This is the Pre-Published Version.

The following publication He, J., & Liu, Z. (2021). Applying the New MODIS-Based Precipitable Water Vapor Retrieval Algorithm Developed in the North Hemisphere to the South Hemisphere. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-12 is available at https://doi.org/10.1109/TGRS.2021.3059876

Applying the New MODIS-based Precipitable Water Vapor Retrieval Algorithm Developed in the North Hemisphere to the South Hemisphere

Jia He, Zhizhao Liu

***Abstract*— A new algorithm to retrieve water vapor from MODIS NIR channels using the ensemble-based empirical regression model, which was developed based on the North Hemisphere (western North America) data, was for the first time applied and validated to the South Hemisphere, mainly the Australia and its surrounding regions. By employing the empirical regression algorithm to retrieve water vapor from MODIS Level 1 reflectance data, the wet bias of MODIS product has been significantly reduced. Validation against GPS water vapor observations over the period 1 January 2017 to 31 December 2019 in and around Australia show that the RMSE of water vapor data obtained from MODIS/Terra has reduced by 58.53% from 5.712 mm to 2.369 mm when using 2-channel ratio transmittance and has reduced by 56.14% to 2.505 mm when using 3-channel ratio transmittance. For the data obtained from MODIS/Aqua, the RMSE has reduced by 49.17% from 5.170 mm to 2.628 mm using 2-channel ratio transmittance and has reduced by 46.60% to 2.761 mm using 3-channel ratio transmittance, respectively. In addition, validations of the retrieved water vapor results over such a large research area (0°-55°S in latitude and 95°-180°E in longitudes) also show no temporal or spatial dependency, implying that the algorithm is homogeneous, accurate, and robust.**

***Index Terms*— MODIS, GPS, Water vapor, Australia**

I. Introduction

Water vapor is one of the most important climate variables and it plays a key role in atmospheric processes, hydrological circulation, weather formation, and climate change [1], [2]. It contributes to about 60% of the natural greenhouse effects and provides the largest positive feedback in predictive models of climate change [3]. Furthermore, water vapor is a parameter often used in remote sensing techniques while observing the Earth's surface [4]. Due to its importance, the Global Climate Observing System (GCOS) declared the total column water vapor as an essential climate variable (ECV) and should meet the observation requirement of stability of 0.3% per decade and no more than 5% measurement uncertainty [5].

A number of different water vapor observation techniques have been developed. However, observing water vapor with high precision is still challenging as it is highly variable both spatially and temporally [6], [7]. Ground-based observations, such as radiosonde and Global Navigation Satellite Systems (GNSS) / Global Positioning System (GPS) data, are usually used as ground-truth in validation analysis because of their high observation precision [8], [9]. The radiosonde has a long record of historical water vapor observations but it only observes twice per day or once daily [8], [10]. The large time difference between the radiosonde observation and satellite overpass may introduce validation uncertainties [10]. GNSS/GPS provides continuous observation in all-weather conditions every hour [9], [11], therefore, the GPS derived water vapor data are employed as reference data in this study. The space-based remote sensing satellite technique is the most cost-effective way of observing water vapor globally but it has lower temporal resolution and larger uncertainty compared to ground-based approach [12]–[14].

The MODerate resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites is probably the most widely used space instrument to observe water vapor with both near-infrared (NIR) and traditional infrared (IR) bands [15]. The operational products of MODIS NIR channels (MOD05 for Terra and MYD05 for Aqua) were calculated from a pre-calculated look-up table generated from radiative transfer models, with an estimated uncertainty of 5% to 10% [15], [16]. Evaluations show that the performance of MODIS NIR products varies from place to place. For instance, validation over Germany shows an overestimation of MODIS against GPS by 7% to 14% [17]. Evaluation of MODIS NIR product against GPS in mainland China shows that the root mean squares error (RMSE) is 5.76 mm [12], while the RMSE between MODIS and radiosonde in Hong Kong is 13.09 mm [18]. Research shows that the RMSE varies from 5.87 to 9.37 mm at two stations in India [19]. Inter-comparison of MODIS against GPS over the US shows that the RMSE is 5.05 mm [14].

Satellite-based remote sensing of water vapor has its

The support from the Key Program of the National Natural Science Foundation of China (No. 41730109) is acknowledged. The supports by the Hong Kong Research Grants Council (RGC) (No. B-Q52W PolyU 152149/16E and B-Q61L PolyU 152222/17E) are acknowledged. The Emerging Frontier Area (EFA) Scheme of Research Institute for Sustainable Urban Development (RISUD) of the Hong Kong Polytechnic University (No. 1-BBWJ) is also acknowledged.

Jia He is with the Department of Land Surveying and Geo-Informatics, as well as the Research Institute for Sustainable Urban Development, Hong Kong Polytechnic University, Hung Hum, Hong Kong (e-mail: gail.he@connect.polyu.hk).

Zhizhao Liu (correspondence author) is with the Department of Land Surveying and Geo-Informatics, as well as the Research Institute for Sustainable Urban Development, Hong Kong Polytechnic University, Hung Hum, Hong Kong (e-mail: lszzliu@polyu.edu.hk).

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

advantages of large coverage and low costs despite its relatively low accuracy. Thus, it has been widely used in the research and application communities. In order to enhance its accuracy, development of improved algorithms for MODIS water vapor retrieval in the NIR channels has been carried out in several studies. By introducing empirical correction coefficients for transmittance calculation, the wet bias of NIR product is reduced and the validation on a global scale shows that the RMSE is between 0.9 and 2 mm [20]. Localized optimization of water vapor measurement in western Iran shows that the retrieval accuracy is improved at the local scale and the RMSE is reduced to 2.702 mm [21]. An empirical regression method has been proposed to improve MODIS NIR water vapor retrieval and the new algorithm can reduce RMSE to 2.243 mm over the western North America and the RMSE of some global stations to 5.946 mm [22].

In the past, most of the work was conducted in the North Hemisphere. Unfortunately, the performance of the algorithm in the Southern Hemisphere has not been well studied. In our previous study, our new algorithm was derived based on GPS water vapor data from western North America [22]. The algorithm was validated to perform well in western North America. However, its performance in the other regions, particularly in the South Hemisphere where the climate is significantly different from western North America, has not been validated.

Australia and its neighboring regions are selected in this study for the following reasons: (1) Australia covers a large landmass and its diverse weather and climate conditions are representative of the South Hemisphere; (2) the Geoscience Australia has developed a dense GNSS network across Australia and high accuracy water vapor products derived from the GNSS data [23] can be used for assessment of our developed algorithm; (3) the Australian climate is particularly sensitive to El Niño-Southern Oscillation (ENSO) and La Niña events, which are strongly related to water vapor distribution [24], [25].

Our previous study have derived an empirical regression model for water vapor retrieval from MODIS data, based on the water vapor data estimated from the western North America GNSS network [22]. In order to evaluate the model's performance in the South Hemisphere, a new set of MODIS water vapor data is first retrieved using the coefficients of the empirical model. The new water vapor dataset is then validated by comparing against the water vapor data derived from 520 GPS stations across the Australia and its surrounding regions such as New Zealand and Pacific Islands. Section 2 provides information on the research area. Section 3 gives a detailed description of the datasets used in this work, including the calculation method of GPS precipitable water vapor (PWV) and algorithm of water vapor retrieval from MODIS NIR channels. Section 4 discusses the validation results of calibrated water vapor data using our algorithm as well as MODIS's operational water vapor product in comparison to the GPS water vapor observations. Section 5 presents the conclusion of this research.

II. Study Area

The research covers the area in Australia and its neighboring regions, with latitudes from 0° S to 55° S, and longitudes from 95° E to 180° E. The area includes a wide variety of landscapes, including the Australian continent, New Zealand, and several islands of the Malay Archipelago and the New Guinea. For the Australia continent, there are tropical rainforests in the north-east, mountains in the south-east, and desert in the center. As the continent is surrounded by the Indian and Pacific oceans, the climate of the research area is significantly influenced by ocean currents, such as the Indian Ocean Dipole and the El Niño-Southern Oscillation [26], causing the rainfall to vary markedly from year to year [24]. The majority part of Northern Australia has a tropical, predominantly summer-rainfall, while the south-west part of Australia has a Mediterranean climate. The south-east has a climate ranging from oceanic to humid-subtropical, and the interior is arid to semi-arid. There are four distinct seasons in this area, which are at opposite times to those in the North Hemisphere. The Malay Archipelago and the New Guinea have a tropical climate. Except in high elevations, most areas have a warm, humid climate throughout the year with seasonal variation related to the northeast monsoon [27]. New Zealand is predominantly oceanic, with mild temperature and humid climate all year round [28].

Water vapor data from a total of 520 GPS stations equipped with GPS receivers and meteorological observation equipment are utilized as reference data in this study. These stations provide continuous atmospheric water vapor observations. The distribution map of the GPS stations is shown in Figure 1.

III. Data Description

Two types of data observed during 2017~2019 are employed in this research, including GPS-derived water vapor data that are used as reference data, and MODIS data Level 1 data on surface reflectance and geolocation, to which our model is applied and from which improved water vapor products are retrieved. The MODIS Level 2 water vapor product is also used for comparison purpose in this study. The descriptions of the data characteristics are listed in Table 1.

In order to ensure enough data to be used in the model validation in the South Hemisphere, the data covers a period of three years from 1 January 2017 to 31 December 2019. Spatially and temporally collocated GPS and MODIS data collected under the cloud-free conditions are used in model validation. To reduce the temporal discrepancies between GPS and MODIS data, only data pairs with a time difference of less than 30 minutes are used in this study.

A. GPS PWV

GPS PWV is calculated based on the propagation delays of the atmosphere [9]. In this study, the zenith tropospheric delay (ZTD) data are obtained from the Asia Pacific Regional Geodetic Project (APRGP) GPS Campaign from Geoscience Australia (ftp://ftp.ga.gov.au/geodesyoutgoing/gnss/products/troposphere/rapid/). It is a project of the Geodetic Reference Frame Working Group of the Regional Committee of the United Nations Global Geospatial Information Management for

Asia and the Pacific (UN-GGIM-AP). The GPS ZTDs were processed using Bernese GNSS Software and the GPS solution was constrained to the ITRF2008 reference frame [23]. These ZTD data are then converted to PWV using the surface pressure, temperature and humidity profiles obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-5 [29]. The ERA-5 data from 00:00, 06:00, 12:00 and 18:00 UTC of each day were employed. The spatial resolution is 0.25° with 37 vertical pressure levels.

To calculate the zenith hydrostatic delay (ZHD) from GNSS stations, the Saastamoinen model [30] is employed:

$$\mathrm{ZHD} = \frac{0.0022768P_0}{(1 - 0.00266\cos(2\varphi) - 0.00028h_0)} \tag{1}$$

where P_0 represents the air pressure at the height of the station; φ is the station latitude, and h_0 is the height of the station.

Because the ZTD is modeled as a sum of ZHD and zenith wet delay (ZWD), the ZWD could be calculated using:

$$\mathrm{ZWD} = \mathrm{ZTD} - \mathrm{ZHD} \tag{2}$$

The ZWD is then converted to PWV [9]:

$$\mathrm{PWV} = \Pi \times \mathrm{ZWD} \tag{3}$$

where Π represents the dimensionless constant of proportionality. It is calculated using [9]:

$$\Pi = \frac{10^5}{461.495(k_3/T_m + k'_2)} \tag{4}$$

where k_3 equals to 3.776×10^5 K^2/hPa; k'_2 equals to 16.52 K/hPa. And T_m is the weighted mean temperature [31]:

$$T_m = \frac{\int_H^\infty e/T}{\int_H^\infty e/T^2} dh \tag{5}$$

where H represents the height of the GPS station; T represents the temperature at height h in degrees Kelvin (T=t+273.15). The e is the water vapor partial pressure, which is calculated using [32], [33]:

$$e = 6.1121(1.0007 + 3.46 \times 10^{-6} \times P) \times RH \times \exp\{\frac{[18.729 - \frac{T - 273.15}{227.3}](T - 273.15)}{T - 15.28}\} \tag{6}$$

where P is the total pressure (unit: hPa), RH represents the relative humidity (unitless). It should be noted that the geopotential height is approximately equal to geometric height. Moreover, a bilinear interpolation procedure is conducted at four surrounding grid points for relative humidity and temperature using the ERA-5 data to calculate PWV from these GPS sites. The Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) function [34] is employed to interpolate or extrapolate the value in geometric heights.

The accuracy of GPS-derived water vapor data has been evaluated with radiosonde water vapor data. The RMSE for the retrieved GPS water vapor against radiosonde observation is 1.48 mm [25]. Because of the high precision and high temporal resolution, the GPS data is suitable to be used as reference data in this study.

B. MODIS NIR PWV

MODIS is a passive whisk-broom scanning imaging spectroradiometer with 36 spectral bands onboard the Terra and Aqua satellites operated by NASA. Five NIR channels are used for water vapor retrieval, with three water vapor absorption channels centered at 905 nm, 936 nm and 940 nm, and two window channels centered at 865 nm and 1240 nm [16]. Detailed descriptions of the MODIS NIR channels are listed in Table 2.

The amount of water vapor in the atmosphere is estimated at each channel through the observation of its effect on absorbing radiance when it is transmitted down to the earth surface and reflected to the sensor [15]. The transmittance is usually estimated by measuring the mean radiance ratio of one absorption channel to one or more window channels. For a ground surface type of which reflectance varies linearly with wavelength, the function using a 2-channel ratio method to calculate the transmittance is defined as:

$$T_i \cong R_i = \frac{L_i}{L_2} \tag{7}$$

where T_i is the transmittance of channel i, which is approximately equal to the reflectance ratio R_i; L_i is the reflectance in absorption channel i (i=17, 18 and 19); L_2 is the reflectance in window channel 2 centered at 865 nm.

For ground surface covers with complex reflectance spectra, two window channels (band 2 and band 5) are required to calculate the transmittance in the absorption channel. The 3-channel ratio method is defined as:

$$T_i \cong R_i = \frac{L_i}{[C_1 L_2 + C_2 L_5]} \tag{8}$$

where the coefficients C_1 and C_2 are prescribed as 0.8 and 0.2, respectively. It is assumed that the reflectance ratio around 1 µm remains the same, or the reflectance ratio varies linearly [35].

Normally, the inverted amount of water vapor can be retrieved through further calculation with either a lookup table generated from the radiative transfer model [15], a regression method [22], or an artificial neural network [36]. In this article, the empirical regression method proposed by He and Liu (2020) is employed instead to calculate water vapor over the research area.

1) Operational Products

In assessing the performance of our new algorithm, the current products from MODIS are also employed and compared with GPS-derived water vapor. The accuracy of the MODIS products will be compared with that of calibrated water vapor data estimated from our new algorithm. The operational algorithm to calculate water vapor from MODIS NIR channels is based on the radiative transfer theory [15]. The relationship between the atmospheric transmission and water vapor content is simulated using High-Resolution Transmission (HITRAN) 2000 [15], [16]. The relationship can be simplified by an exponential formula written as:

$$T_w = exp\left(\alpha - \beta\sqrt{W^*}\right) \tag{9}$$

where T_w is the transmittance of water vapor; α and β are coefficients determined by surface type, and the W^* is water vapor along the sun-surface-sensor (slant) path [16].

2) Improved Water Vapor Retrieval

As the current MODIS product systematically overestimates water vapor values, a new algorithm was developed to retrieve water vapor from MODIS with improved accuracy based on the GPS water vapor data collected in western North America [22]. Spatially and temporally collocated GPS and MODIS data collected under cloud-free conditions were used for model development. Ensemble functions were generated from ten resampling training sets that were generated using bootstrap method, as the multiple training sets could average prediction errors and reduce the bias and variance errors [37]. This algorithm model the relationship between atmospheric transmittance and water vapor content through an empirical regression method. In contrast to operational algorithms, this new approach provides an effective way to estimate water vapor through MODIS Level 1 reflectance data without pre-observed atmospheric profiles. However, the model was previously derived with data obtained from the western North America region. To study its application in other places such as the South Hemisphere, the model and its coefficients are directly employed in this research to estimate the water vapor for Australia and its neighboring regions in the South Hemisphere.

The transmittance is firstly calculated using the Level 1 reflectance data from MODIS absorption channels using both 2-channel and 3-channel ratio methods. Then the transmittance is converted to water vapor on the sun-surface-sensor path using the following equation [22]:

$$T_w = a \exp\,(b\, W_i^{*}) + c \exp\,(d\, W_i^{*}) \qquad (10)$$

where the a, b, c and d are the coefficients determined through the least squares fitting [22]; T_w is the transmittance of the water vapor obtained from MODIS NIR channel i; W_i^* is the slant water vapor content at the channel i. The vertical total precipitable water vapor W_i for the corresponding absorption channel is calculated using:

$$W_i = W_i^{*} / \left(\frac{1}{cos\theta} + \frac{1}{cos\theta_0}\right) \qquad (11)$$

where θ represents the view zenith angle and θ_0 is the solar zenith angle.

Water vapor could be estimated at each absorption channel. As previous studies show, the three absorption channels have different sensitivities under different conditions [15], [22]. To get a more accurate water vapor estimation, the weighted mean value of the three absorption channels is calculated in this research:

$$W = f_1 W_{17} + f_2 W_{18} + f_3 W_{19} \qquad (12)$$

where W_{17}, W_{18} and W_{19} represent the water vapor values estimated from band 17, band 18 and band 19, respectively; f_1, f_2 and f_3 are the corresponding weighting parameters calculated based on the sensitivity in each band:

$$f_i = \frac{\eta_i}{\eta_1 + \eta_2 + \eta_3} \qquad (13)$$

where η_i represents the slope of the graph of transmission versus water vapor at each water vapor absorption band of MODIS.

By employing the above equations, new sets of ensemble members of water vapor are calculated for each absorption channel (W_{17}, W_{18} and W_{19}), and the weighted mean value of the three absorption channels is also estimated as the final calibrated water vapor value. The ensemble median is used to represent the calibrated water vapor for validation analysis.

An example of water vapor distribution observed on 27 December 2019, 0035 UTC from MODIS/Terra is displayed in Figure 2 with a 1 km * 1 km resolution. The operational product MOD05 (Figure 2a) along with the weighted mean value of the absorption channels calculated using the new algorithm with 2-channel ratio transmittance (Figure 2b) and 3-channel ratio transmittance (Figure 2c) is presented. Each data point is the ensemble median of the ensemble members at the corresponding pixels. Generally, the calibrated water vapor calculated using both 2-channel and 3-channel ratio transmittance are systematically smaller than the operational product, especially for area with high water vapor concentration. Furthermore, no stripe features appear in either of the calibrated water vapor maps, indicating that the algorithm is stable for further applications. To further assess the performance of the algorithm, validation against GPS reference water vapor data collected in Australia and its neighboring region is discussed in the following section.

IV. Validation of South Hemisphere MODIS Water Vapor Results

In the performance assessment of the calibrated MODIS PWV in the South Hemisphere, a few statistical metrics are employed. They are the coefficient of determination (R^2), which indicates the relationship strength between the calibrated water vapor and the reference GPS PWV; the mean bias (MB), which shows the systematic difference between the MODIS and GPS water vapor observations; the RMSE, which measures the overall agreement of the two datasets. The metrics are written as:

$$R^2 = \left[\frac{\sum_{i=1}^{n}(PWV_{R_i} - \overline{PWV}_R)(PWV_{O_i} - \overline{PWV}_O)}{\sqrt{\sum_{i=1}^{n}(PWV_{R_i} - \overline{PWV}_R)^2 (PWV_{O_i} - \overline{PWV}_O)^2}}\right]^2 \qquad (14)$$

$$MB = \frac{1}{N}\sum_{i=1}^{n}\left(PWV_{O_i} - PWV_{R_i}\right) \qquad (15)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|PWV_{O_i} - PWV_{R_i}\right|^2} \qquad (16)$$

where the PWV_{R_i} is the reference PWV derived from Australian and its neighboring GPS networks; $\overline{PWV}_R$ is the mean PWV from GPS; PWV_{O_i} is the observed water vapor obtained from MODIS, including the existing MODIS products and the calibrated water vapor data that are calculated based on the coefficients obtained from the North Hemisphere data (western North America) [22].

A. Validation against GPS

The results of validation against GPS PWV observations are shown in Figure 3. It shows that the calibrated MODIS PWV data calculated using the ensemble-based empirical regression method has better accuracy than the MODIS's own operational PWV product. For data obtained from MODIS/Terra shown in the upper panel of Figure 3, the RMSE of PWV data has reduced by 58.53% from 5.712 mm to 2.369 mm using the 2-channel ratio transmittance (Figure 3b) and has reduced by 56.14% to 2.505 mm using the 3-channel ratio transmittance (Figure 3c). For data obtained from MODIS/Aqua shown in the lower panel of Figure 3, the RMSE has reduced by 49.17% from 5.170 mm to 2.628 mm using the 2-channel ratio transmittance (Figure 3e) and has reduced by 46.60% to 2.761 mm using the 3-channel ratio transmittance (Figure 3f). It is worth mentioning that there are several outliers shown in Figure 3. It might be caused by mixed pixels, impact of hazy conditions, or being observed over dark surfaces [15].

The validation results for water vapor estimated from each single absorption band are listed in Table 3. For the MODIS/Terra data, water vapor retrieved from band 18 has the best results, with an RMSE of 2.617 mm and 2.695 mm for 2-channel and 3-channel transmittance, respectively. For the MODIS/Aqua data, water vapor obtained from band 19 has the best agreement with the reference GPS data, with an RMSE of 2.644 mm and 2.703 mm for 2-channel and 3-channel transmittance, respectively. Band 17 has the worst retrieval accuracy among the three absorption channels and its RMSE is larger than the other two absorption channels. The results also confirm that the weighted mean water vapor data of three channels have better accuracy than those retrieved from a single absorption channel.

B. Geographical Dependency

To analyze the geographical dependency of the mean bias, distribution maps of the mean bias between GPS PWV and the calibrated water vapor retrieved using 2-channel ratio and 3-channel ratio transmittance for both Terra and Aqua satellites are shown in Figure 4. The results show that for the PWV retrieved from both satellites have positive biases at most of the GPS stations, indicating that the calibrated MODIS PWV on average are larger than the GPS PWV. This corresponds to the mean biases of all the data points shown in Figure 3. On the other hand, although the GPS stations are not evenly distributed over the research area, generally the mean bias has no visible dependency on the location of GPS stations. It is also worth mentioning that there are a few stations near the equator that have relatively larger negative biases. These errors are likely because of the source error in cloud-mask, reflectance measurement error over water, and the limitation on valid data pairs obtained from those stations [14], [26].

C. Temporal Dependency

To study the temporal dependency of the retrieval accuracy, validation against the reference GPS water vapor data is performed on an annual basis for the period 2017 to 2019. The results are summarized in Table 4. The calibrated PWV shows improvement in all years for both Terra and Aqua platforms. The difference in annual RMSE between GPS PWV and calibrated PWV is in the range of 0.056 mm to 0.277 mm, depending on the use of 2-channel or 3-channel ratio transmittance, and the Terra or Aqua satellite platform. This implies that the water vapor retrieval algorithm basically has no annual dependency. For the MODIS/Terra satellite, the annual RMSE for 2-channel ratio transmittance in 2017, 2018 and 2019 is 2.332 mm, 2.384 mm, and 2.388 mm, respectively; the maximum annual difference is 0.056 mm. The RMSE for 3-channel ratio transmittance in 2017, 2018 and 2019 is 2.441 mm, 2.567 mm, and 2.510 mm, respectively; the maximum annual difference is 0.126 mm. For the MODIS/Aqua satellite, the annual RMSE in 2017, 2018 and 2019 is 2.609 mm, 2.759 mm, and 2.543 mm, respectively, for water vapor estimated using 2-channel ratio transmittance; the maximum annual difference is 0.216 mm. The RMSE is 2.738 mm, 2.927 mm, and 2.650 mm for data calculated using 3-channel ratio transmittance in 2017, 2018 and 2019, respectively; the maximum annual difference is 0.277 mm. The results show that there is no significant annual bias between any two years. The retrieval algorithm is consistent during the whole observation period.

Validation result for each season is also discussed to examine the seasonal variation of the algorithm. The seasons are defined in the following way: September to November are defined as spring months, December to February as summer months, March to May as autumn months, and June to August as winter months. The results in Table 5 show that for the MODIS operational PWV products (MOD05 and MYD05), the retrieval accuracy varies seasonally from 4.208 mm to 6.795 mm for MODIS/Terra product MOD05, while the seasonal RMSE for MYD05 products of MODIS/Aqua varies from 3.777 mm to 6.182 mm. The RMSE values are high in summer and autumn seasons and it is small in winter season. The maximum seasonal difference is 2.587 mm for MODIS/Terra and 2.405 mm for MODIS/Aqua products.

After the calibration, the seasonal RMSE has significantly reduced. For the MODIS/Terra data, the seasonal RMSE varies in the range of 2.087 mm to 2.581 mm using 2-channel ratio transmittance and it is in the range of 2.094 mm to 2.818 mm using 3-channel ratio transmittance. For the MODIS/Aqua data, the seasonal RMSE varies in the range of 1.941 mm to 3.210 mm using 2-channel ratio transmittance and it is in the range of 1.932 mm to 3.445 mm using 3-channel ratio transmittance. The result shows that the maximum seasonal RMSE difference has reduced to 0.724 mm for MODIS/Terra data and 1.513 mm for MODIS/Aqua data. This indicates that the new retrieval algorithm can significantly reduce the seasonal RMSE errors.

The time series of daily mean PWV obtained from GPS, MODIS/Terra, calibrated PWV calculated from Terra using both 2-channel ratio and 3-channel ratio transmittance for 2017-2019 is shown in Figure 5. The GPS PWV data are considered as the baseline for water vapor comparison, as they have proved to have a better accuracy than remote sensing PWV [13], [14].

The results show that the calibrated daily PWV records have a better agreement with GPS PWV compared to the operational MODIS products.

V. Conclusion

Water vapor products from MODIS NIR channel have been widely used in research in many research communities. However, systematic wet biases have been observed. Many different retrieval algorithms have been proposed to improve the MODIS water vapor products. But most of the modifications were at a local scale. We developed a new algorithm to retrieve water vapor using the ensemble-based empirical regression algorithm and it showed to significantly improve MODIS water vapor accuracy in the North Hemisphere [22].

In this study we have systematically evaluate the performance of our algorithm in the South Hemisphere by using water vapor data from 520 ground-based GPS stations in the Australia and its surrounding regions. A comprehensive validation against the GPS reference water vapor dataset reveals that the overall accuracy of the calibrated water vapor records has been greatly improved compared to the operational product. The weighted mean water vapor data obtained from MODIS/Terra has reduced the RMSE by 58.53% from 5.712 mm to 2.369 mm using 2-channel ratio transmittance and has reduced by 56.14% to 2.505 mm using 3-channel ratio transmittance. For the data obtained from MODIS/Aqua, the RMSE has reduced by 49.17% from 5.170 mm to 2.628 mm using 2-channel ratio transmittance and has reduced by 46.60% to 2.761 mm using 3-channel ratio transmittance.

Water vapor data retrieved from the single absorption channels also show improved retrieval accuracy compared to the operational product. For MODIS/Terra, the RMSE for data retrieved from single absorption channel ranges from 2.617 mm to 3.061 mm, while for MODIS/Aqua, the RMSE ranges from 2.644 mm to 3.674 mm. For both platforms, water vapor retrieved from band 17 performs the worst.

The spatial and temporal dependency of retrieval accuracy is also studied in this research. The results show that the mean bias of the water vapor does not show obvious dependence on the stations' location, though the GPS stations are not evenly distributed over the research area. The difference in annual RMSE between GPS PWV and calibrated PWV is in the range of 0.056 mm to 0.277 mm during 2017-2019. This clearly shows the spatial and temporal robustness and homogeneity of the algorithm. In terms of seasonal RMSE error, the maximum seasonal RMSE difference is 2.587 mm for MODIS/Terra MOD05 operational products and 2.405 mm for MODIS/Aqua MYD05 products. After the calibration, the maximum seasonal RMSE difference is reduced to 0.724 mm for MODIS/Terra data and 1.513 mm for MODIS/ Aqua data.

In short, the three-year results clearly show that the new ensemble-based empirical regression model, which was developed based on the North Hemisphere GPS water vapor and MODIS data, is still valid in the South Hemisphere. The model has significantly reduced the error of the MODIS water vapor data collected from 2017 to 2019 for Australia and its surrounding regions. It has a good property of having no temporal or spatial dependency over a large research area. The model is straightforward and the coefficients can be easily applied to areas of interest. It does not require pre-calculated input parameters of atmospheric profiles. Therefore, it is reasonable to conclude that this algorithm provides an effective way to retrieve water vapor globally under cloud-free conditions. It is worth mentioning that although a large number of datasets have been employed in the validation analysis, the number of data points under extremely wet or arid conditions are still limited, which may result in misinterpretation of the performance under these circumstances. Analysis with more datasets observed under these extreme conditions is needed in further studies.

Acknowledgment

The authors would like to thank the research staffs in National Aeronautics and Space Administration (NASA) Goddard Earth Sciences Distributed Active Archive Center (DAAC) (https://search.earthdata.nasa.gov) for providing MODIS/Terra and MODIS/Aqua Level 1 reflectance data and Level 2 water vapor data. They would like to thank the Asia Pacific Regional Geodetic Project (APRGP) GPS Campaign from Geoscience Australia (ftp://ftp.ga.gov.au/geodesy-outgoing/gnss/products/troposphere/rapid/) for the GPS data. The authors would like to thank researchers from European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA-5 reanalysis data (https://cds.climate.copernicus.eu).

References

[1] B. J. Soden, R. T. Wetherald, G. L. Stenchikov, and A. Robock, "Global Cooling After the Eruption of Mount Pinatubo: A Test of Climate Feedback by Water Vapor," *Science*, vol. 296, no. 5568, pp. 727–730, Apr. 2002, doi: 10.1126/science.296.5568.727.

[2] T. R. Karl and K. E. Trenberth, "Modern Global Climate Change," *Science*, vol. 302, no. 5651, pp. 1719–1723, Dec. 2003, doi: 10.1126/science.1090228.

[3] I. M. Held and B. J. Soden, "Water Vapor Feedback and Global Warming," *Annu. Rev. Energy Environ.*, vol. 25, no. 1, pp. 441–475, 2000, doi: 10.1146/annurev.energy.25.1.441.

[4] Z.-L. Li *et al.*, "Satellite-derived land surface temperature: Current status and perspectives," *Remote Sens. Environ.*, vol. 131, pp. 14–37, Apr. 2013, doi: 10.1016/j.rse.2012.12.008.

[5] A. Belward, "The Global Observing System for Climate: Implementation Needs," *Ref. Number GCOS-200*, vol. 315, p. 349, 2016.

[6] B. Tian, B. J. Soden, and X. Wu, "Diurnal cycle of convection, clouds, and water vapor in the tropical upper troposphere: Satellites versus a general circulation model," *J. Geophys. Res. Atmospheres*, vol. 109, no. D10, p. D10101, May 2004, doi: 10.1029/2003JD004117.

[7] J. Wang, A. Dai, and C. Mears, "Global Water Vapor Trend from 1988 to 2011 and Its Diurnal Asymmetry Based on GPS, Radiosonde, and Microwave Satellite Measurements," *J. Clim.*, vol. 29, no. 14, pp. 5205–5222, Apr. 2016, doi: 10.1175/JCLI-D-15-0485.1.

[8] I. Durre, R. S. Vose, and D. B. Wuertz, "Overview of the Integrated Global Radiosonde Archive," *J. Clim.*, vol. 19, no. 1, pp. 53–68, Jan. 2006, doi: 10.1175/JCLI3594.1.

[9] M. Bevis, S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, "GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system," *J. Geophys. Res. Atmospheres*, vol. 97, no. D14, pp. 15787–15801, Oct. 1992, doi: 10.1029/92JD01517.

[10] T. Trent, M. Schröder, and J. Remedios, "GEWEX Water Vapor Assessment: Validation of AIRS Tropospheric Humidity Profiles With

Characterized Radiosonde Soundings," *J. Geophys. Res. Atmospheres*, vol. 124, no. 2, pp. 886–906, Jan. 2019, doi: 10.1029/2018JD028930.

[11] K. L. Holub and S. I. Gutman, "Ground-Based Global Positioning System (GPS) Meteorology Integrated Precipitable Water Vapor (IPW)." National Centers for Environmental Information (NCEI), 2016, Accessed: Dec. 05, 2018. [Online]. Available: https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00962.

[12] K. Gui *et al.*, "Evaluation of radiosonde, MODIS-NIR-Clear, and AERONET precipitable water vapor using IGS ground-based GPS measurements over China," *Atmospheric Res.*, vol. 197, pp. 461–473, Nov. 2017, doi: 10.1016/j.atmosres.2017.07.021.

[13] J. Vaquero-Martínez *et al.*, "Inter-comparison of integrated water vapor from satellite instruments using reference GPS data at the Iberian Peninsula," *Remote Sens. Environ.*, vol. 204, pp. 729–740, Jan. 2018, doi: 10.1016/j.rse.2017.09.028.

[14] J. He and Z. Liu, "Comparison of Satellite-Derived Precipitable Water Vapor Through Near-Infrared Remote Sensing Channels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10252–10262, Dec. 2019, doi: 10.1109/TGRS.2019.2932847.

[15] B.-C. Gao and Y. J. Kaufman, "Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels," *J. Geophys. Res. Atmospheres*, vol. 108, no. D13, p. 4389, 2003, doi: 10.1029/2002JD003023.

[16] Y. J. Kaufman and B. C. Gao, "Remote sensing of water vapor in the near IR from EOS/MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 871–884, Sep. 1992, doi: 10.1109/36.175321.

[17] Z. Li, J.-P. Muller, and P. Cross, "Comparison of precipitable water vapor derived from radiosonde, GPS, and Moderate-Resolution Imaging Spectroradiometer measurements," *J. Geophys. Res. Atmospheres*, vol. 108, no. D20, p. 4651, Oct. 2003, doi: 10.1029/2003JD003372.

[18] Z. Liu, M. S. Wong, J. Nichol, and P. W. Chan, "A multi-sensor study of water vapour from radiosonde, MODIS and AERONET: a case study of Hong Kong," *Int. J. Climatol.*, vol. 33, no. 1, pp. 109–120, Jan. 2013, doi: 10.1002/joc.3412.

[19] A. K. Prasad and R. P. Singh, "Validation of MODIS Terra, AIRS, NCEP/DOE AMIP-II Reanalysis-2, and AERONET Sun photometer derived integrated precipitable water vapor using ground-based GPS receivers over India," *J. Geophys. Res.*, vol. 114, no. D5, p. D05107, Mar. 2009, doi: 10.1029/2008JD011230.

[20] H. Diedrich, R. Preusker, R. Lindstrot, and J. Fischer, "Retrieval of daytime total columnar water vapour from MODIS measurements over land surfaces," *Atmos Meas Tech*, vol. 8, no. 2, pp. 823–836, Feb. 2015, doi: 10.5194/amt-8-823-2015.

[21] M. H. Merrikhpour and M. Rahimzadegan, "Improving the Algorithm of Extracting Regional Total Precipitable Water Vapor Over Land From MODIS Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5889–5898, Oct. 2017, doi: 10.1109/TGRS.2017.2716414.

[22] J. He and Z. Liu, "Water Vapor Retrieval From MODIS NIR Channels Using Ground-Based GPS Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3726–3737, May 2020, doi: 10.1109/TGRS.2019.2962057.

[23] G. Hu, "Report on the analysis of the Asia Pacific Regional Geodetic Project (APRGP) GPS campaign 2016," Geoscience Australia, 2017. doi: 10.11636/Record.2017.017.

[24] L. Ashcroft, J. Gergis, and D. J. Karoly, "Long-term stationarity of El Niño–Southern Oscillation teleconnections in southeastern Australia," *Clim. Dyn.*, vol. 46, no. 9–10, pp. 2991–3006, May 2016, doi: 10.1007/s00382-015-2746-3.

[25] X. Wang *et al.*, "The correlation between GNSS-derived precipitable water vapor and sea surface temperature and its responses to El Niño–Southern Oscillation," *Remote Sens. Environ.*, vol. 216, pp. 1–12, Oct. 2018, doi: 10.1016/j.rse.2018.06.029.

[26] G. Wang and H. H. Hendon, "Sensitivity of Australian Rainfall to Inter–El Niño Variations," *J. Clim.*, vol. 20, no. 16, pp. 4211–4226, Aug. 2007, doi: 10.1175/JCLI4228.1.

[27] P. Sillitoe, "Whether Rain or Shine: Weather Regimes from a New Guinea Perspective," *Oceania*, vol. 64, no. 3, pp. 246–270, 1994, doi: 10.1002/j.1834-4461.1994.tb02467.x.

[28] A. P. Sturman and N. J. Tapper, *The weather and climate of Australia and New Zealand*. Oxford University Press, USA, 1996.

[29] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020, doi: 10.1002/qj.3803.

[30] J. Saastamoinen, "Contributions to the theory of atmospheric refraction," *Bull. Géod. 1946-1975*, vol. 105, no. 1, pp. 279–298, Sep. 1972, doi: 10.1007/BF02521844.

[31] J. L. Davis, T. A. Herring, I. I. Shapiro, A. E. E. Rogers, and G. Elgered, "Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length," *Radio Sci.*, vol. 20, no. 6, pp. 1593–1607, 1985, doi: 10.1029/RS020i006p01593.

[32] A. L. Buck, "New Equations for Computing Vapor Pressure and Enhancement Factor," *J. Appl. Meteorol. Climatol.*, vol. 20, no. 12, pp. 1527–1532, Dec. 1981, doi: 10.1175/1520-0450(1981)020<1527:NEFCVP>2.0.CO;2.

[33] R. J. Ross and W. P. Elliott, "Tropospheric Water Vapor Climatology and Trends over North America: 1973–93," *J. Clim.*, vol. 9, no. 12, pp. 3561–3574, Dec. 1996, doi: 10.1175/1520-0442(1996)009<3561:TWVCAT>2.0.CO;2.

[34] F. N. Fritsch and LLNL, "Piecewise Cubic Interpolation Package," Lawrence Livermore National Lab. (LLNL), Livermore, CA (United States), PCHIP; 002103MNFRM00, Apr. 1982. Accessed: Jan. 15, 2021. [Online]. Available: https://www.osti.gov/biblio/1231005.

[35] S. W. Seemann, E. E. Borbas, J. Li, W. P. Menzel, and L. E. Gumley, "MODIS atmospheric profile retrieval algorithm theoretical basis document," Univ. of Wis.–Madison, Madison, 2006. Accessed: Nov. 24, 2016. [Online]. Available: http://eospso.nasa.gov/sites/default/files/atbd/atbd_mod07.pdf.

[36] B. Stefania, M. Vinia, B. Patrizia, C. Piero, and P. Nazzareno, "Satellite-Based Retrieval of Precipitable Water Vapor Over Land by Using a Neural Network Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3236–3248, Sep. 2011, doi: 10.1109/TGRS.2011.2160184.

[37] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," *SIGKDD Explor Newsl*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.