remote sensing

MDPI

Article

An All-Weather Land Precipitable Water Vapor Retrieval Method Integrating Physical Constraints and Machine Learning

Fang-Cheng Zhou [1,2,3], Guangwei Li [4], Xiaoning Song [5], Xiuzhen Han [1,2,3,*] and Shiyi Chen [1,2,3]

[1] National Satellite Meteorological Center (National Center for Space Weather), Beijing 100081, China; zhoufc@cma.gov.cn (F.-C.Z.); chenshiyi@cma.gov.cn (S.C.)
[2] Innovation Center for FengYun Meteorological Satellite (FYSIC), Beijing 100081, China
[3] Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites/Key Laboratory of Space Weather, China Meteorological Administration, Beijing 100081, China
[4] Hainan Institute of Meteorological Science, Haikou 570203, China; gweylee@gmail.com
[5] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China; songxn@ucas.ac.cn
* Correspondence: hanxz@cma.gov.cn

Highlights

What are the main findings?

- Compared with conventional physically constrained machine learning retrieval methods, the proposed method reduces the RMSE by 16.5%.
- Compared with ground-based SuomiNet site observations, the minimum retrieval RMSE is as low as 2.95 mm.

What are the implications of the main findings?

- The physically constrained machine learning retrieval method for land-based PWV can improve the retrieval accuracy.
- The PWV index derived via reformulation of the radiative transfer equation contributes substantially to the improvement of retrieval accuracy.

Abstract

Atmospheric precipitable water vapor (PWV) is an essential geophysical parameter for weather forecasting, climate change research and hydrological studies. Spaceborne passive microwave sensors support all-weather PWV observation, yet accurate retrieval of land PWV still faces great difficulties owing to the spatiotemporal complexity and variability of land surface emissivity (LSE). This study proposes a novel all-weather land PWV retrieval approach using Fengyun-3D microwave radiation imager (MWRI) observations, which combines physical radiative transfer principles with machine learning techniques. By reorganizing the radiative transfer equation, this study constructs two characteristic factors that can effectively characterize LSE and PWV, and further incorporates them into the machine learning retrieval framework. Three typical machine learning algorithms, namely Random Forest (RF), Adaptive Kernel Extreme Learning Machine (AKELM), and Extreme Gradient Boosting (XGBoost), achieve the optimal retrieval accuracy, with an RMSE of 3.80 mm and an R^2 of 0.91. Multiple validation results indicate that (1) compared with conventional physically constrained machine learning retrieval methods, the proposed method reduces the root mean square error (RMSE) by 16.5%; (2) against ground-based SuomiNet site observations, the minimum retrieval RMSE is as low as 2.95 mm; (3) the proposed method exhibits stable superiority over the Moderate Resolution Imaging Spectroradiometer (MODIS) MYD05 water vapor products with lower RMSE in all experimental cases. This study offers an effective and reliable technical scheme for all-weather land PWV

Check for updates

Academic Editor: Manuel Antón

Received: 15 June 2026
Revised: 20 September 2026
Accepted: 22 September 2026
Published: 7 October 2026

Copyright: © 2026 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.

retrieval, and can facilitate high-precision atmospheric monitoring and related practical operational services.

Keywords: PWV; FY-3D/MWRI; physical constraints; machine learning; all-weather

1. Introduction

Atmospheric total precipitable water vapor (PWV) is a core indicator of vertically integrated atmospheric water content, and is closely associated with precipitation evolution and climate feedback processes. Existing studies [1] have confirmed that sharp rises in PWV generally occur two to three hours ahead of precipitation events. As the most prevalent greenhouse gas, water vapor accounts for approximately 70% of atmospheric radiative absorption and exerts prominent impacts on global climate warming [2]. Accordingly, PWV plays a critical role in weather prediction, climate simulation, and atmospheric correction for remote sensing [3,4], underscoring the need for long-term, high-precision PWV monitoring.

At present, PWV observation techniques are mainly divided into ground- and space-based categories. In situ observational techniques, including radiosondes, Global Navigation Satellite System (GNSS) stations, ground-based microwave radiometers, sun photometers, and Raman lidars [5–9], are capable of acquiring high-precision PWV measurements. Nevertheless, such ground observations are constrained by sparse and uneven spatial coverage, particularly over remote and under-sampled regions such as deserts, oceans, and inland lakes, limiting their ability to provide continuous characterization of regional and global PWV distributions. In contrast, satellites equipped with near-infrared (NIR), thermal infrared (TIR) and microwave sensors are capable of realizing large-scale global PWV observation [10–12].

Representative satellite-based PWV products mainly include the NIR or TIR retrievals from the Moderate Resolution Imaging Spectroradiometer (MODIS) and corresponding products generated from Fengyun (FY) satellite series [13–15]. Xie et al. [16] validated the NIR PWV products derived from FY satellites using observations collected at 369 AERONET sites, and verified that such FY PWV retrievals achieve satisfactory accuracy. Nevertheless, NIR/TIR-based PWV retrievals are invalid under cloudy conditions. By contrast, microwave radiation penetrates cloud layers, overcoming this limitation and enabling all-weather PWV retrieval. A series of microwave sensors, including the Advanced Microwave Scanning Radiometer for EOS (AMSR-E), the Advanced Microwave Scanning Radiometer 2 (AMSR2), the Special Sensor Microwave Imager (SSM/I), the TRMM Microwave Imager (TMI), WindSat, and the MicroWave Radiation Imager (MWRI) onboard the FY-3 series satellites, utilize the water-vapor absorption channel near 23 GHz for PWV retrieval [12,17].

While satellite microwave PWV retrieval over ocean surfaces is well established, with a root mean square error (RMSE) of approximately 1 mm [18–22], accurate land PWV retrieval remains considerably challenging. Heterogeneous land surface conditions, including diverse land cover types, variable vegetation coverage, snow cover, soil moisture, and surface roughness, give rise to complicated microwave radiative interactions. This interference, driven primarily by the strong spatiotemporal variability of land surface temperature (LST) and land surface emissivity (LSE), severely masks the atmospheric water-vapor signal in satellite brightness temperature observations [23–27]. In the radiative transfer equation, the brightness temperature is jointly determined by the atmospheric

transmittance (which carries the PWV signal), LSE, and LST; the two surface terms must, therefore, be decoupled before PWV can be retrieved.

Deeter [17] and Zhou et al. [28] constructed physics-based land PWV retrieval methods based on the radiative transfer equations corresponding to 18 GHz and 23 GHz channels, respectively. These two schemes share a common simplification that neglects the interference of cloud liquid water (CLW). Specifically, Deeter's method eliminates LSE interference by assuming identical LSE polarization differences between the two channels. On this basis, the Zhou-based method further hypothesizes that the 18 GHz and 23 GHz channels share equivalent microwave thermal sampling depths, which validates the rationality of adopting identical LST values for both channels. Du et al. [12] proposed a statistical PWV retrieval algorithm using the Microwave Atmospheric Water Vapor Index (MAWVI, defined as the polarization difference ratio of brightness temperatures at 23.8 GHz and 18.7 GHz), together with surface elevation, LST, and CLW. In their study, LST was estimated using the brightness temperature ratio of 18.7 GHz and 23.8 GHz channels, while CLW was parameterized based on the polarization difference ratio of brightness temperatures at 89 GHz and 36.5 GHz. Kazumori [29] developed a simplified PWV retrieval model that relies solely on MAWVI and discards the LST and CLW correction terms. Although this algorithm achieves a more concise structure, it exhibits inferior accuracy, with an RMSE of 5.8 mm, compared with the 4.7 mm RMSE of the model proposed by Du et al. [12]. These findings demonstrate that incorporating multi-source auxiliary parameters to fully characterize land surface and atmospheric conditions is an effective strategy for improving PWV retrieval accuracy. Among various interference factors, LSE represents one of the most critical and uncertain parameters, severely limiting the accuracy of land PWV retrieval. Therefore, numerous studies have focused on optimizing LSE estimation schemes to mitigate land surface interference. Gao et al. [30] estimated LSE by establishing empirical relationships between polarization ratios and LSE for forested and non-forested land types. Ji and Shi [31] showed that the LSE polarization difference ratio between 18.7 and 23.8 GHz ($\Delta\varepsilon_{18.7}/\Delta\varepsilon_{23.8}$) can be parameterized by the microwave vegetation index together with the polarization ratios at 6.9 GHz, 10.65 GHz, 18.7 GHz, and 36.5 GHz. On this basis, Ji et al. [32] further optimized the $\Delta\varepsilon_{18.7}/\Delta\varepsilon_{23.8}$ estimation model by removing the microwave vegetation index and the 36.5 GHz polarization ratio from the predictor set, and additionally corrected for terrain effects using DEM data; the RMSE of the retrieved global 0.25–gridded TPW was thereby reduced to 3.45 mm.

However, excessive input parameters will introduce strong nonlinear characteristics into retrieval models and may further increase estimation errors. In this context, machine learning (ML) methods, such as random forest (RF) and deep learning algorithms, possess powerful capabilities to fit complex multi-factor correlations from massive datasets, thus showing great applicability to nonlinear PWV retrieval tasks [33]. For instance, Gao et al. [34] combined the physical radiative transfer equation with neural networks to retrieve PWV using AMSR2 data. Xia et al. [35] adopted automatic ML (AutoML) based on AMSR2 observations and obtained an RMSE of 3.1 mm. In addition, Jiang et al. [36] utilized back propagation neural networks (BPNN) and achieved an RMSE of 3.25 mm for descending-orbit PWV retrieval.

The other persistent challenge lies in the accurate characterization of LST—a key factor governing evapotranspiration and further modulating PWV variations—at the coarse spatial resolution of passive microwave sensors. Gao et al. [34] excluded LST from their physical framework but incorporated in situ SuomiNet LST measurements into their neural network model. By comparison, Xia et al. [35] and Jones et al. [37] adopted the vertically polarized brightness temperature at 36.5 GHz as an effective proxy for cloudy-sky LST estimation.

Existing studies have demonstrated that the MWRI onboard the FY-3D satellite is capable of all-weather land PWV retrieval. Nevertheless, dedicated studies and operational PWV products based on FY-3/MWRI remain scarce. More importantly, most existing physically constrained ML methods do not select their input predictors strictly in accordance with the stepwise derivation of the radiative transfer equation, which restricts further accuracy improvement and weakens the physical interpretability of the retrieval. In this study, FY-3D/MWRI observations serve as the primary input and are combined with multi-source auxiliary datasets; SuomiNet Global Positioning System (GPS) PWV observations provide the ground truth. Multiple ML algorithms are compared and screened, and a land PWV retrieval algorithm with physically constrained ML is finally established.

2. Materials and Methods

2.1. Materials

Multiple datasets were adopted in this study for model construction and performance evaluation. First, brightness temperature observations acquired by the FY-3D/MWRI at five typical frequency channels (10.65, 18.7, 23.8, 36.5, and 89 GHz) with both vertical and horizontal polarizations were used to provide all-weather passive microwave radiative information. Second, the MODIS MYD05_L2 infrared atmospheric water vapor product and the AMSR2 land total precipitable water product were employed for cross-comparison and quantitative evaluation. Third, high-precision ground-based GPS-PWV measurements from 127 SuomiNet stations served as the reference for model training and validation.

Multiple auxiliary variables were additionally incorporated to constrain surface and atmospheric conditions, including the GEBCO_2024 digital elevation model (DEM) to correct terrain-induced atmospheric variations and MODIS MCD12Q1 land cover data classified by the International Geosphere-Biosphere Programme (IGBP) scheme to characterize heterogeneous surface features. Geographical coordinates (latitude and longitude) and the normalized sine term of the day of year (DOY) were also included to represent spatial and seasonal variations. All datasets were spatially resampled to a unified $0.1^\circ \times 0.1^\circ$ geographic grid and temporally matched within a ± 30 min window to guarantee rigorous spatiotemporal consistency for model training and validation.

2.1.1. FY-3D MWRI Brightness Temperature Data

The FY-3D satellite is an afternoon polar-orbiting satellite belonging to China's FY satellite series, which carries the MWRI as its core microwave detection payload. The MWRI is equipped with ten observation channels covering five frequency bands (10.65, 18.7, 23.8, 36.5, and 89.0 GHz), with both horizontal and vertical polarization observations available. It achieves a swath width of approximately 1400 km at an incidence angle of 53°, with two daily overpasses corresponding to the descending orbit (02:00 local solar time) and ascending orbit (14:00 local solar time). In this study, MWRI Level-1 brightness temperature data with a native spatial resolution of 10 km during 2018–2019 were acquired from the National Satellite Meteorological Center (NSMC), China Meteorological Administration. All original data were resampled to a uniform 0.1° spatial resolution using cubic spline interpolation for subsequent model processing.

2.1.2. MYD05_L2 PWV Data and AMSR2 PWV Data

MODIS MYD05_L2 Collection 6.1 (V61) is a standard PWV product derived from infrared observations by the MODIS aboard the Aqua satellite. With a nominal spatial resolution of 5 km, it retrieves column-integrated PWV under clear-sky conditions using physically based infrared radiative transfer algorithms and is widely used as a reference for atmospheric water vapor validation. The AMSR2 land PWV product provides verti-

cally integrated water vapor amounts exclusively over land surfaces. In this study, both MYD05_L2 V61 and AMSR2 PWV data were used solely for cross-validation, not for model training. Data for four representative dates (15 January, 15 April, 15 July, and 15 October 2019) were acquired from the NASA LAADS Web and the JAXA AMSR data portal. Owing to the near-coincident overpass times of FY-3D, Aqua/MODIS, and AMSR2, no temporal interpolation was performed. The MYD05_L2 dataset was spatially resampled from its original 5 km resolution to a consistent 0.1° grid via cubic spline interpolation to match the spatial framework used throughout this work.

2.1.3. SuomiNet PWV Data

Ground-based GPS-PWV observations from the SuomiNet North America Network were used as the high-precision benchmark to quantitatively evaluate the accuracy of FY-3D/MWRI PWV retrievals. As a globally distributed GPS observation network, SuomiNet provides continuous, spatiotemporally dense column-integrated PWV measurements with a well-documented accuracy of better than 2 mm [38]. Hourly SuomiNet PWV datasets used in this study were acquired from the University Corporation for Atmospheric Research (UCAR). To match the spatial framework of satellite observations, all in situ PWV measurements were gridded into a unified 0.1° × 0.1° latitude–longitude grid according to their geographic coordinates (Figure 1). For each MWRI brightness temperature pixel, the SuomiNet PWV record with the closest temporal match to the satellite overpass was selected as the ground truth. This strict spatiotemporal matching scheme effectively reduces temporal mismatches between in situ and remote sensing observations, thereby ensuring reliable and credible validation of the proposed retrieval algorithm.

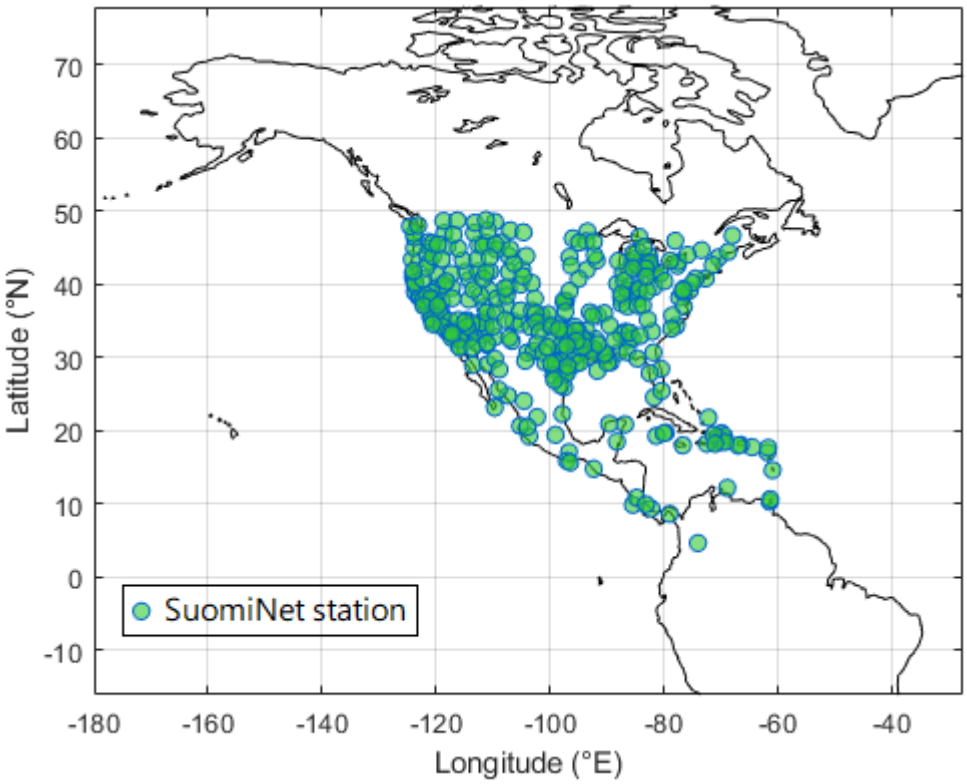


Figure 1. Geographic distribution of SuomiNet stations.

It should be noted that a 0.1° grid cell represents an areal average within the satellite footprint, whereas a SuomiNet station provides a point measurement of the column-integrated water vapor; in addition, the ±30 min matching window may introduce further

uncertainty during rapid weather changes. Compared with LST, however, PWV varies considerably more smoothly in space, so that the point-to-pixel mismatch error tends to be limited.

2.2. Methods

2.2.1. Physical Basis

Based on the radiation transfer equation, the brightness temperature at frequency i can be expressed as follows:

$$T_{b,i,p} = T_{a,i}^{\uparrow} + \tau_i \cdot \varepsilon_{i,p} \cdot T_S + (1 - \varepsilon_{i,p}) \tau_i \cdot T_{a,i}^{\downarrow} \quad (1)$$

where $T_{b,i,p}$ denotes the brightness temperatures at frequency i with polarization p (p stands for the horizontal or vertical polarization, hereafter referred as h or v); τ_i is the atmospheric transmittances at frequency i; $T_{a,i}^{\uparrow}$ and $T_{a,i}^{\downarrow}$ are the upwelling and downwelling effective radiating temperatures (K); $\varepsilon_{i,p}$ refers to the LSE at frequency i with polarization p; T_s represents the LST (K). This equation neglects cosmic background radiation and atmospheric scattering effects, while inherently accounting for the angular dependence of sensor observation geometry.

In Equation (1), τ_i, $T_{a,i}^{\uparrow}$ and $T_{a,i}^{\downarrow}$, which represent the atmospheric effect, have been proved that can be estimated from PWV [39,40] and described as follows:

$$\tau_i = f(PWV) \quad (2)$$

$$T_{a,i}^{\uparrow} \approx T_{a,i}^{\downarrow} = f\left(PWV^2, PWV\right) \quad (3)$$

Notably, Equations (2) and (3) are established under the zero cloud liquid water (CLW) assumption, which is rationalized by two key considerations. First, Zhou et al. [39] revealed that a CLW uncertainty of 0.1 mm would cause a 1.5 mm bias in PWV retrieval, while actual atmospheric CLW values are generally far below 0.1 mm under conventional atmospheric conditions. Second, Xia et al. [35] verified that CLW exerts a negligible influence on ML-based PWV retrieval, further supporting the feasibility of excluding CLW from the model formulation.

Based on Equations (1)–(3), PWV can be expressed as a function of $T_{b,i}$, ε_i, and T_S:

$$PWV = f(T_{b,i,p}, \varepsilon_{i,p}, T_S) \quad (4)$$

PWV distribution exhibits prominent vertical heterogeneity that is strongly modulated by terrain elevation [34]. Specifically, the effective atmospheric column height for water vapor decreases as altitude increases. Combined with the vertical stratification characteristic of atmospheric water vapor—which is predominantly concentrated in the lower troposphere and gradually diminishes at higher altitudes—high-altitude regions generally present considerably lower PWV contents compared with lowland areas. Land use type also serves as a critical modulating factor for near-surface PWV variations [12]. It regulates surface evapotranspiration partitioning and surface energy balance dynamics, thereby substantially affecting local water vapor accumulation. Furthermore, although latitude, longitude, and DOY do not directly participate in the physical formation of PWV, they provide essential spatiotemporal constraints for machine learning modeling. Given the evident spatial and seasonal regularity of PWV distribution, the incorporation of these auxiliary variables can effectively enhance the retrieval accuracy and generalization capability

of the ML-based model. Accordingly, the comprehensive PWV retrieval relationship can be further formulated as follows:

$$PWV = f(T_{b,i,p}, \varepsilon_{i,p}, T_S, H, LU, Lat, Lon, DOY) \quad (5)$$

where H is DEM, LU denotes the land use types, and Lat and Lon stand for the latitude and longitude, separately.

Equation (5) provides the physical foundation for all-weather land PWV retrieval using passive microwave brightness temperature observations. In this formulation, the variable $T_{b,i,p}$ incorporates not only original single-channel brightness temperature values but also multi-channel brightness temperature indices derived from combinatorial observations. Accurate pixel-scale estimation of LSE remains a challenging task for passive microwave retrieval. Instead of conducting explicit LSE retrieval, this study adopts empirically derived brightness temperature indices as effective surrogate parameters to characterize LSE variations. Meanwhile, several physically meaningful PWV indices summarized from previous studies are also integrated as core model inputs. The quantitative physical mechanisms linking these brightness temperature indices to LSE and PWV variations are detailed in the following sections.

2.2.2. Collection of Input Parameters

Gao et al. [34] proposed single- and multi-variable regression schemes to estimate horizontal LSE based on the polarization ratio (PR, means $T_{b,i,h}/T_{b,i,v}$) for forested and non-forested land types. For forested areas, the 18.7 GHz LSE can be well parameterized via a second-order polynomial function of the 18.7 GHz PR. For non-forested regions, by contrast, a multivariate scheme is more applicable, in which the 18.7 GHz LSE is optimally estimated using the combined PR observations at 6.9 GHz and 18.7 GHz:

$$\varepsilon_{18.7h,fore} = 1.0038 \times PR_{18.7}^2 - 0.1226 \times PR_{18.7} + 0.0799 \quad (6)$$

$$\varepsilon_{18.7h,nonfore} = -1.0482 - 0.5229 \times PR_{6.9} + 2.5255 \times PR_{18.7} \quad (7)$$

where $\varepsilon_{18.7h,fore}$ and $\varepsilon_{18.7h,nonfore}$ represent the horizontal LSE in the forested and non-forested areas, while $PR_{18.7}$ and $PR_{6.9}$ represent the polarization ratios at 18.7 GHz and 6.9 GHz. Nevertheless, the FY-3D/MWRI sensor does not carry a 6.9 GHz observation channel, which restricts the direct application of the original multivariate parameterization scheme. Therefore, only the 18.7 GHz PR and its quadratic term ($PR_{18.7}^2$ and $PR_{18.7}$) adopted as model inputs. Furthermore, the horizontal polarization brightness temperature at 18.7 GHz ($T_{b,18.7h}$), which serves as a fundamental variable in Equation (5), is also incorporated into the input feature set to constrain LSE variations.

Zhou et al. [28] developed a land PWV retrieval algorithm based on passive microwave observations that does not require LST inputs. Their model was established based on three core physical assumptions. First, microwave radiation at two adjacent frequency channels possesses nearly identical penetration capabilities, indicating that the corresponding LST values are equivalent for both channels. Second, the LSE values at these two channels are also approximated to be equal. Based on the radiative transfer equation (Equation (1)) proposed in this study, for two designated channels i and j with consistent LST and LSE conditions, the LST-related term can be isolated and eliminated by transforming and combining the radiative transfer equations of the two channels, yielding a simplified merged equation:

$$T_{b,i} \cdot \tau_j - T_{b,j} \cdot \tau_i + T_{a,j} \cdot \tau_i - T_{a,i} \cdot \tau_j = (1 - \varepsilon)\tau_i \cdot \tau_j (T_{a,i} - T_{a,j}) \quad (8)$$

where ε denotes the unified LSE of the two channels, and T_a represents the upwelling or downwelling effective radiating temperatures of the atmosphere. Previous studies [41] based on simulated data revealed that τ can be expressed by a linear function of PWV, with an R^2 of 0.99. Similarly, other researchers [28,39] found that the two terms $T_{a,j} \cdot \tau_i - T_{a,i} \cdot \tau_j$ and $\tau_i \cdot \tau_j(T_{a,i} - T_{a,j})$ in Equation (8) can also be approximated by linear relationships with PWV, yielding a R^2 of 0.98. Accordingly, this study adopts the third core assumption that both terms vary linearly with PWV. Under this constraint, Equation (8) can be rewritten as:

$$\begin{aligned} T_{b,i}\cdot(\alpha_1\cdot PWV + \alpha_2) - T_{b,j}\cdot(\alpha_3\cdot PWV + \alpha_4) + (\alpha_5\cdot PWV + \alpha_6) \\ = (1-\varepsilon)(\alpha_7\cdot PWV + \alpha_8) \end{aligned} \tag{9}$$

Zhou et al. [28] validated, via simulation experiments, that the horizontally polarized brightness temperatures at 18.7 GHz and 23.8 GHz fully satisfy the above three physical assumptions. On this basis, Equation (9) can be further rearranged to derive the analytical PWV expression:

$$PWV_{zhou} = \frac{\beta_1 + \beta_2\cdot\varepsilon_h + \beta_3\cdot T_{b,18.7h} + \beta_4\cdot T_{b,23.8h}}{\beta_5 + \beta_6\cdot\varepsilon_h + \beta_7\cdot T_{b,18.7h} + \beta_8\cdot T_{b,23.8h}} \tag{10}$$

where βi (i = 1, . . ., 8) are empirical coefficients, and ε_h stands for LSE at horizontal polarization for both the 18.7 GHz and 23.8 GHz channels. In this study, PWV_{zhou} is introduced as one of the model input features. The βi coefficients are assigned according to the calibrated values provided in the original literature [28], and the unified LSE value is fixed at 0.85 for calculation.

Additionally, Du et al. [12] and Jones et al. [37] proposed a brightness temperature index to estimate PWV over land, namely the Microwave Atmospheric Water Vapor Index (MAWVI), which is defined based on the polarization difference ratio of brightness temperatures at 23.8 GHz and 18.7 GHz:

$$MAWVI = \Delta T_{b,23.8} / \Delta T_{b,18.7} \tag{11}$$

$$\Delta T_{b,23.8} = T_{b,23.8v} - T_{b,23.8h} \tag{12}$$

$$\Delta T_{b,18.7} = T_{b,18.7v} - T_{b,18.7h} \tag{13}$$

Nevertheless, Xia et al. [35] demonstrated that MAWVI-based PWV retrievals perform optimally over bare soil but exhibit limitations in vegetated regions, where the polarization difference in LSE is minimal. To mitigate this deficiency and improve retrieval performance across complex land cover types, two auxiliary parameters proposed by Jones et al. [37], including the vegetation transmissivity factor (Fh) and the 18.7 GHz polarization ratio ($PR_{18.7}$) for open water fraction characterization, were integrated into the retrieval framework. The calculation formula for Fh is expressed as:

$$Fh = T_{b,23.8h} / T_{b,18.7h} \tag{14}$$

Accordingly, MAWVI and Fh are incorporated as input features. By combining all the aforementioned physical variables and auxiliary environmental covariates, Equation (5) can be further expanded to establish the final PWV retrieval function:

$$PWV = f\begin{pmatrix} T_{b,18.7h}, PR_{18.7}^2, PR_{18.7}, PWV_{zhou}, \\ MAWVI, Fh, T_s, H, LU, Lat, Lon, DOY \end{pmatrix} \tag{15}$$

For the T_s term in Equation (15), this study adopts the LST parameterization scheme proposed by Holmes et al. [42]. This method retrieves LST solely using the vertically

polarized brightness temperature at 36.5 GHz ($T_{b,36.5h}$), with a reliable estimation error of approximately 3.5 K for low-vegetation regions and 2.5 K for forested areas. Benefiting from its simplicity and robustness, this algorithm has been widely applied in relevant research, with more than 300 citations in the Web of Science. Therefore, $T_{b,36.5h}$ is selected as the input feature to effectively characterize surface thermal conditions for PWV retrieval.

2.2.3. Methodology Development

Equation (15) summarizes a comprehensive set of physical and environmental predictors that jointly dominate land PWV variations based on passive microwave radiative transfer derivation. Nevertheless, explicitly constructing a unified analytical retrieval function from such multi-dimensional and highly nonlinear parameters remains extremely challenging. To address this issue, this study employs ML algorithms to achieve high-precision PWV estimation. Benefiting from the powerful capability of modeling complex nonlinear relationships and fusing multi-source auxiliary variables, ML models are well-suited for heterogeneous land surface PWV retrieval and can effectively improve model accuracy and robustness.

Three typical and high-performance ML algorithms, namely Random Forest (RF), Adaptive Kernel Extreme Learning Machine (AKELM), and Extreme Gradient Boosting (XGBoost), were adopted for PWV retrieval, each possessing unique superiority in nonlinear regression tasks. The RF algorithm exhibits strong adaptability to complex datasets and can automatically capture high-order feature interactions and implicit nonlinear mappings between input variables and PWV. Unlike iterative neural network frameworks, AKELM randomly initializes the weights between the input and hidden layers and analytically solves the output-layer weights without iterative optimization, thereby achieving considerably higher training efficiency than conventional neural networks. By contrast, XGBoost excels at learning intricate nonlinear dependencies between features and PWV, delivering stable and accurate retrieval performance even for datasets with limited or moderate sample sizes.

To determine the optimal retrieval model, systematic hyperparameter tuning was implemented for all three algorithms. For RF, the number of decision trees was tuned from 50 to 500, and the minimum number of leaf samples ranged from 1 to 50. For XGBoost, two core hyperparameters, including the number of estimators (50–1000) and learning rate (0.01–0.3), were independently optimized. Unlike conventional ELM, AKELM can adaptively optimize the number of hidden-layer neurons by stopping the node increment process when reaching the predefined error threshold. In this study, we initialized the number of hidden neurons as 100 and kept the regularization coefficient fixed at 0.1, which facilitates stable parameter initialization and promotes stable model convergence during adaptive learning. All datasets collected in 2018 were used for model training, and a 10-fold cross-validation strategy was adopted to evaluate model performance in terms of R^2 and RMSE.

3. Results and Discussion

3.1. Performances of Three ML Methodology

Table 1 summarizes the optimal retrieval accuracy (RMSE and R^2) of the three machine learning algorithms after hyperparameter optimization, demonstrating clear performance discrepancies among the adopted models. The optimized RF model, configured with 50 decision trees and a minimum leaf sample size of 1, yielded an RMSE of 4.86 mm and an R^2 of 0.71. The XGBoost model adopted 200 estimators and a learning rate of 0.1 and achieved the best performance, with an RMSE of 3.80 mm and an R^2 of 0.91. By comparison, the AKELM model produced relatively larger estimation errors, with an RMSE of 5.90 mm

and an R^2 of 0.56. Overall, XGBoost outperforms RF and AKELM in both error magnitude and fitting capabilities. Accordingly, the XGBoost algorithm is selected as the optimal model for subsequent PWV prediction and quantitative evaluation in this study.

Table 1. RMSE and R^2 of PWV estimations using three ML methods.

Methodology	Best Parameters	RMSE (mm)	R^2
RF	Trees: 50 Leaves: 1	4.86	0.71
XGBoost	Trees: 200 Learning Rate: 0.1	3.80	0.91
AKELM	Hidden Neurons: 100 Regularization Coefficient: 0.1	5.90	0.56

3.2. Performance Comparison Between the Proposed Model and the Xia's Model

Xia et al. [35] developed a ML-based retrieval algorithm for land-based PWV, including the MAVWI, Ts, the ratio of polarization differences at 36.5 GHz and 89 GHz (to characterize the impact of CLW), H, satellite orbit type, Lat, Lon, Fh, $PR_{18.7}$, LU, and DOY. Experimental validation shows that the RMSE over the study area is 3.1 mm. Further analysis using the SHapley Additive exPlanations method reveals that the CLW-related index exerts the least influence on PWV retrieval results. The Xia model [35] was selected as the primary benchmark for two reasons: (i) it is a representative of the same methodological family (physically constrained machine learning land PWV retrieval) and (ii) its feature set substantially overlaps with that of the proposed method (MAWVI, Ts, H, Fh, $PR_{18.7}$, LU, Lat, Lon, DOY).

Compared with the Xia model, the present study constructs new characteristic estimators for LSE and PWV on the basis of refined radiative transfer assumptions, so as to further improve the estimation accuracy of land surface PWV (denoted as the Xia model and the proposed Zhou-based model). To quantitatively and objectively compare the performance of the two physically constrained retrieval schemes, high-precision SuomiNet GPS-PWV observations were adopted as the ground truth reference. Specifically, all 2018 datasets were used for model training and 10-fold cross-validation to evaluate in-sample accuracy, while independent 2019 datasets were employed as a test set to verify the out-of-sample generalization capability of each model.

We first establish the standalone physical retrieval as the baseline to quantify the contribution of the machine learning component. Figure 2 presents a direct comparison of the standalone PWV_zhou retrievals (the analytical model of Equation (10) without ML) against SuomiNet GPS observations. To ensure the fairest baseline, the polynomial relationship of the standalone index was recalibrated on the 2018 training data; after recalibration, the standalone PWV_zhou achieves an RMSE of approximately 8.69 mm with an R^2 of only 0.34 in the 2019 validation—far above the accuracy of the complete physically constrained ML model (R^2 = 0.86, RMSE = 4.90 mm, XGBoost). This comparison demonstrates that the standalone physical index alone cannot fully decouple the surface and atmospheric contributions, and that the machine learning component, anchored by the physical constraints, reduces the RMSE by nearly 43.6% relative to the recalibrated standalone baseline.

The cross-validation results in Figure 3a,c indicate that the two models achieve comparable accuracy during the training stage. The Xia model produces an R^2 of 0.90, an RMSE of 4.09 mm, and a mean absolute error (MAE) of 2.93 mm, whereas the proposed Zhou-based model achieves a slightly better performance, with an R^2 of 0.91, an RMSE of 3.80 mm, and a MAE of 2.71 mm. More significant performance differences are observed in the inde-

pendent validation using 2019 datasets, as illustrated in Figure 3b,d. The proposed model increases the R^2 from 0.80 to 0.86 (a 7.5% improvement), reduces the RMSE from 5.87 mm to 4.90 mm (a 16.5% reduction), and decreases the MAE from 4.34 mm to 3.63 mm (a 16.4% reduction). These consistent improvements in multiple error metrics demonstrate that the newly proposed model possesses superior retrieval accuracy and stronger spatiotemporal generalization robustness compared with the conventional Xia model.

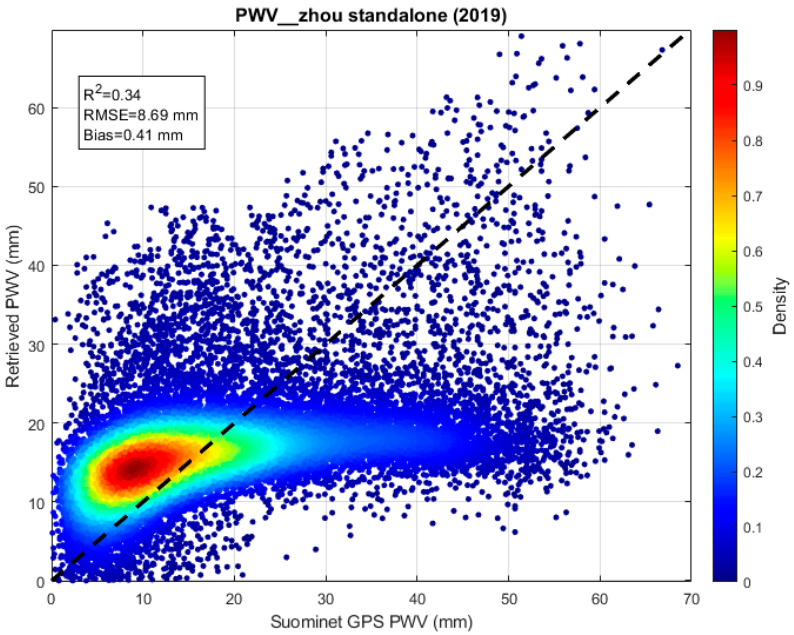


Figure 2. Scatter density maps of the recalibrated standalone PWV_zhou versus GPS PWV (2019).

3.3. Feature Importance Values Analysis

Considering the potential collinearity among the input variables, we further performed a variance inflation factor (VIF) analysis and a data-driven feature selection (Table 2). Several features exhibit VIF values well above the commonly used threshold of 10, including the homologous 18.7 GHz terms ($PR^2_{18.7}$: 411; $PR_{18.7}$: 374; $T_{b,18.7h}$: 51) and, to a lesser extent, T_s (11.4). This is expected, since $PR^2_{18.7}$ is the squared term of $PR_{18.7}$ and both are derived from the same 18.7 GHz brightness temperatures. By contrast, the two water-vapor-related indices, PWV_zhou and MAWVI, are statistically independent (VIF = 1). Three configurations were then compared by 10-fold cross-validation on the 2018 training set: (A) all 12 features (RMSE = 3.80 mm); (B) greedy backward elimination of the VIF > 10 candidates with a +0.05 mm tolerance, which removes $T_{b,18.7h}$ and $PR^2_{18.7}$ while retaining $PR_{18.7}$ and T_s (10 features, RMSE = 3.83 mm); and (C) one-off removal of all VIF > 10 features (8 features, RMSE = 4.01 mm). Scheme C confirms that a purely VIF-driven elimination incurs a measurable accuracy loss, because several high-VIF features still carry non-redundant information that the tree-based model can exploit. Since the A–B difference (0.03 mm) lies within the cross-validation variability, the full 12-feature model (Scheme A) is retained as the primary model, consistent with the radiative-transfer derivation of Equation (15), with Scheme B serving solely as a robustness check.

The comparison and validation results above confirm that the proposed Zhou-based model achieves more accurate PWV retrieval than the Xia model at both the training (10-fold cross-validation) and the independent validation stages. To further investigate the mechanism underlying this improvement, we quantitatively assessed the contributions of the newly introduced LSE and PWV estimation operators using SHAP analysis based on the exact TreeSHAP algorithm.

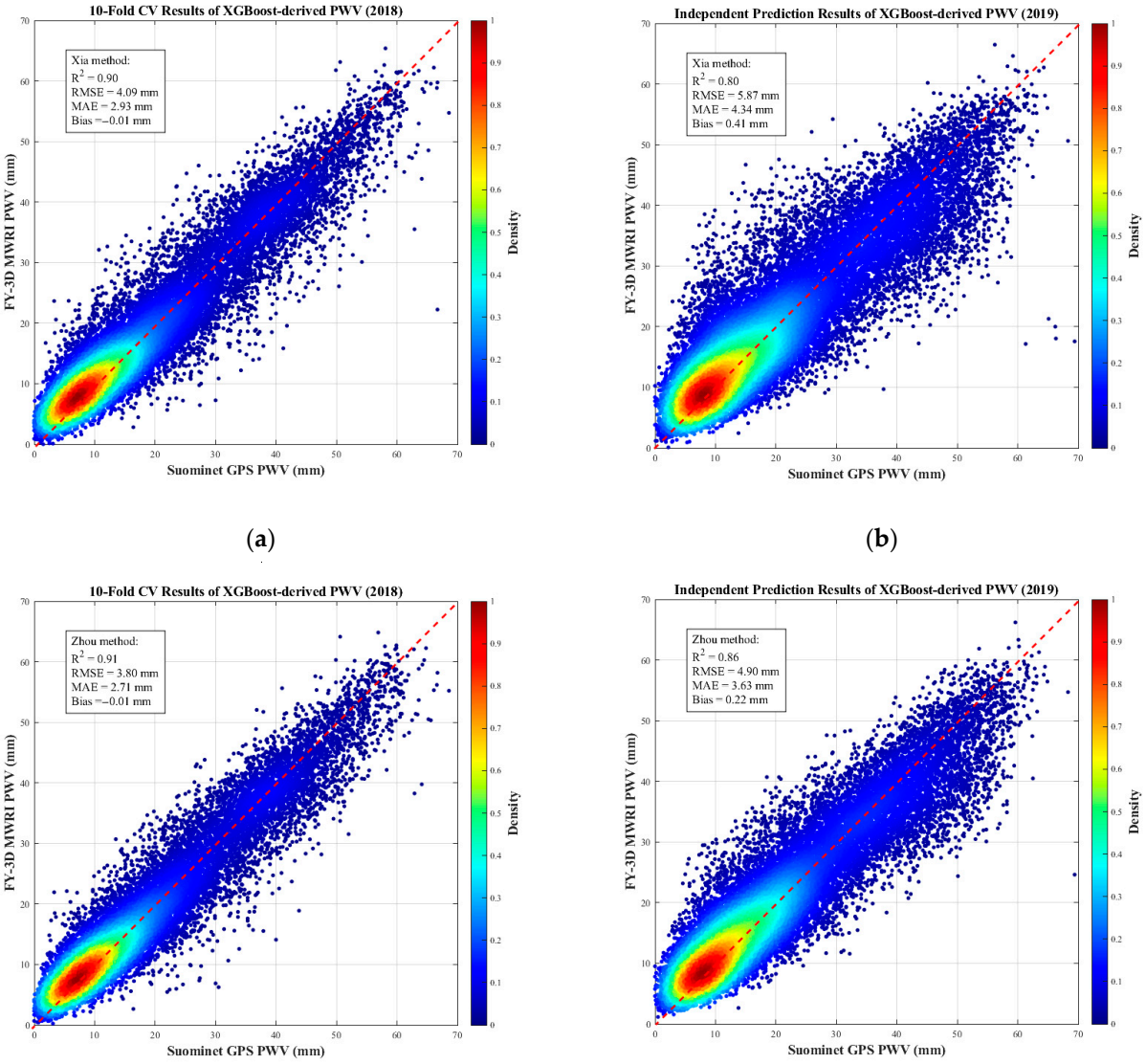
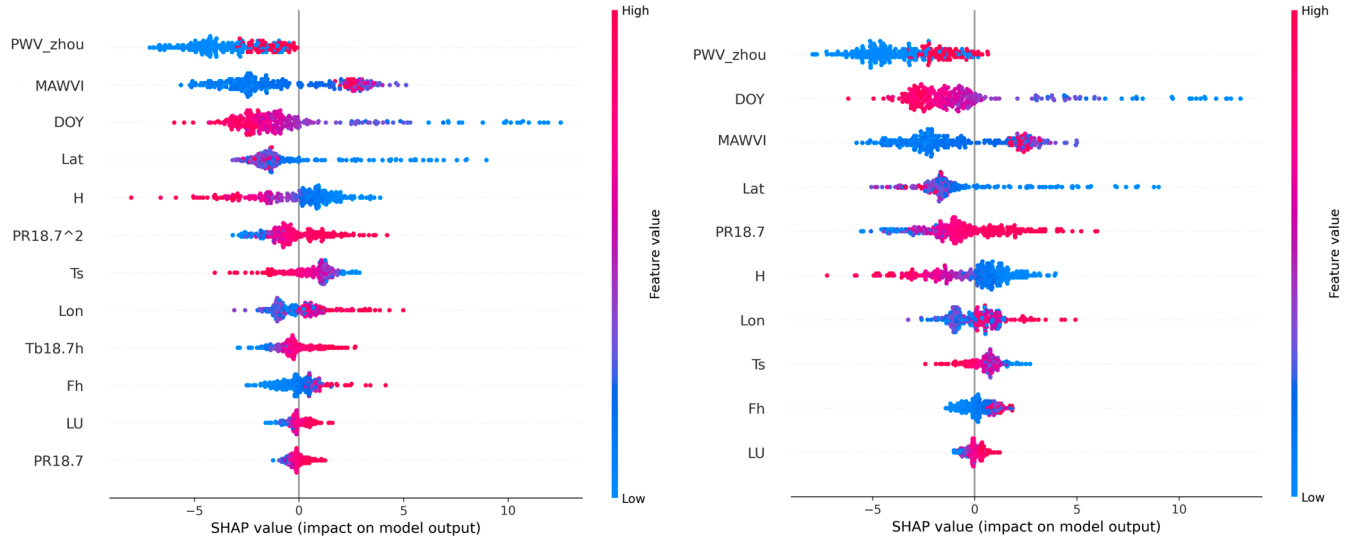


Figure 3. Scatter density maps of XGBoost-retrieved PWV versus GPS-derived PWV: (**a**) Scatter density plot of the Xia model with 10-fold cross-validation on the 2018 training set; (**b**) scatter density plot of the Xia model for independent validation on the 2019 test set; (**c**) scatter density plot of the Zhou model with 10-fold cross-validation on the 2018 training set; (**d**) scatter density plot of the Zhou model for independent validation on the 2019 test set.

Table 2. The variance inflation factor of 12 input data.

Input Data	VIF
$T_{b,18.7h}$	51
$PR^2_{18.7}$	411
$PR_{18.7}$	374
PWV_{zhou}	1
MAWVI	1
Fh	5.2
T_s	11.4
H	1.3
LU	1.3
Lat	2.3
Lon	1.7
DOY	1.2

To keep the feature attribution consistent with the model actually used for retrieval, the analysis is presented for the primary 12-feature model (Scheme A) in Figure 4a, with the reduced 10-feature model (Scheme B) provided in Figure 4b as a robustness check. In Figure 4a, the overall descending ranking is PWV_{zhou} (labeled PWV_zhou in the figure) > MAWVI > DOY > Lat > H > $PR^2_{18.7}$ (labeled PR18.7^2 in the figure) > Ts > Lon > $T_{b18.7h}$ (labeled Tb18.7h in the figure) > Fh > LU > $PR_{18.7}$ (labeled PR18.7 in the figure). Notably, the newly added PWV estimation operator (PWV_{zhou}) ranks first among all twelve input features, while PR18.7^2, a core parameter of the LSE estimation operator, ranks sixth. Both therefore rank among the most influential inputs of the Zhou model, indicating that the two newly introduced operators play a leading role in improving the retrieval performance.



(**a**) (**b**)

Figure 4. SHAP feature importance for the primary 12-feature model (**a**) and the reduced 10-feature model (**b**).

Because the strong collinearity within the 18.7 GHz group makes the individual importance of $PR_{18.7}$ and its squared term partly interchangeable, $PR_{18.7}$ is ranked last in the full 12-feature model. When the collinear squared term is removed (Figure 4b, 10-feature model), the importance of $PR_{18.7}$ rises from twelfth to fifth, whereas the set of leading contributors—PWV_{zhou}, MAWVI, DOY and Lat—remains unchanged; only the relative ordering of MAWVI and DOY is slightly reversed between the two configurations. This confirms that the attribution of the Zhou model is robust to the treatment of collinearity, and that the low individual ranking of $PR_{18.7}$ reflects its redundancy with $PR^2_{18.7}$ rather than a lack of physical information.

Further analysis indicates that conventional auxiliary variables, including DEM, latitude, and longitude, mainly characterize the macroscopic spatiotemporal patterns of PWV distribution but cannot directly reflect the intrinsic physical coupling between water vapor variations and microwave radiative transfer processes. In contrast, the newly introduced PWV and LSE estimation operators effectively compensate for this deficiency. The integration of these physically constrained parameters complements traditional spatiotemporal covariates, enabling the model to simultaneously capture large-scale PWV spatiotemporal variations and subtle radiative transfer signatures. This dual constraint significantly enhances model adaptability and retrieval accuracy under complex surface heterogeneity and diverse atmospheric water vapor conditions.

3.4. PWV Distribution

Figure 5 illustrates the all-weather spatial distributions of PWV across the study area on four representative dates in 2019 (15 January, 15 April, 15 July, and 15 October). The left column presents in situ PWV measurements from SuomiNet GPS stations, the middle column displays the PWV retrievals derived from the proposed FY-3D/MWRI algorithm, and the right column shows the corresponding AMSR2 PWV product distributions. Obvious seasonal variations in PWV are captured consistently by all datasets. In January, mid- and high-latitude regions of North America exhibit generally low PWV values, while relatively high PWV concentrations are concentrated near 20°N. In April, PWV levels increase significantly around 40°N, with maximum values ranging from 40 mm to 50 mm over southeastern Mexico, Panama, and surrounding areas. July corresponds to the humid season with widespread high PWV across the entire study domain, especially within the longitude band of 80°W–100°W. High PWV values of 40–50 mm dominate the central and eastern United States, the Gulf of Mexico coast, and the Caribbean coastal regions. In October, PWV decreases over the central U.S., while the eastern and southern United States, the Gulf of Mexico coastline, and the Caribbean rim still maintain relatively high PWV values of 40–50 mm.

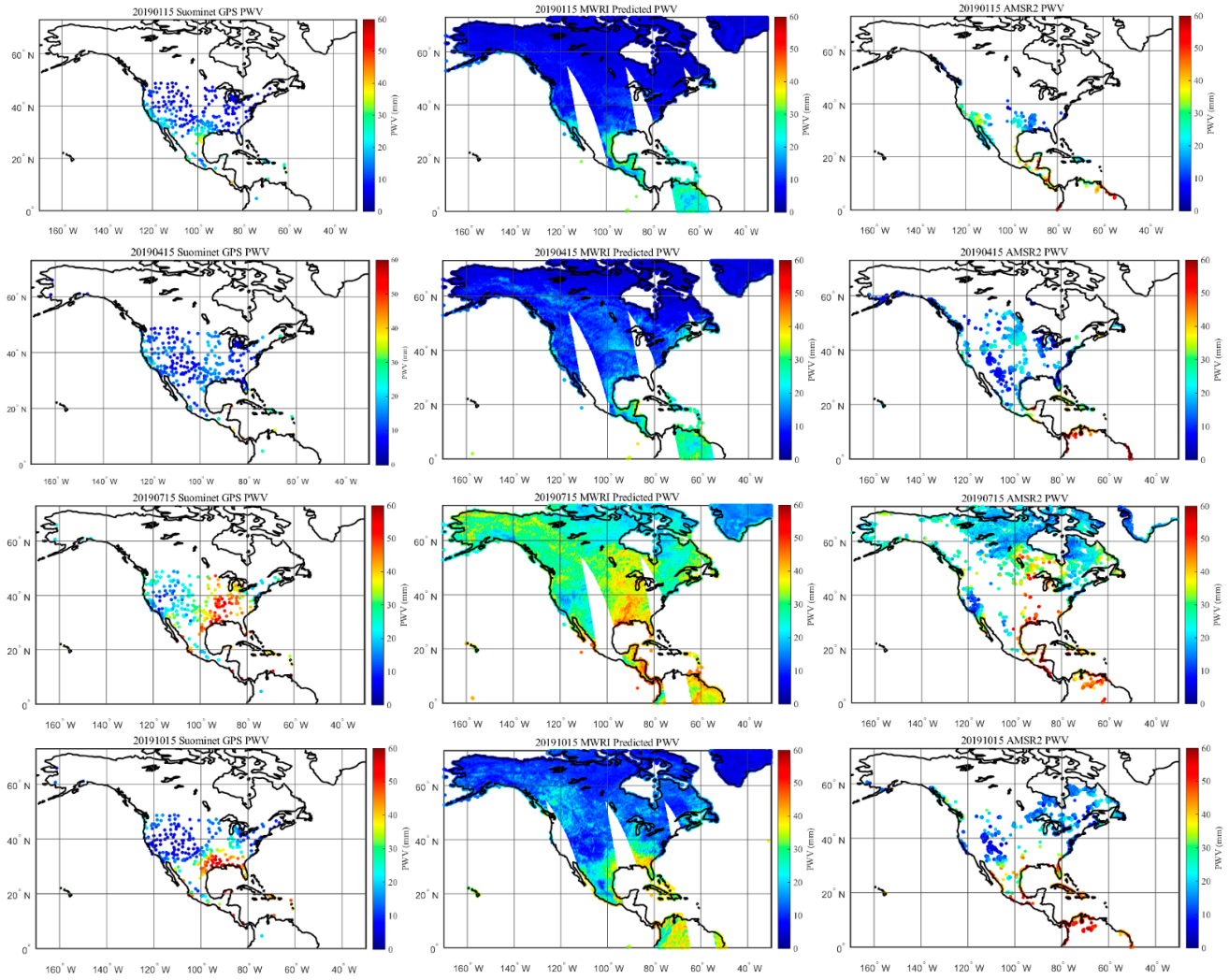


Figure 5. Distribution and scatter maps of PWV over the study area on the 15th day of January, April, July, and October 2019. (The left column depicts observations from SuomiNet GPS stations, the middle column illustrates retrieval results from FY-3D MWRI, and the right column shows the spatial distribution of AMSR2 PWV.).

Notably, although the AMSR2 PWV product shows good consistency with SuomiNet GPS observations, it suffers from severe limitations in spatial coverage. Specifically, valid AMSR2 pixels on 15 January are even sparser than discrete GPS station records, which greatly restricts its applicability for fine-scale and continuous spatial PWV analysis. In comparison, the proposed FY-3D/MWRI retrieval method achieves reliable accuracy while delivering superior spatial continuity and full coverage, demonstrating prominent advantages for all-weather regional PWV monitoring.

Figure 6 further quantitatively compares the retrieval accuracy of the proposed FY-3D/MWRI-based PWV algorithm and the MODIS MYD05_L2 product against SuomiNet GPS reference data across four seasonal dates. The scatter plots take GPS-observed PWV as the horizontal axis and satellite-retrieved PWV as the vertical axis, where red circles represent MWRI retrievals and blue triangles denote MODIS MYD05_L2 estimates. The overall results demonstrate that the proposed algorithm outperforms the MODIS infrared PWV product under diverse seasonal conditions.

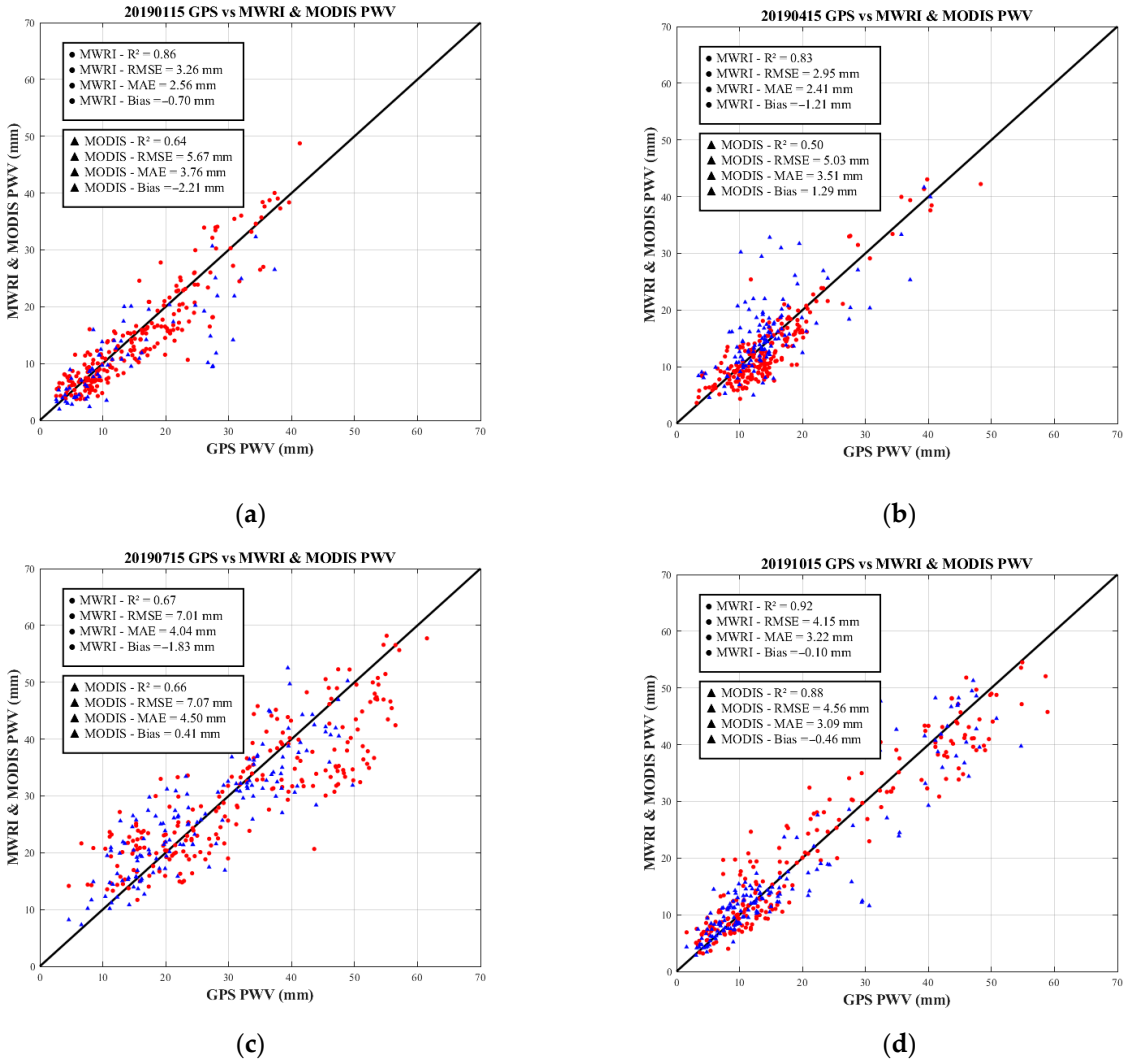


Figure 6. Scatter plots for the comparison of PWV from GPS site observations, MWRI retrievals, and the MODIS PWV product. The horizontal axis represents PWV from GPS stations, and the vertical axis denotes MWRI-retrieved PWV and MODIS PWV, respectively: (**a**) 15 January; (**b**) 15 April; (**c**) 15 July; (**d**) 15 October.

In winter and spring (15 January and 15 April), the proposed model achieves remarkable accuracy improvements. The MWRI retrievals yield RMSE values of 3.26 mm and

2.95 mm with R^2 values of 0.86 and 0.83, respectively. By contrast, the MODIS MYD05 product presents considerably larger errors, with RMSE values of 5.67 mm and 5.03 mm and lower R^2 values of 0.64 and 0.50. In summer and autumn (15 July and 15 October), the proposed method still exhibits stable and superior performance. The MWRI-based PWV retrievals obtain RMSE values of 7.01 mm and 4.15 mm and R^2 values of 0.67 and 0.92, while the corresponding MODIS product yields slightly higher RMSEs (7.07 mm and 4.56 mm) and lower R^2 (0.66 and 0.88). These cross-product comparisons fully validate the reliability and superiority of the proposed physically constrained ML-based PWV retrieval approach.

4. Discussion

4.1. Physical Interpretation of the Results

The physically derived indices PWV_{zhou} and MAWVI, obtained by eliminating the surface terms from the radiative transfer equations, are consistently identified as leading contributors in both the twelve- and ten-feature SHAP analyses (Section 3.3). This indicates that the analytical derivation of the LSE/PWV characteristic operators from the radiative transfer equations—rather than the mere stacking of physically meaningful predictors—is the key factor behind the accuracy improvement over the conventional model.

4.2. Comparison with Existing Methods

Compared with the Xia model (R^2 = 0.80, RMSE = 5.87 mm on the 2019 independent validation), the proposed method raises the R^2 to 0.86 and reduces the RMSE by 16.5%, to 4.90 mm. Relative to the MODIS MYD05 product, it achieves a lower RMSE on all four representative dates and, more importantly, delivers spatially continuous all-weather retrievals where the infrared product is compromised by cloud-induced gaps (Figure 5). Over the clear-sky common pixels, the accuracy gap narrows in summer (Section 3.4)—a behaviour that reflects the physical saturation of the microwave signal at high PWV rather than a deficiency of the machine learning framework.

5. Conclusions

The primary challenge in all-weather retrieval of land PWV using passive microwave remote sensing lies in the complexity and uncertainty of LSE over heterogeneous land surfaces. This study develops a machine learning-based all-weather land PWV retrieval method constrained by physical models. By introducing reasonable assumptions and mathematical transformations to the radiative transfer equation, we construct physically meaningful PWV indices. Building on previous research, multiple PWV brightness temperature indices, as well as LST and LSE indicators, are incorporated as input variables, leveraging the strong capability of machine learning algorithms to handle highly nonlinear relationships and improve the accuracy of land PWV retrieval.

Most existing physically constrained machine learning retrieval methods fail to select input predictors strictly in accordance with stepwise derivation of the radiative transfer equation, which limits further accuracy promotion and weakens the physical interpretability of retrieval models. To fill this research gap, this study optimizes the input variable selection strategy. In terms of parameter configuration, physically derived PWV characteristic factors are introduced into the input system as core predictive variables, forming one of the major innovations of this research. Meanwhile, we screen out representative indicative indices to characterize LST and LSE based on physical theories. Furthermore, widely recognized microwave brightness temperature indices closely correlated with PWV, such as MAWVI and Fh, together with auxiliary environmental factors including land use type, elevation, geographic coordinates, and Julian day, are adopted to impose solid physical constraints on the established retrieval algorithm.

For the selection of the machine learning model, three algorithms—RF, XGBoost, and AKELM—are compared. The results show that XGBoost outperforms the other two methods in terms of both R^2 and RMSE and is thus selected as the retrieval model in this study. The model is trained using 2018 data and validated using 2019 data, with SuomiNet GPS observations serving as the ground truth. The results indicate that the R^2 of the training set reaches 0.91, while the R^2 of the validation set is 0.86. Compared with the model proposed by Xia et al., the performance of the Zhou model proposed in this study is significantly improved: R^2 increases from 0.80 (Xia model) to 0.86 (Zhou model), marking a 7.5% accuracy enhancement; RMSE decreases from 5.87 mm to 4.90 mm, representing a 16.5% accuracy improvement. The importance ranking of input parameters reveals that the physically derived PWV index proposed in this study ranks first, followed by the MAWVI, which explains why the proposed model outperforms Xia et al.'s model in accuracy. Further daily validation is conducted on 15 January, 15 April, 15 July, and 15 October 2019, with comparisons against the corresponding PWV products from AMSR2 and MODIS. The results show that the AMSR2 PWV product has too few valid pixels to demonstrate the advantages of passive microwave remote sensing in all-weather observation. Meanwhile, the accuracy of the MODIS PWV product is consistently lower than that of the algorithm proposed in this study.

During model construction, this study conducts progressive derivation and formula transformation based on the radiative transfer equation, selects appropriate indicative indices, and fully incorporates existing research achievements in this field. Multiple validation experiments demonstrate that the proposed method outperforms conventional approaches in terms of retrieval accuracy and model robustness, and also achieves higher accuracy compared with mainstream global land PWV products. However, this study is subject to regional limitations, as the model was trained and validated solely on data from the SuomiNet North America network. Future work will incorporate observations from a wider distribution of sites to retrain the model and thereby improve its accuracy over a broader area. This study provides a reliable new paradigm for all-weather land PWV retrieval via microwave remote sensing under physical model constraints.

Author Contributions: Conceptualization, F.-C.Z.; methodology, F.-C.Z. and G.L.; software, F.-C.Z.; validation, X.S.; investigation, F.-C.Z.; data curation, F.-C.Z.; writing—original draft preparation, F.-C.Z.; writing—review and editing, S.C.; visualization, F.-C.Z.; supervision, X.H.; funding acquisition, F.-C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of China grants U2542208, U2242211, and the China Meteorological Administration Innovation and Development Project (Grant No. CMA2026ZD01).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Yang, J.; Yao, Y.; Xu, C.; Cao, N. Analysis of the Correlation Between PWV and Actual Rainfall. *J. Geomat.* **2016**, *41*, 18–21. (In Chinese)
2. Trenberth, K.E.; Fasullo, J.; Smith, L. Trends and variability in column-integrated atmospheric water vapor. *Clim. Dyn.* **2005**, *24*, 741–758. [CrossRef]
3. Li, X.; Dick, G.; Lu, C.; Ge, M.; Nilsson, T.; Ning, T.; Wickert, J.; Schuh, H. Multi-GNSS meteorology: Real-time retrieving of atmospheric water vapor from BeiDou, Galileo, GLONASS, and GPS observations. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6385–6393. [CrossRef]

4. Hanssen, R.F.; Weckwerth, T.M.; Zebker, H.A.; Klees, R. High-resolution water vapor mapping from interferometric radar measurements. *Science* **1999**, *283*, 1297–1299. [CrossRef] [PubMed]
5. Yao, Y.; Shan, L.; Zhao, Q. Establishing a method of short-term rainfall forecasting based on GNSS-derived PWV and its application. *Sci. Rep.* **2017**, *7*, 12465. [CrossRef] [PubMed]
6. Berezin, I.A.; Timofeyev, Y.M.; Virolainen, Y.A.; Volkova, K.A. Comparison of ground-based microwave measurements of precipitable water vapor with radiosounding data. *Atmos. Ocean. Opt.* **2016**, *29*, 274–281. [CrossRef]
7. Torres, B.; Cachorro, V.E.; Toledano, C.; Ortiz de Galisteo, J.P.; Berjon, A.; de Frutos, A.M.; Bennouna, Y.; Laulainen, N. Precipitable water vapor characterization in the Gulf of Cadiz region (southwestern Spain) based on Sun photometer, GPS, and radiosonde data. *J. Geophys. Res. Atmos.* **2010**, *115*, D18103. [CrossRef]
8. Froidevaux, M.; Higgins, C.W.; Simeonov, V.; Ristori, P.; Pardyjak, E.; Serikov, I.; Calhoun, R.; van den Bergh, H.; Parlange, M.B. A Raman lidar to measure water vapor in the atmospheric boundary layer. *Adv. Water Resour.* **2013**, *51*, 345–356. [CrossRef]
9. Bevis, M.; Businger, S.; Herring, T.A.; Rocken, C.; Anthes, R.A.; Ware, R.H. GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System. *J. Geophys. Res. Atmos.* **1992**, *97*, 15787–15801. [CrossRef]
10. Gui, K.; Che, H.Z.; Chen, Q.L.; Zeng, Z.L.; Liu, H.Z.; Wang, Y.Q.; Zheng, Y.; Sun, T.Z.; Liao, T.T.; Wang, H.; et al. Evaluation of radiosonde, MODIS-NIR-Clear, and AERONET precipitable water vapor using IGS ground-based GPS measurements over China. *Atmos. Res.* **2017**, *197*, 461–473. [CrossRef]
11. Liu, H.; Tang, S.; Hu, J.; Zhang, S.; Deng, X. An improved physical split-window algorithm for precipitable water vapor retrieval exploiting the water vapor channel observations. *Remote Sens. Environ.* **2017**, *194*, 366–378. [CrossRef]
12. Du, J.; Kimball, J.S.; Jones, L.A. Satellite microwave retrieval of total precipitable water vapor and surface air temperature over land from AMSR2. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2520–2531. [CrossRef]
13. Gao, B.C.; Kaufman, Y.J. Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels. *J. Geophys. Res. Atmos.* **2003**, *108*, 4389. [CrossRef]
14. Wang, L.; Hu, X.Q.; Xu, N.; Chen, L. Water Vapor Retrievals from Near-infrared Channels of the Advanced Medium Resolution Spectral Imager Instrument onboard the Fengyun-3D Satellite. *Adv. Atmos. Sci.* **2021**, *38*, 1351–1366. [CrossRef]
15. Hu, J.Y.; Tang, S.H.; Liu, H.L.; Min, M. An Operational Precipitable Water Vapor Retrieval Algorithm for Fengyun-2F/VISSR Using a Modified Three-Band Physical Split-Window Method. *J. Meteorol. Res.* **2019**, *33*, 276–288. [CrossRef]
16. Xie, Y.Q.; Li, Z.Q.; Hou, W.Z.; Guang, J.; Ma, Y.; Wang, Y.Y.; Wang, S.H.; Yang, D. Validation of FY-3D MERSI-2 Precipitable Water Vapor (PWV) Datasets Using Ground-Based PWV Data from AERONET. *Remote Sens.* **2021**, *13*, 3246. [CrossRef]
17. Deeter, M.N. A new satellite retrieval method for precipitable water vapor over land and ocean. *Geophys. Res. Lett.* **2007**, *34*, L02815. [CrossRef]
18. Schluessel, P.; Emery, W.J. Atmospheric water vapour over oceans from SSM/I measurements. *Int. J. Remote Sens.* **1990**, *11*, 753–766. [CrossRef]
19. Wentz, F.J.; Meissner, T. *AMSR Ocean Algorithm, Version 2, Algorithm Theoretical Basis Document*; Tech. Rep. 121599A-1; Remote Sensing Systems: Santa Rosa, CA, USA, 2000.
20. Basili, P.; Bonafoni, S.; Mattioli, V.; Ciotti, P.; Pierdicca, N. Mapping the atmospheric water vapor by integrating microwave radiometer and GPS measurements. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1657–1665. [CrossRef]
21. Wentz, F.J. A well-calibrated ocean algorithm for Special Sensor Microwave Imager. *J. Geophys. Res.* **1997**, *102*, 8703–8718. [CrossRef]
22. Mangalsinh, R.N.; Varma, A.K. Improved Algorithms for the Measurement of Total Precipitable Water and Cloud Liquid Water from SARAL Microwave Radiometer Observations. *Mar. Geod.* **2019**, *42*, 367–381. [CrossRef]
23. Boukabara, S.A.; Garrett, K.; Chen, W. Global coverage of total precipitable water using a microwave variational algorithm. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3608–3621. [CrossRef]
24. Prakash, S.; Norouzi, H.; Azarderakhsh, M.; Blake, R.; Prigent, C.; Khanbilvardi, R. Estimation of consistent global microwave land surface emissivity from AMSR-E and AMSR2 observations. *J. Appl. Meteorol. Climatol.* **2018**, *57*, 907–919. [CrossRef]
25. Prigent, C.; Liang, P.; Tian, Y.; Aires, F.; Moncet, J.-L.; Boukabara, S.A. Evaluation of modeled microwave land surface emissivities with satellite-based estimates. *J. Geophys. Res. Atmos.* **2015**, *120*, 2706–2718. [CrossRef]
26. Weng, F.Z.; Liu, Q.H. Satellite data assimilation in numerical weather prediction models. Part I: Forward radiative transfer and Jacobian modeling in cloudy atmospheres. *J. Atmos. Sci.* **2003**, *60*, 2633–2646. [CrossRef]
27. Karbou, F.; Gérard, É.; Rabier, F. Microwave land emissivity and skin temperature for AMSU-A and -B assimilation over land. *Q. J. R. Meteorol. Soc.* **2006**, *132*, 2333–2355. [CrossRef]
28. Zhou, F.C.; Song, X.; Leng, P.; Wu, H.; Tang, B.H. An Algorithm for Retrieving Precipitable Water Vapor over Land Based on Passive Microwave Satellite Data. *Adv. Meteorol.* **2016**, *2016*, 4126393. [CrossRef]
29. Kazumori, M. Precipitable Water Vapor Retrieval Over Land from GCOM-W/AMSR2 and its Application to Numerical Weather Prediction. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6663–6666.

30. Gao, H.; Fu, R.; Dickinson, R.E.; Negron Juarez, R.I. A Practical Method for Retrieving Land Surface Temperature from AMSR-E Over the Amazon Forest. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 193–199. [CrossRef]
31. Ji, D.; Shi, J. Water Vapor Retrieval Over Cloud Cover Area on Land Using AMSR-E and MODIS. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3105–3116. [CrossRef]
32. Ji, D.; Shi, J.; Xiong, C.; Wang, T.; Zhang, Y. A Total Precipitable Water Retrieval Method Over Land Using the Combination of Passive Microwave and Optical Remote Sensing. *Remote Sens. Environ.* **2017**, *191*, 313–327. [CrossRef]
33. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
34. Gao, Z.; Jiang, N.; Xu, Y.; Xu, T.; Liu, Y. Precipitable Water Vapor Retrieval Over Land from GCOM-W/AMSR2 Based on a New Integrated Method. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 51384. [CrossRef]
35. Xia, X.; Fu, D.; Shao, W.; Jiang, R.; Wu, S.; Zhang, P.; Yang, D.; Xia, X. Retrieving Precipitable Water Vapor Over Land from Satellite Passive Microwave Radiometer Measurements Using Automated Machine Learning. *Geophys. Res. Lett.* **2023**, *50*, e2023GL105197. [CrossRef]
36. Jiang, N.; Xu, Y.; Xu, T.; Li, S.; Gao, Z. Land Water Vapor Retrieval for AMSR2 Using a Deep Learning Method. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5803011. [CrossRef]
37. Jones, L.A.; Ferguson, C.R.; Kimball, J.S.; Zhang, K.; Chan, S.T.K.; McDonald, K.C.; Njoku, E.G.; Wood, E.F. Satellite Microwave Remote Sensing of Daily Land Surface Air Temperature Minima and Maxima from AMSR-E. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 111–123. [CrossRef]
38. Ware, R.H.; Fulker, D.W.; Stein, S.A.; Anderson, D.N.; Avery, S.K.; Clark, R.D.; Droegemeier, K.K.; Kuettner, 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]
39. Zhou, F.C.; Li, Z.L.; Wu, H.; Tang, B.H.; Tang, R.L.; Song, X.N.; Yan, G.J. Retrieving K-Band Instantaneous Microwave Land Surface Emissivity Based on Passive Microwave Brightness Temperature and Atmospheric Precipitable Water Vapor Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *12*, 5608–5617. [CrossRef]
40. Han, X.J.; Duan, S.B.; Li, Z.L. Atmospheric correction for retrieving ground brightness temperature at commonly-used passive microwave frequencies. *Opt. Express* **2017**, *25*, A36–A46. [CrossRef] [PubMed]
41. Liu, Z.L.; Wu, H.; Tang, B.H.; Qiu, S.; Li, Z.L. Atmospheric Corrections of Passive Microwave Data for Estimating Land Surface Temperature. *Opt. Express* **2013**, *21*, 15654–15663. [CrossRef] [PubMed]
42. Holmes, T.R.H.; De Jeu, R.A.M.; Owe, M.; Dolman, A.J. Land Surface Temperature from Ka Band (37 GHz) Passive Microwave Observations. *J. Geophys. Res. Atmos.* **2009**, *114*, D04113. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.