m*		
MOD35		Every day	1000 *m*		

Table 2
Summary of NIR bands of the MODIS sensor utilized for PWV observations.

Band number	Center (nm)	Width (nm)	Spatial resolution (m)	Description
2	865	40	250	Window Band
5	1240	20	500	Window Band
17	905	30	1000	Absorption Band
18	936	10	1000	Absorption Band
19	940	20	1000	Absorption Band

Global Positioning System (GPS) (Bevis et al., 1994; Li et al., 2015; Rocken et al., 1997), radiosonde (McCarthy et al., 2009; Miloshevich et al., 2006; Ross and Elliot, 2001), and microwave radiometer (Fragkos et al., 2019). The most common ground-based observation techniques are GPS and radiosonde. The radiosonde-based PWV observations, limited by weather conditions, exhibit a poor temporal resolution of one or two measurements per day. (Li et al., 2003; Ross and Elliot, 2001; Vaquero-Martínez et al., 2018). In contrast, GPS can provide continuous all-weather atmospheric PWV observations at a high time resolution e.g. hourly (Bevis et al., 1994; Vaquero-Martínez et al., 2018). GPS-derived

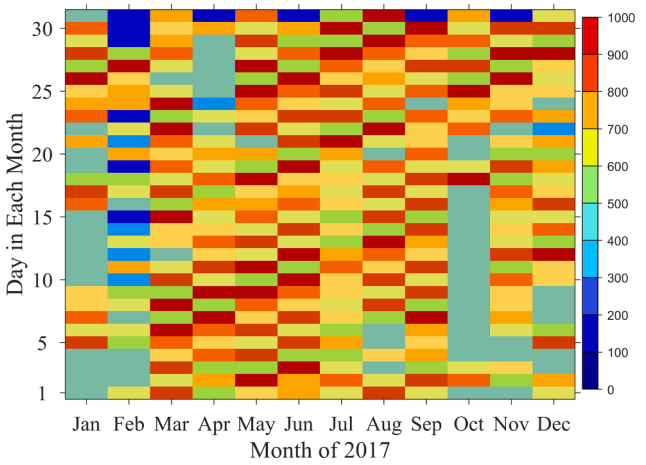


Fig. 2. The number of the matched MODIS–GPS data pairs in each day in each month in 2017 in Australia and China used for model training.

PWV measurements have been utilized as in-situ reference PWV in many satellite-based PWV studies (Vaquero-Martínez et al., 2018; Xu and Liu, 2021a).

In addition to in-situ PWV observations, Earth-orbiting remote sensing satellites are a main technique to monitor atmospheric PWV distribution with a large or even global coverage. The satellite-measured PWV can be retrieved from different spectral wavelength channels. Among them, the near-infrared (NIR) bands are much more perceptive to water vapor in the atmospheric boundary layer where most PWV exists (King et al., 2003). The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (Justice et al., 1998) can provide daily near-global PWV observations using NIR bands (Gao and Kaufman, 2003; Kaufman and Gao, 1992). Three PWV absorption bands (i.e. 905, 936, and 940 nm) and two nearby non-absorption bands (i.e. 865 and 1240 nm) are utilized (Gao and Kaufman, 2003; Kaufman and Gao, 1992). Both 2-band and 3-band ratio approaches are utilized in deriving PWV estimates from MODIS NIR measurements, which rely upon an atmospheric radiative transfer model and look-up tables (Gao and Kaufman, 2003; Kaufman and Gao, 1992). The official MODIS NIR water vapor products, i.e. MOD05 for MODIS/Terra and MYD05 for MODIS/Aqua, are generated over land, ocean, and clouds at a space resolution of 1000 m (Gao and Kaufman, 2003; Kaufman and Gao, 1992).

However, the official MODIS NIR PWV products exhibit a large retrieval error when cloud is present, as NIR observations cannot penetrate the clouds (He and Liu, 2019; Xu and Liu, 2022). The comparison between MOD05 PWV and GPS PWV showed a root-mean-square error (RMSE) of 13.07 mm in cloudy sky conditions, much larger than the retrieval error in clear sky conditions (RMSE = 5.48 mm) (He and Liu, 2019). On account of their degraded accuracy in the existence of clouds, the PWV estimates from MODIS NIR measurements cannot be used in all weather conditions (He and Liu, 2019). Most current studies utilize MODIS NIR confident-clear PWV products only, which largely limits the application of MODIS NIR PWV observations in climate studies (Ccoica-Lopez et al., 2019; Li and Long, 2020). In addition, the measurement performance of MODIS NIR PWV retrievals showed a remarkable dependence on several variables – location, cloud, season, and solar zenith angle (Vaquero-Martínez et al., 2018, 2017). Such influence factors also affect PWV retrieval accuracy of NIR observations from other satellites, such as Sentinel-3 NIR PWV products (Xu and Liu, 2021a). Therefore, an enhanced water vapor retrieval approach is greatly needed to further improve the PWV retrieval accuracy of NIR water vapor observations, especially under cloudy sky conditions.

Several water vapor retrieval algorithms have been developed to enhance the retrieval accuracy of PWV estimates from MODIS NIR observations. Ma et al. (2022) developed an enhanced water vapor

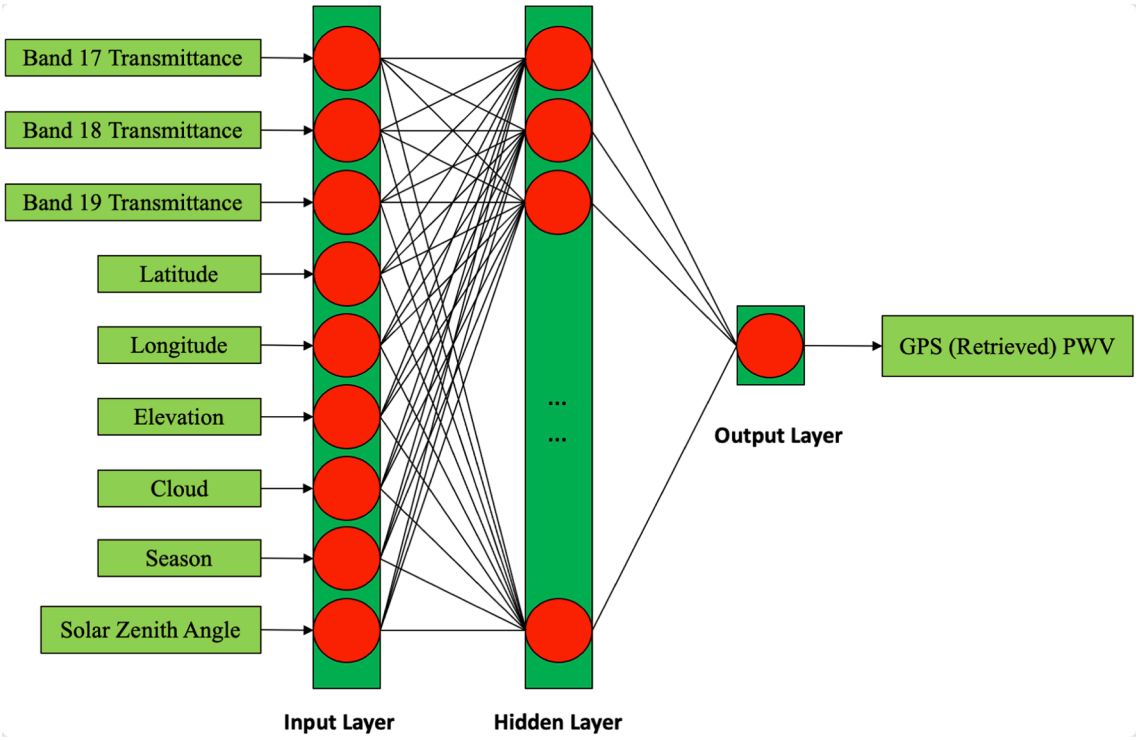


Fig. 3. The schematic of the BPNN-based method for retrieving all-weather PWV estimates from MODIS NIR measurements.

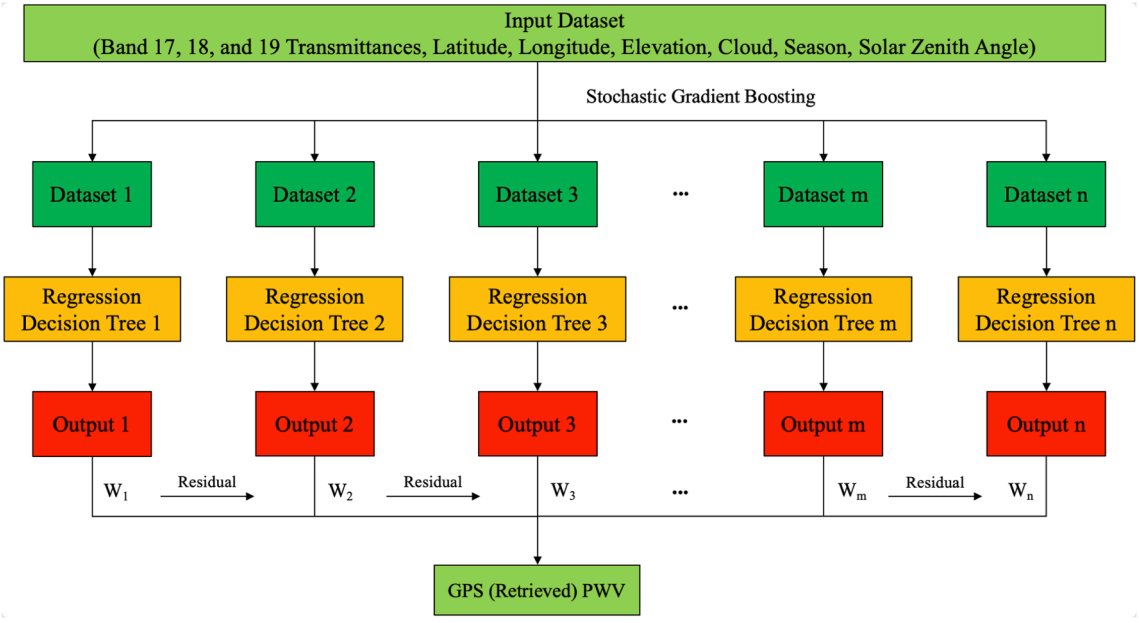


Fig. 4. The schematic of the GBDT-based method for retrieving all-weather PWV estimates from MODIS NIR measurements.

retrieval method for MODIS NIR bands relying upon a Back Propagation Neural Network (BPNN) considering land surfaces. The new PWV result reduced the confident-clear RMSE of MOD05 PWV by 66.32 % for 2-band ratio approach and by 68.67 % for 3-band ratio approach in the North America area (Ma et al., 2022). Yet this retrieval algorithm used the same dataset for model development and validation, namely-one-year data in 2020. The work in He and Liu (2020) developed an ensemble-based water vapor retrieval algorithm to retrieve PWV from MODIS NIR observations in cloud-free conditions (i.e. confident clear sky conditions), showing a reduction in RMSE of MODIS/Terra NIR PWV of 22.48 % for 2-band ratio method and 21.69 % for 3-band ratio method in the North America region. In addition to enhanced retrieval algorithms for MODIS NIR measurements, an improved PWV retrieval algorithm for Sentinel-3A NIR radiance observations was developed for confident clear condition (Xu and Liu, 2021b). The results show that the RMSE of Sentinel-3A NIR PWV versus GPS PWV in confident clear conditions drops 10.68 % ~ 12.94 % in Australia (Xu and Liu, 2021b). Nevertheless, the current existing improved PWV retrieval algorithms are built on the confident-clear sky conditions only, and they cannot address the problem of poor PWV accuracy in cloudy-sky and all-weather conditions. The NIR PWV retrieval algorithms have not taken the location, cloud, season, and solar zenith angle into account in

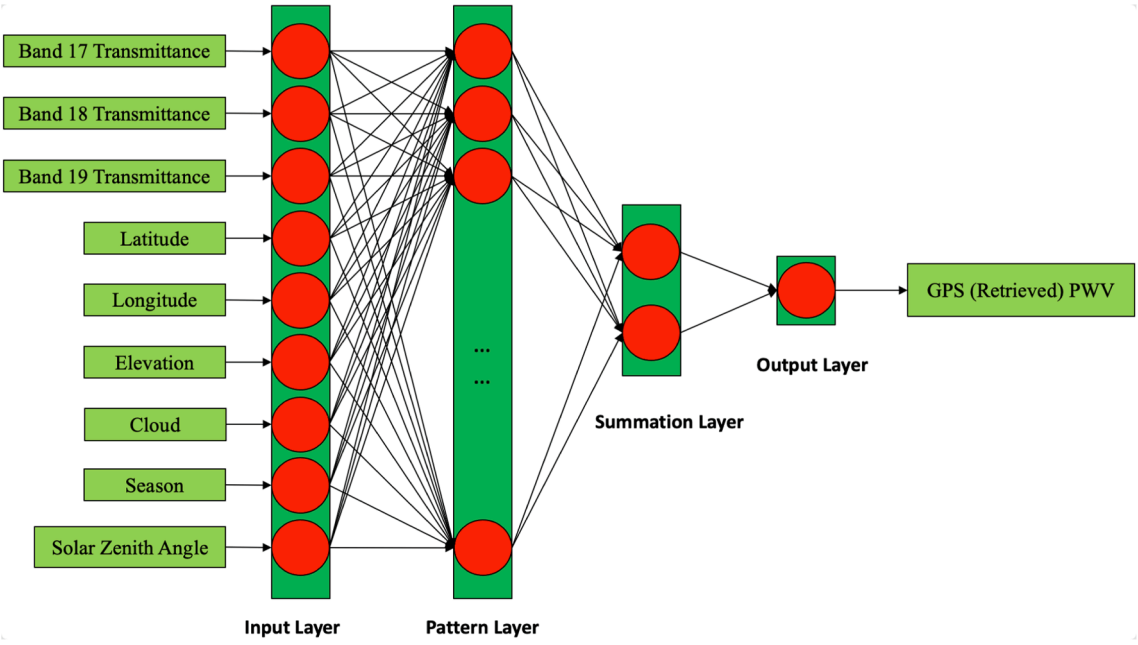


Fig. 5. The schematic of the GRNN-based method for retrieving all-weather PWV estimates from MODIS NIR measurements.

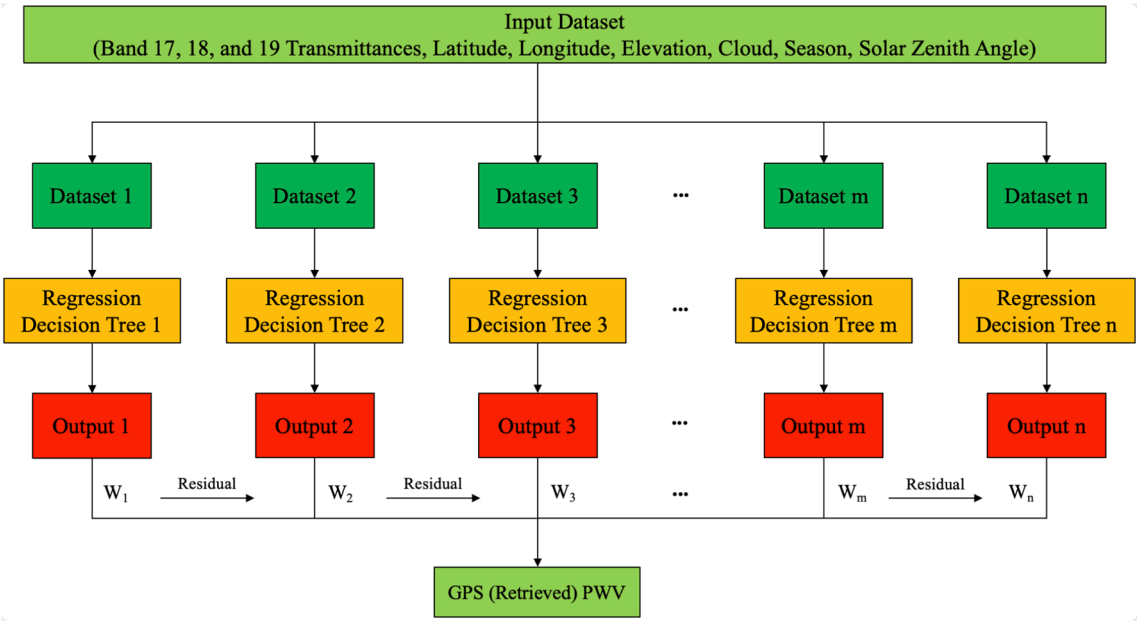


Fig. 6. The schematic of the XGBoost-based method for retrieving all-weather PWV estimates from MODIS NIR measurements.

retrieving PWV from satellite NIR measurements, though they are in strong association with the performance of satellite-sensed NIR PWV retrieval (Vaquero-Martínez et al., 2018, 2017; Xu and Liu, 2021a). To the best of our knowledge, no study has reported retrieving all-weather PWV estimates from satellite-based NIR observations with an enhanced retrieval accuracy.

In this work, we develop four new PWV retrieval algorithms based on machine learning methods to enhance the PWV retrieval accuracy from MODIS NIR channels under all weather conditions. The methods consist of BPNN (Rumelhart et al., 1986), Gradient Boosting Decision Tree (GBDT) (Friedman, 2002), Generalized Regression Neural Network (GRNN) (Specht, 1991), and eXtreme Gradient Boosting (XGBoost) (Chen et al., 2015). The input data of the retrieval algorithms include water vapor transmittance from three PWV absorption bands (i.e. 905, 936, and 940 nm), location (latitude, longitude, and elevation), cloud, season, and solar zenith angle. In the model training procedure, the in-situ one-year PWV estimates observed in 2017 from 453 GPS sites in Australia and 214 GPS sites in China are employed as water vapor data in the target output layer. The validation of the retrieval methods is performed using independent PWV data measured in 2018–2019. To the best of our knowledge, this work is the first one to retrieve all-weather PWV estimates from satellite-based NIR observations considering multiple influence parameters – location, cloud, season, and solar zenith angle.

The main objectives of this research are as follows: (1) enhancing the accuracy of PWV retrieval from MODIS NIR measurements in all-weather conditions and also in each different weather condition (i.e. confident clear, probably clear, probably cloudy, and confident cloudy);

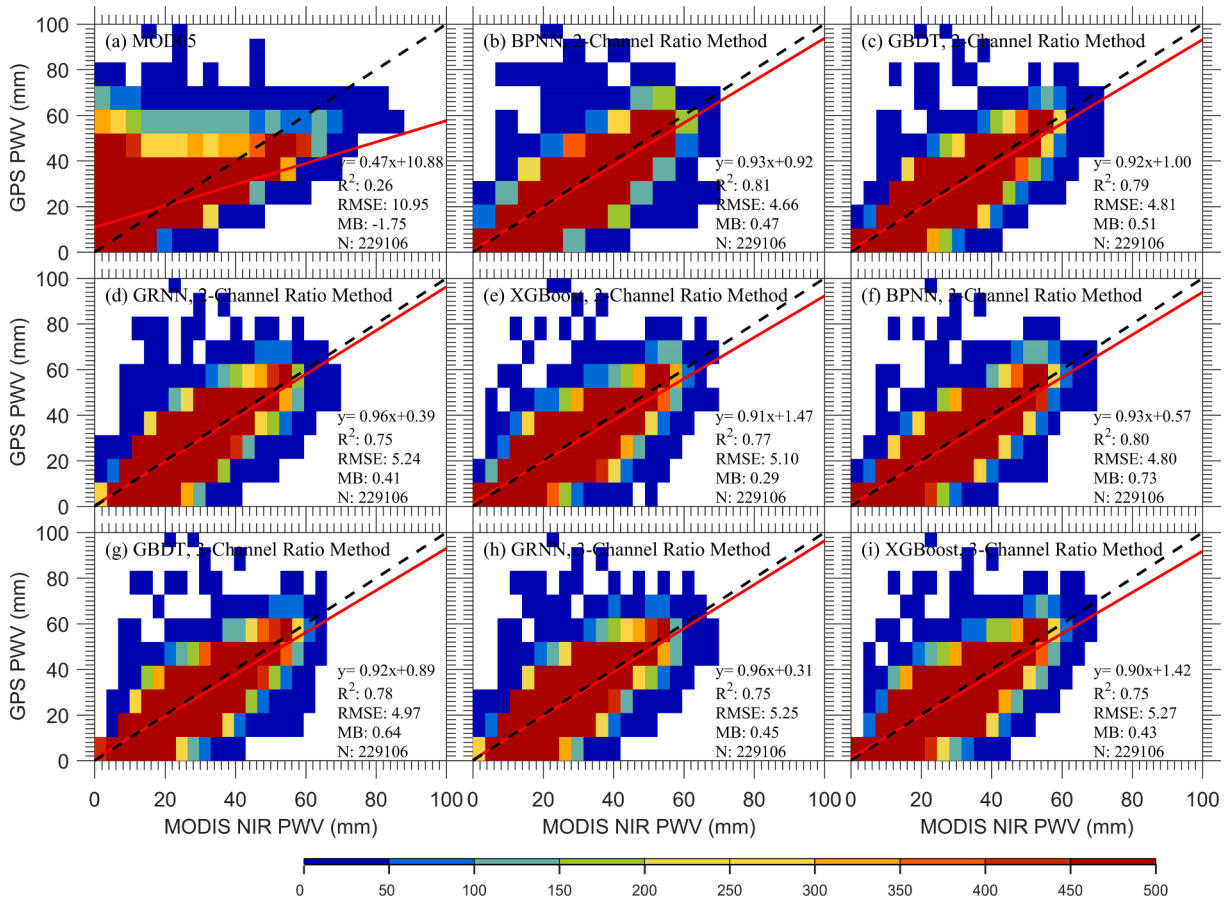


Fig. 7. Comparison of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in Australia under all weather conditions. The red line is the linear regression result between PWV from MODIS NIR and GPS observations. The dashed black line is the 1:1 line utilized as a reference. The color bar indicates the number of PWV pairs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2) comparing the performances of the four machine learning based PWV retrieval algorithms in MODIS NIR PWV estimation. It is expected that this work can provide a new approach to retrieving satellite PWV (NIR bands and other bands) with an enhanced accuracy under all-weather conditions by considering multiple dependence factors – location, cloud, season, and solar zenith angle. The GPS and MODIS data, used for model development and verification, are described in Section 2. In Section 3, the four machine learning based water vapor retrieval algorithms for MODIS NIR measurements are presented in detail. Section 4 lists the validation results of the new PWV retrievals with respect to ground-based GPS PWV observations. The discussion of the evaluation results is listed in Section 5. The major findings of this research are summarized in Section 6.

2. Data and Pre-processing

Fig. 1 indicates the research region and the spatial locations of 453 in-situ GPS sites in Australia and 214 in-situ GPS sites in China, which are employed for model development and validation in this work. The research region covers latitudes between 10°41′ S and 43°39′ S and longitudes between 112°57′ E and 153°45′ E in Australia, and latitudes in the range of 16°42′ N to 53°33′ N and longitudes in the range of 73°40′ E to 135°03′ E in China.

The Australian climate is usually humid (tropical region) and temperate (mid-latitude region), which is a representative climate in the Southern Hemisphere region. Australia has four distinct seasons – spring (September, October, and November), summer (December, January, and February), autumn (March, April, and May), and winter (June, July, and August). The climate of China presents varying characteristics from place to place, including the arid, cold, polar, and temperate climates (Feng and Du, 2020). It is representative of the climate in the Northern Hemisphere region, with four distinct seasons opposite to those of Australia, i.e. spring (March, April, and May), summer (June, July, and August), autumn (September, October, and November), and winter (December, January, and February). Both Australia and China, representative regions in the Southern and Northern Hemispheres, are selected for model development and assessment.

In this work, two types of data sets were utilized for model training and validation: ground-based GPS data and satellite-based MODIS data. The data were collected between January 1, 2017 and December 31, 2019 over both Australia and China. Table 1 lists the detailed characteristics of the GPS and MODIS data.

2.1. Ground-based GPS data

The PWV observations estimated from in-situ GPS sites are considered as ground PWV truth for model construction and verification. The Zenith Total Delay (ZTD) products, obtained from 453 GPS sites of the Geoscience Australia (Pigram, 2012) and from 214 GPS sites of the Crustal Movement Observation Network of China (CMONOC) (Wang and Zhang, 2001), were utilized to derive in-situ PWV data. This GPS PWV retrieval is processed utilizing the approach developed by Bevis et al. (1992). The GPS ZTD measurements include two distinct components, i.e. Zenith Hydrostatic Delay (ZHD) and Zenith Wet Delay (ZWD). PWV can be retrieved based on the GPS ZWD data when the water vapor weighted mean temperature is known (Yuan et al., 2014). In this work,

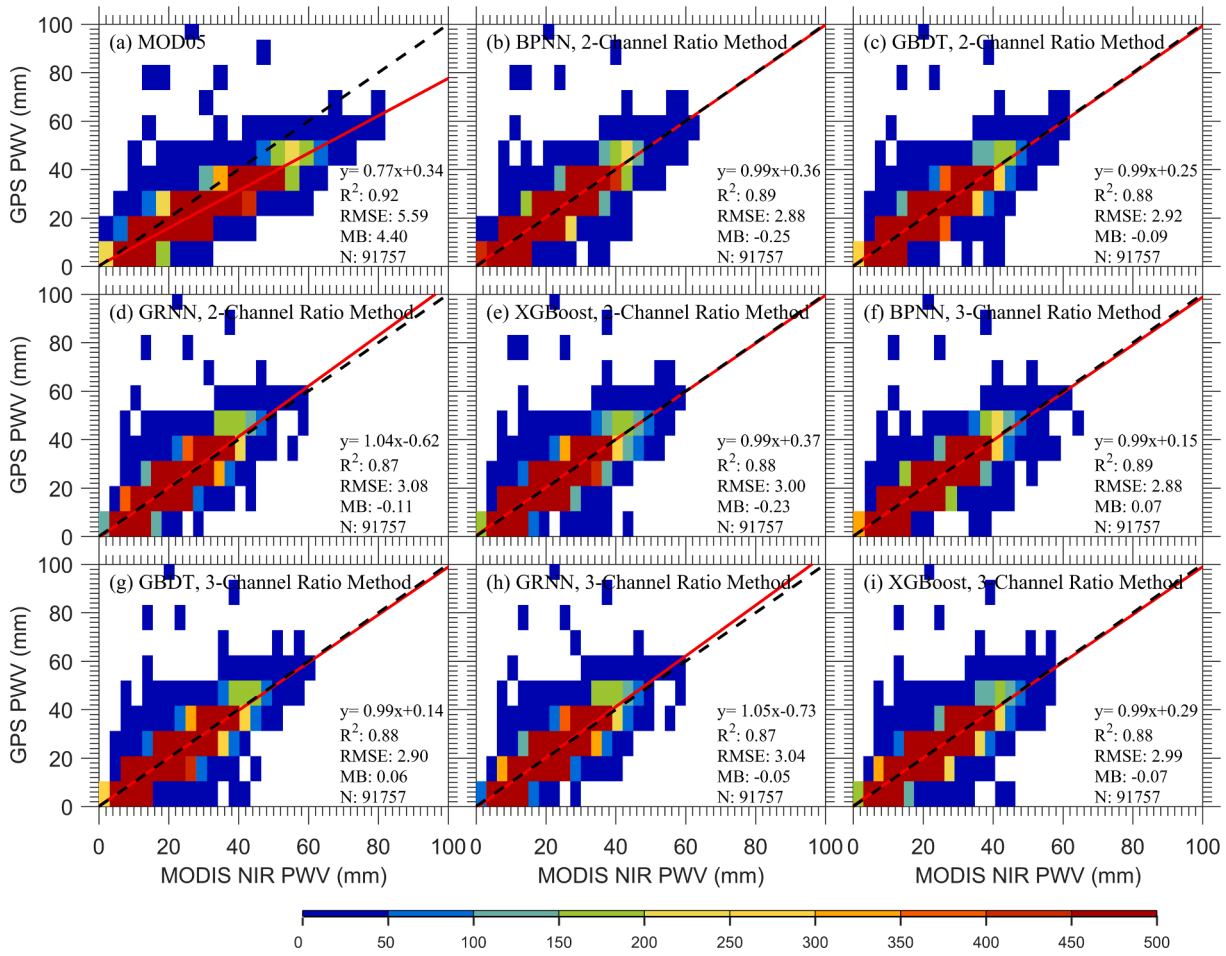


Fig. 8. Comparison of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in Australia under confident clear conditions. The red line is the linear regression result between PWV from MODIS NIR and GPS observations. The dashed black line is the 1:1 line utilized as a reference. The color bar indicates the number of PWV pairs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the atmospheric pressure and surface temperature observations from ERA5 reanalysis (Hersbach et al., 2020) were utilized to convert the GPS ZTD into the GPS PWV. The GPS-derived PWV data were obtained at a time resolution of one hour.

As displayed in Table 1, the in-situ GPS PWV data retrieved in 2017 from both Australia and China were used to develop the four machine learning based PWV retrieval algorithms. The GPS-based PWV measurements in 2018–2019 were utilized to validate the performance of these machine learning-based retrieval models.

2.2. Satellite-based MODIS data

The MODIS sensor is a hyperspectral radiometer on board both Terra and Aqua satellites (Salomonson et al., 1989). It can provide daily near-global PWV measurements using NIR channels (Gao and Kaufman, 2003; Kaufman and Gao, 1992). The PWV are estimated from MODIS NIR observations at five bands: three PWV absorption channels (i.e. 905, 936, and 940 nm) and two nearby window channels (i.e. 865 and 1240 nm) (Gao and Kaufman, 2003; Kaufman and Gao, 1992) (see Table 2).

Four types of MODIS products from the Terra satellite were utilized (see Table 1). The MODIS/Terra data products include the MOD021KM (radiance product), MOD03 (geolocation product), MOD05 (PWV product), and MOD35 (cloud mask product). The latitude, longitude, elevation, and solar zenith angle information, used as the input to the retrieval models, were obtained from the geolocation product, i.e. MOD03. The MOD35 product was employed to provide the cloud information of each MODIS pixel in the model construction, which classifies the sky weather condition into four groups, namely, confident clear, probably clear, probably cloudy, and confident cloudy (Platnick et al., 2003). The MODIS/Terra data obtained in 2017 over Australia and China were used for model training. In the model verification procedure, the MODIS/Terra products measured from January 1, 2018 to December 31, 2019 in Australia and China were utilized.

2.3. MODIS and GPS PWV pairing criteria

In order to perform the model training and validation, it is important to match the MODIS data with the GPS data. Two main collocation criteria were utilized in this paper: (1) the spatial distance between MODIS and GPS has to be the lowest, and this distance is no longer than 10 km; (2) the temporal difference between measurements of MODIS and GPS is required to be shorter than 30 min.

3. Methods

3.1. Physical basis

The radiance at the downward-looking satellite MODIS instrument in NIR bands can be calculated using a simplified form (Gao and Kaufman, 2003; Kaufman and Gao, 1992):

$$L_{sensor}(\lambda) = L_{path}(\lambda) + L_{sun}(\lambda)T(\lambda)\rho(\lambda) \quad (1)$$

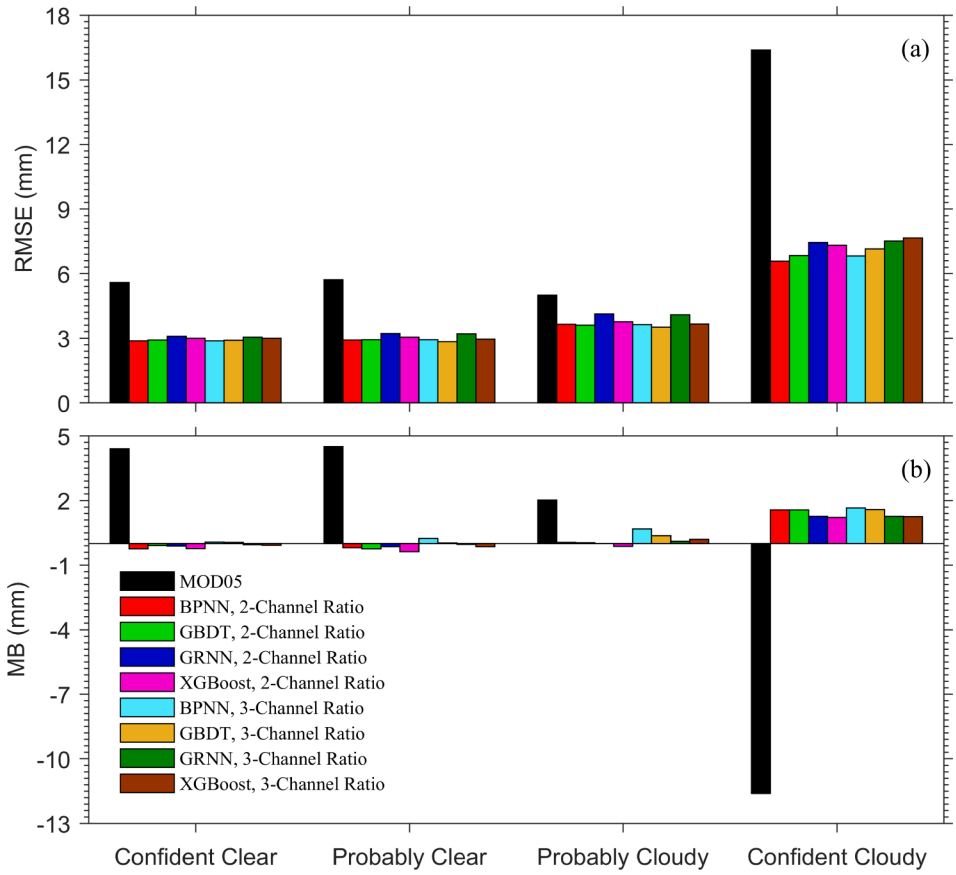


Fig. 9. Comparison of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in Australia under each weather condition. The weather condition is defined based on the cloud-mask flags from the official MOD35 product.

where $L_{sensor}(\lambda)$ is the measured radiance at the satellite sensor; λ is the wavelength in spectral bands; $L_{path}(\lambda)$ is the scattered radiance measurements in the sun-surface-sensor path; $L_{sun}(\lambda)$ is the solar radiance observations at the top of atmosphere (TOA); $T(\lambda)$ is the total atmospheric transmittance; and $\rho(\lambda)$ is the surface reflectance.

Note that the effect of the aerosol optical depth is usually small in the NIR bands. Hence, the scattered radiance observations (i.e. $L_{path}(\lambda)$) can be ignored in Eq. (1) (Kaufman and Gao, 1992) and it can be approximated as:

$$\rho^*(\lambda) = T(\lambda)\rho(\lambda) \tag{2}$$

where $\rho^*(\lambda)$ is the TOA reflectance, defined as $L_{sensor}(\lambda)/L_{sun}(\lambda)$.

At a known spectral wavelength, the reflectance in the surface varies greatly with different types of surface (Gao and Kaufman, 2003; Kaufman and Gao, 1992). It is not possible to obtain PWV transmittance using the radiance observations of individual NIR absorption band of the MODIS instrument (Gao and Kaufman, 2003; Kaufman and Gao, 1992). A differential absorption algorithm has been developed to determine the atmospheric water vapor transmittance relying upon the ratio of the radiance between one water vapor absorption band with one or two non-absorption bands (Schläpfer et al., 1998). If the surface reflectance remains stable with the changing wavelength, the atmospheric water vapor transmittance of the individual absorption band can be estimated from a 2-band ratio of one absorption band to a nearby non-absorption band (Gao and Kaufman, 2003; Kaufman and Gao, 1992). It is calculated as:

$$T_i = \frac{\rho_i^*}{\rho_2^*} \tag{3}$$

where i is the absorption channel 17 (905 nm), 18 (936 nm), and 19 (940 nm) of the MODIS instrument; T_i is the atmospheric water vapor transmittance in the absorption channel i; ρ_i^* is the TOA reflectance in the absorption band i; ρ_2^* is the TOA reflectance in the non-absorption band 2 of the MODIS instrument (865 nm).

If the surface reflectance changes linearly with the varying wavelength, the atmospheric water vapor transmittance of the individual absorption band can be computed from a 3-band ratio of one absorption band to two nearby non-absorption bands (Gao and Kaufman, 2003; Kaufman and Gao, 1992). It is written as:

$$T_i = \frac{\rho_i^*}{(0.8\rho_2^* + 0.2\rho_5^*)} \tag{4}$$

where ρ_5^* is the TOA reflectance in the window channel 5 of the MODIS instrument (1240 nm).

3.2. Water vapor retrieval from MODIS NIR bands using Machine learning

The four machine learning algorithms, i.e. BPNN, GBDT, GRNN, and XGBoost, were utilized to retrieve the all-weather PWV estimation from MODIS NIR observations. In the four retrieval models, the latitude, longitude, elevation, cloud, season, and solar zenith angle information was included, which are linked to the retrieval performance of the satellite NIR PWV observations (Vaquero-Martínez et al., 2018, 2017; Xu and Liu, 2021a).

The retrieval of PWV is an inverse problem relating the PWV to the transmittance, latitude, longitude, elevation, cloud, season, and solar zenith angle. PWV can be determined from each MODIS absorption band, and the PWV values from the three absorption bands of MODIS are different due to their different sensitivities though they are in the same

Table 3
The seasonal comparison between the satellite-based MODIS NIR PWV retrievals against the reference ground-based GPS PWV retrievals in 2018–2019 in Australia under all weather conditions as well as confident clear conditions.

		All Weather						Confident Clear					
		Slope	Offset	R^2	RMSE (mm)	MB (mm)	RMSE reduction	Slope	Offset	R^2	RMSE (mm)	MB (mm)	RMSE reduction
Spring	MOD05	0.43	10.11	0.22	9.23	−1.82	–	0.78	0.44	0.89	4.58	3.44	–
	BPNN, 2-Channel Ratio	0.90	1.01	0.74	4.47	0.78	51.57 %	0.96	0.53	0.85	2.80	0.09	38.86 %
	GBDT, 2-Channel Ratio	0.89	1.08	0.71	4.69	0.91	49.19 %	0.96	0.35	0.85	2.82	0.29	38.43 %
	GRNN, 2-Channel Ratio	0.94	0.26	0.67	4.92	0.82	46.70 %	1.01	−0.56	0.83	2.93	0.40	36.03 %
	XGBoost, 2-Channel Ratio	0.86	1.75	0.67	5.03	0.73	45.50 %	0.96	0.38	0.84	2.84	0.15	37.99 %
	BPNN, 3-Channel Ratio	0.91	0.50	0.73	4.52	0.98	51.03 %	0.95	0.28	0.85	2.79	0.38	39.08 %
	GBDT, 3-Channel Ratio	0.88	1.15	0.69	4.82	0.95	47.78 %	0.95	0.40	0.85	2.84	0.36	37.99 %
	GRNN, 3-Channel Ratio	0.94	0.21	0.67	4.91	0.84	46.80 %	1.02	−0.64	0.84	2.89	0.42	36.90 %
	XGBoost, 3-Channel Ratio	0.85	1.74	0.66	5.12	0.84	44.53 %	0.95	0.47	0.84	2.91	0.29	36.46 %
Summer	MOD05	0.35	17.16	0.16	13.78	−2.53	–	0.77	0.86	0.90	6.59	5.40	–
	BPNN, 2-Channel Ratio	0.94	1.00	0.77	5.58	0.55	59.51 %	0.99	0.67	0.86	3.36	−0.48	49.01 %
	GBDT, 2-Channel Ratio	0.92	1.40	0.76	5.74	0.63	58.35 %	0.99	0.45	0.85	3.44	−0.32	47.80 %
	GRNN, 2-Channel Ratio	0.98	0.11	0.69	6.39	0.51	53.63 %	1.07	−1.15	0.84	3.63	−0.31	44.92 %
	XGBoost, 2-Channel Ratio	0.91	2.01	0.74	6.01	0.36	56.39 %	0.99	0.65	0.84	3.55	−0.48	46.13 %
	BPNN, 3-Channel Ratio	0.94	0.57	0.75	5.82	1.02	57.76 %	1.00	−0.07	0.86	3.32	0.00	49.62 %
	GBDT, 3-Channel Ratio	0.92	1.19	0.74	5.97	0.84	56.68 %	0.99	0.33	0.85	3.40	−0.17	48.41 %
	GRNN, 3-Channel Ratio	0.98	−0.06	0.69	6.43	0.58	53.34 %	1.07	−1.35	0.84	3.60	−0.22	45.37 %
	XGBoost, 3-Channel Ratio	0.90	2.11	0.71	6.31	0.55	54.21 %	0.99	0.59	0.84	3.51	−0.31	46.74 %
Autumn	MOD05	0.37	13.15	0.18	11.63	−1.65	–	0.77	0.00	0.91	6.26	5.30	–
	BPNN, 2-Channel Ratio	0.90	1.65	0.79	4.75	0.44	59.16 %	0.98	0.66	0.88	2.82	−0.39	54.95 %
	GBDT, 2-Channel Ratio	0.90	1.66	0.77	4.89	0.39	57.95 %	0.97	0.71	0.88	2.85	−0.27	54.47 %
	GRNN, 2-Channel Ratio	0.94	0.80	0.73	5.27	0.43	54.69 %	1.04	−0.4	0.86	3.07	−0.35	50.96 %
	XGBoost, 2-Channel Ratio	0.88	2.33	0.74	5.22	0.11	55.12 %	0.97	0.96	0.87	2.99	−0.43	52.24 %
	BPNN, 3-Channel Ratio	0.92	1.02	0.78	4.84	0.68	58.38 %	0.97	0.51	0.88	2.84	−0.07	54.63 %
	GBDT, 3-Channel Ratio	0.91	1.29	0.76	4.98	0.58	57.18 %	0.97	0.49	0.88	2.82	−0.05	54.95 %
	GRNN, 3-Channel Ratio	0.94	0.67	0.72	5.29	0.50	54.51 %	1.05	−0.62	0.86	3.02	−0.26	51.76 %
	XGBoost, 3-Channel Ratio	0.89	2.03	0.73	5.34	0.30	54.08 %	0.97	0.74	0.87	2.96	−0.24	52.72 %
Winter	MOD05	0.37	8.30	0.17	7.22	−0.82	–	0.75	0.46	0.84	4.11	3.09	–
	BPNN, 2-Channel Ratio	0.94	0.72	0.72	3.26	0.01	54.85 %	0.97	0.43	0.83	2.21	−0.16	46.23 %
	GBDT, 2-Channel Ratio	0.95	0.63	0.72	3.30	0.03	54.29 %	0.96	0.43	0.83	2.20	−0.02	46.47 %
	GRNN, 2-Channel Ratio	1.02	0.00	0.65	3.69	−0.25	48.89 %	1.05	−0.44	0.82	2.31	−0.16	43.80 %
	XGBoost, 2-Channel Ratio	0.92	1.06	0.68	3.53	−0.09	51.11 %	0.96	0.53	0.83	2.25	−0.10	45.26 %
	BPNN, 3-Channel Ratio	0.96	0.35	0.70	3.42	0.13	52.63 %	0.96	0.48	0.82	2.30	−0.02	44.04 %
	GBDT, 3-Channel Ratio	0.95	0.48	0.69	3.45	0.11	52.22 %	0.96	0.30	0.83	2.20	0.14	46.47 %
	GRNN, 3-Channel Ratio	1.03	−0.12	0.65	3.69	−0.24	48.89 %	1.06	−0.58	0.82	2.27	−0.13	44.77 %
	XGBoost, 3-Channel Ratio	0.92	1.03	0.66	3.66	−0.02	49.31 %	0.96	0.38	0.83	2.21	0.04	46.23 %

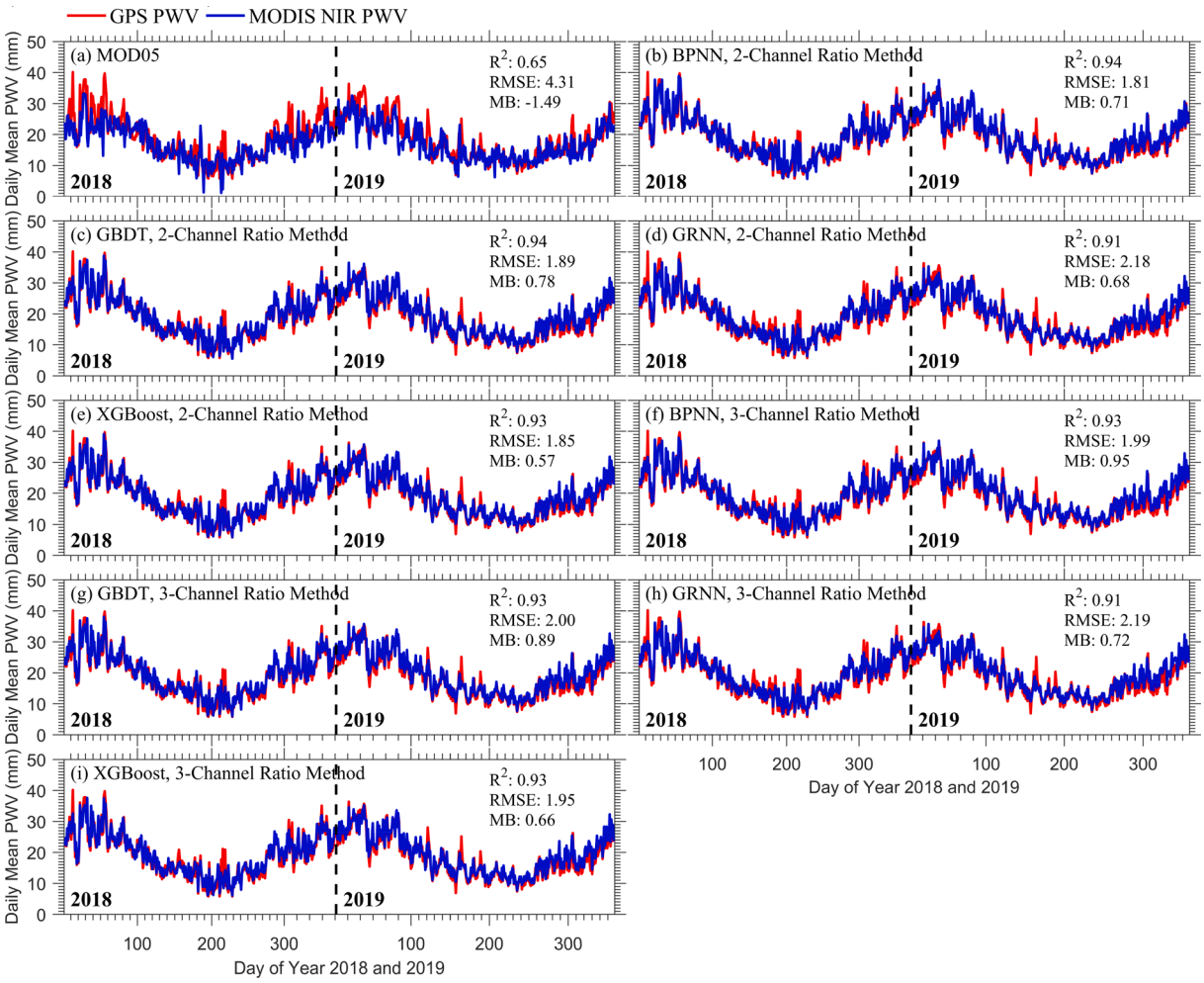


Fig. 10. Time-series variation of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in Australia under all weather conditions.

atmospheric conditions (Gao and Kaufman, 2003; Kaufman and Gao, 1992). The MODIS NIR water vapor estimates are retrieved using all the three MODIS NIR absorption bands in this paper. It is defined as:

$$W = F(T_{17}, T_{18}, T_{19}, lat, lon, ele, cld, sen, sza) \quad (5)$$

where W is the PWV from MODIS NIR measurements; T_{17}, T_{18}, and T_{19} are the water vapor transmittance at the absorption band 17 (wavelength 905 nm), 18 (936 nm), and 19 (940 nm) of the MODIS instrument, respectively; *lat* is the latitude in the unit of degree; *lon* is the longitude in the unit of degree; *ele* is the elevation in the unit of meter; *cld* is the cloud information; *sen* is the seasonal information; *sza* is the solar zenith angle in the unit of degree. In this equation, the cloud information of the retrieval models is denoted using numbers "1", "2", "3", and "4", which represent the confident clear, probably clear, probably cloudy, and confident cloudy condition, respectively. The seasonal information is characterized using numbers "1" (spring), "2" (summer), "3" (autumn), and "4" (winter).

The function F of this work is defined utilizing BPNN, GBDT, GRNN, and XGBoost. In the model training procedure, the in-situ high-accuracy PWV data, collected in 2017 from 453 GPS sites located in Australia and 214 GPS sites situated in China, were utilized. Fig. 2 presents the number of matched data pairs between MODIS and GPS data in each day in each month in 2017, which are utilized for training the four machine learning based water vapor retrieval methods.

In the optimization of the parameters of the retrieval methods, a 5-fold cross validation approach is utilized. The detailed description of the MODIS NIR all-weather PWV retrieval from BPNN, GBDT, GRNN, and XGBoost approaches can be found in the following subsections.

3.2.1. BPNN

The BPNN model was developed by Rumelhart et al. (1986) as a solution to the problem of classification and regression. It comprises three-type layers, i.e. the input, hidden, and output layers. In BPNN, the neuron number in the input (output) layer is equal to the size of the input (output) learning sample data. The number of the hidden layer and the number of neurons in each hidden layer need to be defined. In this work, we only utilized one hidden layer in the BPNN-based water vapor retrieval approach, as a one-hidden-layer neural network is generally enough to reduce the error caused by the bounded continuous functions (Hornik et al., 1989). The number of neurons of this hidden layer of the BPNN method was set to the integer power of 2, i.e. 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024 (Hornik et al., 1989).

Each neuron in the input or hidden layers can be directly connected to each of neurons in the next hidden or output layer through the activation functions. We selected the hyperbolic tangent function as the activation function between neurons of the input layer and neurons of the hidden layer. It is written as:

$$f(x) = \frac{2}{1 + exp(-2x)} - 1 \quad (6)$$

The linear function was selected as the activation function between neurons of the hidden layer and neurons of the output layer. It is

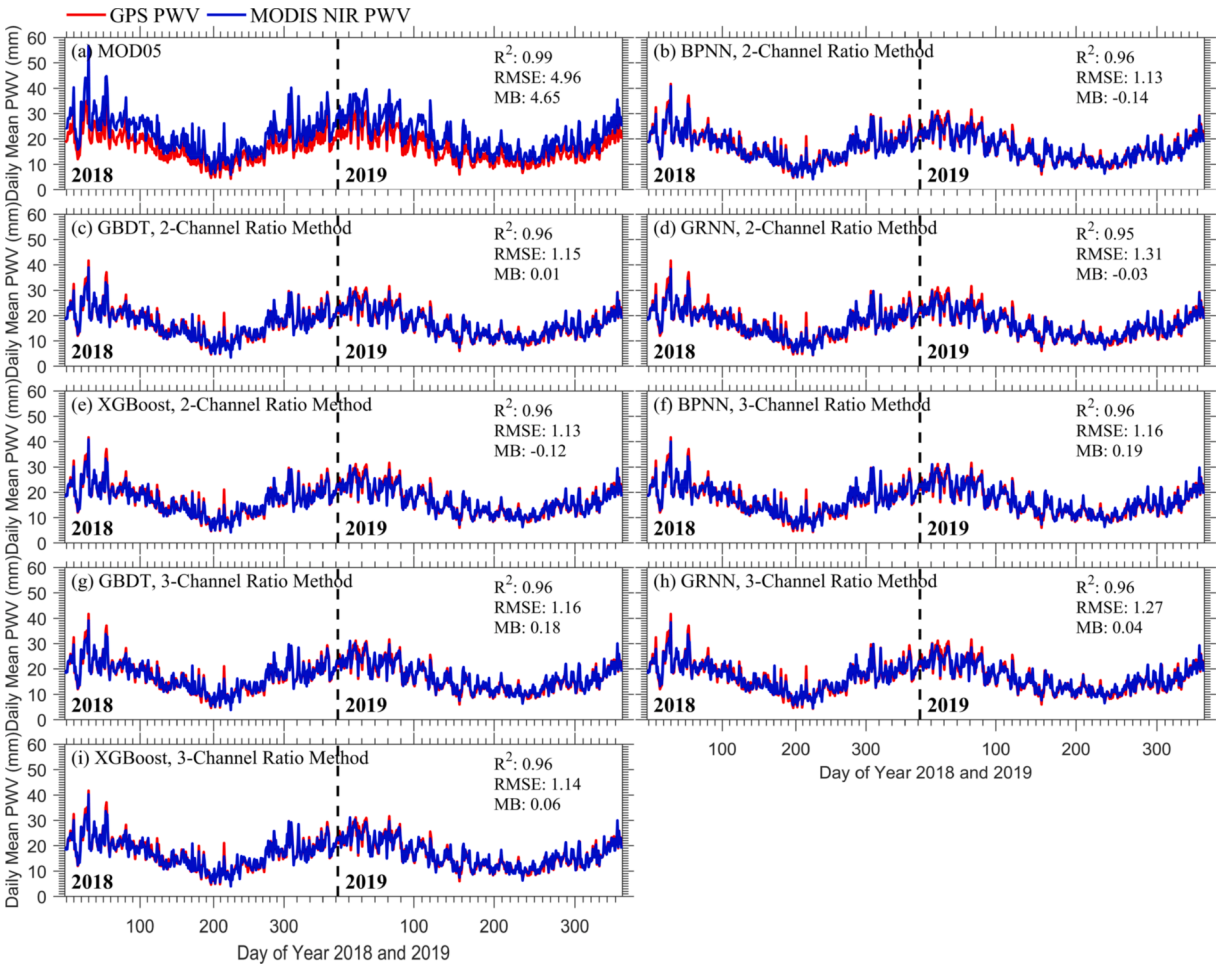


Fig. 11. Time-series variation of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in Australia under confident clear conditions.

determined as:

$$g(x) = x \tag{7}$$

The final output result of the BPNN-based retrieval approach can be calculated as:

$$W(\boldsymbol{X}) = g(\boldsymbol{\beta}_2 \bullet f(\boldsymbol{\beta}_1 \bullet \boldsymbol{X} + \boldsymbol{c}_1) + \boldsymbol{c}_2) \tag{8}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are weight matrices; $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are bias matrices; $\boldsymbol{X}$ is the variable vector in the input layer; $\boldsymbol{W(X)}$ is the final PWV retrieval result in the output layer.

The structure of the all-weather PWV retrieval from MODIS NIR bands using the BPNN algorithm is displayed in Fig. 3. The input data of the BPNN model consist of water vapor transmittances from three absorption bands, latitude, longitude, elevation, cloud, season, and solar zenith angle, and hence, the input layer in the BPNN-based retrieval approach has 9 neurons. The output of the BPNN model is the GPS-derived PWV data in the training process or the MODIS-derived NIR PWV data in the retrieval process. The optimal neuron number in the hidden layer was computed to 64 in this BPNN-based retrieval method.

3.2.2. GBDT

GBDT is a tree-based assemble method that can be utilized for both classification and regression problems. It was first proposed by Friedman in 2002 (Friedman, 2002). In GBDT model, the tree is fit based on the residual of the former tree, reducing the biases in the training process.

Fig. 4 lists the structure of the GBDT-based all-weather water vapor retrieval algorithm using MODIS NIR observations. The three-absorption-channel transmittances, latitude, longitude, elevation, cloud, season, and solar zenith angle data are included as the input source in the GBDT-based retrieval method. The output of the GBDT approach is the GPS-retrieved PWV data in the training stage or the MODIS-retrieved NIR PWV data in the retrieval stage. The final output result of the GBDT-based retrieval model is the weighted mean result estimated from the output results of each regression tree. It can be written as:

$$W(\boldsymbol{X}) = \sum_{j=1}^{n} \boldsymbol{W}_j(\boldsymbol{X})\beta_j \tag{9}$$

where n is the quantity of regression trees; $\boldsymbol{W}_j\boldsymbol{(X)}$ is the output PWV result in the regression tree j; β_j is the weighted mean coefficient in the regression tree j; $\boldsymbol{X}$ is the input parameters; $\boldsymbol{W(X)}$ is the final output PWV result in the GBDT algorithm.

In the GBDT model training, we set the number of regression trees in the range of 5 to 500 at a step of 5. The maximum depth of each regression tree was set to in the range between 1 and 50 at a step of 1. In this research, the number of regression trees was optimized to 200, with an optimal maximum depth of 10.

3.2.3. GRNN

The GRNN approach, first developed by Specht (1991), has been considered a widely used method in regression, especially in non-linear regression problems. The GRNN method consists of four layers: the input, pattern, summation, and output layers. The neuron number in the input or pattern layer is equivalent to the size of each input training

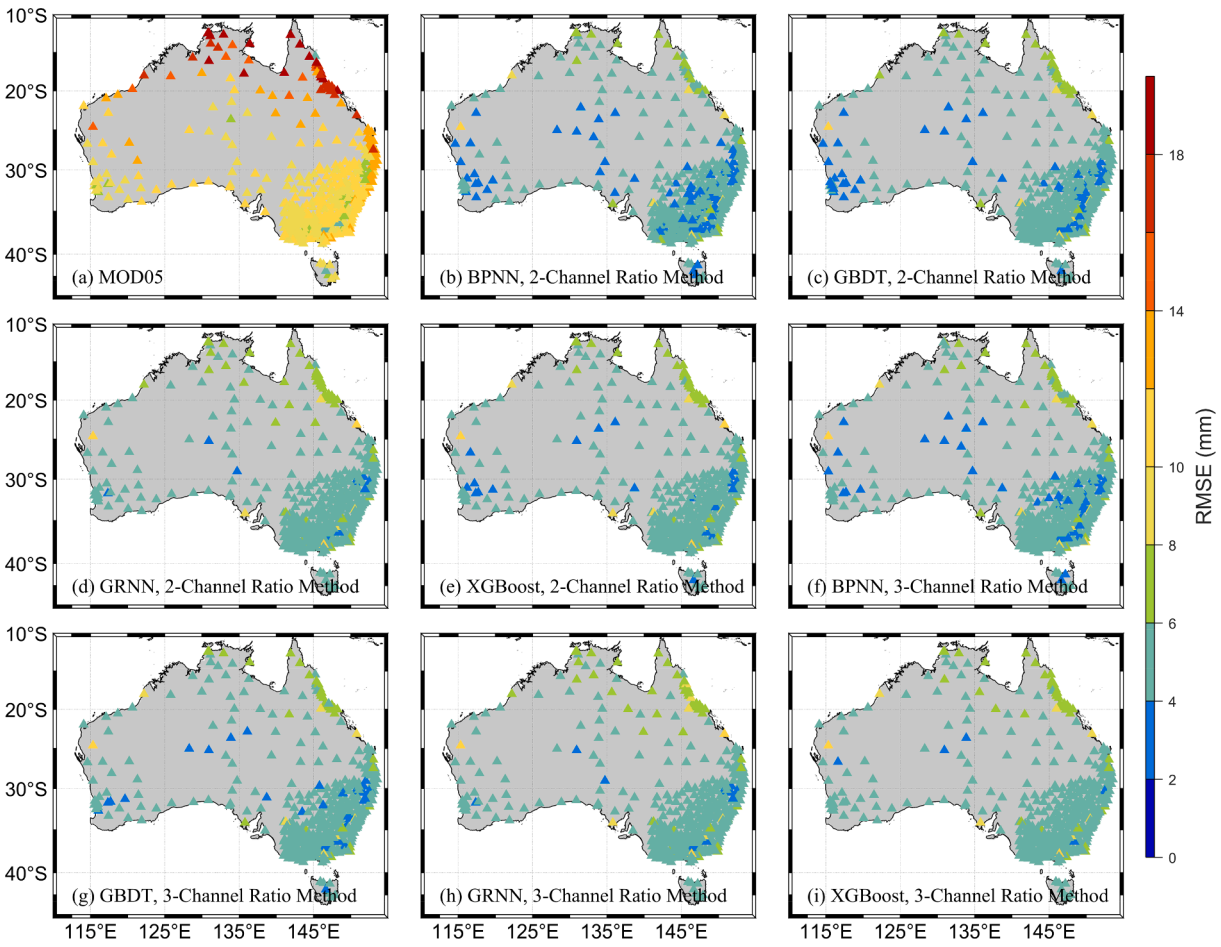


Fig. 12. Annual RMSE map for each in-situ GPS site for satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in Australia under all weather conditions.

sample data. The GRNN model has two types of summation neurons in the summation layer, which are respectively employed to obtain the arithmetic summation result and the weighted summation result from the output results of all neurons in the pattern layer. The neurons of the output layer are equivalent to the size of each output learning sample data.

The GRNN-based structure for estimating all-weather PWV from MODIS NIR observations is presented in Fig. 5. There are nine neurons in both input and pattern layers associated with the water vapor transmittances from three absorption channels, latitude, longitude, elevation, cloud, season, and solar zenith angle. The output layer includes one neuron only, which is the GPS-estimated PWV data in the training phase or the MODIS-estimated NIR PWV data in the retrieval phase. The activation function of the pattern layer in the GRNN model can be calculated as:

$$W_k = exp\left\{ -\frac{(\boldsymbol{X}-\boldsymbol{X}_k)^T(\boldsymbol{X}-\boldsymbol{X}_k)}{2\sigma^2} \right\} \tag{10}$$

where W_k is the output result of the neuron k in the pattern layer; $\boldsymbol{X}_k$ is the k-th training sample vector; $\boldsymbol{X}$ is the testing sample vector; σ is the spread coefficient.

In the GRNN-based retrieval algorithm, only the spread coefficient (i.e. σ) needs to be defined. We defined a set of spread parameter values in the range between 0.01 and 2 at a step of 0.01. The final optimal spread coefficient (σ) was calculated to 0.10.

3.2.4. XGBoost

The XGBoost is a scalable end-to-end tree boosting approach proposed by Chen et al. (2015). It is a new implementation for the GBDT method developed by Friedman (2002). The XGBoost model can be employed for addressing both regression and classification problems. In contrast to the GBDT algorithm that only performs a first-order Taylor expansion on the error component, the XGBoost algorithm can perform a second-order Taylor expansion on the error component. The XGBoost method can also perform parallel computing automatically, whereas the GBDT approach can perform single-line calculation only. Based on the additive training strategies, the XGBoost can combine all the predictions from a set of weak learners into a strong learner. The final output in the XGBoost model is the weighted mean of the output of each regression tree. It can be written as:

$$W(\boldsymbol{X}) = \sum_{j=1}^{n} \boldsymbol{W}_j(\boldsymbol{X})\beta_j \tag{11}$$

where n is the number of regression trees, $\boldsymbol{W_j(X)}$ is the output PWV result in the regression tree j, β_j is the weighted mean coefficient in the regression tree j, $\boldsymbol{X}$ is the input variables, and $\boldsymbol{W(X)}$ is the final output PWV result in the XGBoost algorithm.

Fig. 6 shows the structure of the all-weather PWV estimation from MODIS NIR channels based on the XGBoost method. In XGBoost, the water vapor transmittance from three absorption channels, latitude, longitude, elevation, cloud, season, and solar zenith angle are included as the input data. There is only one output data source that is the GPS PWV data in the training state or the MODIS NIR PWV data in the retrieval state.

For model parameter setting, we selected the gradient boosted tree (gbtree) booster as the general approach in the XGBoost method in this

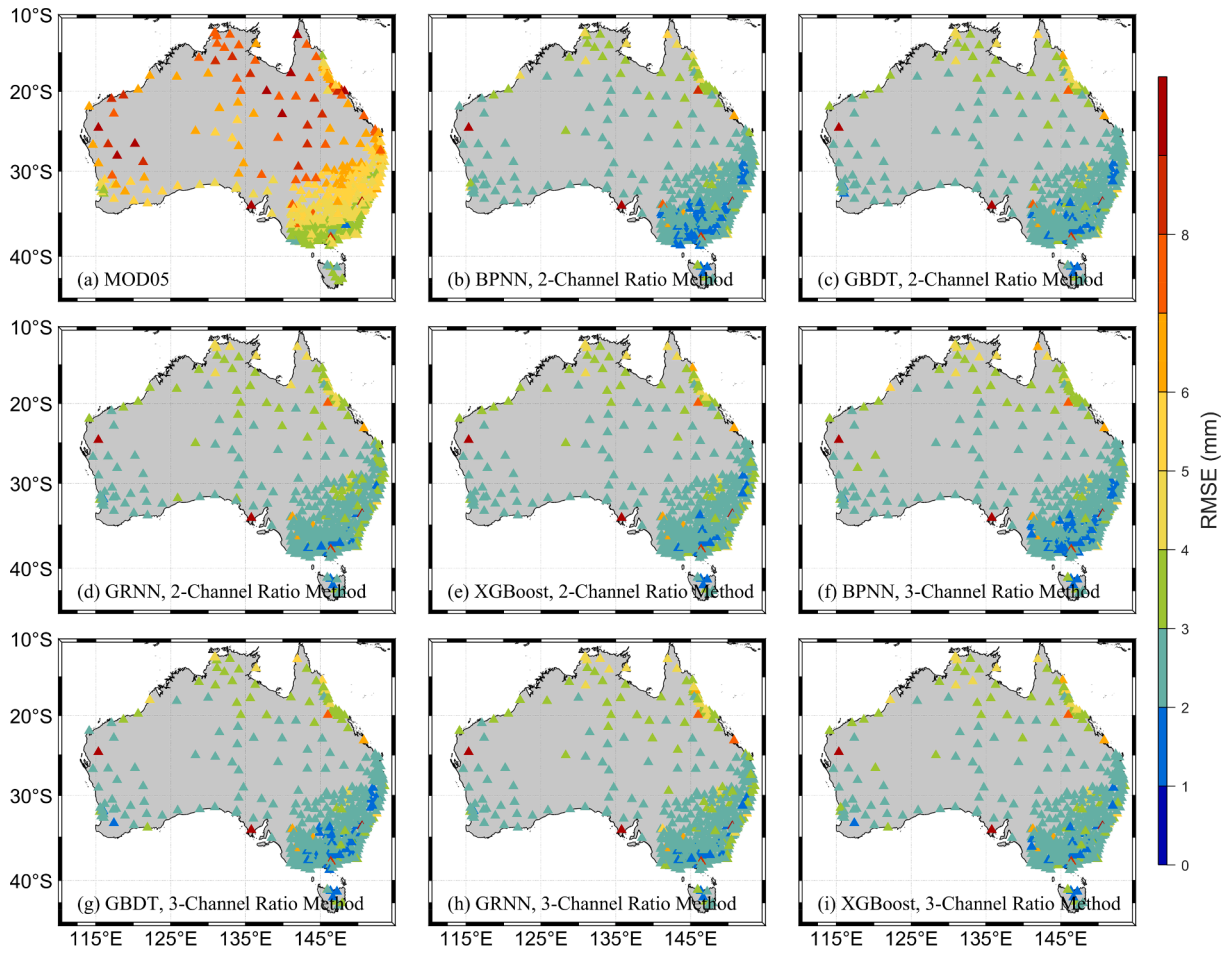


Fig. 13. Annual RMSE map for each in-situ GPS site for satellite-based MODIS-based NIR PWV observations against ground-based GPS-based PWV observations in 2018–2019 in Australia, in confident clear conditions.

study. The number of regression trees of the XGBoost-based retrieval algorithm was optimally computed to 200 from a set of numbers between 5 and 500 at a step of 5. We utilized a set of numbers in the range of 1 to 50 at a step of 1 to estimate the optimal maximum depth of each regression tree. The optimal maximum depth for the XGBoost model was determined to 12.

3.3. Model performance assessment

In this study, we utilized three evaluation metrics to assess the performance of the water vapor estimation from MODIS NIR bands using the four water vapor retrieval methods (i.e. BPNN, GBDT, GRNN, and XGBoost). The water vapor data measured from in-situ GPS stations were used as reference PWV values in the validation process. The three statistical metrics include the correlation coefficient (R^2), RMSE, and mean bias (MB). The R^2 is used to provide the strength information between the MODIS-measured PWV data against the reference GPS PWV data. It can be calculated as:

$$R^2 = \left[\frac{\sum_{i=1}^{N}(PWV_M - \overline{PWV}_M)(PWV_R - \overline{PWV}_R)}{\sqrt{\sum_{i=1}^{N}(PWV_M - \overline{PWV}_M)^2 (PWV_R - \overline{PWV}_R)^2}} \right]^2 \quad (12)$$

The RMSE is used to indicate the difference information between the matched PWV data sets and it can be written as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(PWV_M - PWV_R)^2} \quad (13)$$

The MB is used to show the underestimation or the overestimation information of the MODIS-based PWV measurements versus the GPS PWV measurements. It can be defined as:

$$MB = \frac{1}{N}\sum_{i=1}^{N}(PWV_M - PWV_R) \quad (14)$$

In the above equations, PWV_M is the PWV estimates retrieved from MODIS NIR observations, $\overline{PWV}_M$ is the averaged PWV estimates retrieved from MODIS NIR observations, PWV_R is the PWV estimates retrieved from in-situ GPS observations, $\overline{PWV}_R$ is the averaged PWV estimates retrieved from in-situ GPS observations, and N is the total quantity of matched MODIS–GPS data points.

4. Results

To study the performance of the four retrieval methods, we first utilized the GPS and MODIS data sets in 2017 in Australia and China to train the four retrieval methods. Then we applied the developed retrieval approaches to retrieve the new MODIS NIR all-weather PWV estimates from 2018 to 2019 in Australia and China. The new all-weather PWV retrievals were evaluated with respect to ground-based GPS reference PWV data.

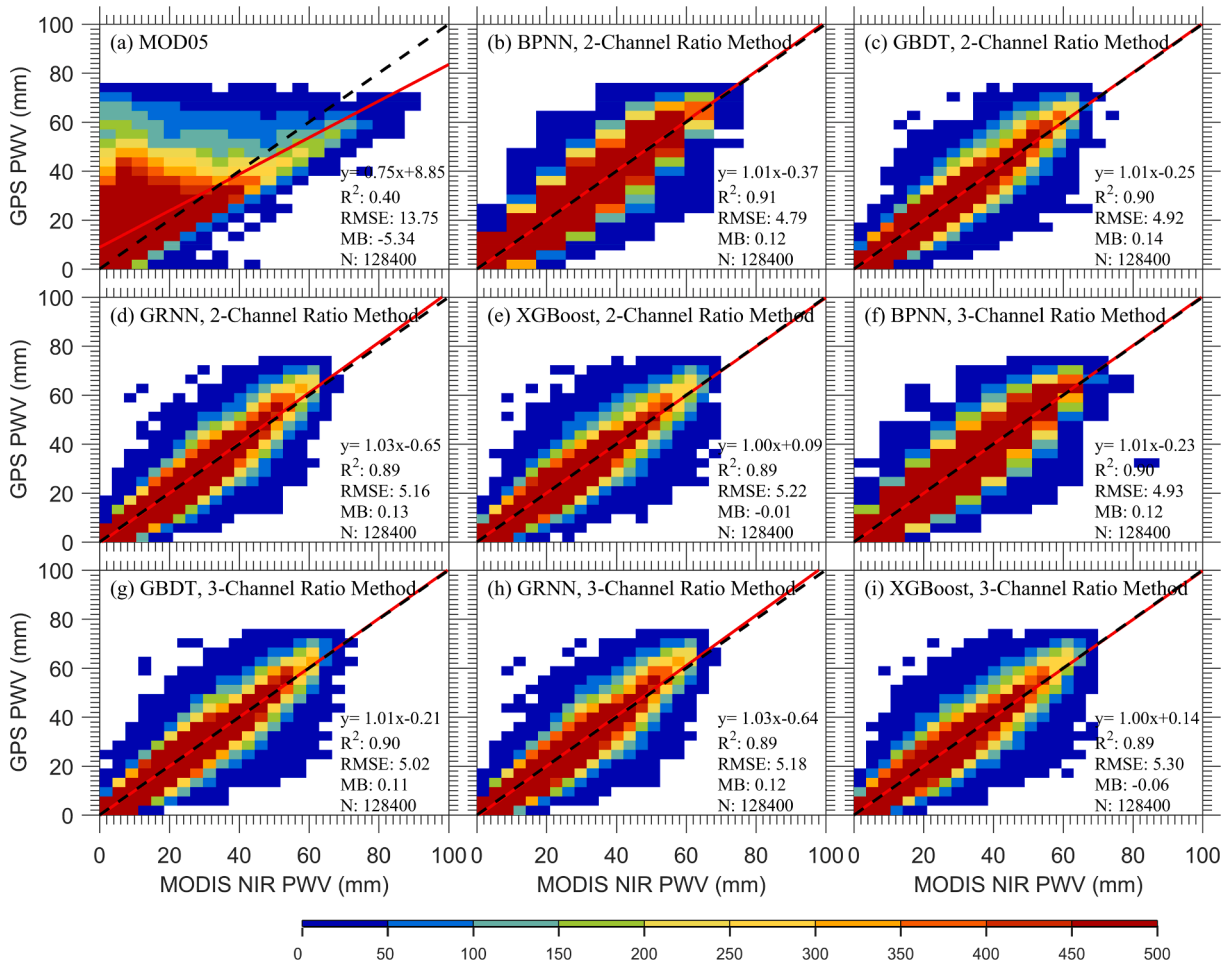


Fig. 14. Comparison of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in China under all weather conditions. The red line is the linear regression result between PWV from MODIS NIR and GPS observations. The dashed black line is the 1:1 line utilized as a reference. The color bar indicates the number of PWV pairs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Comparison of satellite-based MODIS NIR PWV data versus ground-based GPS PWV data in Australia

4.1.1. Statistical analysis

Fig. 7 displays the general validation results between MODIS-based PWV and GPS-based PWV in 2018–2019 in Australia under all-weather conditions. The new PWV retrieval algorithms can greatly enhance the retrieval performance of the all-weather PWV estimates from MODIS NIR measurements, exhibiting larger R^2 and lower RMSE compared to the official MOD05 product. The official MOD05 product cannot provide good quality water vapor measurements under all weather conditions, due to the impact of clouds. The R^2 between MOD05 PWV and GPS PWV was 0.26, and the RMSE between them was 10.95 mm.

For the new MODIS-retrieved NIR PWV estimates, the correlation was in the range of 0.75 to 0.81 for 2-band ratio transmittance and in the range of 0.75 to 0.80 for 3-band ratio transmittance. The RMSE reduced 57.44 % from 10.95 mm to 4.66 mm for BPNN, 56.07 % to 4.81 mm for GBDT, 52.15 % to 5.24 mm for GRNN, and 53.42 % to 5.10 mm for XGBoost, when the 2-band ratio transmittance data were utilized. When the 3-band ratio transmittance data were utilized, the RMSE dropped 56.16 % to 4.80 mm, 54.61 % to 4.97 mm, 52.05 % to 5.25 mm, and 51.87 % to 5.27 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. The new all-weather PWV result retrieved using 2-band ratio method performed better than 3-band ratio method. The BPNN-estimated PWV retrievals presented the best PWV retrieval accuracy for both 2-band and 3-band ratio approaches. The official MOD05 PWV product exhibited a MB of −1.75 mm (underestimated) in all weather conditions, while all the new all-weather PWV estimates showed an overestimation trend with positive MBs (MB = 0.29–0.73).

In confident-clear conditions (see Fig. 8), all the new MODIS-retrieved NIR PWV data showed slightly lower correlation than that of the original MOD05 PWV product. With the use of the machine learning based retrieval approaches, the reduction in RMSE was in the range between 44.90 % and 48.48 % for 2-channel ratio method and in the range between 45.62 % and 48.48 % for 3-channel ratio method. Among them, the PWV data derived with the BPNN method showed the highest RMSE reduction and the lowest RMSE of 2.88 mm, while the PWV data derived with the GRNN method had the smallest RMSE reduction and the largest RMSE of 3.08 mm. All the new MODIS-estimated PWV retrievals showed smaller MB values than the official MOD05 PWV product (MB = 4.40 mm).

4.1.2. Cloud analysis

The evaluation results between the PWV observations from MODIS and GPS instruments under each different weather condition is displayed in Fig. 9. The weather condition is defined utilizing the official MODIS/Terra cloud-mask product, i.e. MOD35 (Platnick et al., 2003). The results suggest that all the new PWV data exhibited better agreement with reference GPS PWV data than the official MODIS PWV products under each type of cloud condition, i.e. confident clear, probably clear, probably cloudy, and confident cloudy conditions.

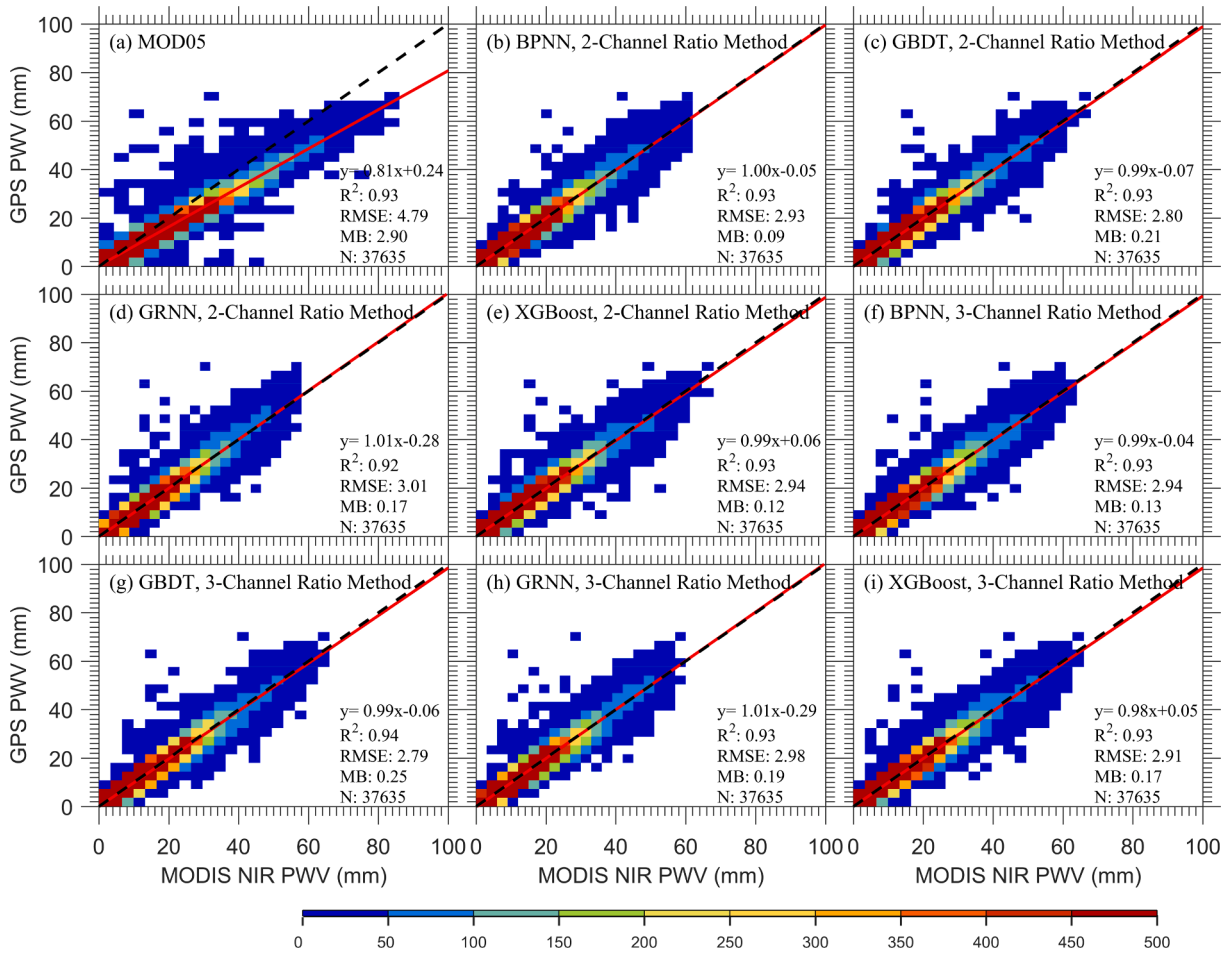


Fig. 15. Comparison of satellite-based MODIS-based NIR PWV observations against ground-based GPS-based PWV observations in 2018–2019 in China under confident clear conditions. The red line is the linear regression result between PWV from MODIS NIR and GPS observations. The dashed black line is the 1:1 line utilized as a reference. The color bar indicates the number of PWV pairs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As shown in Figs. 8 and 9, the RMSE reduction was in the range of 44.90 % ~ 48.48 % for 2-band ratio method and in the range of 45.62 % ~ 48.48 % for 3-band ratio method, when the PWV data were retrieved under confident clear conditions. In probably clear conditions, the RMSE reduced 49.13 % from 5.72 mm to 2.91 mm for BPNN, 48.78 % to 2.93 mm for GBDT, 43.71 % to 3.22 mm for GRNN, 46.85 % to 3.04 mm for XGBoost for the PWV data retrieved with 2-band ratio transmittance, and for the PWV data retrieved with 3-band ratio transmittance it dropped 48.78 % to 2.93 mm, 50.35 % to 2.84 mm, 44.06 % to 3.20 mm, and 48.25 % to 2.96 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. Under probably cloudy conditions, the new GBDT-based PWV result estimated from 3-band ratio approach presented the best PWV retrieval performance with the largest RMSE reduction of 29.60 % from 5.00 mm to 3.52 mm. The GRNN-estimated PWV retrievals calculated using 2-band ratio transmittance showed the poorest PWV retrieval performance with the smallest RMSE reduction of 17.60 % from 5.00 mm to 4.12 mm. The new confident-cloudy PWV result exhibited a reduction in RMSE in the range of 54.55 % (GRNN) to 59.91 % (BPNN) for 2-band ratio approach and in the range of 53.33 % (XGBoost) to 58.39 % (BPNN) for 3-band ratio approach, compared with the official MOD05 PWV product (RMSE = 16.39 mm). In terms of RMSE reduction, the new PWV result showed the largest improvement in PWV accuracy under confident-cloudy conditions, reducing the impact of clouds on MODIS NIR PWV estimates. All the new MODIS-derived NIR PWV estimates had smaller MB values than the official MOD05 PWV retrievals under each weather condition.

4.1.3. Seasonal analysis

Table 3 lists the seasonal comparison results between MODIS-based PWV and GPS-based PWV between January 1, 2018 and December 31, 2019 in Australia under both all-weather and confident-clear conditions. In each season, all the new MODIS-estimated NIR PWV results showed better PWV retrieval accuracy with smaller RMSE values than the official MOD05 PWV product.

In all-weather conditions, the seasonal RMSE in spring dropped 51.57 % from 9.23 mm to 4.47 mm for BPNN, 49.19 % to 4.69 mm for GBDT, 46.70 % to 4.92 mm for GRNN, and 45.50 % to 5.03 mm for XGBoost for 2-band ratio approach, and for 3-band ratio approach it reduced 51.03 % to 4.52 mm, 47.78 % to 4.82 mm, 46.80 % to 4.91 mm, and 44.53 % to 5.12 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. In summer, the seasonal RMSE reduced 53.63 % (GRNN) ~ 59.51 % (BPNN) for 2-band ratio transmittance and 53.34 % (GRNN) ~ 57.76 % (BPNN) for 3-band ratio transmittance. The seasonal RMSE reduction was in the range of 54.69 % (GRNN) ~ 59.16 % (BPNN) and 54.08 % (XGBoost) ~ 58.38 % (BPNN) in autumn, respectively. In winter, the seasonal RMSE reduced 54.85 % from 7.22 mm to 3.26 mm for BPNN, 54.29 % to 3.30 mm for GBDT, 48.89 % to 3.69 mm for GRNN, and 51.11 % to 3.53 mm for XGBoost for the PWV data estimated using 2-band ratio method. For the PWV data estimated using 3-band ratio method, the new all-weather PWV result presented a seasonal RMSE reduction of 52.63 % to 3.42 mm, 52.22 % to 3.45 mm, 48.89 % to 3.69 mm, and 49.31 % to 3.66 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. All the new MODIS-retrieved NIR PWV estimates

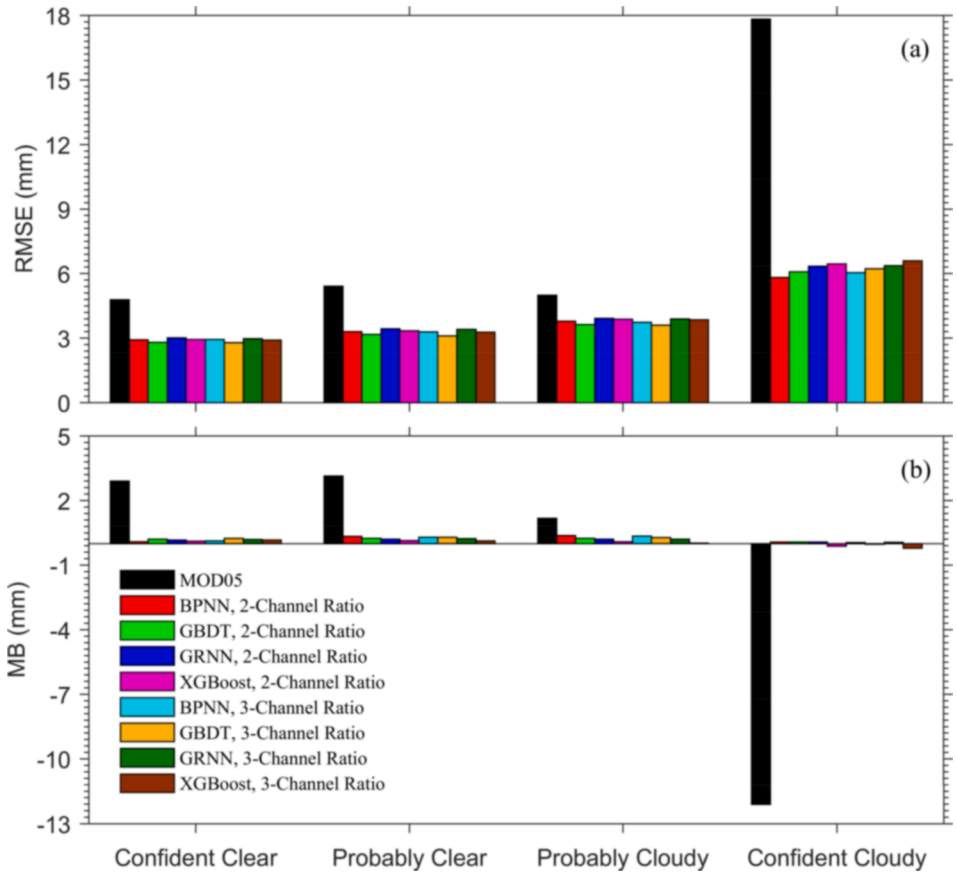


Fig. 16. Comparison of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in China, under different weather conditions. The weather condition is defined based on the cloud-mask flags from the official MOD35 product.

exhibited smaller MB in all four seasons, compared to official MOD05 PWV retrievals.

When the PWV data were retrieved under confident clear conditions, the seasonal reduction in RMSE for BPNN was 38.86 % from 4.58 mm to 2.80 mm in spring, 49.01 % from 6.59 mm to 3.36 mm in summer, 54.95 % from 6.26 mm to 2.82 mm in autumn, and 46.23 % from 4.11 mm to 2.21 mm in winter for 2-band ratio method, and for 3-band ratio method it was 39.08 % to 2.79 mm, 49.62 % to 3.32 mm, 54.63 % to 2.84 mm, and 44.04 % to 2.30 mm in spring, summer, autumn, and winter, respectively. For GBDT, the RMSE dropped 38.43 % (spring) ~ 54.47 % (autumn) for 2-band ratio transmittance and 37.99 % (spring) ~ 54.95 % (autumn) for 3-band ratio transmittance. The GRNN-based PWV estimates showed a seasonal RMSE reduction of 36.03 % (spring) ~ 50.96 % (autumn) and 36.90 % (spring) ~ 51.76 % (autumn) for 2-band ratio approach and 3-band ratio approach, respectively. For XGBoost, the largest RMSE reduction was 52.72 % from 6.26 mm to 2.96 mm in autumn, while the smallest RMSE reduction was 36.46 % from 4.58 mm to 2.91 mm in spring. In all four seasons, all the new PWV retrievals had smaller MBs compared with the official MOD05 product.

In both all-weather and confident-clear conditions, all the new PWV result exhibited larger enhancement in MODIS NIR PWV accuracy in summer and autumn, compared to spring and winter.

4.1.4. Time-series analysis

In Fig. 10, the time-series comparison result between the daily averaged all-weather PWV values estimated from GPS, MOD05, BPNN, GBDT, GRNN, and XGBoost is presented. All the new MODIS-derived daily mean PWV data records agreed better than the official MODIS PWV product, when compared with in-situ daily mean PWV data from GPS. The daily mean all-weather PWV values from the official MOD05 product tended to underestimate the PWV values due to the effect of clouds (MB = -1.49 mm). With the employment of the new retrieval models, the daily-scale MB values between MODIS NIR PWV and GPS PWV were reduced to the range between 0.57 mm and 0.78 mm for 2-band ratio approach and to the range between 0.66 mm and 0.95 mm for 3-band ratio approach, showing an overestimation trend of PWV in a daily basis. The new PWV estimates exhibited the daily-scale RMSE in the range of 1.81 mm ~ 2.18 mm for 2-band ratio transmittance and in the range of 1.95 mm ~ 2.19 mm for 3-band ratio transmittance, much lower than the official MOD05 product (RMSE = 4.31 mm).

In confident clear conditions (see Fig. 11), the new PWV result performed better on a daily basis, showing a daily-scale RMSE in the range between 1.13 mm and 1.31 mm and a daily-scale MB in the range between −0.14 mm and 0.19 mm. Both RMSE and MB of new PWV estimates were much smaller than the official MOD05 product (RMSE = 4.96 mm and MB = 4.65 mm) in both 2018 and 2019. However, the new PWV retrievals had a correlation coefficient R^2 in the range of 0.95 ~ 0.96 in a daily basis, slightly smaller than the official MOD05 PWV product (R^2 = 0.99).

4.1.5. Station-wise analysis

Under all-weather conditions, the station-wise RMSE of the official MOD05 PWV product at most GPS stations was above 8 mm as shown in Fig. 12. All the new MODIS-estimated PWV data showed the all-weather station-wise RMSE below 8 mm among most GPS stations, which implies that the retrieval performance of MODIS NIR PWV can be enhanced using the four PWV retrieval methods under all weather conditions.

Under confident-clear conditions, the station-wise RMSE of the MOD05 product was in the range of 4 mm to 10 mm at most GPS sites (see Fig. 13). With the use of the retrieval models, the station-wise RMSE values were reduced to below 4 mm at most GPS stations in confident clear conditions. It is thus concluded that all the new MODIS-retrieved

Table 4
The seasonal comparison between the satellite-based MODIS NIR PWV against the ground-based GPS reference PWV in 2018–2019 in China under all weather and confident clear conditions.

		All Weather						Confident Clear					
		Slope	Offset	R^2	RMSE (mm)	MB (mm)	RMSE reduction	Slope	Offset	R^2	RMSE (mm)	MB (mm)	RMSE reduction
Spring	MOD05	0.76	7.29	0.37	11.41	−4.38	–	0.83	0.08	0.89	4.02	2.25	–
	BPNN, 2-Channel Ratio	1.05	−0.52	0.87	4.71	−0.25	58.72 %	1.03	−0.25	0.90	2.75	−0.11	31.59 %
	GBDT, 2-Channel Ratio	1.02	−0.23	0.86	4.87	−0.15	57.32 %	0.98	0.06	0.91	2.64	0.20	34.33 %
	GRNN, 2-Channel Ratio	1.07	−0.79	0.85	5.06	−0.33	55.65 %	1.05	−0.55	0.90	2.79	−0.05	30.60 %
	XGBoost, 2-Channel Ratio	1.01	0.06	0.84	5.16	−0.25	54.78 %	0.98	0.11	0.90	2.78	0.15	30.85 %
	BPNN, 3-Channel Ratio	1.04	−0.24	0.86	4.80	−0.37	57.93 %	1.02	−0.18	0.90	2.73	−0.07	32.09 %
	GBDT, 3-Channel Ratio	1.02	−0.19	0.86	4.91	−0.20	56.97 %	0.97	0.08	0.91	2.63	0.23	34.58 %
	GRNN, 3-Channel Ratio	1.07	−0.79	0.85	5.07	−0.33	55.57 %	1.05	−0.54	0.90	2.77	−0.01	31.09 %
	XGBoost, 3-Channel Ratio	1.01	0.16	0.84	5.22	−0.33	54.25 %	0.98	0.11	0.90	2.74	0.16	31.84 %
Summer	MOD05	0.36	25.39	0.15	20.96	−10.20	–	0.80	0.08	0.92	7.22	5.77	–
	BPNN, 2-Channel Ratio	1.02	−0.53	0.85	6.15	−0.21	70.66 %	0.99	−0.10	0.90	3.74	0.40	48.20 %
	GBDT, 2-Channel Ratio	1.02	−0.39	0.84	6.28	−0.24	70.04 %	0.99	−0.02	0.90	3.64	0.37	49.58 %
	GRNN, 2-Channel Ratio	1.06	−1.80	0.83	6.63	−0.13	68.37 %	1.03	−1.25	0.88	4.05	0.54	43.91 %
	XGBoost, 2-Channel Ratio	1.00	0.44	0.83	6.60	−0.48	68.51 %	0.98	0.27	0.90	3.78	0.21	47.65 %
	BPNN, 3-Channel Ratio	1.02	−0.33	0.84	6.43	−0.22	69.32 %	1.00	−0.49	0.89	3.83	0.49	46.95 %
	GBDT, 3-Channel Ratio	1.02	−0.29	0.83	6.47	−0.35	69.13 %	0.99	−0.15	0.91	3.61	0.48	50.00 %
	GRNN, 3-Channel Ratio	1.06	−1.80	0.83	6.66	−0.15	68.23 %	1.03	−1.30	0.89	4.01	0.56	44.46 %
	XGBoost, 3-Channel Ratio	1.00	0.51	0.82	6.76	−0.55	67.75 %	0.98	0.07	0.90	3.75	0.35	48.06 %
Autumn	MOD05	0.74	7.98	0.38	12.10	−4.62	–	0.83	0.00	0.92	4.12	2.48	–
	BPNN, 2-Channel Ratio	0.96	0.15	0.89	4.62	0.50	61.82 %	1.00	0.14	0.91	2.82	−0.11	31.55 %
	GBDT, 2-Channel Ratio	0.95	0.52	0.88	4.80	0.43	60.33 %	1.00	0.14	0.92	2.67	−0.10	35.19 %
	GRNN, 2-Channel Ratio	0.97	0.03	0.87	4.89	0.46	59.59 %	0.99	0.14	0.92	2.81	−0.01	31.80 %
	XGBoost, 2-Channel Ratio	0.93	1.02	0.86	5.16	0.26	57.36 %	0.98	0.44	0.91	2.86	−0.22	30.58 %
	BPNN, 3-Channel Ratio	0.95	0.41	0.89	4.71	0.54	61.07 %	0.98	0.32	0.92	2.80	−0.10	32.04 %
	GBDT, 3-Channel Ratio	0.94	0.59	0.88	4.86	0.44	59.83 %	0.99	0.20	0.92	2.66	−0.10	35.44 %
	GRNN, 3-Channel Ratio	0.97	0.04	0.87	4.90	0.46	59.50 %	0.99	0.13	0.92	2.77	0.00	32.77 %
	XGBoost, 3-Channel Ratio	0.93	1.04	0.86	5.21	0.27	56.94 %	0.98	0.44	0.92	2.82	−0.19	31.55 %
Winter	MOD05	0.84	3.10	0.47	6.33	−2.05	–	0.75	0.60	0.84	3.04	1.28	–
	BPNN, 2-Channel Ratio	0.99	−0.43	0.86	3.02	0.54	52.29 %	0.97	−0.25	0.85	2.18	0.42	28.29 %
	GBDT, 2-Channel Ratio	0.98	−0.46	0.86	3.02	0.61	52.29 %	0.99	−0.47	0.88	1.99	0.53	34.54 %
	GRNN, 2-Channel Ratio	1.02	−0.82	0.82	3.45	0.62	45.50 %	1.02	−0.59	0.86	2.09	0.44	31.25 %
	XGBoost, 2-Channel Ratio	0.97	−0.20	0.84	3.23	0.48	48.97 %	0.98	−0.32	0.87	2.07	0.49	31.91 %
	BPNN, 3-Channel Ratio	0.98	−0.40	0.86	3.12	0.62	50.71 %	0.98	−0.33	0.85	2.20	0.46	27.63 %
	GBDT, 3-Channel Ratio	0.98	−0.40	0.85	3.14	0.60	50.39 %	0.97	−0.39	0.88	2.01	0.57	33.88 %
	GRNN, 3-Channel Ratio	1.02	−0.80	0.82	3.47	0.62	45.18 %	1.02	−0.59	0.87	2.07	0.43	31.91 %
	XGBoost, 3-Channel Ratio	0.97	−0.17	0.83	3.31	0.45	47.71 %	0.97	−0.34	0.87	2.08	0.53	31.58 %

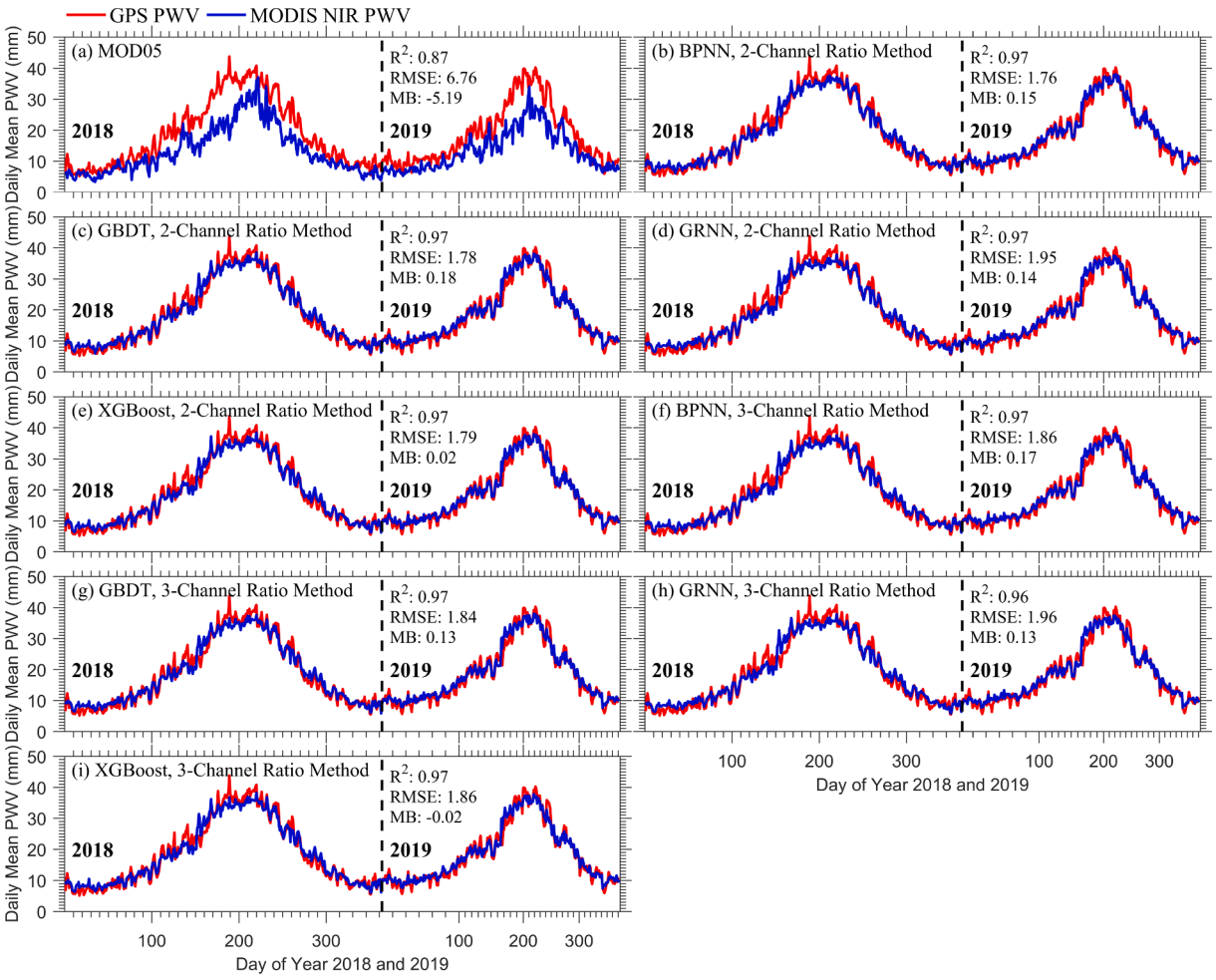


Fig. 17. Time-series variation of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in China under all weather conditions.

PWV result agreed better with reference GPS-derived PWV data in most stations in confident clear conditions.

4.2. Comparison of satellite-based MODIS NIR PWV data versus ground-based GPS PWV data in China

4.2.1. Statistical analysis

The results in Fig. 14 suggest that all the new MODIS NIR PWV data showed better retrieval accuracy than the official MOD05 product, when compared with in-situ GPS PWV data under all-weather conditions in 2018–2019 in China. The MOD05 product had an RMSE of 13.75 mm, a mean bias of −5.34 mm, and a R^2 of 0.40 under all weather conditions. With the use of the retrieval methods, the correlation coefficient R^2 of the new PWV data was in the range between 0.89 and 0.91 for 2-channel ratio method and between 0.89 and 0.90 for 3-channel ratio method. For RMSE, it reduced 65.16 % from 13.75 mm to 4.79 mm for BPNN, 64.22 % to 4.92 mm for GBDT, 62.47 % to 5.16 mm for GRNN, and 62.04 % to 5.22 mm for XGBoost for 2-band ratio transmittance. For 3-band ratio transmittance, the RMSE has reduced 64.15 % to 4.93 mm, 63.49 % to 5.02 mm, 62.33 % to 5.18 mm, and 61.45 % to 5.30 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. The BPNN-based PWV result showed the highest PWV retrieval accuracy (RMSE = 4.79 mm and 4.93 mm), whereas the XGBoost-based PWV result showed the lowest PWV retrieval accuracy (RMSE = 5.22 mm and 5.30 mm). The MB values of MODIS NIR PWV were reduced to the range of −0.06 mm to 0.14 mm, which were much lower than the official MOD05 PWV product (MB = -5.34 mm).

Fig. 15 showed the results obtained under confident-clear conditions. The R^2 of new PWV results was in the range of 0.92 ~ 0.93 for 2-band ratio approach and 0.93 ~ 0.94 for 3-band ratio approach, which were comparable to the official MOD05 product (R^2 = 0.93). For the PWV data derived with 2-band ratio transmittance, the RMSE reduced 38.83 % from 4.79 mm to 2.93 mm for BPNN, 41.54 % to 2.80 mm for GBDT, 37.16 % to 3.01 mm for GRNN, and 38.62 % to 2.94 mm for XGBoost. For 3-band ratio approach, the RMSE dropped 38.62 % to 2.94 mm, 41.75 % to 2.79 mm, 37.79 % to 2.98 mm, and 39.25 % to 2.91 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. The new confident-clear PWV results showed the MB values in the range between 0.09 mm and 0.21 mm for 2-band ratio transmittance and in the range between 0.13 mm and 0.25 mm for 3-band ratio transmittance. All of them were much smaller than those of the MOD05 product (MB = 2.90 mm).

4.2.2. Cloud analysis

Fig. 16 indicates that all the new MODIS PWV retrievals performed better than the MODIS product when compared with in-situ GPS reference PWV in all the four weather conditions, i.e. confident clear, probably clear, probably cloudy, and confident cloudy.

For confident-clear conditions, the RMSE dropped 37.16 % ~ 41.54 % for 2-band ratio method and 37.79 % ~ 41.75 % for 3-band ratio method, with results displayed in Figs. 15 and 16. Under probably clear conditions, the new PWV result showed a reduction in RMSE in the range of 36.72 % to 41.51 % and 37.08 % to 42.62 % for 2-band and 3-band ratio approaches, respectively. Particularly, the GBDT-derived PWV from 3-band ratio transmittance showed the highest PWV

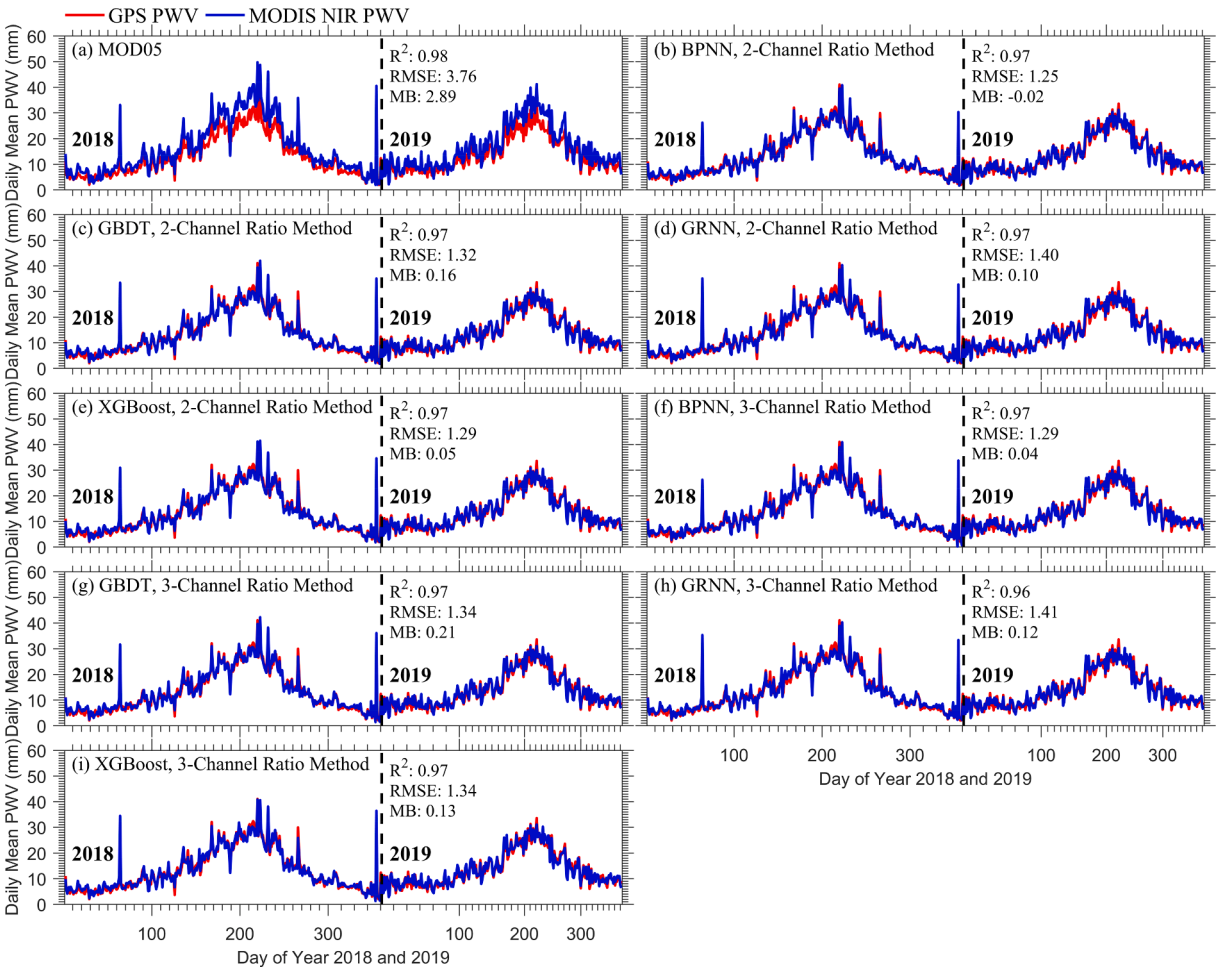


Fig. 18. Time-series variation of satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations in 2018–2019 in China under confident clear conditions.

retrieval performance (RMSE = 3.11 mm), while the GRNN-derived PWV calculated using 2-band ratio transmittance exhibited the worst PWV performance (RMSE = 3.43 mm). For probably cloudy conditions, the reduction in RMSE was in the range between 21.44 % and 27.25 % for 2-channel ratio transmittance and in the range between 21.84 % and 27.86 % for 3-channel ratio transmittance. Under confident cloudy conditions, the RMSE reduced 67.36 % from 17.83 mm to 5.82 mm for BPNN, 65.90 % to 6.08 mm for GBDT, 64.44 % to 6.34 mm for GRNN, and 63.83 % to 6.45 mm for XGBoost for the PWV data retrieved with 2-band ratio transmittance. For the PWV data retrieved with 3-band ratio transmittance, the RMSE reduced 66.12 % to 6.04 mm, 65.06 % to 6.23 mm, 64.27 % to 6.37 mm, and 62.98 % to 6.60 mm for BPNN, GBDT, GRNN, and XGBoost, respectively.

In each different weather condition, the new PWV estimates showed a smaller MB value than the official MOD05 product, for both 2-band and 3-band ratio approaches. The new PWV results showed a large accuracy enhancement under confident-cloudy conditions, indicating the capability of new algorithms to reduce cloud effect on NIR PWV retrieval.

4.2.3. Seasonal analysis

In Table 4, all the new MODIS NIR PWV data showed a seasonal improvement in the retrieval accuracy in all-weather and confident-clear conditions, with smaller seasonal RMSE and MB against in-situ GPS reference PWV data.

In all-weather conditions, the seasonal reduction in RMSE was 54.78 % (XGBoost) ~ 58.72 % (BPNN) for 2-band ratio transmittance and 54.25 % (XGBoost) ~ 57.93 % (BPNN) for 3-band ratio transmittance in spring. In summer, the seasonal RMSE reduction was in the range of 68.36 % (GRNN) to 70.66 % (BPNN) and in the range of 67.75 % (XGBoost) to 69.32 % (BPNN) for 2-band and 3-band ratio approaches, respectively. The seasonal RMSE in autumn dropped 61.82 % from 12.10 mm to 4.62 mm for BPNN, 60.33 % to 4.80 mm for GBDT, 59.59 % to 4.89 mm for GRNN, and 57.36 % to 5.16 mm for XGBoost for 2-band ratio method, and for 3-band ratio method it reduced 61.07 % to 4.71 mm, 59.83 % to 4.86 mm, 59.50 % to 4.90 mm, and 56.94 % to 5.21 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. In winter, the seasonal RMSE reduced 45.50 % (GRNN) ~ 52.29 % (BPNN and GBDT) for 2-band ratio method and 45.18 % (GRNN) ~ 50.71 % (BPNN) for 3-band ratio method.

Similarly, under confident clear conditions, the seasonal reduction in RMSE was in the range between 30.60 % (GRNN) and 34.33 % (GBDT) for 2-channel ratio method and in the range between 31.09 % (GRNN) and 34.58 % (GBDT) for 3-channel ratio method in spring. The seasonal RMSE in summer dropped 43.91 % (GRNN) ~ 49.58 % (GBDT) and 44.46 % (GRNN) ~ 50.00 % (GBDT) for 2-band and 3-band ratio methods, respectively. The seasonal RMSE reduced 30.58 % (XGBoost) ~ 35.19 % (GBDT) for 2-band ratio transmittance and 31.55 % (XGBoost) ~ 35.44 % (GBDT) for 3-band ratio transmittance in autumn. In winter, the seasonal RMSE dropped 28.29 % from 3.04 mm to 2.18 mm for BPNN, 34.54 % to 1.99 mm for GBDT, 31.25 % to 2.09 mm for GRNN, and 31.91 % to 2.07 mm for XGBoost for the PWV data calculated using 2-band ratio transmittance. For the PWV data calculated using 3-band ratio transmittance, it has reduced 27.63 % to 2.20 mm,

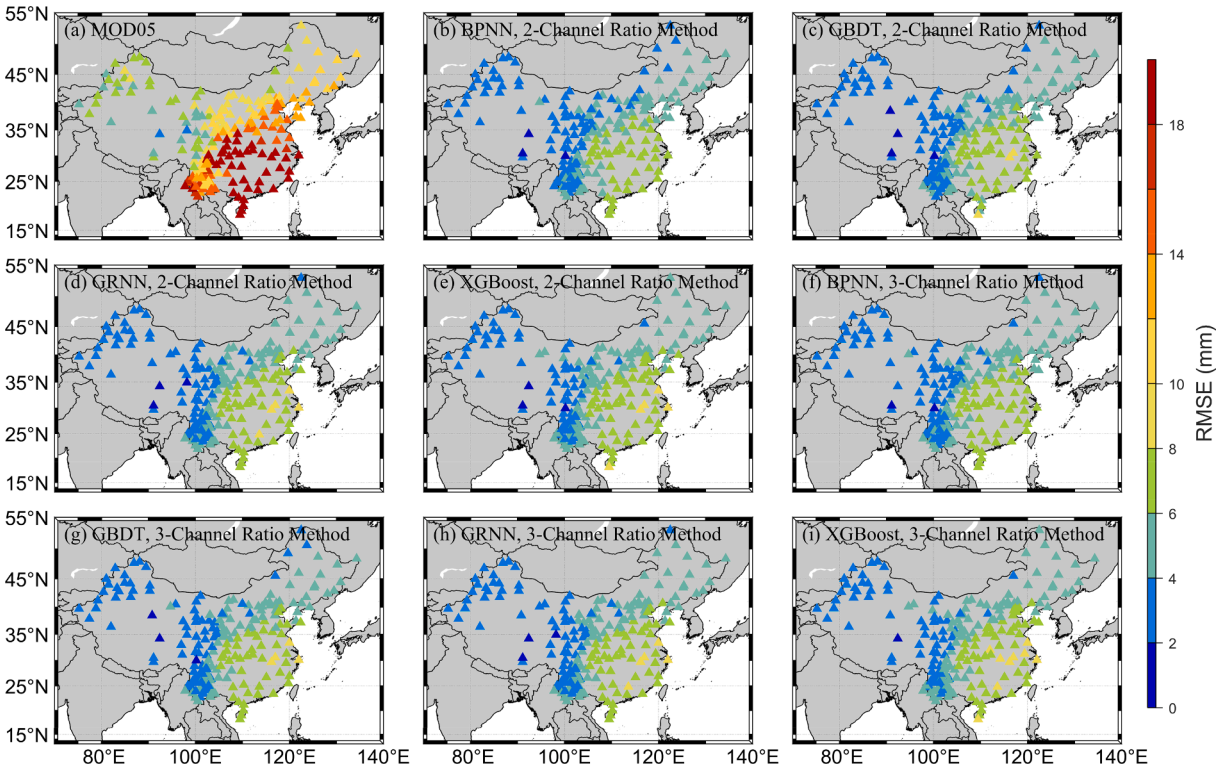


Fig. 19. Annual RMSE map for satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations at all GPS stations in 2018–2019 in China under all weather conditions.

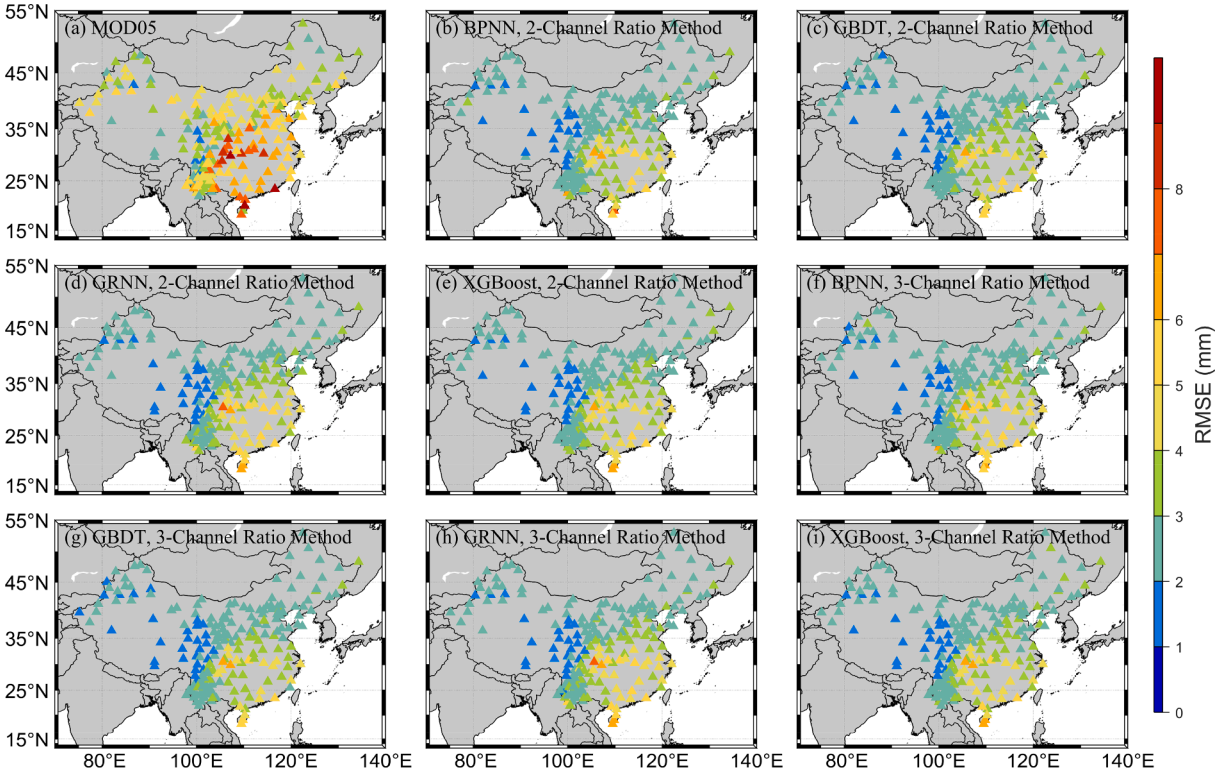


Fig. 20. Annual RMSE map for satellite-based MODIS NIR PWV observations against ground-based GPS PWV observations at all GPS stations in 2018–2019 in China under confident clear conditions.

33.88 % to 2.01 mm, 31.91 % to 2.07 mm, and 31.58 % to 2.08 mm for BPNN, GBDT, GRNN, and XGBoost, respectively.

For both all-weather and confident-clear conditions, the new PWV estimates exhibited larger accuracy improvement in summer and autumn than spring and winter.

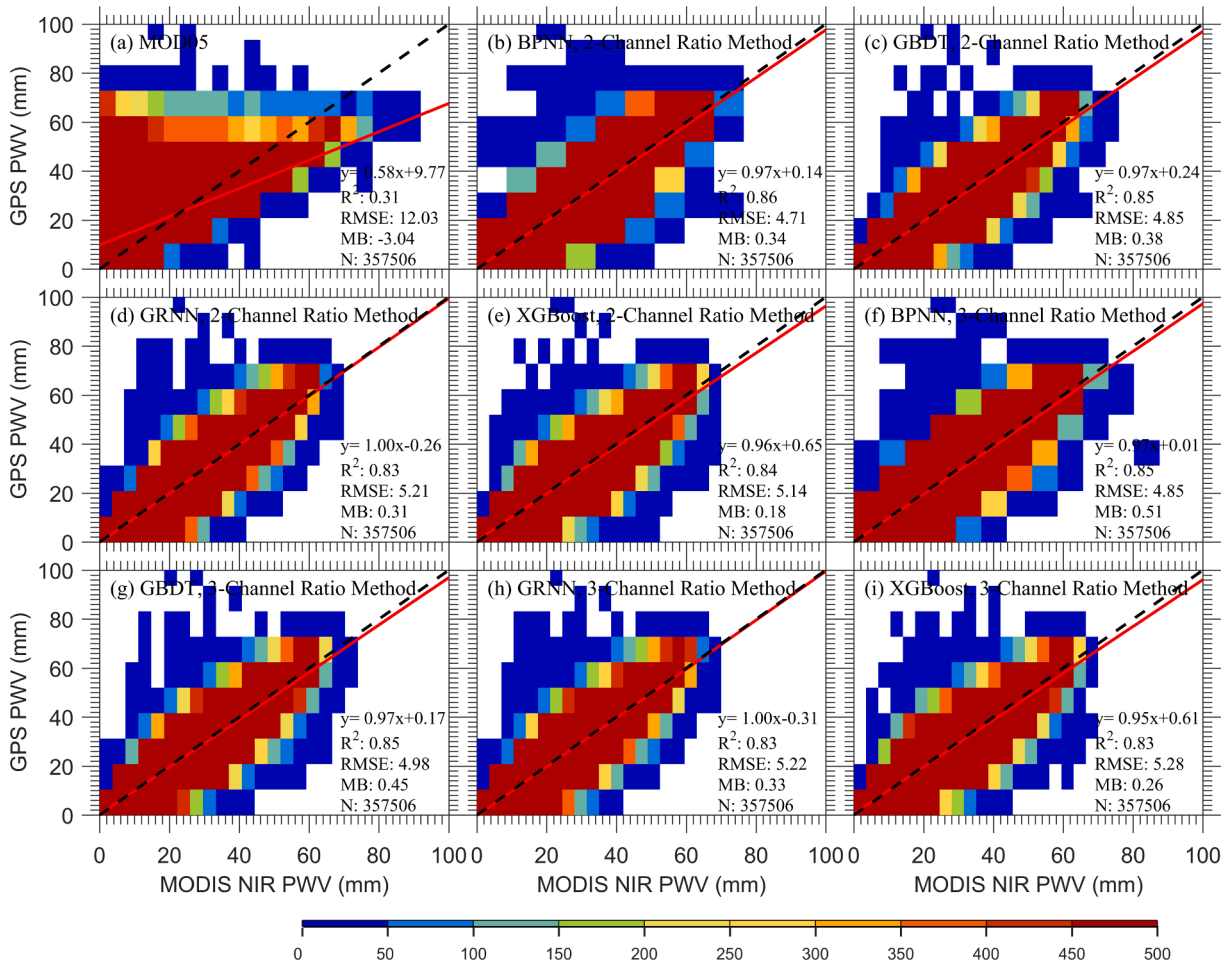


Fig. 21. Comparison of satellite-based MODIS NIR PWV against ground-based GPS PWV in 2018–2019 in both Australia and China under all weather conditions. The red line is the linear regression result between PWV from MODIS NIR and GPS observations. The dashed black line is the 1:1 line utilized as a reference. The color bar indicates the number of PWV pairs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.4. Time-series analysis

The daily time-series comparison result between the PWV observations from MODIS and GPS instruments is presented in Figs. 17 and 18. In all-weather and confident-clear conditions, the daily new PWV result showed a better agreements with that of in-situ GPS PWV in both 2018 and in 2019. This implies that the retrieval methods are temporally independent and perform well in different years.

Under all-weather conditions (see Fig. 17), the official MOD05 PWV product showed a daily R^2 of 0.87, a daily RMSE of 6.76 mm, and a daily MB of −5.19 mm with respect to GPS daily averaged reference PWV. The new PWV result exhibited daily R^2 in the range of 0.96 ~ 0.97, daily RMSE in the range of 1.76 mm ~ 1.96 mm, and daily MB in the range of −0.02 mm ~ 0.18 mm. It can be seen the performance of the new PWV data is much better than that of the MODIS PWV product.

The result under confident clear conditions was shown in Fig. 18. By employing the new retrieval methods, the daily RMSE values were reduced to the range between 1.25 mm and 1.41 mm, and the MB values were reduced to the range between −0.02 mm to 0.21 mm. Both RMSE and MB were much smaller than those of the official MOD05 confident-clear PWV product (RMSE = 3.76 mm and MB = 2.89 mm). The correlation coefficient R^2 of the new daily mean PWV under confident-clear condition was in the range between 0.96 and 0.97, slightly smaller than R^2 of the official MOD05 product 0.98.

4.2.5. Station-wise analysis

In Fig. 19, the station-wise RMSE values of the MOD05 product were above 8 mm at most stations under all-weather conditions. The new MODIS PWV result at most sites showed a station-wise RMSE in the range between 2 mm and 8 mm. This implies that the new algorithms improved MODIS NIR PWV retrievals at most stations under all-weather conditions. In confident-clear conditions (see Fig. 20), the official MOD05 product presented a station-wise RMSE in the range between 4 mm and 10 mm at most sites. Their station-wise RMSE values were reduced to the range between 1 mm and 4 mm at most sites after using the retrieval algorithms developed in this research.

4.3. Comparison of satellite-based MODIS NIR PWV data versus ground-based GPS PWV data in both Australia and China

The all-weather comparison analysis result between MODIS PWV and GPS PWV in 2018–2019 in both Australia and China is included in Fig. 21. The new PWV result showed the correlation coefficient R^2 in the range of 0.83 ~ 0.86 for 2-band ratio approach and in the range of 0.83 ~ 0.85 for 3-band ratio approach, much better than the official MOD05 product (R^2 = 0.31). For RMSE, it reduced 60.85 % from 12.03 mm to 4.71 mm for BPNN, 59.68 % to 4.85 mm for GBDT, 56.69 % to 5.21 mm for GRNN, and 57.27 % to 5.14 mm for XGBoost for the PWV data retrieved with 2-band ratio transmittance. For the PWV data retrieved with 3-band ratio transmittance, the RMSE dropped 59.68 % to 4.85 mm, 58.60 % to 4.98 mm, 56.61 % to 5.22 mm, and 56.11 % to 5.28 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. Their MB values were reduced to the range of 0.18 mm ~ 0.38 mm for 2-band ratio approach and to the range of 0.26 mm ~ 0.51 mm for 3-band ratio approach, which were much smaller than that of the official MOD05

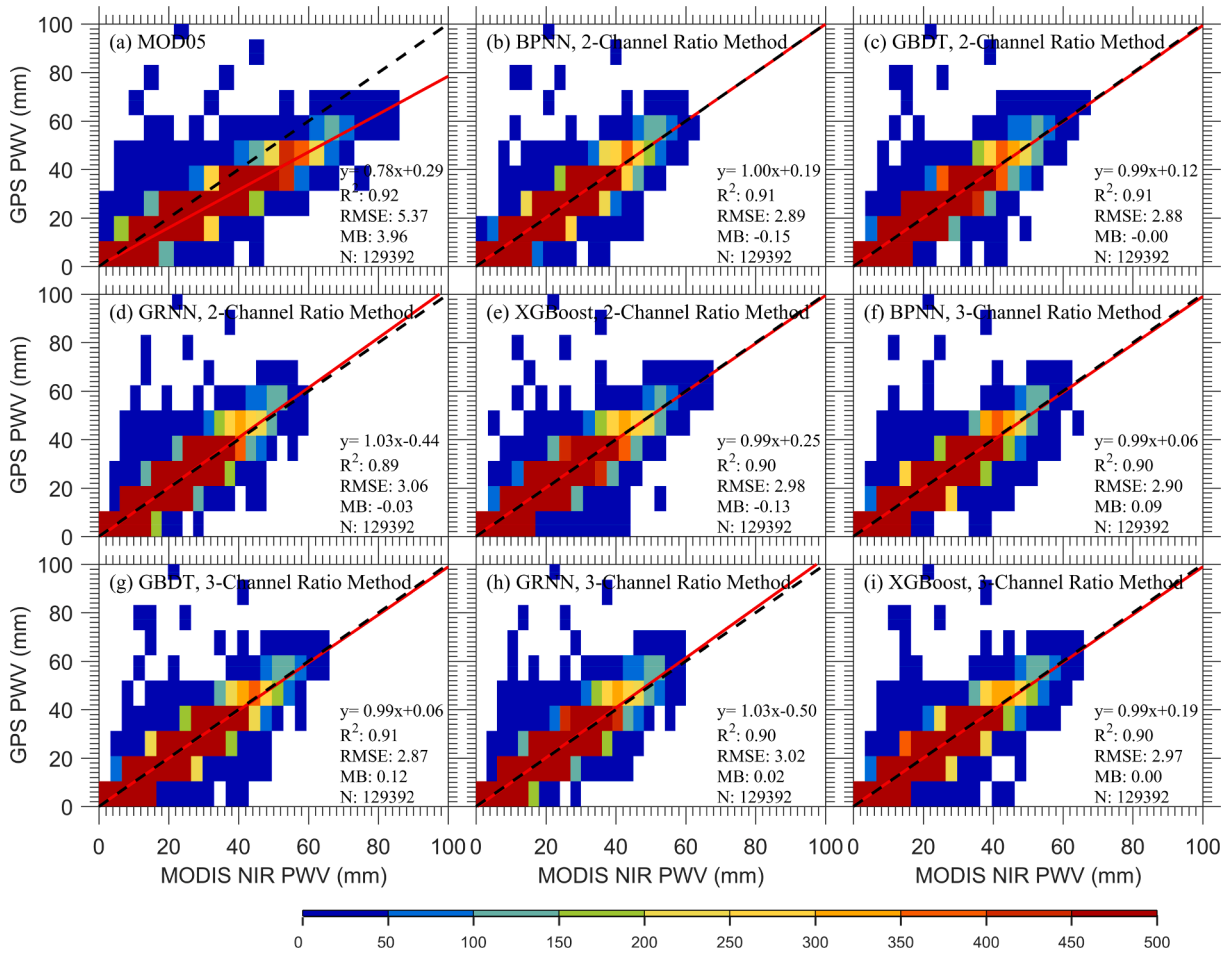


Fig. 22. Comparison of satellite-based MODIS NIR PWV against ground-based GPS PWV observations in 2018–2019 in both Australia and China under confident clear conditions. The red line is the linear regression result between PWV from MODIS NIR and GPS observations. The dashed black line is the 1:1 line utilized as a reference. The color bar indicates the number of PWV pairs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

product (MB = -3.04 mm). The BPNN-based PWV calculated with 2-band ratio transmittance had the best PWV retrieval performance, showing the highest R^2 and the smallest RMSE against in-situ GPS reference PWV.

In Fig. 22, the result obtained under confident clear conditions is shown. R^2 of new PWV estimates was 0.89 ~ 0.91 for 2-channel ratio method and 0.90 ~ 0.91 for 3-channel ratio method, which were comparable to the MOD05 product (R^2 = 0.92). The RMSE reduced 46.18 % from 5.37 mm to 2.89 mm for BPNN, by 46.37 % to 2.88 mm for GBDT, by 43.02 % to 3.06 mm for GRNN, and by 44.51 % to 2.98 mm for XGBoost for 2-band ratio transmittance. For 3-band ratio transmittance, the RMSE reduced 46.00 % to 2.90 mm, 46.55 % to 2.87 mm, 43.76 % to 3.02 mm, and 44.69 % to 2.97 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. Their MB values were reduced to around 0, much smaller than the MB in official MOD05 PWV product (MB = 3.96 mm).

Fig. 23 showed that under probably clear conditions the new PWV result with RMSE in the range of 2.99 mm ~ 3.27 mm for 2-band ratio approach and in the range of 2.91 mm ~ 3.25 mm for 3-band ratio method, indicating a reduction in RMSE of 42.12 % ~ 47.08 % and 42.48 % ~ 48.50 % for 2-band and 3-band ratio transmittances, respectively. Under probably cloudy conditions, the RMSE dropped 19.00 % ~ 27.60 % for 2-band ratio transmittance and 19.60 % ~ 29.00 % for 3-band ratio transmittance. Under confident-cloudy the RMSE reduced 63.70 % from 17.19 mm to 6.24 mm for BPNN, 62.19 % to 6.50 mm for GBDT, 59.40 % to 6.98 mm for GRNN, and 59.63 % to 6.94 mm for XGBoost for 2-band ratio approach. For 3-band ratio approach, the new PWV retrievals reduced RMSE by 62.30 % to 6.48 mm, by 60.67 % to 6.76 mm, by 59.16 % to 7.02 mm, and by 58.12 % to 7.20 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. The MB values of new PWV estimates were reduced to the range between −1 mm and 1 mm under different weather conditions, smaller than those of the MOD05 product. The retrieval methods exhibited a larger RMSE reduction (larger accuracy enhancement) in confident cloudy conditions.

Table 5 displays the seasonal evaluation results between MODIS PWV and GPS PWV in 2018–2019 over both Australia and China, under all-weather conditions and confident-clear conditions. For both all-weather and confident-clear conditions, the new PWV result exhibited larger improvement (larger RMSE reduction) in MODIS NIR PWV accuracy in summer and autumn than spring and winter.

For all weather conditions, the seasonal RMSE reduction was in the range of 49.85 % (XGBoost) ~ 54.99 % (BPNN) in spring, 60.84 % (GRNN) ~ 65.01 % (BPNN) in summer, 56.01 % (XGBoost) ~ 60.17 % (BPNN) in autumn, and 47.76 % (GRNN) ~ 53.98 % (BPNN) in winter for the PWV data retrieved with 2-band ratio method. For the PWV data retrieved with 3-band ratio method, the seasonal RMSE reduction was 49.06 % (XGBoost) ~ 54.29 % (BPNN), 60.59 % (GRNN) ~ 63.50 % (BPNN), 55.17 % (XGBoost) ~ 59.41 % (BPNN), and 47.76 % (GRNN) ~ 52.10 % (BPNN) in spring, summer, autumn, and winter, respectively.

Under confident clear conditions, the new PWV result presented a seasonal RMSE reduction in the range of 34.55 % ~ 37.27 % (spring), 44.66 % ~ 48.81 % (summer), 47.17 % ~ 50.71 % (autumn), and 41.86 % ~ 44.44 % (winter) for 2-band ratio transmittance and 35.23 % ~

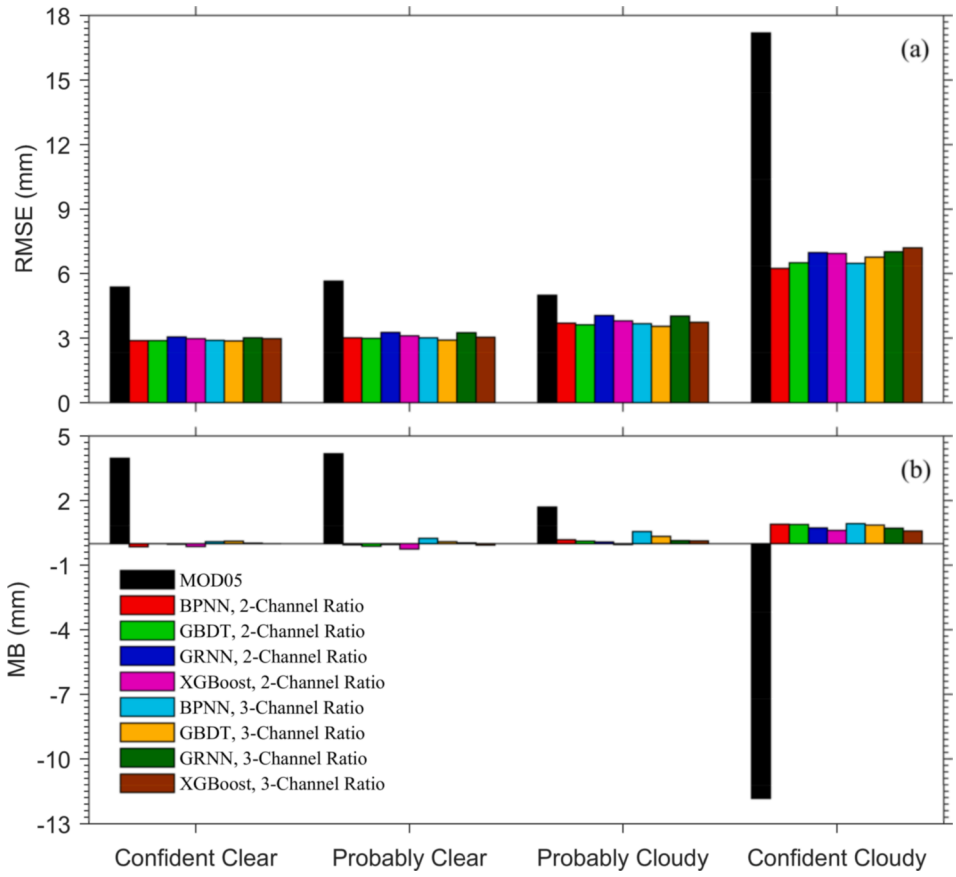


Fig. 23. Comparison of satellite-based MODIS NIR PWV against ground-based GPS PWV in 2018–2019 in both Australia and China under different weather conditions. The weather condition is defined based on the cloud-mask flags from the official MOD35 product.

37.05 % (spring), 45.10 % ~ 48.81 % (summer), 48.06 % ~ 51.06 % (autumn), and 41.34 % ~ 44.44 % (winter) for 3-band ratio transmittance.

5. Discussion

The main objective of this research is to enhance the accuracy of the MODIS NIR PWV retrieval under all-weather conditions utilizing four novel water vapor retrieval algorithms based on machine learning methods, i.e. BPNN, GBDT, GRNN, and XGBoost. The band 17 estimated transmittance (905 nm), band 18 estimated transmittance (936 nm), band 19 estimated transmittance (940 nm), latitude, longitude, elevation, cloud, season, and solar zenith angle were utilized as input parameters to the machine learning models. The in-situ PWV data measured in 2017 from 453 GPS sites in Australia and 214 GPS sites in China were used as the target output PWV result in the model training procedure. In the model validation procedure, we utilized GPS and MODIS data collected from 2018 to 2019, which are independent of the training data of 2017.

The official MOD05 product does not provide good quality water vapor retrievals under all weather conditions. This is because the NIR observations from the MODIS instrument are notably affected by the existence of clouds (He and Liu, 2019; Kaufman and Gao, 1992). In our work, the correlation coefficient R^2 of the MOD05 PWV product was found to be below 0.4, and the RMSE of the MOD05 PWV product was found to be above 10 mm in all weather conditions.

The new all-weather PWV result had R^2 values in the range of 0.83 ~ 0.86 and RMSE in the range of 4.71 mm ~ 5.28 mm, reducing RMSE by 56.69 % (GRNN) ~ 60.85 % (BPNN) for 2-band ratio approach and 56.11 % (XGBoost) ~ 59.68 % (BPNN) for 3-band ratio approach. This implies that the retrieval algorithms can significantly reduce the impact of clouds on retrieving MODIS NIR all-weather PWV. The new BPNN-based PWV had the best PWV accuracy, for both 2-band and 3-band ratio transmittances.

The new PWV estimates performed better than the MOD05 product in each different weather condition. For confident-clear conditions, the RMSE reduced 46.18 % from 5.37 mm to 2.89 mm for BPNN, 46.37 % to 2.88 mm for GBDT, 43.02 % to 3.06 mm for GRNN, and 44.51 % to 2.98 mm for XGBoost for 2-band ratio transmittance. For 3-band ratio transmittance the RMSE dropped 46.00 % to 2.90 mm, 46.55 % to 2.87 mm, 43.76 % to 3.02 mm, and 44.69 % to 2.97 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. Under all weather conditions, the retrieval accuracy of new PWV estimates (RMSE = 4.71 mm ~ 5.28 mm) is still better than that of the official MOD05 PWV product in confident-clear conditions (RMSE = 5.37 mm), implying that the four novel water vapor retrieval models can obtain good quality MODIS PWV under all-weather conditions, outperforming the official MOD05 confident-clear PWV product.

In probably clear conditions, the new PWV result showed an RMSE reduction of 42.12 % ~ 47.08 % for 2-band ratio transmittance and 42.48 % ~ 48.50 % for 3-band ratio transmittance. The RMSE reduction of probably cloudy conditions was in the range between 19.00 % and 27.60 % for 2-band ratio transmittance and in the range between 19.60 % and 29.00 % for 3-band ratio transmittance. For confident cloudy conditions, the new PWV, calculated with 2-band ratio transmittance, reduced RMSE by 63.70 % to 6.24 mm for BPNN, by 62.19 % to 6.50 mm for GBDT, by 59.40 % to 6.98 mm for GRNN, and by 59.63 % to 6.94 mm for XGBoost, compared with the MOD05 product (RMSE = 17.19 mm). For the new PWV result retrieved with 3-band ratio transmittance, the RMSE dropped 62.30 % to 6.48 mm, 60.67 % to 6.76 mm, 59.16 % to 7.02 mm, and 58.12 % to 7.20 mm for BPNN, GBDT, GRNN, and XGBoost, respectively. The algorithms exhibited higher accuracy

Table 5
The seasonal comparison between the satellite-based MODIS NIR PWV retrievals against the reference ground-based GPS PWV retrievals in 2018–2019 in both Australia and China under both all weather conditions and confident clear conditions.

		All Weather						Confident Clear					
		Slope	Offset	R^2	RMSE (mm)	MB (mm)	RMSE reduction	Slope	Offset	R^2	RMSE (mm)	MB (mm)	RMSE reduction
Spring	MOD05	0.57	8.68	0.29	10.13	−2.80	–	0.80	0.29	0.89	4.40	3.03	–
	BPNN, 2-Channel Ratio	0.98	−0.01	0.81	4.56	0.39	54.99 %	0.99	0.17	0.87	2.78	0.02	36.82 %
	GBDT, 2-Channel Ratio	0.96	0.15	0.79	4.76	0.50	53.01 %	0.97	0.21	0.88	2.76	0.26	37.27 %
	GRNN, 2-Channel Ratio	1.01	−0.57	0.77	4.97	0.38	50.94 %	1.02	−0.52	0.86	2.88	0.25	34.55 %
	XGBoost, 2-Channel Ratio	0.94	0.61	0.76	5.08	0.35	49.85 %	0.97	0.25	0.87	2.82	0.15	35.91 %
	BPNN, 3-Channel Ratio	0.98	−0.14	0.80	4.63	0.46	54.29 %	0.98	0.08	0.87	2.77	0.22	37.05 %
	GBDT, 3-Channel Ratio	0.96	0.19	0.78	4.86	0.50	52.02 %	0.96	0.24	0.88	2.77	0.32	37.05 %
	GRNN, 3-Channel Ratio	1.01	−0.59	0.77	4.97	0.39	50.94 %	1.02	−0.56	0.87	2.85	0.27	35.23 %
	XGBoost, 3-Channel Ratio	0.94	0.64	0.76	5.16	0.39	49.06 %	0.96	0.30	0.87	2.85	0.24	35.23 %
Summer	MOD05	0.36	19.68	0.15	16.52	−5.09	–	0.78	0.58	0.91	6.74	5.48	–
	BPNN, 2-Channel Ratio	0.99	0.03	0.82	5.78	0.30	65.01 %	0.98	0.63	0.87	3.45	−0.28	48.81 %
	GBDT, 2-Channel Ratio	0.98	0.26	0.81	5.92	0.34	64.16 %	0.99	0.46	0.87	3.49	−0.16	48.22 %
	GRNN, 2-Channel Ratio	1.02	−0.93	0.78	6.47	0.30	60.84 %	1.05	−0.89	0.85	3.73	−0.12	44.66 %
	XGBoost, 2-Channel Ratio	0.96	0.92	0.80	6.21	0.08	62.41 %	0.98	0.68	0.86	3.60	−0.32	46.59 %
	BPNN, 3-Channel Ratio	0.99	−0.34	0.81	6.03	0.61	63.50 %	1.00	−0.09	0.87	3.45	0.11	48.81 %
	GBDT, 3-Channel Ratio	0.98	0.06	0.80	6.14	0.45	62.83 %	0.99	0.32	0.87	3.45	−0.02	48.81 %
	GRNN, 3-Channel Ratio	1.03	−1.06	0.78	6.51	0.34	60.59 %	1.05	−1.05	0.85	3.70	−0.04	45.10 %
	XGBoost, 3-Channel Ratio	0.96	0.91	0.78	6.46	0.18	60.90 %	0.98	0.56	0.86	3.57	−0.16	47.03 %
Autumn	MOD05	0.50	10.94	0.26	11.80	−2.71	–	0.78	0.16	0.92	5.66	4.39	–
	BPNN, 2-Channel Ratio	0.93	0.87	0.84	4.70	0.46	60.17 %	0.99	0.38	0.90	2.82	−0.30	50.18 %
	GBDT, 2-Channel Ratio	0.93	1.06	0.83	4.86	0.41	58.81 %	0.99	0.41	0.90	2.79	−0.22	50.71 %
	GRNN, 2-Channel Ratio	0.96	0.38	0.80	5.14	0.44	56.44 %	1.02	−0.12	0.89	2.99	−0.24	47.17 %
	XGBoost, 2-Channel Ratio	0.91	1.66	0.80	5.19	0.16	56.02 %	0.98	0.68	0.89	2.95	−0.36	47.88 %
	BPNN, 3-Channel Ratio	0.93	0.68	0.83	4.79	0.63	59.41 %	0.98	0.41	0.90	2.83	−0.08	50.00 %
	GBDT, 3-Channel Ratio	0.93	0.90	0.82	4.94	0.53	58.14 %	0.98	0.34	0.91	2.77	−0.06	51.06 %
	GRNN, 3-Channel Ratio	0.96	0.32	0.80	5.16	0.48	56.27 %	1.03	−0.22	0.89	2.94	−0.17	48.06 %
	XGBoost, 3-Channel Ratio	0.91	1.50	0.80	5.29	0.29	55.17 %	0.98	0.58	0.90	2.91	−0.22	48.59 %
Winter	MOD05	0.55	5.65	0.32	6.91	−1.27	–	0.75	0.53	0.86	3.87	2.62	–
	BPNN, 2-Channel Ratio	0.98	0.05	0.81	3.18	0.20	53.98 %	0.99	0.06	0.86	2.20	−0.01	43.15 %
	GBDT, 2-Channel Ratio	0.98	−0.01	0.80	3.20	0.24	53.69 %	0.99	−0.02	0.86	2.15	0.12	44.44 %
	GRNN, 2-Channel Ratio	1.04	−0.49	0.75	3.61	0.06	47.76 %	1.06	−0.57	0.85	2.25	0.00	41.86 %
	XGBoost, 2-Channel Ratio	0.96	0.33	0.78	3.42	0.12	50.51 %	0.99	0.09	0.86	2.20	0.05	43.15 %
	BPNN, 3-Channel Ratio	0.98	−0.10	0.79	3.31	0.31	52.10 %	0.98	0.07	0.85	2.27	0.10	41.34 %
	GBDT, 3-Channel Ratio	0.98	−0.04	0.79	3.34	0.29	51.66 %	0.98	−0.06	0.87	2.15	0.25	44.44 %
	GRNN, 3-Channel Ratio	1.04	−0.53	0.75	3.61	0.07	47.76 %	1.06	−0.64	0.86	2.22	0.02	42.64 %
	XGBoost, 3-Channel Ratio	0.96	0.31	0.76	3.54	0.15	48.77 %	0.98	0.00	0.86	2.18	0.17	43.67 %

enhancement (larger RMSE reduction) in retrieving MODIS NIR water vapor estimates under confident cloudy conditions.

The new PWV showed a better accuracy with respect to in-situ GPS reference PWV in all the seasons of the year. For all-weather conditions, the seasonal reduction in RMSE was 49.85 % ~ 54.99 % in spring, 60.84 % ~ 65.01 % in summer, 56.01 % ~ 60.17 % in autumn, and 47.76 % ~ 53.98 % in winter for 2-band ratio approach, and for 3-band ratio approach it was 49.06 % ~ 54.29 %, 60.59 % ~ 63.50 %, 55.17 % ~ 59.41 %, and 47.76 % ~ 52.10 % in spring, summer, autumn, and winter, respectively. In confident-clear conditions, the seasonal RMSE dropped 34.55 % ~ 37.27 % (spring), 44.66 % ~ 48.81 % (summer), 47.17 % ~ 50.71 % (autumn), and 41.86 % ~ 44.44 % (winter) for 2-band ratio transmittance and 35.23 % ~ 37.05 % (spring), 45.10 % ~ 48.81 % (summer), 48.06 % ~ 51.06 % (autumn), and 41.34 % ~ 44.44 % (winter) for 3-band ratio transmittance. The four machine learning retrieval models showed a smaller seasonal PWV accuracy improvement in spring and winter, compared to summer and autumn. The discrepancy in accuracy enhancement of spring and winter when compared to summer and autumn could be due to the magnitude of PWV values. In spring and winter (i.e. dry atmospheric conditions), the PWV values are generally smaller than those in summer and winter (i.e. wet atmospheric conditions). Lower PWV values can result in smaller RMSE, as shown in Xu and Liu (2021a). Therefore in spring and winter, the machine learning based PWV retrieval approaches could have a smaller accuracy enhancement, i.e. smaller RMSE reduction. In addition, the new PWV exhibited smaller daily-scale and station-wise RMSE with respect to in-situ GPS reference PWV, compared with the official MOD05 PWV product.

Inter-comparison showed that under all-weather conditions, the new algorithms had similar performance in both Australia (RMSE reduction 51.87 % ~ 57.44 %) and in China (RMSE reduction 61.45 % ~ 65.16 %). For confident-clear conditions, the RMSE of the new PWV result reduced 44.90 % ~ 48.48 % in Australia and 37.16 % ~ 41.75 % in China. This indicated that the performance of the retrieval models in Australia (representative of the Southern Hemisphere) is comparable to that in China (representative of the Northern Hemisphere). In addition, both new daily averaged PWV showed a better agreement with in-situ GPS daily mean PWV under all-weather and confident-clear conditions in 2018 and 2019, showing that the four new algorithms are not dependent on the time period. This implies that the retrieval algorithms are not dependent on the geolocations or observation periods.

An enhanced MODIS NIR PWV retrieval algorithm showed an RMSE reduction of 22.48 % for 2-band ratio approach and 21.69 % for 3-band ratio approach under confident clear conditions in the North America region, compared with the official MOD05 product (RMSE = 7.67 mm) (He and Liu, 2020). In the recent work conducted by Ma et al. (2022), an improved PWV retrieval algorithm was developed for MODIS NIR channels, which is based on the BPNN approach considering land surfaces. The retrieval algorithm reduced the RMSE of MOD05 confident-clear PWV by 66.32 % for 2-band ratio transmittance and by 68.67 % for 3-band ratio transmittance based on the data in 2020 in the North America region. Both He and Liu (2020) and Ma et al. (2022) focused on confident-clear conditions to enhance the MODIS NIR PWV retrieval accuracy. The retrieval algorithms in our work show a much larger RMSE reduction under confident clear conditions than the retrieval algorithm developed by He and Liu (2020). This work is for the first time, to our knowledge, to retrieve MODIS NIR all-weather PWV estimates considering multiple influence parameters – location, cloud, season, and solar zenith angle, superior to the official MOD05 confident-clear PWV product.

The four new machine learning based PWV retrieval algorithms provide a more accurate way to retrieve the PWV data from MODIS NIR channels, as the retrieval models have considered multiple influence factors such as location, cloud, season, and solar zenith angle. The new all-weather PWV estimates exhibited a better retrieval accuracy than the official MOD05 confident-clear PWV product.

6. Conclusion

In this research, we have developed four novel water vapor retrieval algorithms based on machine learning methods to derive MODIS NIR PWV data under all-weather conditions. The new MODIS-derived NIR PWV retrievals are evaluated utilizing reference GPS-derived PWV data in 2018–2019, which are independent of the training data of 2017. The main conclusions are given as follows.

1. The new all-weather PWV estimates exhibit a better agreement with ground-based GPS-estimated PWV observations, showing $R^2 = 0.83 \sim 0.86$, RMSE = 4.71 mm ~ 5.28 mm, and MB = 0.18 mm ~ 0.51 mm. This could be because the retrieval algorithms have considered multiple dependence factors that affect the MODIS NIR PWV measurement accuracy in the models.
2. The all-weather PWV data estimated using BPNN algorithm and 2-band ratio transmittance have the highest performance, whereas using XGBoost algorithm and 3-band ratio transmittance has the worst performance.
3. The new PWV results agree better than the original PWV with in-situ GPS PWV data in each different weather condition, exhibiting a reduction in RMSE of 43.02 % ~ 46.55 % (confident clear), 42.12 % ~ 48.50 % (probably clear), 19.00 % ~ 29.00 % (probably cloudy), and 58.12 % ~ 63.70 % (confident cloudy).
4. The retrieval algorithms can improve the seasonal performance of PWV estimation from MODIS NIR measurements, with a higher accuracy enhancement in summer and autumn.
5. The variation trend of daily mean water vapor of the new PWV estimates agrees better with that of in-situ GPS-retrieved PWV estimates. The station-wise RMSE of MODIS-measured NIR water vapor is generally reduced to below 8 mm (all-weather) and below 4 mm (confident clear) at most stations.
6. The retrieval methods present little spatial or temporal dependence, which have great potentials to be applied to other global regions as well as other temporal periods.

CRediT authorship contribution statement

Jiafei Xu: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Zhizhao Liu:** Supervision, Conceptualization, Methodology, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgment

The work described in this paper was partially supported by the grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU/RGC 15221620/B-Q80Q and PolyU/RGC 15211919/B-Q73B). The support by the Emerging Frontier Area (EFA) Scheme of Research Institute for Sustainable Urban Development (RISUD) of The Hong Kong Polytechnic University under Grant 1-BBWJ is also acknowledged. The authors would like to thank the two anonymous reviewers and the editorial team for their precious time and invaluable comments that help improve the quality of this manuscript. The authors gratefully acknowledge the support by the National Aeronautics and Space Administration (NASA) Level-1 and Atmosphere Archive and Distribution System (LAADS)

Distributed Active Archive Center (DAAC) for providing MODIS data. The authors would like to thank the Geoscience Australia and the Crustal Movement Observation Network of China (CMONOC) for providing GPS data. The European Centre for Medium-Range Weather Forecasts (ECMWF) is acknowledged for providing atmospheric pressure and surface temperature data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2022.103050.

References

Bevis, M., Businger, S., Herring, T.A., Rocken, C., Anthes, R.A., Ware, R.H., 1992. GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. J. Geophys. Res. Atmospheres 97, 15787–15801. https://doi.org/10.1029/92JD01517.

Bevis, M., Businger, S., Chiswell, S., Herring, T., Anthes, R., Rocken, C., Ware, R., 1994. Gps Meteorology - Mapping Zenith Wet Delays Onto Precipitable Water. J. Appl. Meteorol. 33, 379–386. https://doi.org/10.1175/1520-0450(1994)033<0379:GMMZWD>2.0.CO;2.

Ccoica-Lopez, K.L., Pasapera-Gonzales, J.J., Jimenez, J.C., 2019. Spatio-Temporal Variability of the Precipitable Water Vapor over Peru through MODIS and ERA-Interim Time Series. Atmosphere 10, 192. https://doi.org/10.3390/atmos10040192.

Chavez, P.S., 1988. An Improved Dark-Object Subtraction Technique for Atmospheric Scattering Correction of Multispectral Data. Remote Sens. Environ. 24 (3), 459–479.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., 2015. Xgboost: extreme gradient boosting. R Package Version 04–2 (1), 1–4.

Chen, Z., Xu, J., Zhang, H., 2019. Evaluation of HJ-1A/B CCD Surface Reflectance Products Using the VNIR and MODIS-Based Atmospheric Correction Approaches. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 437–449. https://doi.org/10.1109/JSTARS.2018.2875263.

Feng, Y., Du, S., 2020. CLIMATE CHANGES AND LANDSCAPE RESPONSES OF CHINA DURING THE PAST 40 YEARS (1979–2018) UNDER KÖPPEN-GEIGER CLIMATE CLASSIFICATION, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Presented at the XXIV ISPRS Congress, Commission III (Volume V-3-2020) - 2020 edition, Copernicus GmbH, pp. 731–737. 10.5194/isprs-annals-V-3-2020-731-2020.

Fragkos, K., Antonescu, B., Giles, D.M., Ene, D., Boldeanu, M., Efstathiou, G.A., Belegante, L., Nicolae, D., 2019. Assessment of the total precipitable water from a sun photometer, microwave radiometer and radiosondes at a continental site in southeastern Europe. Atmospheric Meas. Tech. 12, 1979–1997. https://doi.org/10.5194/amt-12-1979-2019.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. Nonlinear Methods and Data Mining 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.

Gao, B.-C., Kaufman, Y.J., 2003. Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels. J. Geophys. Res. Atmospheres 108 (D13), n/a–n/a.

He, J., Liu, Z., 2019. Comparison of Satellite-Derived Precipitable Water Vapor Through Near-Infrared Remote Sensing Channels. IEEE Trans. Geosci. Remote Sens. 57 (12), 10252–10262.

He, J., Liu, Z., 2020. Water Vapor Retrieval From MODIS NIR Channels Using Ground-Based GPS Data. IEEE Trans. Geosci. Remote Sens. 58, 3726–3737. https://doi.org/10.1109/TGRS.2019.2962057.

Held, I.M., Soden, B.J., 2000. Water Vapor Feedback and Global Warming. Annu. Rev. Energy Environ. 25, 441–475. https://doi.org/10.1146/annurev.energy.25.1.441.

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

Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Nakajima, T., Lavenu, F., Jankowiak, I., Smirnov, A., 1998. AERONET—A federated instrument network and data archive for aerosol characterization. Remote Sens. Environ. 66 (1), 1–16.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8.

Ichoku, C., Levy, R., Kaufman, Y.J., Remer, L.A., Li, R.-R., Martins, V.J., Holben, B.N., Abuhassan, N., Slutsker, I., Eck, T.F., 2002. Analysis of the performance characteristics of the five-channel Microtops II Sun photometer for measuring aerosol optical thickness and precipitable water vapor. J. Geophys. Res. Atmospheres 107, AAC-5.

Justice, C.O., Vermote, E., Townshend, J.R.G., Defries, R., Roy, D.P., Hall, D.K., Salomonson, V.V., Privette, J.L., Riggs, G., Strahler, A., Lucht, W., Myneni, R.B., Knyazikhin, Y., Running, S.W., Nemani, R.R., Wan, Z.M., Huete, A.R., van Leeuwen, W., Wolfe, R.E., Giglio, L., Muller, J.P., Lewis, P., Barnsley, M.J., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. IEEE Trans. Geosci. Remote Sens. 36, 1228–1249. https://doi.org/10.1109/36.701075.

Kaufman, Y.J., Gao, B.-C., 1992. Remote sensing of water vapor in the near IR from EOS/MODIS. IEEE Trans. Geosci. Remote Sens. 30, 871–884.

King, M.D., Kaufman, Y.J., Menzel, W.P., Tanré, D., 1992. Remote Sensing of Cloud, Aerosol, and Water Vapor Properties from the Moderate Resolution Imaging Spectrometer (MODIS). IEEE Trans. Geosci. Remote Sens. 30, 2–27. https://doi.org/10.1109/36.124212.

King, M.D., Menzel, W.P., Kaufman, Y.J., Tanré, D., Gao, B.-C., Platnick, S., Ackerman, S. A., Remer, L.A., Pincus, R., Hubanks, P.A., 2003. Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS. IEEE Trans. Geosci. Remote Sens. 41, 442–458.

Li, X., Dick, G., Lu, C., Ge, M., Nilsson, T., Ning, T., Wickert, J., Schuh, H., 2015. Multi-GNSS Meteorology: Real-Time Retrieving of Atmospheric Water Vapor From BeiDou, Galileo, GLONASS, and GPS Observations. IEEE Trans. Geosci. Remote Sens. 53, 6385–6393. https://doi.org/10.1109/TGRS.2015.2438395.

Li, X., Long, D., 2020. An improvement in accuracy and spatiotemporal continuity of the MODIS precipitable water vapor product based on a data fusion approach. Remote Sens. Environ. 248, 111966 https://doi.org/10.1016/j.rse.2020.111966.

Li, Z., Muller, J.-P., Cross, P., 2003. Comparison of precipitable water vapor derived from radiosonde, GPS, and Moderate-Resolution Imaging Spectroradiometer measurements. J. Geophys. Res, Atmospheres, p. 108.

Liu, Z., Wong, M.S., Nichol, J., Chan, P.W., 2013. A multi-sensor study of water vapour from radiosonde, MODIS and AERONET: a case study of Hong Kong. Int. J. Climatol. 33, 109–120. https://doi.org/10.1002/joc.3412.

Ma, X., Yao, Y., Zhang, B., Qin, Y., Zhang, Q., Zhu, H., 2022. An Improved MODIS NIR PWV Retrieval Algorithm Based on an Artificial Neural Network Considering the Land-Cover Types. IEEE Trans. Geosci. Remote Sens. 60, 1–12. https://doi.org/10.1109/TGRS.2022.3170078.

McCarthy, M.P., Thorne, P.W., Titchner, H.A., 2009. An analysis of tropospheric humidity trends from radiosondes. J. Clim. 22, 5820–5838. https://doi.org/10.1175/2009JCLI2879.1.

Miloshevich, L.M., Vomel, H., Whiteman, D.N., Lesht, B.M., Schmidlin, F.J., Russo, F., 2006. Absolute accuracy of water vapor measurements from six operational radiosonde types launched during AWEX-G and implications for AIRS validation. J. Geophys. Res.-Atmospheres 111, D09S10. https://doi.org/10.1029/2005JD006083.

Muller, C.J., Back, L.E., O'Gorman, P.A., Emanuel, K.A., 2009. A model for the relationship between tropical precipitation and column water vapor. Geophys. Res. Lett. 36 https://doi.org/10.1029/2009GL039667.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., Zhang, H., 2013. Anthropogenic and natural radiative forcing. Clim. Change 2013 Phys. Sci. Basis Contrib. Work. Group Fifth Assess. Rep. Intergov. Panel Clim. Change 659–740.

Pigram, C., 2012. Geoscience Australia – a multi disciplined agency. Episodes J. Int. Geosci. 35, 524–525. 10.18814/epiiugs/2012/v35i4/009.

Platnick, S., King, M.D., Ackerman, S.A., Menzel, W.P., Baum, B.A., Riedi, J.C., Frey, R. A., 2003. The MODIS cloud products: Algorithms and examples from Terra. IEEE Trans. Geosci. Remote Sens. 41, 459–473. https://doi.org/10.1109/TGRS.2002.808301.

Rocken, C., Hove, T.V., Ware, R., 1997. Near real-time GPS sensing of atmospheric water vapor. Geophys. Res. Lett. 24, 3221–3224. https://doi.org/10.1029/97GL03312.

Ross, R.J., Elliot, W.P., 2001. Radiosonde-based Northern Hemisphere tropospheric water vapor trends. J. Clim. 14, 1602–1612. https://doi.org/10.1175/1520-0442(2001)014<1602:RBNHTW>2.0.CO;2.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0.

Salomonson, V.V., Barnes, W.L., Maymon, P.W., Montgomery, H.E., Ostrow, H., 1989. MODIS: advanced facility instrument for studies of the Earth as a system. IEEE Trans. Geosci. Remote Sens. 27, 145–153. https://doi.org/10.1109/36.20292.

Schläpfer, D., Borel, C.C., Keller, J., Itten, K.I., 1998. Atmospheric Precorrected Differential Absorption Technique to Retrieve Columnar Water Vapor. Remote Sens. Environ. 65, 353–366. https://doi.org/10.1016/S0034-4257(98)00044-3.

Schneider, T., O'Gorman, P.A., Levine, X.J., 2010. Water Vapor and the Dynamics of Climate Changes. Rev. Geophys. 48 https://doi.org/10.1029/2009RG000302.

Sherwood, S.C., Roca, R., Weckwerth, T.M., Andronova, N.G., 2010. Tropospheric water vapor, convection, and climate. Rev. Geophys. 48 https://doi.org/10.1029/2009RG000301.

Soden, B.J., Wetherald, R.T., Stenchikov, G.L., Robock, A., 2002. Global cooling after the eruption of Mount Pinatubo: A test of climate feedback by water vapor. Science 296, 727–730.

Specht, D.F., 1991. A general regression neural network. IEEE Trans. Neural Netw. 2, 568–576. https://doi.org/10.1109/72.97934.

Stocker, T.F., Intergovernmental Panel on Climate Change (Eds.), 2013. Climate change 2013: the physical science basis ; summary for policymakers, a report of Working Group I of the IPCC, technical summary, a report accepted by Working Group I of the IPCC but not approved in detail and frequently asked questions ; part of the Working Group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change, New York.

Trenberth, K.E., Fasullo, J., Smith, L., 2005. Trends and variability in column-integrated atmospheric water vapor. Clim. Dyn. 24, 741–758.

Vaquero-Martínez, J., Antón, M., de Galisteo, J.P.O., Cachorro, V.E., Costa, M.J., Román, R., Bennouna, Y.S., 2017. Validation of MODIS integrated water vapor

product against reference GPS data at the Iberian Peninsula. Int. J. Appl. Earth Obs. Geoinformation 63, 214–221.

Vaquero-Martínez, J., Antón, M., de Galisteo, J.P.O., Cachorro, V.E., Álvarez-Zapatero, P., Román, R., Loyola, D., Costa, M.J., Wang, H., Abad, G.G., 2018. Inter-comparison of integrated water vapor from satellite instruments using reference GPS data at the Iberian Peninsula. Remote Sens. Environ. 204, 729–740.

Wang, R., Gentine, P., Yin, J., Chen, L., Chen, J., Li, L., 2021. Long-term relative decline in evapotranspiration with increasing runoff on fractional land surfaces. Hydrol. Earth Syst. Sci. 25, 3805–3818. https://doi.org/10.5194/hess-25-3805-2021.

Wang, Q., Zhang, P., 2001. The Initial Result of Crust Movement Observation Network of China : GPS-derived Velocity Field (1998–2001). AGU Fall Meet. Abstr. 1, 0220.

Xu, J., Liu, Z., 2021a. The First Validation of Sentinel-3 OLCI Integrated Water Vapor Products Using Reference GPS Data in Mainland China. IEEE Trans. Geosci. Remote Sens. 1–17 https://doi.org/10.1109/TGRS.2021.3099168.

Xu, J., Liu, Z., 2021b. Radiance-based retrieval of total water vapor content from sentinel-3A OLCI NIR channels using ground-based GPS measurements. Int. J. Appl. Earth Obs. Geoinformation 104, 102586. https://doi.org/10.1016/j.jag.2021.102586.

Xu, J., Liu, Z., 2022. A Linear Regression of Differential PWV Calibration Model to Improve the Accuracy of MODIS NIR All-weather PWV Products Based on Ground-based GPS PWV Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 1–24 https://doi.org/10.1109/JSTARS.2022.3204823.

Ye, H., Fetzer, E.J., Wong, S., Behrangi, A., Olsen, E.T., Cohen, J., Lambrigtsen, B.H., Chen, L., 2014. Impact of increased water vapor on precipitation efficiency over northern Eurasia. Geophys. Res. Lett. 41, 2941–2947. https://doi.org/10.1002/2014GL059830.

Yuan, Y., Zhang, K., Rohm, W., Choy, S., Norman, R., Wang, C.-S., 2014. Real-time retrieval of precipitable water vapor from GPS precise point positioning. J. Geophys. Res. Atmospheres 119, 10044–10057. https://doi.org/10.1002/2014JD021486.