Columnar Water Vapor Retrieval by Using Data from the Polarized Scanning Atmospheric Corrector (PSAC) Onboard HJ-2 A/B Satellites

Yanqing Xie, Weizhen Hou, Zhengqiang Li, Sifeng Zhu, Zhenhai Liu, Jin Hong, Yan Ma, Cheng Fan, Jie Guang, Benyong Yang, Xuefeng Lei, Honglian Huang, Xiaobing Sun, Xiao Liu, Ying Zhang, Maoxin Song, Peng Zou and Yanli Qiao

Abstract: As the latest members of Chinese Environmental Protection and Disaster Monitoring Satellite Constellation, the first two of HuanjingJianzai-2 (HJ-2) series satellites were launched on 27 September 2020 by China and are usually abbreviated as HJ-2 A/B satellites. The polarized scanning atmospheric corrector (PSAC) is one of main sensors onboard HJ-2 A/B satellites, which is mainly used to monitor atmospheric components such as water vapor and aerosols. In this study, a columnar water vapor (CWV) retrieval algorithm using two bands (865 and 910 nm) is developed for PSAC. The validation results of PSAC CWV data based on ground-based CWV data derived from AERONET show that PSAC CWV data has a high accuracy, and all statistical parameters of PSAC CWV data are better than those of Moderate-resolution Imaging Spectroradiometer (MODIS) CWV data released by NASA. Overall, there is no obvious overestimation or underestimation in PSAC CWV data. The root mean square error (RMSE), mean absolute error (MAE), relative error (RE), and percentage of CWV data with error within ±0.10% of PSAC CWV data are 0.17 cm, 0.13 cm, 0.08, and 78.19%, respectively. Compared with MODIS CWV data, PSAC CWV data shows a 71% decrease in RMSE, a 73% decrease in MAE, a 71% decrease in RE, and a 372% increase in PER10. In addition, the results of day-to-day comparisons between PSAC CWV data and AERONET data show that PSAC CWV data can effectively characterize the change trend of CWV.

Keywords: water vapor retrieval; columnar water vapor (CWV); polarized scanning atmospheric corrector (PSAC); HJ-2 A/B satellites

1. Introduction

Water vapor is one of the components of the atmosphere that is mainly distributed in the troposphere, and its proportion in the atmosphere usually varies in the range of 0–4%. Its main sources are evaporation of water and transpiration of plants. The water vapor content can be characterized by the columnar water vapor (CWV) with the unit of cm, which is the total amount of water vapor contained in a vertical column per unit cross-sectional area from the ground to the top of the atmosphere (TOA). Although water...
vapor accounts for only a small fraction of the atmosphere, it plays an important role in climate change and weather systems [1–4]. In addition, water vapor also has an important impact on the accuracy of quantitative remote sensing such as remote sensing inversion of atmospheric parameters and atmospheric correction of remote sensing data [5–7]. The spatial distribution of water vapor has significant spatial and temporal differences due to its long-range transport, differences in emission sources, and short lifetime. Therefore, it is of great importance to monitor water vapor on a large scale. Because of the very large spatial coverage of remote sensing data obtained by Earth observation satellites, large-scale monitoring of water vapor using remote sensing data is now the dominant approach.

The Huanjingjianzai-2 (HJ-2) series of satellites are the latest members of the Chinese Environmental Protection and Disaster Monitoring Satellite Constellation developed by China, and the first two (HJ-2A and HJ-2B) HJ-2 series satellites were launched on 27 September 2020. Each satellite is equipped with four imaging sensors, including the wide view charge-coupled device (CCD) camera, the hyper-spectral imager (HSI), the infrared multispectral scanner (IRMSS), and the polarized scanning atmospheric corrector (PSAC) [8]. The first three sensors are mainly used for surface observations, and PSAC is mainly used to monitor atmospheric components such as water vapor and aerosols. In order to remove the atmospheric interference to the data of the first three main sensors to extend their application, it is necessary to perform atmospheric correction on these data according to the atmospheric parameters including CWV [9,10]. Because of the large temporal differences in distribution of water vapor, it is essential to use CWV data synchronous with HJ-2 satellites rather than CWV data from other satellites for high-precision atmospheric correction. In addition, since the observation time of different satellites is different, the CWV data derived from HJ-2 satellites is useful for the studies on weather forecasting and climate change. Therefore, it is of great significance to develop CWV data using PSAC data.

The current remote sensing retrieval algorithms for water vapor can be classified into four categories based on the wavelength of the data used, namely, visible algorithm, near-infrared (NIR) algorithm, thermal infrared (TIR) algorithm, and microwave (MW) algorithm [11–14]. Each of these algorithms has its own scope of application. Among them, the visible algorithm, TIR algorithm, and MW algorithm can be used not only for CWV retrieval over the land, but also for CWV retrieval over the ocean, while the NIR algorithm is only applicable to areas with large surface albedo, such as land and the sun-glint region of the ocean. Due to the use of data in the visible-NIR spectral range, the visible algorithm and NIR algorithm are only suitable for water vapor retrieval in the daytime, while the TIR algorithm and MW algorithm can be used for water vapor retrieval at any time. In addition, the MW algorithm can be used for water vapor detection both over cloud-free areas and over areas covered by clouds because of the strong penetration ability of microwaves, while the other three algorithms are only applicable to areas not affected by clouds. In addition to the above methods based on remote sensing image data, there are still many other methods for monitoring water vapor content. For example, radio occultation methods [15,16] and methods for retrieving CWV by using data observed by ground stations including global positioning system (GPS) receiver, sun photometer, and radiosonde [17,18]. In general, the accuracy of CWV data derived from ground-based observations is higher than that of CWV data derived from satellite data, but ground-based observations have difficulty monitoring water vapor on a large scale due to the limitation of the number of ground-based stations.

All the above-mentioned remote sensing retrieval algorithms for water vapor have been used in the development of water vapor products for multiple sensors. For example, the visible algorithm has been used for water vapor retrieval for sensors such as the Ozone Monitoring Instrument (OMI) and the Global Ozone Monitoring Experience 2 (GOME-2) [14,19], the NIR algorithm has been applied to the water vapor retrieval for sensors such as the Moderate-resolution Imaging Spectroradiometer (MODIS) and the Medium Resolution Spectral Imager (MERSI)-2 [20,21], the TIR algorithm has been used for water vapor retrieval for sensors such as the Spinning Enhanced Visible and Infrared Imager (SEVIRI) and the Advanced Very High Resolution Radiometer (AVHRR) [12,22], and the
MW algorithm has been applied to water retrieval for sensors such as the Microwave Radiation Imager (MWRI) and the calibration microwave radiometer (CMR) [13,23]. On the whole, among the four algorithms, the NIR algorithm has the highest accuracy and the MW algorithm has the strongest applicability. In this work, the NIR algorithm is selected for water vapor retrieval of PSAC according to the band setting of PSAC.

In this study, we use the NIR algorithm to retrieve CWV from PSAC onboard HJ-2 satellites and evaluate the accuracy of retrieval algorithm based on ground-based data. The retrieval results can be used for atmospheric correction of data from other sensors onboard HJ-2 satellites and provide data support for studies related to weather forecasting and climate change. The paper is structured as follows. The data used are introduced in Section 2 and the CWV retrieval algorithm is presented in Section 3. Then, the CWV retrieval results and analysis are shown in Section 4. After that, the limitations of this study are discussed in Section 5. Finally, the paper is summarized in Section 6.

2. Data
2.1. PSAC Data

The main purpose of HJ-2 series satellites is for monitoring and early warning of environmental pollution and disasters, which is operated by China Centre for Resources Satellite Data and Application (CCRSDA. Available online: http://www.cresda.com (accessed on 18 January 2022)). HJ-2A and HJ-2B are the first two of the HJ-2 series, and both satellites are polar orbiting satellites with an orbital altitude of 645 km. PSAC onboard HJ-2 satellite is mainly used for atmospheric monitoring. It has a spatial resolution of about 6 km and an imaging width of more than 800 km. PSAC has nine spectral channels with polarization detection capability, including two NIR channels (865 and 910 nm) dedicated to CWV retrieval. The spectral response functions (SRF) of two bands of PSAC used for CWV retrieval are shown in Figure 1, and more characteristics of them can be found in Table 1. Since CCRSDA only provides us with PSAC data from March 2021 to August 2021 during the in-orbit testing period, the time period chosen for this study was these 6 months.

![Figure 1. The SRFs of two bands of PSAC used for CWV retrieval. The first line in the legend (blue line) represents the water vapor absorption transmittance (WVAT) of the atmosphere when the view zenith angle (VZA) and solar zenith angle (SZA) are equal to 0 and CWV is equal to 1 cm. The second line (yellow line) represents the WVAT when VZA and SZA are equal to 0 and CWV is equal to 3 cm. The next two are SRFs of the two bands of PSAC used for CWV retrieval.](http://www.cresda.com)
Table 1. The specific band settings of the polarized scanning atmospheric corrector (PSAC).

<table>
<thead>
<tr>
<th>No.</th>
<th>Center Wavelength</th>
<th>Spectral Range</th>
<th>Spectral Width</th>
<th>Spatial Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 1</td>
<td>0.410 μm</td>
<td>0.400–0.420 μm</td>
<td>20 nm</td>
<td>6 km</td>
</tr>
<tr>
<td>Band 2</td>
<td>0.443 μm</td>
<td>0.433–0.453 μm</td>
<td>20 nm</td>
<td>6 km</td>
</tr>
<tr>
<td>Band 3</td>
<td>0.555 μm</td>
<td>0.545–0.565 μm</td>
<td>20 nm</td>
<td>6 km</td>
</tr>
<tr>
<td>Band 4</td>
<td>0.670 μm</td>
<td>0.660–0.680 μm</td>
<td>20 nm</td>
<td>6 km</td>
</tr>
<tr>
<td>Band 5</td>
<td>0.865 μm</td>
<td>0.845–0.885 μm</td>
<td>40 nm</td>
<td>6 km</td>
</tr>
<tr>
<td>Band 6</td>
<td>0.910 μm</td>
<td>0.900–0.920 μm</td>
<td>20 nm</td>
<td>6 km</td>
</tr>
<tr>
<td>Band 7</td>
<td>1.380 μm</td>
<td>1.360–1.400 μm</td>
<td>40 nm</td>
<td>6 km</td>
</tr>
<tr>
<td>Band 8</td>
<td>1.610 μm</td>
<td>1.580–1.640 μm</td>
<td>60 nm</td>
<td>6 km</td>
</tr>
</tbody>
</table>

2.2. Ground-Based CWV Data

The aerosol robotic network (AERONET) is a global ground-based aerosol observation network, which has been in operation for nearly 25 years [24]. Although it is specially established for observing aerosols rather than water vapor, it can also provide high-quality CWV data. The uncertainty of AERONET CWV data is about 10% [25,26]. Due to the continuity and high accuracy of AERONET CWV data, it is widely used for quality assessment of remote sensing CWV data [27]. In this paper, the last version (Version 3.0) of Level 1.5 AERONET data after cloud mask and quality control were used for the validation of remote sensing CWV data [28]. A total of 350 sites were used in this work, and their spatial distribution is shown in Figure 2. The AERONET CWV data mentioned above are all available from the Internet (https://aeronet.gsfc.nasa.gov/, last accessed on 23 January 2022).

Figure 2. Spatial distribution of AERONET sites used in this study. Each point represents a site.

3. Method

3.1. Water Vapor Retrieval

As mentioned before, the NIR water vapor retrieval algorithm is only applicable to sensors with both the water vapor absorption band and band not affected by water vapor absorption in the spectral range of 850–1250 nm [11]. Therefore, bands 5 and 6 (865 and 910 nm) of PSAC are bands selected for water vapor retrieval. Among them, band 5 is located in the spectral range of the atmospheric window and is basically not affected by gas absorption, while band 6 is strongly absorbed by the atmosphere, almost all of which is caused by water vapor.

For NIR bands, the apparent reflectance observed by the satellite (\(R_{TOA}\)) at the top of the atmosphere can be expressed as [29,30]:

\[
R_{TOA} = T_S \left( \rho_0 + T_I T_f R_{surf} \right),
\]

(1)
where \( T_g \) is the total gaseous absorption transmittance, \( \rho_0 \) denotes the atmospheric path reflectance caused by the atmosphere, \( R_{surf} \) is surface reflectance, \( T_{↑} \) and \( T_{↓} \) represent the total transmittance (direct + diffuse) upward and downward, respectively.

Due to their relatively long and close wavelengths, the difference between path radiation and atmospheric transmittance of bands 5 and 6 can be ignored. In addition, the difference in surface reflectance between the two bands is very small. For the above reasons, it can be considered that the apparent reflectance of band 5 is equal to that of band 6 without considering water vapor absorption [21]. Since the effect of gases other than water vapor on the absorption of radiation observed by PSAC in bands 5 and 6 is negligible, the water vapor absorption transmittance of band 6 can be calculated by dividing the apparent reflectance of band 6 by that of band 5.

\[
T_{g}^{910 \text{ nm}} = \frac{R_{TOA}^{910 \text{ nm}}}{R_{TOA}^{865 \text{ nm}}},
\]

where \( T_{g}^{910 \text{ nm}} \) is the water vapor absorption transmittance of band 6, \( R_{TOA}^{865 \text{ nm}} \) and \( R_{TOA}^{910 \text{ nm}} \) are the TOA reflectance of bands 5 and 6, respectively.

Previous studies have shown that the water vapor absorption transmittance of the NIR channel is strongly correlated with CWV. This relationship can be written as [31,32]:

\[
V_{CWV} = A \cdot (\ln(T_g))^2 + B \cdot \ln(T_g) + C,
\]

\[
L = \frac{1}{\cos(\theta_s)} + \frac{1}{\cos(\theta_v)},
\]

where \( V_{CWV} \) denotes CWV, \( T_g \) represents the water vapor absorption transmittance of NIR band, \( \theta_s \) is the solar zenith angle, \( \theta_v \) is the satellite zenith angle, and \( A, B, \) and \( C \) are three constants related to the characteristics of the NIR band, respectively.

In this paper, the selected PSAC data was divided into two parts. Among them, PSAC data in East Asia (18–53°N, 73–130°E) was used to determine above three parameters (\( A, B, \) and \( C \)), and the remaining PSAC data was used to test the effectiveness of the water vapor retrieval algorithm. For the convenience of description, the region used to determine parameters of algorithm model is referred to as region 1, and the region used to test the algorithm is referred to as region 2. There are 38 AERONET sites in region 1 and 312 AERONET sites in region 2. In order to construct water vapor retrieval model for PSAC, we first matched PSAC L1 data containing TOA reflectance data, solar zenith angle data, and satellite zenith angle data with AERONET CWV data within region 1. In this paper, we use the temporal mean of AERONET PWV data within half an hour of observation time of PSAC to match the cloudless pixels where the AERONET site is located. Note that the TOA reflectance of band 2 and the TOA reflectance difference between adjacent pixels of bands 2 and 7 are used to judge whether a pixel is a cloud pixel. The specific judgment criteria are described in Section 3.2. There are a total of 839 valid matches.

After that, we can calculate the natural logarithm of water vapor absorption transmittance of band 6 of PSAC (i.e., \( \ln(T_{g}^{910 \text{ nm}}) \)) and the slanted water vapor content in the path of the sun-ground-satellite (i.e., \( V_{CWV} \cdot L \)) of each valid match based on the corresponding AERONET CWV data, solar zenith angle data, satellite zenith angle data, and TOA reflectance of bands 5 and 6. Finally, we can substitute the above matching results into Equation (3) to establish 839 equations containing only three unknowns (\( A, B, \) and \( C \)), and determine the values of \( A, B, \) and \( C \) by using the least squares method. The scatter plot of the slanted water vapor content in the path of the sun-ground-satellite and natural logarithm of water vapor absorption transmittance of band 6 of PSAC of the above matching results is shown in Figure 3.


\[ y = 13.944x^2 - 4.867x - 0.049 \]
\[ R^2 = 0.9836 \]

Figure 3. Scatter plot of matching results between the slanted water vapor content in the path of the sun-ground-satellite (i.e., \( V_{\text{CWV}} * L \)) and the natural logarithm of the water vapor absorption transmittance of band 6 of PSAC (i.e., \( \ln T_g^{910 \text{ nm}} \)) in East Asia (18°–53°N, 73°–130°E). Each blue solid circle represents a matching result of AERONET data and PSAC data. The red solid line represents the fitting equation of the natural logarithm of the water vapor absorption transmittance of band 6 and the slanted water vapor content in the path of the sun-ground-satellite.

As shown in Figure 3, there is a significant correlation \( (R^2 = 0.98) \) between the slanted water vapor content in the path of the sun-ground-satellite \( (V_{\text{CWV}} * L) \) and natural logarithm of water vapor absorption transmittance of band 6 of PSAC \( (\ln T_g^{910 \text{ nm}}) \). This means that it is feasible to retrieve water vapor based on the data from bands 5 and 6 of PSAC. For band 6 of PSAC, A, B, and C in Equation (3) are equal to 13.944, −4.867, and −0.049, respectively. Therefore, the semi-empirical equation for retrieving water vapor content can be expressed as:

\[ V_{\text{CWV}} = \frac{13.944 \left( \ln T_g^{910 \text{ nm}} \right)^2 - 4.867 \ln T_g^{910 \text{ nm}} - 0.049}{L}. \]  

(5)

3.2. Cloud Mask

The NIR CWV retrieval algorithm is only applicable to the pixels not affected by clouds, so it is necessary to remove the cloud-influenced part of the PSAC CWV retrieval results to ensure the quality of PSAC CWV data. Two methods are used to identify cloud pixels. If a pixel is identified as a cloud pixel by any of the methods, the pixel is removed as a cloud pixel.

1. The TOA reflectance of band 2 (443 nm) and band 7 (1380 nm) is used for the cloud identification [33,34]. For a pixel that needs to be detected, if the TOA reflectance of band 2 of the pixel is greater than 0.4, the pixel would be regarded as a cloud pixel. Next, we identify cloud pixels based on the TOA reflectance of band 7. For a pixel that needs to be detected, it is first assumed that the pixel is not affected by clouds. Then we can calculate the CWV of the pixel based on Equation (5). Next, the water vapor absorption transmittance of band 7 can be calculated using the Second Simulation of the Satellite Signal in the Solar Spectrum (6SV) model based on the calculated CWV [35]. Since the TOA reflectance of band 7 is less than 1 for cloud-free pixels when water vapor content is equal to 0, so the actual observed TOA reflectance of band 7 should be less than the value of 1 multiplied by
the water vapor absorption transmittance calculated above. In other words, if a pixel is not affected by clouds, the actual observed TOA reflectance of band 7 of the pixel is less than the water vapor absorption transmittance of band 7 of the pixel calculated based on the above method. Therefore, if the actual observed TOA reflectance of band 7 of a pixel is not smaller than the water vapor absorption transmittance of band 7 of the pixel calculated based on the above method, the pixel would be regarded as a cloud pixel.

(2) The TOA reflectance difference between adjacent pixels of band 2 and band 7 is used for the cloud identification [5,36]. For a pixel that needs to be detected, the standard deviation of apparent reflectance of it and its nearest 8 pixels is used to characterize the difference. The specific calculation method is given in Equation (6). If the calculated standard deviation of band 2 is greater than 0.01, or the standard deviation of band 7 is greater than 0.005, the pixel would be considered as a cloud pixel.

\[
\text{STD} = \sqrt{\frac{\sum_{i=1}^{9} (R_{\text{TOA}}^i - \overline{R_{\text{TOA}}})^2}{9}},
\]

where \(R_{\text{TOA}}^i\) is the TOA reflectance with a different serial number, \(\overline{R_{\text{TOA}}}\) is the average value of all 9 TOA reflectance participating in the calculation, and STD represents the calculated standard deviation.

Note that the above thresholds (i.e., 0.4, 0.01, and 0.005) were set with reference to the threshold used by the cloud mask algorithms of the Advanced Himawari Imager (AHI) and MODIS that use similar bands to PSAC. Since the cloud mask is not the focus of this paper, the requirement for the cloud mask algorithm in this paper is that most cloud pixels can be detected. Examples of cloud mask results based on the above methods are shown in Figure 4. By analyzing Figure 4, it can be seen that the cloud mask methods used in this work are basically effective in identifying cloudy pixels.

Figure 4. Examples of PSAC false color composite images and cloud mask results. (a) PSAC false color composite image of example 1. (b) PSAC cloud mask result of example 1. (c) PSAC false color composite image of example 2. (d) PSAC cloud mask result of example 2.

4. Results

In this section, the quality of PSAC CWV data is evaluated in two ways, including overall validation results and analysis of day-to-day variation in CWV.
4.1. Overall Validation Results

To clearly understand the advantages and disadvantages of PSAC CWV data, we not only used AERONET data to assess the quality of PSAC CWV data but compared its accuracy with that of widely used and fully validated MODIS CWV data derived from MODIS onboard Terra (MOD05) [11,37,38]. Existing validation results show that MOD05 has a high correlation with ground-based CWV data, but MOD05 has a significant overestimation [39,40]. Since the accuracy of MOD05 derived from Terra/MODIS and MYD05 derived from MODIS/Aqua is close [40], MOYD05 is not used together with MOD05. The above MODIS CWV data can be downloaded free of charge from the Internet (https://ladsweb.modaps.eosdis.nasa.gov/, last accessed on 26 January 2022). A total of five parameters were used for quality assessment of remote sensing CWV data, including root mean square error (RMSE), mean absolute error (MAE), mean bias (MB), relative error (RE), and percentage of CWV data with error within $\pm (0.05 + 0.1 \times CWV_{AERONET})$ (PER10). The calculation equations of RMSE, MAE, MB, and RE are as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (V_{\text{cwv}}^i - V_{\text{cwv}}^{'i})^2}{N}},$$  \hfill (7)

$$\text{MAE} = \frac{\sum_{i=1}^{N} |V_{\text{cwv}}^i - V_{\text{cwv}}^{'i}|}{N},$$  \hfill (8)

$$\text{MB} = \frac{\sum_{i=1}^{N} (V_{\text{cwv}}^i - V_{\text{cwv}}^{'i})}{N},$$  \hfill (9)

$$\text{RE} = \frac{\sum_{i=1}^{N} |V_{\text{cwv}}^i - V_{\text{cwv}}^{'i}|}{\sum_{i=1}^{N} V_{\text{cwv}}^{'i}}$$  \hfill (10)

where $N$ is the number of effective matching results between AERONET CWV data and remote sensing CWV data, $V_{\text{cwv}}^i$ is the value of remote sensing CWV data of the $i$th matching result, and $V_{\text{cwv}}^{'i}$ is the value of ground-based CWV data corresponding to $V_{\text{cwv}}^i$.

In this work, we used the temporal mean of ground-based observations within half an hour of the satellite observation time to match the spatial average of remote sensing CWV data within the range of $9 \times 9$ km centered on the ground-based site [27,39]. The scatter plots of the matching results between PSAC CWV data, MODIS CWV data, and AERONET data are shown in Figure 5. From Figure 5, it can be found that there is no obvious overestimation or underestimation in PSAC CWV data (MB = $-0.03$ cm). However, there is a large overestimation in the MODIS CWV dataset (MB = $0.47$ cm), which is consistent with the existing quality assessment results of MODIS CWV data [39,40]. The RMSE, MAE, RE, and PER10 of PSAC CWV data are $0.17$ cm, $0.13$ cm, $0.08$, and $78.19\%$, respectively. The RMSE, MAE, RE, and PER10 of MODIS CWV data are $0.59$ cm, $0.48$ cm, $0.28$, and $16.55\%$, respectively. Compared with MODIS CWV data, PSAC CWV data shows a $71\%$ decrease in RMSE, a $73\%$ decrease in MAE, a $71\%$ decrease in RE. By analyzing the values of above statistical parameters, it can be found that, whether the MAE and RMSE used to characterize absolute error or the RE used to characterize relative error is selected for quality assessment of remote sensing CWV data, the accuracy of PSAC CWV data is significantly better than that of MODIS CWV data.

The obvious overestimation of MODIS CWV data is the main reason why the absolute error and relative error of MODIS CWV data are obviously worse than those of PSAC CWV data. In addition, due to the big systematic error of MODIS CWV data, the PER10 of PSAC CWV data (78.19%) is significantly larger than that of MODIS CWV data (16.55%). Since the PER10 of PSAC CWV data is greater than $68.27\%$, it can be considered that the expected error of PSAC CWV data is less than $\pm (0.05 + 0.1 \times CWV_{AERONET})$. In summary, it can be concluded that PSAC CWV data has no large systematic deviation and has a high accuracy, which is higher than that of MODIS CWV data.
4.2. Analysis of Day-to-Day Variation in CWV

As shown in Figure 5, the day-to-day comparisons between PSAC CWV data and AERONET CWV data were performed at four AERONET sites in different regions to evaluate the time consistency between them. The selected sites are NEON_LENO (88.16°W, 31.85°N), SEDE_BOKER (34.78°E, 30.86°N), Tomsk (85.05°E, 56.48°N), and Ussuriysk (132.16°E, 43.70°N), respectively. In general, AERONET CWV data and PSAC CWV data have very high consistency, and more than 90% of them having a difference of less than 0.2 cm. This indicates that PSAC CWV data can effectively characterize the daily variation in CWV.

Figure 5. Scatter plots of the matching results between PSAC CWV data, MODIS CWV data, and AERONET data. (a) PSAC CWV data. (b) MODIS CWV data.

Figure 6. Day-to-day comparisons between PSAC CWV data and AERONET CWV data. (a) NEON_LENO. (b) SEDE_BOKER. (c) Tomsk. (d) Ussuriysk.

5. Discussion

Compared with the widely used water vapor retrieval algorithm based on look-up table, the semi-empirical algorithm developed in this study has an extremely high retrieval efficiency because it does not need to use the radiative transfer model such as the MODerate resolution atmospheri TRANsmission (MODTRAN) computer code [41] to build the look-up table before retrieval and retrieve CWV by querying the look-up table.

Figure 6. Cont.
5. Discussion

Compared with the widely used water vapor retrieval algorithm based on look-up table, the semi-empirical algorithm developed in this study has an extremely high retrieval efficiency because it does not need to use the radiative transfer model such as the MODerate resolution atmospheric TRANsmission (MODTRAN) computer code [41] to build the look-up table before retrieval and retrieve CWV by querying the look-up table [20]. Additionally, our algorithm has very strong portability. Although it is only used for PSAC at present, it can be easily used for CWV retrieval of sensors such as MODIS and MERSI-2 that have both a water vapor absorption band and a band unaffected by water vapor absorption in the spectral range of 850–1250 nm [11].

Although PSAC CWV data developed in this work has a high accuracy and can effectively characterize the change trend of CWV, it still has some limitations. First, because the water vapor retrieval algorithm developed in this study needs to use historical PSAC data and AERONET data to determine the parameters of the algorithm model (A, B, and C in Equation (3)), it cannot be used for CWV retrieval immediately after the launch of HJ-2 satellites. However, HJ-2 satellites will undergo nearly one year of in-orbit testing before operational operation, a large amount of data will be accumulated during this period, which is sufficient for building the algorithm model. Therefore, it is feasible to use it as an operational algorithm.

In addition, as shown in Figure 6, the retrieval error of PSAC CWV data generally increases with the increase in CWV. This is unavoidable mainly because the sensitivity of band 6 to water vapor gradually decreases with the increasing water vapor content [42]. Although the sensitivity of all NIR bands to water vapor decreases with increasing water vapor content, the sensitivity of different water vapor absorption channels to water vapor is different. This means that the retrieval error of remote sensing CWV data can be reduced by using a band that is more sensitive to water vapor at high water vapor content compared to band 6. For example, MODIS and MERSI-2 are equipped with three NIR water vapor absorption bands to improve the accuracy of CWV retrieval results. Therefore, to improve
the accuracy of CWV retrieval results at high water vapor content, two or more water vapor absorption bands are necessary for the subsequent PSACs.

![Image of the distribution of the absolute error of PSAC CWV retrieval results from March 2021 to August 2021. The sloping red line is the linear regression line. The red points are the means for specific ranges of absolute error, and the vertical red lines are the mean ± 2σ (standard deviation) of absolute error in a certain range. The color of a point is determined by the density of the point where it is located.](image_url)

**Figure 7.** The distribution of the absolute error of PSAC CWV retrieval results from March 2021 to August 2021. The sloping red line is the linear regression line. The red points are the means for specific ranges of absolute error, and the vertical red lines are the mean ± 2σ (standard deviation) of absolute error in a certain range. The color of a point is determined by the density of the point where it is located.

6. Conclusions

In this study, we developed a water vapor retrieval algorithm using two bands (865 and 910 nm) for PSAC onboard HJ-2 satellites, which is constructed based on the high correlation between the water vapor absorption transmittance of band 6 at 910 nm and water vapor content. The semi-empirical algorithm consists of three parts: (1) calculating the water vapor absorption transmittance of band 6 of PSAC; (2) constructing the relationship between the water vapor absorption transmittance of band 6 and CWV; (3) inverting CWV based on the constructed relationship. Since it can directly calculate CWV based on PSAC data and does not need to build the look-up table before retrieval and retrieve CWV by querying the look-up table, it has a very high retrieval efficiency.

To clearly assess the quality of PSAC CWV data, we not only use AERONET data to validate PSAC CWV data but compare its accuracy with that of widely used MODIS CWV data. The validation results based on ground-based CWV data from 312 AERONET sites show that PSAC CWV has a very high accuracy, and its accuracy is significantly better than that of MODIS CWV data released by NASA. The RMSE, MAE, RE, and PER10 of PSAC CWV data are 0.17 cm, 0.13 cm, 0.08, and 78.19%, respectively. Compared with MODIS CWV data, PSAC CWV data shows a 71% decrease in RMSE, a 73% decrease in MAE, a 71% decrease in RE, and a 372% increase in PER10. In addition, time-series comparisons between PSAC CWV data and AERONET data were performed to evaluate the time consistency between them. The results show that PSAC CWV data can effectively characterize the change trend of CWV.

**Author Contributions:** This work was carried out in collaboration with all the authors. Conceptualization, Y.X., W.H., Z.L. (Zhengqiang Li), Z.L. (Zhenhai Liu), J.H., B.Y., X.S. and Y.Q.; methodology, Y.X., W.H., Z.L. (Zhengqiang Li), Y.M., C.F. and J.G.; data acquisition, Y.X., W.H., S.Z., X.L. (Xuefeng Lei), H.H., X.L. (Xiao Liu), Y.Z., M.S. and P.Z.; writing and revision, Y.X., W.H., Z.L. (Zhengqiang Li), S.Z., Y.Z. and Z.L. (Zhenhai Liu). All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No. 41871269), the National Outstanding Youth Foundation of China (Grant No. 41925019), the National Natural Science Foundation of China (Grant No. 41701407), and the K. C. Wong Education Foundation (Grant No. GJTD-2018-15).

**Institutional Review Board Statement:** Not applicable.
Informed Consent Statement: Not applicable.


Acknowledgments: The authors thank Jun Lin, Zhongzheng Hu and Lanlan Fan of China Centre for Resources Satellite Data and Application for providing PSAC L1 data during the in-orbit test of the HJ-2 A/B satellites, and Zhaoguang Bai, Jun Zhu and Qiang Cong of DFH Satellite Co. Ltd. for helpful discussions on the PSAC instrument. The ground-based CWV data used to validate the remote sensing CWV data were obtained from 350 AERONET sites. We would also like to thank the PI(s), Co-I(s) and their staff for their efforts in establishing and maintaining the AERONET sites. In addition, we thank NASA for providing us with MODIS water vapor data.

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

References
1. Lyngwa, R.V.; Nayak, M.A. Atmospheric river linked to extreme rainfall events over Kerala in August 2018. Atmos. Res. 2021, 253, 13. [CrossRef]


