Comparisons of Performance Using Data Assimilation and Data Fusion Approaches in Acquiring Precipitable Water Vapor: A Case Study of a Western United States of America Area

Zhaohui Xiong 1,2, Jizhang Sang 1*, Xiaogong Sun 2, Bao Zhang 1 and Junyu Li 1

1 School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; cehui_xiong@whu.edu.cn (Z.X.); sggzb@whu.edu.cn (B.Z.); junyu_li@whu.edu.cn (J.L.)
2 Chinese Academy of Meteorological Sciences, Beijing 100081, China; xgsun@cma.gov.cn

* Correspondence: jzhsang@sgg.whu.edu.cn

Received: 4 September 2020; Accepted: 18 October 2020; Published: 21 October 2020

Abstract: There are two main types of methods available to obtain precipitable water vapor (PWV) with high accuracy. One is to assimilate observations into a numerical weather prediction (NWP) model, for example, the Weather Research and Forecasting (WRF) model, to improve the accuracy of meteorological parameters, and then obtain the PWV with improved accuracy. The other is the direct fusion of multi-source PWV products. Regarding the two approaches, we conduct a comparison experiment on the West Coast of the United States of America with the data from May 2018, in which the WRF data assimilation (DA) system is used to assimilate the Global Navigation Satellite System (GNSS) PWV, while the method by Zhang et al. to fuse the GNSS PWV, ERA5 PWV and MODIS (moderate-resolution imaging spectroradiometer) PWV. As a result, four groups of PWV products are generated: the assimilated GNSS PWV, the unassimilated GNSS PWV, PWV from the fusion of the GNSS PWV and ECWMF (European Centre for Medium-Range Weather Forecasts) ERA5 (ECWMF Reanalysis 5) PWV, and PWV from the fusion of the GNSS PWV, ERA5 PWV and MODIS PWV. Experiments show that the data assimilation based on the WRF model (WRFDA) and adopted fusion method can generate PWV products with similar accuracy (1.47 mm vs. 1.52 mm). Assimilating the GNSS PWV into the WRF model slightly improves the accuracy of the inverted PWV by 0.18 mm. The fusion of the MODIS PWV, GNSS PWV and ERA5 PWV results in a higher accuracy than the fusion of GNSS PWV and ERA5 PWV by a margin of 0.35 mm. In addition, the inland canyon topography appears to have an influence on the inversion accuracy of both the methods.

Keywords: data fusion; WRF data assimilation; precipitable water vapor (PWV)

1. Introduction

Water vapor makes up less than 4% of the atmosphere’s mass, but it plays an important role in atmospheric processes at all scales. When passing through the neutral atmosphere, electromagnetic waves will be affected by the water vapor, resulting in velocity delay and path bending. Precipitable water vapor (PWV) is the most commonly used term to express the amount of atmospheric water vapor. It is defined as the total atmospheric water vapor contained in a vertical column of unit cross-sectional area extending between any two specified levels and commonly expressed in terms of the height to which that water substance would stand if completely condensed and collected in a vessel of the same unit cross section [1,2]. Obtaining high-precision precipitable water vapor is conducive to not only the description of the atmospheric processes, but also the high-precision inversion and interpretation of Earth observations. There are many different sources of PWV products, such as numerical weather
prediction (NWP) models, for example, ECWMF (European Centre for Medium-Range Weather Forecasts), NCEP (National Centers for Environmental Prediction), geodetic observations based on electromagnetic waves, for example, GNSS (Global Navigation Satellite System), and satellite remote sensing data, for example, MODIS (moderate-resolution imaging spectroradiometer). PWV from different sources have different precision and spatial-temporal resolutions. The GNSS PWV has the advantages of high temporal resolution, high accuracy and weather-proof, but its spatial resolution is sparse. The PWV generated by NWP models has high resolution, but its accuracy is generally lower than that of the GNSS PWV. In order to obtain PWV products with both high precision and high spatial-temporal resolution, many data fusion methods have been proposed. They can be divided into two categories. The first one is the data assimilation (DA) method, in which the observed PWV data is assimilated into an NWP model, for example, the DA based on the Weather Research and Forecasting (WRF) model. As a result, the accuracy of the predicted meteorological parameters including temperature, pressure and relative humidity is improved, as well as the accuracy of the inverted PWV. The other is to directly use mathematical methods to generate high-precision PWV products by fusing multi-source PWV data.

The data assimilation based on an NWP model can generate the water vapor products with improved accuracy, and improve the weather forecast accuracy. One of the NWP models is the WRF model, which was developed and maintained by Mesoscale and Microscale Meteorology (MMM) Laboratory of the National Center for Atmospheric Research (NCAR). It uses numerical weather prediction products (such as NCEP FNL (final), etc.) as its background field, and actual observations (such as GNSS PWV/zenith total delay (ZTD)) as assimilation data source to improve the background field accuracy. The ZTD can be divided into two part, one is the zenith hydrostatic delay (ZHD), another is the zenith wet delay (ZWD). Sharifi et al. [3] conducted a GNSS PWV assimilation experiment based on the WRF model in northern Iran and found that the assimilation of GNSS PWV could improve rainfall forecast accuracy. Mateus et al. [4] studied the assimilation of InSAR (interferometric synthetic aperture radar) PWV into the WRF model and found that the bias of the PWV forecast was corrected after the assimilation, and the precipitation accuracy of small or moderate rain was improved even 9 h after the assimilation. Saito et al. [5] assimilated GNSS PWV into the Japan Meteorological Agency (JMA) operational mesoscale model in Tokyo, and it showed some improvement in weather forecast. Other studies include the assimilation of ZTD into NWP models [6–9]. On a wider front on the water-related applications, the data assimilation is attracting significant attention, for example in [10–12]. Wang et al. [10] improved real-time forecasting of water quality indicators with a combination of process-based models and data assimilation techniques. Wang et al. [11] took both the spatial and temporal relations into consideration to develop models of correcting water level outputs in the Singapore Regional Model (SRM) and its coarse version (SRMC). Karri and Babovic [12] enhanced tidal prediction performance through the use of the data assimilation tool OpenDA in which the ensemble Kalman filter is integrated into a hydrodynamic model.

In addition to the data assimilation approach, high-precision PWV can also be obtained by the fusion of multi-source PWV products. Li [13] fused GNSS PWV and MODIS PWV with inverse distance weighting method to produce a 1 km × 1 km PWV product. Alshawaf et al. [14] fused InSAR PWV and GNSS PWV by a geostatistical data fusion approach based on the method of fixed-rank Kriging. Zhang et al. [2] fused ECWMF ERA5 (ECWMF Reanalysis 5) PWV, MODIS PWV and GNSS PWV using a spherical cap harmonic model and Helmert variance component estimation to generate high-precision PWV products. These approaches are all for the PWV product fusion. Other approaches with neural networks and neural networks-chaos theory combination have been developed to improve the forecast accuracy of current speeds in the Danish Oresund Strait [15] with neural networks, and the accuracy of tidal prediction at measurement stations in the Singapore Regional Model domain with time-delay neural networks combined with chaos theory [16].

In this paper, the data assimilation based on the WRF model (WRFDA) and the method of Zhang et al. [2] are used to process the PWV data in the western United States of America, and various
results are compared. The assessment of these two methods by comparing the differences in the PWV products could help data assimilation researchers to better understand new fusion methods and possibly obtain more accurate assimilation data sources; on the other hand, it could help data fusion researchers to better understand the data assimilation approach and use data of various sources better in the fusion.

In the following, the research area and the multi-source PWV data are presented first. The WRFDA approach and the data fusion method by Zhang et al. [2] are briefly discussed in Section 3. Section 4 presents the results, followed by the conclusion in Section 5.

2. Research Area and Multi-Source Data

2.1. Research Area

The research area is shown in Figure 1, with latitude from 33° N to 38° N and longitude from 118° W to 123° W, covering southern California and part of Nevada. The area contains Central Valley stretching through the middle of California from just north of Sacramento to Los Angeles. The Sierra Nevada in eastern California contains a large number of peaks over 3962 m of altitude, including Mount Whitney (4421 m). There are more than 20 GNSS stations, two radiosonde stations and abundant MODIS data in this range, which is well suited for experimenting the numerical DA and data fusion methods. In addition, the topography of the study area is very undulating, which brings some challenges to the experiment.

2.2. Multi-Source Data

The data used in this paper include reanalysis data from the numerical weather prediction model (e.g., ECWMF, NCEP), the GNSS PWV and the MODIS PWV. In the data assimilation experiment, because the default background field error file provided by WRFDA was generated based on NCEP data, we choose NCEP FNL (final) Operational Global Analysis Data as the background data with a temporal resolution of 6 h and a spatial resolution of 1° × 1°. The reanalysis data is produced by the Global Data Assimilation System (GDAS). In the data fusion experiment, we use the PWV data provided by ECWMF ERA5, which has a temporal resolution of 1 h and a spatial resolution of about 31 km.

The GNSS PWV data used in the experiment were derived from SuomiNet Network, with a time resolution of 30 min. The specific GNSS PWV processing and acquisition method can be referred to Zhang et al. [2].

The MODIS PWV includes NIR (near infrared) PWV and IR (infrared) PWV, with resolutions of 1 km and 5 km, respectively. The accuracy of the NIR PWV is higher than that of the IR PWV. The MODIS PWV is vulnerable to cloud and rain, and its quality is relatively poor against GNSS PWV. Therefore, it is necessary to assess its quality. First, the quality control information provided with the MODIS PWV is used to screen out the PWV data under cloudy and sunny weather. Then, negative PWV and abnormally large PWV are deleted. Detailed data processing methods are presented in Zhang et al. [2]. Altitude information of the MODIS PWV is obtained by interpolating STRMDEM (Shuttle Radar Topography Mission Digital Elevation Model) Version 4.1 data which has a resolution of 250 m [17,18].

In this paper, data from two radiosonde stations is also used. The data of the station near VAN6 is in the IGRA2 (Integrated Global Radiosonde Archive) radiosonde data files, while that of the station near P224 is provided by the University of Wyoming.
Figure 1. (a) Research area, the triangle is for the Global Navigation Satellite System (GNSS) station, and the star for the radiosonde station; (b) a map of the USA showing the research area within the square.
3. Method

3.1. WRF Data Assimilation

In the experiments of this paper, WRF Version 3.7 and WRFDA 3-D variational model are adopted to assimilate the GNSS PWV provided by SuomiNet data [19]. Considering that the accuracy of the GNSS PWV is generally at 1–2 mm level, those with accuracy lower than 2 mm are excluded. The background data is from the NCEP FNL (final) Operational Global Analysis whose time resolution is 6 h and spatial resolution is 1° × 1°. The pressure of the top layer is set to 50 hPa. There are 10 layers in the planetary boundary layer (PBL). In the vertical direction, 37 isobaric layers are formed, and all the altitudes in the experiments are converted into the normal height system. The flowchart of the WRFDA experiment is shown in Figure 2.

![Flowchart of WRFDA experiment](image)

The background field would be processed by the WPS (WRF preprocess system) and real.exe. The GNSS PWV is written in the Little_R format. The default background data error file be.dat of WRFDA is used. The MODIS PWV is not assimilated into the WRF model due to its poor accuracy. When the GNSS PWV file, background data and background error information file are prepared, we run WRFDA. The results are then processed by NCAR (National Centre for Atmospheric Research) Command Language (NCL) [20], and the PWV after the assimilation is obtained by using the following equation:

\[ PWV = \Pi \times ZWD \]

where \( \Pi \) is a dimensionless conversion factor, and it is calculated by Equation (2):

\[ \Pi = \frac{10^6}{\rho_{\text{water}} R_w (k/T_m + k')} \]  

(2)

where \( \rho_{\text{water}} \) is the density of water, \( R_w = 461 \) (J × kg\(^{-1}\) × K\(^{-1}\)) the water vapor ratio constant, \( k = (3.776 \pm 0.014) \times 10^5 \) K\(^2\) × hPa, \( k' = 16.48 \) K × hPa\(^{-1}\) and \( T_m \) is the weighted average temperature computed with the method by Yao et al. [21]. The ZWD in Equation (1) is calculated by the method presented in Vedel and Huang [22].

\[ ZWD = \sum p \times q \times (k_1 + k_2 \frac{T}{T}) \times \frac{h}{T} \]

(3)
where $p$ is the pressure in Pa, $q$ the specific humidity in g/g, $T$ the temperature in K, $k_1 = 2.21 \times 10^{-7}$ K/Pa, $k_2 = 3.73 \times 10^{-3}$ K$^2$/Pa, and $h$ the layer height in m.

It is noted that the assimilation results are unrelated to the physical parameter setting of the WRF [23]. However, it would still be useful to examine whether the grid resolution used in the assimilation experiment has any significant effect on the results. As such, we set four grid resolutions of 5000 m, 7500 m, 10,000 m and 12,500 m, respectively, along the east-west and south-north directions with the center at (35.5° N, 120.5° W), and the grid numbers are thus $112 \times 112$, $74 \times 74$, $56 \times 56$ and $46 \times 46$, respectively. Radiosonde data is treated as true value. Since the radiosonde is launched only at 0:00 UTC and 12:00 UTC daily, the experiments can only be made at those times (by the way, the radiosonde data at 12:00 UTC on 6 May and 0:00 UTC on 7 May is missing). With this setup, we can obtain four PWV result sets with respect to the four grid resolutions at the radiosonde locations.

In this article, bias means the average of the data, STD means the standard deviation, and RMS means the root mean square. Bias, STD and RMS are used as the accuracy measures which are computed by Equations (4)–(6), respectively.

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^{N} (\tilde{x}_i - x_i)$$ \hspace{1cm} (4)

$$\text{STD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (D_i - u)^2}; \; D_i = (\tilde{x}_i - x_i); \; u = \frac{1}{N} \sum_{i=1}^{N} D_i$$ \hspace{1cm} (5)

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\tilde{x}_i - x_i)^2}$$ \hspace{1cm} (6)

where $\tilde{x}_i$ is the reference value, $x_i$ denotes the value estimated by the model and $N$ is the number of observations.

The differences between the inverted and reference PWV values for the four grid resolutions, given in Table 1, appear only at the second decimal place, which indicates that the grid resolution has little influence on the accuracy of the inverted PWV.

| Resolution (m) | 5000 | 7500 | 10,000 | 12,500 |
|---------------|------|------|--------|--------|
| Bias          | 1.14 | 1.19 | 1.20   | 1.15   |
| STD           | 1.57 | 1.55 | 1.57   | 1.60   |
| RMS           | 1.91 | 1.93 | 1.94   | 1.94   |

Secondly, we also checked the temperature ($T$), pressure ($P$) and relative humidity ($RH$) of the experimental results by using the radiosonde data, and the differences are listed in Table 2. The differences between the maximum and minimum RMS values of the four groups were 0.01 K, 0.02 hPa and 0.1%, respectively, for $T$, $P$ and $RH$. It can be seen that the resolution of the assimilation grid has little influence on inverted temperature, pressure and relative humidity.
Table 2. Differences between the WRFDA output temperature ($T$), pressure ($P$) and relative humidity ($RH$) with respect to the grid resolution and radiosonde data.

| Resolution (m) | 5000 | 7500 | 10,000 | 12,500 |
|---------------|------|------|--------|--------|
| $T$ (K) Bias  | −0.08| −0.08| −0.08  | −0.09  |
| STD           | 1.74 | 1.73 | 1.73   | 1.72   |
| RMS           | 1.74 | 1.73 | 1.73   | 1.73   |
| $P$ (hPa) Bias| 0.56 | 0.57 | 0.56   | 0.55   |
| STD           | 0.59 | 0.60 | 0.63   | 0.60   |
| RMS           | 0.81 | 0.83 | 0.84   | 0.82   |
| $RH$ (%) Bias | 1.82 | 1.81 | 1.83   | 1.80   |
| STD           | 13.38| 13.34| 13.36  | 13.34  |
| RMS           | 13.49| 13.45| 13.47  | 13.46  |

To sum up, within the experimental area, the resolution of the assimilation grid has little influence on the accuracy of the inverted $PWV$, temperature, pressure and relative humidity from the assimilation results. Since the grid resolution of 5000 m results in the minimum RMS for the inverted $PWV$, in the subsequent experiments, we set the grid resolution to 5000 m in the assimilation.

3.2. Data Fusion

We use the method by Zhang et al. [2] in this paper to perform data fusion experiments since the combined $PWV$ obtained by this method can not only reflect the features of the individual $PWV$ maps but also suppress regional deviations. The core idea is to use a spherical cap harmonic (SCH) model to describe the $PWV$ field on the sphere and to use the Helmert variance component estimation (HVCE) to determine the weights for data of different sources.

Before fusion, data from different sources need to be corrected for systematic differences. We use the GNSS $PWV$ data to calibrate the ERA5 $PWV$ and MODIS $PWV$ data since the accuracy of the GNSS $PWV$ is as high as 1–2 mm [13,24–26]. The way to determine the systematic bias is as follows. First, an SCH model is used to fit the MODIS (or ERA5) $PWV$, and then the fitted model is used to calculate the $PWV$ at the GNSS stations. We take the average of the differences between GNSS $PWV$ and MODIS (or ERA5) $PWV$ as the systematic bias.

The SCH model is of the following form:

$$PWV = \sum_{k=0}^{N} \sum_{m=0}^{k} a_k^{(m)} \left( \frac{r}{a} \right) P_m^{n_k(m)}(\cos \theta) \cdot \{g_{bk}^m \cos(m\lambda) + h_k^m \sin(m\lambda)\}$$  (7)

where $a = 6,378,137$ m is the radius of the Earth, $r, \theta, \lambda$ are the geocentric distance, co-latitude and longitude of a point on the Earth's surface in the spherical coordinate system. The $PWV$ data will be expanded onto the sphere of radius equal to $r$. $P_m^{n_k(m)}$ is the associative Legendre function of the first kind of boundary value problem. $n_k(m)$ and $m$ are the order and degree of the SCH model, respectively. $k$ is an integer used to label the order of $n_k(m)$. According to boundary conditions proposed by Haines [27], $m$ needs to be a positive integer, $n_k(m)$ is real and $k \geq m \geq 0$. $N$ is the highest order of the SCH model expansion. $g_{bk}^m$ and $h_k^m$ are the model coefficients representing the amplitude of the harmonic which are to be determined. When $m$ is equal to 0, $\sin(m\lambda)$ is equal to 0, it is impossible to determine the value of $h_k^0$. Thus, there are $(N + 1)^2$ coefficients in total to be determined.

As data fusion involves data from different sources with different accuracy, they should be weighted appropriately. The a priori information on the data accuracy is usually not well known, the posterior method would be needed to determine the weight relation of various data [28]. A common one is the variance component estimation technique, which not only can estimate the variances of observations with different accuracy, but also provides a regularization method [29]. Here the Helmert variance component estimation method is used to determine the weights of $PWV$ from different sources.
The data fusion flowchart is shown in Figure 3. As shown, we first use the spherical cap harmonic function to fit ERA5 (MODIS) PWV, and then the function is used to obtain the interpolated PWV at GNSS stations. The systematic bias is now determined as the average of the differences between GNSS PWV and interpolated ERA5 (MODIS) PWV, and it is used to correct the ERA5 (MODIS) PWV. Next, we determine the weights for PWV of different sources by HVCE. Finally, the PWV data of different sources is fused with the SCH model, in which the model coefficients are estimated. It is noted that the coordinate transformation, the calculations of \( n_k (m) \) and the Legendre polynomials are important in the use of the SCH model to obtain fused PWV.

![Data Fusion Flowchart](image)

**Figure 3.** The flowchart of spherical cap harmonic (SCH) and Helmert variance component estimation (HVCE) data fusion method proposed by Zhang et al. (2019).

The treatment of the uncertainty associated with different data sources in Zhang et al. (2019)’s method includes two aspects. One is regarding the average difference between GNSS PWV and ERA5 (MODIS) PWV as systematic bias; another is to use HVCE to determine weights for PWV data of different sources.

It is now possible to perform the data fusion experiments. In this paper, the ERA5 PWV and GNSS PWV are fused by a SCH model of order 7, and the order is 8 for the fusion of the ERA5 PWV, GNSS PWV and MODIS PWV.

4. Results

Due to the limited resolution of NCEP FNL (final) Operational Global Analysis data, its combination with the GNSS PWV in the experimental area would be rank deficit if using a high-order model, and inaccurate if using a low-order model. Therefore, the GNSS PWV and ECWMF REA5 data are fused. In addition, in the original text of Zhang et al. [2], MODIS PWV is also fused with the GNSS PWV and ERA5 PWV. As a result, we set out to perform four experiments. The first one is to assimilate the GNSS PWV into the WRF model, and the result is labeled as DA_PWV. Compared with DA_PWV, the PWV obtained from the WPS and real.exe is labeled as None_PWV which has no assimilation involved. Fuse_PWV is the result from the fusion of the GNSS PWV and ERA5 PWV, while M_Fuse_PWV is that of the GNSS PWV, ERA5 PWV and MODIS PWV.

Because there is no overlap time period between the fused PWV and radiosonde PWV data, some of the high accuracy GNSS PWV is used as the reference to examine the fusion performance. We choose the GNSS PWV data at stations P729, P056 and LUTZ, which are marked with a red box in Figure 1, as reference, and they are excluded in the assimilation and fusion experiments. The assimilation experiment is carried out for the same time period as of the fusion experiment throughout May 2018.

For the None_PWV and DA_PWV results, the bilinear interpolation and altitude correction are applied to obtain the PWV at the three reference stations. The temperature, pressure and humidity information are also obtained from the results of the respective experiments. The M_Fuse_PWV and Fuse_PWV at the three stations are computed from their fusion models, respectively. The comparison
between the experiment results and the reference data is shown in Figure 4, and the bias, STD and RMS values are given in Table 3.

![Figure 4](image_url)

**Figure 4.** Comparison of four groups of results with true values. (The Fuse\_PWV is the result from the fusion of the GNSS PWV and ERA5 PWV by Zhang et al., while M\_Fuse\_PWV that of the GNSS PWV, ERA5 PWV and MODIS PWV. DA\_PWV is the result from assimilating the GNSS PWV into the WRF mode. Compared with DA\_PWV, the PWV obtained from the WPS and real.exe is labeled as None\_PWV which has no assimilation involved. The x-axis is the time. 2018122.1755 is read as “at 17:55 on DOY 122 of 2018”, and so on).
Table 3. Differences between four PWV result sets and GNSS PWV at three reference stations.

|                  | None_PWV | DA_PWV |                  | None_PWV | DA_PWV |
|------------------|----------|--------|------------------|----------|--------|
|                  | P729     | P056   | LUTZ All         | P729     | P056   |
| Bias             | -0.66    | -1.21  | 0.24 -0.54 -0.43 | -0.77    | -0.39  | -0.25 |
| STD              | 1.20     | 1.95   | 1.11 1.57 0.95   | 1.87     | 1.16   | 1.46  |
| RMS              | 1.35     | 2.27   | 1.12 1.65 1.02   | 1.99     | 1.21   | 1.47  |

|                  | Fuse_PWV | M_Fuse_PWV |                  | Fuse_PWV | M_Fuse_PWV |                  |
|------------------|----------|------------|------------------|----------|------------|------------------|
|                  | P729     | P056       | LUTZ All         | P729     | P056       | LUTZ All         |
| Bias             | -0.46    | -1.99      | -0.81 -1.09 0.09 | 0.11     | -0.41      | -0.07            |
| STD              | 0.71     | 1.04       | 0.75 1.06 1.38   | 1.16     | 0.88       | 1.17             |
| RMS              | 0.84     | 2.23       | 1.09 1.52 1.36   | 1.15     | 0.96       | 1.17             |

The three reference stations are widely separated in the study area, with LUTZ in the upper left corner, P729 in the lower right corner and P056 in the middle right edge of an intermountain valley. The fused or assimilation-inverted results at the three stations are all close to their reference, which shows the reliability of the results.

For the P729 station, the difference between the None_PWV group and DA_PWV group in the period from days of year (DOY) 132 to 150 is large, while the differences during other assimilation time periods are small. The M_Fuse_PWV is close to the Fuse_PWV on the whole, but the differences in the period between DOY 130 and 132 are obvious. For this station, the Fuse_PWV result is slightly smaller than the reference, and the M_Fuse_PWV result is larger than the reference.

The Fuse_PWV is the result from the fusion of the GNSS PWV and ERA5 PWV by Zhang et al. (2019), while M_Fuse_PWV that of the GNSS PWV, ERA5 PWV and MODIS PWV. DA_PWV is the result from assimilating the GNSS PWV into the WRF mode. Compared with DA_PWV, the PWV obtained from the WPS and real.exe is labeled as None_PWV which has no assimilation involved. The x-axis is the time. Additionally, 2,018,122.1755 is read as “at 17:55 on DOY 122 of 2018”, and so on.

For the P056 station, the Fuse_PWV result is also smaller than the reference, while the M_Fuse_PWV one is larger than the reference, and it is closer to the reference than the Fuse_PWV group. The None_PWV group is very close to the DA_PWV group. Since the absolute value of bias of None_PWV is smaller than Fuse_PWV, the closeness of the None_PWV to the reference is better than that of the Fuse_PWV group.

For the P729 station, the minimum absolute bias occurs in the M_Fuse_PWV group, and the maximum and minimum RMS values are in the M_Fuse_PWV group and Fuse_PWV group, respectively. The maximum and minimum values of RMS at LUTZ station are in DA_PWV and M_Fuse_PWV, respectively. For the P056 station, the maximum RMS is in the None_PWV group, and the minimum in the M_Fuse_PWV group.

Because of the existence of the bias, the RMS is a more important measure of the result accuracy, with the minimum RMS indicating the best, and the maximum the worst. From the overall RMS value given in the last column, the M_Fuse_PWV result is the best among the four groups. With respect to the individual stations, both P056 and LUTZ have their minimum RMS values with M_Fuse_PWV, but P729’s minimum RMS is with Fuse_PWV, which is significantly smaller than the station’s maximum RMS unexpectedly from M_Fuse_PWV. This shows that, although there are only three reference stations, the performance in terms of the RMS is not uniform with the data processing methods.

In general, we may conclude that the background field accuracy can be improved when the GNSS PWV is assimilated into the NWP model. The absolute values of bias, STD and RMS in the DA_PWV group are less than those in the None_PWV group. For the two PWV fusion products, the accuracy of the M_Fuse_PWV is higher than that of the Fuse_PWV, suggesting that the fusion of data from more sources is beneficial to the accuracy of resulted PWV. The overall RMS values of the Fuse_PWV group and M_Fuse_PWV group are all smaller than that of the None_PWV group, which indicates
that, to some extent, appropriate data fusion can achieve better PWV field than that of the WRF model without any assimilation.

The overall RMS of the DA_PWV is close to that of the Fuse_PWV but has a large difference with that of the M_Fuse_PWV. In particular, for the two stations P729 and LUTZ located in the coastal area, the two-source Fuse_PWV results are slightly better than those of the assimilation group (DA_PWV), but for the station P056 in the inland valley, the DA_PWV results are better than the Fuse_PWV results. There are two possible explanations. First, before the fusion, the GNSS PWVs have been used to correct the ERA5 PWV for the system difference, while in the assimilation, the system difference between GNSS PWVs and the background field is difficult to correct. Second, the ECWMF ERA5 product has a better spatial resolution (0.25° × 0.25°) than the NCEP FNL (1° × 1°). However, the same correction is applied to all stations to eliminate the system difference between ERA5 PWV and GNSS PWV before the fusion, which may result in a relatively large residual deviation for stations in areas of large topographical fluctuations, and this in turn would affect the fusion accuracy for these stations.

5. Conclusions

This paper presents a comparison study about the accuracy of the assimilated and fused PWV products for a west coast area of the United States of America. The former is obtained from the assimilation of the GNSS PWV into the WRF model, and the latter by the fusion of the GNSS, ERA5 and MODIS PWV data using the method of Zhang et al. [2]. For the study area, high-precision PWV products are obtained from both approaches. The results show that the high precision could be achieved by either adjusting meteorological parameters or directly fusing multi-source data, with accuracy of 1.47 mm by the former and 1.52 mm by the latter. It is seen that the assimilation has reduced the RMS of PWV products from 1.65 to 1.47, a significant improvement in the PWV accuracy, although the temperature, pressure and relative humidity have little change. It also shows that, the fusion of data from three sources including the MODIS PWV clearly outperforms the fusion without the MODIS PWV with the RMS values of 1.17 and 1.52, respectively, suggesting more data sources are helpful to achieve better accuracy.

The other finding from the accuracy comparison at the three reference stations is that, the accuracy at the P056 station located in the inland valley is worse than those at the P729 and LUTZ stations along the coast, indicating that complex topography may affect the performance of the applied data assimilation and data fusion methods.

Author Contributions: Conceptualization, Z.X.; methodology, Z.X. and B.Z.; software, Z.X. and B.Z.; validation, Z.X. and J.L.; writing—original draft preparation, Z.X.; writing—review and editing, J.S.; supervision, X.S.; project administration, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Abbreviations

| Abbreviations | Full Name |
|---------------|-----------|
| DA            | Data assimilation |
| DOY           | Days of year |
| ECWMF         | European Centre for Medium-Range Weather Forecasts |
| ERA5          | ECWMF Reanalysis 5 |
| FNL           | Final |
| GDAS          | Global Data Assimilation System |
| GNSS          | Global Navigation Satellite System |
| HVCE          | Helmert variance component estimation |
| IGRA          | Integrated Global Radiosonde Archive |
| InSAR         | Interferometric synthetic aperture radar |
IR  Infrared  
JMA  Japan Meteorological Agency  
MODIS  Moderate-resolution imaging spectroradiometer  
MMM  Mesoscale and Microscale Meteorology  
NCAR  National Center for Atmospheric Research  
NCEP  National Centers for Environmental Prediction  
NCL  NCAR Command Language  
NIR  Near infrared  
NWP  Numerical weather prediction  
PBL  Planetary boundary layer  
PWV  Precipitable water vapor  
RH  Relative humidity  
RMS  Root mean square  
SCH  Spherical cap harmonic  
SRM  Singapore Regional Model  
SRMC  Singapore Regional Model Coarse  
STD  Standard deviation  
STRMDEM  Shuttle Radar Topography Mission Digital Elevation Model  
WPS  WRF preprocess system  
WRF  Weather Research and Forecasting  
ZHD  Zenith hydrostatic delay  
ZTD  Zenith total delay  
ZWD  Zenith wet delay  

References  
1. American Meteorological Society (AMS). *Glossary of Meteorology*, 2nd ed.; American Meteorological Society (AMS): Boston, MA, USA, 2000.  
2. Zhang, B.; Yao, Y.; Xin, L.; Xu, X. Precipitable water vapor fusion: An approach based on spherical cap harmonic analysis and Helmert variance component estimation. *J. Geod.* 2019, 93, 2605–2620. [CrossRef]  
3. Sharifi, M.A.; Azadi, M.; Khaniani, A.S. Numerical simulation of rainfall with assimilation of conventional and GPS observations over north of Iran. *Ann. Geophys.* 2016, 59, 0322.  
4. Mateus, P.; Tomé, R.; Nico, G.; Catalão, J. Three-Dimensional Variational Assimilation of InSAR PWV Using the WRFDA Model. *IEEE Trans. Geosci. Remote Sens.* 2016, 54, 7323–7330. [CrossRef]  
5. Saito, K.; Shoji, Y.; Origuchi, S.; Duc, L. GPS PWV assimilation with the JMA nonhydrostatic 4DVAR and cloud resolving ensemble forecast for the 2008 August Tokyo metropolitan area local heavy rainfalls. In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. III)*; Springer: Cham, Switzerland, 2017; pp. 383–404.  
6. Faccani, C.; Ferretti, R.; Vespe, F. GPS PW assimilation into MM5 with the nudging technique. *Phys. Chem. Earth Part A* 2001, 26, 481–485. [CrossRef]  
7. Pacione, R.; Sciarretta, C.; Faccani, C.; Ferretti, R.; Vespe, F. GPS PW assimilation into MM5 with the nudging technique. *Adv. Geosci.* 2005, 2, 73–79. [CrossRef]  
8. Bennitt, G.V.; Jupp, A. Operational Assimilation of GPS Zenith Total Delay Observations into the Met Office Numerical Weather Prediction Models. *Mon. Weather Rev.* 2012, 140, 2706–2719. [CrossRef]  
9. Lindskog, M.; Ridal, M.; Thorsteinsson, S.; Ning, T. Data assimilation of GNSS zenith total delays from a Nordic processing centre. *Atmos. Chem. Phys.* 2017, 17, 13983–13998. [CrossRef]  
10. Wang, X.; Babovic, V.; Li, X. Application of spatial-temporal error correction in updating hydrodynamic model. *J. Hydro-Environ. Res.* 2017, 16, 45–57. [CrossRef]  
11. Wang, X.; Zhang, J.; Babovic, V. Improving real-time forecasting of water quality indicators with combination of process-based models and data assimilation technique. *Ecol. Indic.* 2016, 66, 428–439. [CrossRef]  
12. Karri, R.; Vladan, B. Enhanced predictions of tides and surges through data assimilation. *Int. J. Eng.* 2017, 30, 23–29.  
13. Li, Z. Production of regional 1 km x 1 km water vapor fields through the integration of GPS and MODIS data. *Am. J. Public Hyg.* 2004, 19, 1443–1450.
14. Alshawaf, F.; Hinz, S.; Mayer, M.; Meyer, F.J. Constructing accurate maps of atmospheric water vapor by combining interferometric synthetic aperture radar and GNSS observations. *J. Geophys. Res.* 2015, 120, 1391–1403. [CrossRef]

15. Babovic, V.; Cañizares, R.; Jensen, H.R.; Klinting, A. Neural networks as routine for error updating of numerical models. *J. Hydraul. Eng.* 2001, 127, 181–193. [CrossRef]

16. Sun, Y.; Babovic, V.; Chan, E.S. Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. *J. Hydrol.* 2010, 395, 109–116. [CrossRef]

17. Reuter, H.I.; Nelson, A.; Jarvis, A. An evaluation of void-filling interpolation methods for SRTM data. *Int. J. Geogr. Inf. Sci.* 2007, 21, 983–1008. [CrossRef]

18. Sun, Y.; Babovic, V.; Chan, E.S. Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. *J. Hydrol.* 2010, 395, 109–116. [CrossRef]

19. Ware, R.H.; Fulker, D.W.; Stein, S.A.; Anderson, D.N.; Avery, S.K.; Clark, R.D.; Droegemeier, K.K.; Kuetttner, J.P.; Minster, J.B.; Sorooshian, S. SuomiNet: A real-time national GPS network for atmospheric research and education. *Bull. Am. Meteorol. Soc.* 2000, 81, 677–694. [CrossRef]

20. UCAR/NCAR/CISL/VETS: The NCAR Command Language (NCL, Version 6.1.2), UCAR/NCAR/CISL/VETS, Boulder, Colorado, 2013. Available online: http://www.ncl.ucar.edu (accessed on 9 October 2019).

21. Yao, Y.; Zhang, B.; Xu, C.; Yan, F. Improved one/multi-parameter models that consider seasonal and geographic variations for estimating weighted mean temperature in ground-based GPS meteorology. *J. Geod.* 2014, 88, 273–282. [CrossRef]

22. Vedel, H.; Huang, X.Y.; Haase, J.; Ge, M.; Calais, E. Impact of GPS zenith tropospheric delay data on precipitation forecasts in Mediterranean France and Spain. *Geophys. Res. Lett.* 2004, 31, L02102. [CrossRef]

23. Xiong, Z.; Zhang, B.; Yao, Y. Comparisons between the WRF data assimilation and the GNSS tomography technique in retrieving 3-D wet refractivity fields in Hong Kong. *Ann. Geophys.* 2019, 37, 25–36. [CrossRef]

24. Chen, S.H.; Zhao, Z.; Haase, J.S.; Chen, A.; Vandenberghhe, F. A Study of the Characteristics and Assimilation of Retrieved MODIS Total Precipitable Water Data in Severe Weather Simulations. *Mon. Weather Rev.* 2007, 136, 9052–9062. [CrossRef]

25. Prasad, A.K.; Singh, R.P. 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. Atmos.* 2009, 114, D05107. [CrossRef]

26. Chang, L.; Gao, G.; Jin, S.; He, X.; Xiao, R.; Guo, L. Calibration and Evaluation of Precipitable Water Vapor from MODIS Infrared Observations at Night. *IEEE Trans. Geosci. Remote Sens.* 2014, 53, 2612–2620. [CrossRef]

27. Haines, G.V. Spherical cap harmonic analysis. *J. Geophys. Res. Solid Earth* 1985, 90, 2583–2591. [CrossRef]

28. Koch, K.; Kusche, J. Regularization of geopotential determination from satellite data by variance components. *J. Geod.* 2002, 76, 259–268. [CrossRef]

29. Xu, P.; Shen, Y.; Fukuda, Y.; Liu, Y. Variance Component Estimation in Linear Inverse Ill-posed Models. *J. Geod.* 2006, 80, 69–81. [CrossRef]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).