



Exploring the capability of surface-observed spectral irradiance for remote sensing of precipitable water vapor amount under all-sky conditions

Pradeep Khatri¹, Tamio Takamura², and Hitoshi Irie²

¹Department of Science and Engineering for Sustainable Innovation, Faculty of Science and Engineering, Soka University, Hachioji-shi, Tokyo, Japan

²Center for Environmental Remote Sensing, Chiba University, Chiba, Japan

Correspondence: Pradeep Khatri (pradeep@soka.ac.jp)

Received: 20 August 2025 – Discussion started: 17 October 2025

Revised: 30 November 2025 – Accepted: 15 December 2025 – Published: 13 January 2026

Abstract. Precipitable water vapor (PWV) is a key component of Earth's climate and hydrological systems, yet its accurate and continuous observation under varying sky conditions remains challenging. This study demonstrates the strong potential of surface-based spectral irradiance measurements for PWV retrieval across a range of atmospheric conditions using deep neural network (DNN) models trained on water vapor absorption bands. Global, direct, and diffuse spectral irradiances observed at water vapor absorption bands of 929.0–997.3, 800.9–840.5, and 708.1–744.6 nm by a spectroradiometer (MS-700; EKO Instruments Co., Ltd., Japan) equipped with a rotating shadow-band system were used as test data, while PWV observed by a microwave radiometer (MP-1500; Radiometrics Corporation, USA) served as reference data for model training and validation. Models incorporating global, direct, and diffuse irradiances achieved the highest accuracy, exhibiting minimal errors and closely capturing seasonal PWV variations. Notably, even models using only global irradiance – an easier and more accessible measurement – maintained high predictive performance, with low errors and robust seasonal tracking. In contrast, models trained solely on clear-sky direct irradiance with substantially fewer samples than those under all-sky conditions showed relatively higher errors and weaker generalization, underscoring the importance of data volume and diversity in DNN models. These results highlight the effectiveness of spectral irradiance-based approaches for continuous PWV estimation across a range of atmospheric conditions. Future research should incorporate additional spectral bands sen-

sitive to constituents like aerosols and ozone to expand retrieval capability.

1 Introduction

Atmospheric water vapor plays an essential and multifaceted role in the Earth's climate system. It is not only a key driver of weather phenomena such as cloud formation and precipitation, but also a central player in atmospheric thermodynamics and energy transport (Held and Soden, 2000; Trenberth et al., 2009). As a greenhouse gas, it absorbs and emits infrared radiation, significantly enhancing the Earth's natural greenhouse effect. Although it is easily phase-changeable gas compared to carbon dioxide or methane, its abundance and strong infrared absorption characteristics make it the most influential greenhouse gas in terms of its contribution to natural greenhouse effect. Water vapor contributes between 41 % and 67 % of the total natural greenhouse effect, far surpassing the contributions of other gases (Kiehl and Trenberth, 1997; Schmidt et al., 2010). However, unlike carbon dioxide and methane, water vapor is primarily a feedback rather than a direct radiative forcing, as its concentration is controlled by temperature: as global temperatures rise due to anthropogenic greenhouse gases, more water evaporates, increasing atmospheric humidity and enhancing warming further (Dessler, 2013).

The spatial and temporal distribution of water vapor varies dramatically due to its sensitivity to surface temperature, at-

mospheric circulation, and the phase changes of water. This variability affects not only the distribution of latent heat and cloud formation, but also radiative balance, atmospheric stability, and severe weather development. For example, increased atmospheric moisture contributes to more intense precipitation events, a trend that has been observed and is projected to continue under global warming (Trenberth et al., 2003). These characteristics make precipitable water vapor (PWV) – the total column amount of water vapor in the atmosphere above a given location – an important diagnostic variable for weather forecasting, hydrological modelling, and climate research.

Accurate observation of PWV is essential for initializing and validating numerical weather prediction (NWP) models and global climate models (GCMs). High-resolution PWV data can improve short-term precipitation forecasts, especially during convective weather events, by better representing moisture availability and transport in the lower and mid-troposphere (Van Baelen et al., 2011; Li et al., 2020; Muller et al., 2009). Furthermore, trends in PWV are used as indicators of climate change and can provide insights into shifts in the hydrological cycle, including the intensification of droughts and extreme rainfall events (Allan et al., 2014). Thus, continuous and accurate monitoring of water vapor is fundamental not only for scientific understanding of the atmosphere, but also for practical applications such as agriculture, disaster preparedness, and water resource management.

Commonly used observational techniques for atmospheric water vapor – such as radiosondes, microwave radiometers (MWRs), and satellite instruments – each offer distinct strengths and limitations regarding spatial and temporal resolution, vertical coverage, and observational accuracy. Radiosondes, which are balloon-borne instruments widely regarded as the standard for in situ atmospheric profiling, provide high-vertical-resolution measurements of temperature, pressure, and humidity from the surface to the stratosphere. However, they are typically launched only twice daily at synoptic hours (00:00 and 12:00 UTC) and have limited spatial coverage, particularly over oceans and in developing regions (Dirksen et al., 2014; Seidel et al., 2009). Ground-based MWRs enable continuous, high-temporal-resolution monitoring of atmospheric temperature and humidity profiles under clear and cloudy skies; however, retrieval accuracy is commonly degraded during precipitation (Araki et al., 2015; Xu, 2024). These instruments are valuable for capturing short-term variability in atmospheric moisture, especially within the planetary boundary layer; however, their relatively high cost and calibration requirements restrict widespread deployment (Löhnert and Maier, 2012). Satellite-based passive remote sensing instruments, such as the Moderate Resolution Imaging Spectroradiometer (MODIS), the Atmospheric Infrared Sounder (AIRS), and the Infrared Atmospheric Sounding Interferometer (IASI), offer extensive global coverage and long-term observational consistency, which are essential for large-scale climate studies and op-

erational weather forecasting. Nonetheless, they suffer from coarse temporal resolution due to their orbit characteristics and are prone to retrieval errors under cloudy and precipitating conditions, especially when using infrared channels (Schröder et al., 2019). More recently, active sensing techniques such as Global Navigation Satellite System Radio Occultation (GNSS-RO) have become increasingly important, offering high-vertical-resolution humidity profiles with minimal sensitivity to cloud cover. Despite these advantages, the spatial and temporal sampling of GNSS-RO remains relatively sparse compared to passive satellite observations (Anthes et al., 2008). In parallel, the International GNSS Service (IGS) Real-Time Pilot Project has enhanced global access to high-precision GNSS data streams, enabling near-real-time estimation of PWV through precise point positioning (PPP) techniques at dense ground-based networks worldwide (Zhang et al., 2019). Collectively, these observational platforms form the foundation of current atmospheric water vapor monitoring capabilities; however, their limitations – particularly in cloudy-sky conditions – underscore the need for improved and robust retrieval techniques that can function across all sky conditions.

Recently, sun photometers and multi-wavelength radiometers have also become crucial instruments for remotely sensing atmospheric water vapor from surface due to their ability to measure the intensity of direct solar radiation across different wavelengths. Atmospheric water vapor absorbs and scatters solar radiation at specific wavelengths, which leads to detectable variations in the amount of radiation reaching the surface, making these instruments highly effective for PWV retrieval (Campanelli et al., 2014). For example, instruments such as sun photometers or sky radiometers, which measure direct and diffuse radiation at different wavelengths, are operated in established networks such as AERONET (Holben et al., 1998) and SKYNET (Nakajima et al., 2020). These instruments are capable of inferring PWV by analysing transmittance at wavelengths, particularly around 940 nm, where water vapor has prominent absorption features (Campanelli et al., 2014). Similar retrieval procedures have been applied to infer PWV values using direct irradiances observed by multi-wavelength radiometers, such as Multi-Filter Rotating Shadowband Radiometers (MFRSR) (Alexandrov et al., 2009) and grating spectroradiometers (Qiao et al., 2023) as well. Moreover, the accurate retrieval of PWV from these instruments depends on effective calibration techniques as well as accurate quantification of aerosol optical thickness (AOT) at water vapor-absorbing wavelengths. AOT is typically measured at non-absorbing wavelengths and then interpolated to estimate the AOT at the 940 nm wavelength (Pérez-Ramírez et al., 2014). In addition to them, the multi-axis differential optical absorption spectroscopy (MAX-DOAS) technique – used in networks such as A-SKY (Irie et al., 2011; Mizobuchi et al., 2025) and NDACC (De Mazière et al., 2018) – provides continuous PWV observations by analysing scattered solar radiation.

Although these surface-based remote sensing technologies have proven effective, they are generally limited to clear-sky observations. To move beyond this constraint, the present study focuses on retrieving PWV under all-sky conditions. By examining whether surface-based spectral irradiances contain sufficient information to estimate PWV even in the presence of clouds, we aim to assess the feasibility of retrieving PWV under all-sky conditions from surface-observed spectral irradiances.

Extending the remote sensing capabilities of ground-based instruments to retrieve atmospheric water vapor under all-sky conditions (both clear and cloud sky conditions) represents a significant advancement in atmospheric monitoring with several possible long-term benefits. Global irradiance, comprising both direct and diffuse components, can be reliably measured using inexpensive, robust sensors. These spectral irradiances in and within water vapor absorption bands exhibit sensitivity to changes in columnar water vapor and other atmospheric constituents, making them highly effective for retrieving PWV across a wide range of atmospheric conditions. Furthermore, the low cost, ease of deployment, and high temporal resolution of these instruments support their use in both operational networks and research campaigns. Despite this potential, comprehensive methodologies for this task remain underdeveloped. This is primarily due to the non-linear and ill-posed nature of the retrieval problem, the sensitivity to cloud properties, and the computational burden associated with traditional inversion techniques. As a promising alternative, machine learning (ML) approaches offer data-driven solutions capable of capturing complex, non-linear relationships between spectral irradiance and atmospheric water vapor, without the need for explicit physical modelling or iterative inversion (Zheng et al., 2021). Traditional DOAS-type and fully radiative transfer-based retrievals usually operate within selected wavelength windows, require careful baseline determination, and often assume clear-sky, direct-beam conditions, which limits their applicability under cloudy or diffuse-light scenarios (Irie et al., 2011). In contrast, once a machine-learning model – such as a deep neural network (DNN) – is trained, it can directly map measured spectra to PWV with negligible computational cost, enabling high-temporal-resolution retrievals. By capturing complex, non-linear relationships in the spectral data, ML models also reduce reliance on detailed physical assumptions, offering a flexible, scalable, and robust solution compared with conventional DOAS or radiative-transfer-based methods.

2 Data

This study uses data collected at Chiba (35.625° N, 140.104° E), Japan from 2012 to 2018, except for 2014 due to data unavailability, of two different types, as described below.

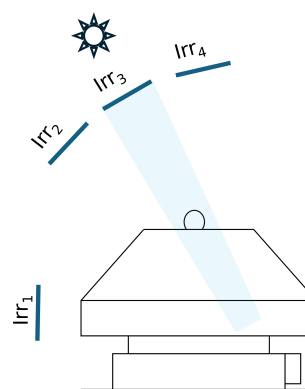


Figure 1. Observation sequence of the MS-700 spectroradiometer equipped with a rotating shadow-band, measuring total horizontal global irradiance (Irr_1), partially shaded irradiances (Irr_2 and Irr_4 ; 8.6° on either side of the Sun), and fully blocking the Sun's direct beam (Irr_3) during each observation cycle.

2.1 Spectral irradiances

This study uses global, direct, and diffuse spectral irradiances observed by a spectroradiometer (MS-700, manufactured by EKO Instruments Co. Ltd., Japan) equipped with a rotating shadow-band system. The spectroradiometer measures spectral global irradiances by using a rotating shadow-band at predefined slant angles along the rotation axis (Khatri et al., 2012; Takamura and Khatri, 2021). The observation sequence of the instrument is illustrated in Fig. 1. The basic concept is similar to that of the Multi-Filter Rotating Shadow-band Radiometer (MFRSR) (Harrison et al., 1994); however, unlike the MFRSR, the MS-700 is a spectroradiometer capable of measuring spectral irradiance over a continuous wavelength range from 300 to 1050 nm with wavelength interval of 3.3 nm. Detailed information about the observation system is provided in Takamura and Khatri (2021). Briefly, the shadow-band-equipped spectroradiometer performs measurements at four different shadow-band positions during each observation cycle, as shown in Fig. 1. Initially, the shadow-band is positioned below the horizontal plane of the observation sensor to measure total spectral irradiance (Irr_1). It then rotates to a position 8.6° behind the center of the sun to measure spectral irradiance with partial shading (Irr_2). Next, the shadow-band moves to fully block the sun's direct beam at the center position, measuring spectral irradiance with complete solar obstruction (Irr_3). Finally, the shadow-band rotates to a position 8.6° ahead of the sun to acquire another partially shaded spectral irradiance measurement (Irr_4). Through this sequence, spectral irradiances corresponding to four distinct shadow-band positions are obtained in a single scan. These spectral irradiances, in units of $W m^{-2} \mu m^{-1}$, are automatically generated by the observational software from the raw signals after applying the factory calibration constants.

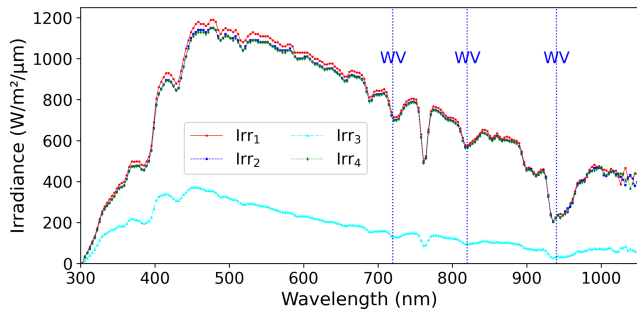


Figure 2. Observational data example from MS-700 for 1 May 2018 (12:00 JST), with vertical lines marking the central wavelengths of water vapor absorption.

Figure 2 presents an example of Irr_1 , Irr_2 , Irr_3 , and Irr_4 measurements collected over Chiba, Japan, at 12:00 JST on 1 May 2018, along with indications of the central wavelengths of major water vapor absorption bands. It is important to note that Irr_1 , Irr_2 , Irr_3 , and Irr_4 are measurements obtained using a horizontally mounted sensor. Consequently, these observed irradiance values are subject to cosine errors, which refer to the deviation of a radiometric sensor's angular response from the ideal cosine law (Lambert's cosine law). This law describes how irradiance should vary as a function of the angle of incoming light relative to the sensor's normal. Cosine error correction factors – typically derived from laboratory-based angular response measurements or field calibrations using solar tracking systems – are usually applied to correct these deviations and improve the accuracy of irradiance measurements, particularly at large solar zenith angles. Such corrections are crucial when using physics-based retrieval methods, which generally assume idealized input conditions. However, in supervised ML-based approaches, explicit correction for cosine errors is not necessarily required, as the ML model can implicitly learn and adjust for these systematic deviations during the training process. To simplify the retrieval procedure and minimize computational burden, this study directly utilizes the observation signals without applying cosine error corrections to the data. The cosine-affected global, diffuse, and direct irradiances used in this study can be expressed as follows (Takamura and Khatri, 2021):

$$\text{spGHI}' = Irr_1 \quad (1)$$

$$\text{spDHI}' = Irr_3 - \left(Irr_1 - \frac{Irr_2 + Irr_4}{2} \right) \quad (2)$$

$$\text{spDNI}' = \text{spGHI}' - \text{spDHI}' \quad (3)$$

where spGHI' , spDHI' and spDNI' represent global horizontal irradiance, global diffuse irradiance and direct normal irradiance, respectively, without applying any correction factors to observation data.

2.2 Precipitable water vapor content (PWV)

The PWV data used in this study were obtained from observations made by a MWR (Model: MP-1500), developed by Radiometrics Corporation, USA. This passive radiometer operates in the K-band frequency range (22–30 GHz). It measures natural thermal emissions from the atmosphere at selected narrowband frequencies, primarily centred around the 22.235 GHz water vapor absorption line. This instrument is capable of retrieving high-resolution vertical profile of water vapor and PWV under all-weather conditions. Its ability to perform accurate PWV measurement makes it highly suitable for both operational and research applications. Ground-based MWRs, such as MP-1500 and its successors (e.g., MP 3000A) have been widely used in meteorological and climate studies, particularly for evaluating boundary layer processes, validating satellite-based retrievals, and supporting numerical weather prediction models (e.g., Bartsevich et al., 2024; Padmanabhan et al., 2009; Sengupta et al., 2004).

3 Methodology

ML approaches have become increasingly important in atmospheric remote sensing for inferring geophysical quantities from complex, nonlinear input data. These data-driven models learn statistical relationships from observational datasets without relying explicitly on physical parameterizations, making them well-suited for problems where traditional inversion is challenging (Zheng et al., 2021). Several major ML techniques have been widely applied in environmental and remote sensing research. Linear regression models provide interpretable baselines but are limited by their inability to model nonlinear dependencies. Support vector machines (SVMs) with kernel functions can capture more complex relationships, but scale poorly with large datasets and are sensitive to kernel selection (Mountrakis et al., 2011). Decision tree-based ensemble methods, including random forests and gradient boosting machines (GBMs), are popular for their robustness, feature importance estimation, and relatively strong performance on tabular datasets (Belgiu and Drăguț, 2016). However, these ensemble methods can have limitations in their ability to capture complex spectral correlations and nonlinearities, which can be critical when working with high-dimensional atmospheric data with multiple features, such as spectral irradiance measurements at different wavelengths.

Deep learning approaches, particularly deep neural networks (DNNs), provide a powerful alternative with significant advantages over the aforementioned ML methods in the context of this study. Spectral irradiance data are high-dimensional and exhibit complex interactions across wavelengths due to absorption, scattering, and cloud modulation. Unlike ensemble methods, which treat each decision independently, DNNs are capable of learning hierarchical representations of data through multiple layers of neurons, al-

lowing them to capture intricate nonlinear relationships between input features and output targets (Lecun et al., 2015). Hence, DNNs are particularly suited for applications like atmospheric remote sensing, where capturing the full complexity of the input data is crucial.

The DNN employed in this study is a fully connected multilayer perceptron (MLP) network, designed to map input spectral irradiance data to the target PWV. The input data consist of measurements of spectral irradiance across multiple wavelengths, day number of year (DOY), and solar zenith angle calculated from local time and latitude and longitude of observation site. The input data were first pre-processed and normalized before being passed through the network. The model architecture consists of several hidden layers, where each layer applies a linear transformation followed by a non-linear activation function. At each layer, the output is calculated as

$$a^{(l)} = f(w^{(l)} a^{(l-1)} + b^{(l)}) \quad (4)$$

where, $w^{(l)}$ and $b^{(l)}$ represent the weights and biases for the l th layer, while $a^{(l)}$ is the output activation and $f(\cdot)$ the activation function. We used Rectified Linear Unit (ReLU) (Krizhevsky et al., 2017) activation function. ReLU is one of the most commonly used activation functions in DNN modelling because it is simple, computationally efficient, and helps prevent gradients from becoming too small during training (Krizhevsky et al., 2017). This process continues through multiple layers, with each layer learning increasingly abstract features of the input data. The final output layer generates the predicted PWV value.

We used data from 2012 to 2016, excluding 2014, of all sky conditions as model inputs. These data were divided into three subsets: training, validation, and test sets. Initially, 60 % of the data was allocated to the training set, which was used to fit the model. The remaining 40 % was temporarily set aside and subsequently split evenly into validation and test sets, each comprising 20 % of the total data. The validation set was used during training to monitor model performance and tune hyperparameters, while the test set was held out entirely from the training process and used only for final performance evaluation.

The model was trained using the mean squared error (MSE) loss function as

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=0}^N (\hat{y}_i - y_i)^2 \quad (5)$$

where, \hat{y}_i and y_i represent the predicted and true PWV values for the i th sample, respectively, \mathcal{L} the loss function, θ , the model parameters (weights and biases), and N the number of training samples. The network parameters (weights and biases) were updated using the Adam optimizer (Kingma and Ba, 2015), which adaptively adjusts learning rates for each parameter based on first- and second-order moment estimates of the gradients, offering efficient and robust convergence in deep learning tasks.

To enhance generalization and mitigate overfitting, the training process incorporated three callback mechanisms available in the Keras deep learning framework (Chollet, 2015): EarlyStopping, which monitored the validation loss (val_loss) and halted training if it failed to improve for 10 consecutive epochs (where an epoch represents one complete pass through the entire training dataset) while restoring the best-performing model weights; ModelCheckpoint, which saved the model whenever a new minimum validation loss was observed to ensure retention of the optimal model; and ReduceLROnPlateau, which reduced the learning rate by a factor of 0.3 if the validation loss plateaued for 10 epochs, facilitating better convergence during later training stages. The model was trained for a maximum of 300 epochs using mini-batch gradient descent, an optimization approach in which the training data are divided into small subsets (mini-batches) to update model parameters after each batch (Bottou, 2010). The mini-batches were supplied by train_generator, a custom data generator that efficiently yields batches of training samples. Simultaneously, validation data were provided via val_generator, another custom generator designed to supply validation samples in batches during training, to enable real-time monitoring of model performance during training.

The trained DNN model was subsequently applied to unseen spectral irradiance data, i.e., data not used during the training process, from 2017 and 2018 to estimate PWV values. These estimates were then compared with ground-truth PWV measurements obtained from a MWR to evaluate the model's predictive performance. For this analysis, only data corresponding to solar zenith angles (SZA) less than 75° were used. Additionally, as part of the input data quality control, any negative values, if present, were excluded from both the model training and evaluation processes.

It is worth mentioning that the training data for DNN model derived from MWR observations are subject to several well-characterized sources of error – such as solar interference, site-specific retrieval coefficient mismatches, contamination from water on the radiometer window, calibration issues, and radio frequency interference. Although the radiometer system incorporates mitigation mechanisms (e.g., temporal averaging, dew sensors, quality control flags), these error sources may still introduce both random and systematic uncertainties into the data. As the primary objective of this study is to explore the capability of surface-observed spectral irradiances for remote sensing of atmospheric water vapor under diverse sky conditions, the effects of these observational uncertainties on the trained models are not explicitly quantified.

In this study, MWR data were used because they were available at the site as a reliable PWV reference. However, the method can operate without relying specifically on MWR measurements. The model can be trained using PWV data from other standard reference sources, such as radiosonde observations or GNSS-based PWV retrievals, depending on data availability at a given location. Although larger and

more diverse training datasets generally improve model robustness, the four-year dataset used here reflects data availability rather than a strict requirement of the method. The necessary training period can be adapted to local atmospheric variability and operational needs.

Furthermore, the model learns statistical relationships between spectral irradiances and PWV, which can be influenced by local atmospheric and surface conditions. Relocating the instrument to a site with substantially different conditions may reduce accuracy. However, the model can be adapted to new locations through retraining with local reference PWV data or, more efficiently, via transfer learning (Pan and Yang, 2010; Weiss et al., 2016), in which a pre-trained model is fine-tuned using a relatively small amount of site-specific data. Recent studies (e.g., Chen et al., 2025; Dong et al., 2024; Gupta et al., 2024) show that such adaptation strategies effectively preserve retrieval performance when transferring ML-based remote-sensing models across sites.

Finally, although modern spectroradiometers are generally stable, minor instrumental drifts can occur over time. Our retrieval framework assumes that the input spectral irradiances are properly calibrated, and any systematic changes in instrument response should ideally be handled at the calibration or quality-control level. In practical operation, however, the DNN learns robust relationships between spectral irradiances and PWV that can remain largely stable even when small residual drifts occur. In such cases, periodic fine-tuning with a limited amount of recent reference data may be typically sufficient to adjust for these gradual shifts and maintain retrieval accuracy.

4 Results and Discussion

As shown in Fig. 2, three major water vapor absorption bands are present within the visible to near-infrared spectral range. To evaluate their relevance for PWV retrieval under all-sky conditions (both cloudy and clear), we developed separate predictive models for each absorption band, as well as a combined model incorporating all bands. Furthermore, given that global spectral irradiance measurements are more widely available – particularly when shadow-band systems are not employed – we assessed the model’s practicality and robustness using only global spectral irradiance data. In addition, we tested the model using direct spectral irradiance data under clear-sky conditions to establish a benchmark for comparison with the all-sky and global-only retrievals.

Table 1 summarizes the wavelengths used in this study for each absorption band.

4.1 Modelling using spectral global, direct, and diffuse irradiances together

Our analyses demonstrated that incorporating a certain degree of feature engineering into the input data, i.e., generat-

Table 1. Water Vapor Absorption Bands Used in This Study.

Band Name	Wavelength Range (Total Number of Wavelengths)
940 nm centered absorption band	929.0–997.3 nm (22)
820 nm centered absorption band	800.9–840.5 nm (13)
720 nm centered absorption band	708.1–744.6 nm (12)

ing additional informative features from existing input data, significantly improves model performance. Accordingly, we prepared the input features as follows: for each wavelength, we computed the ratio of direct to diffuse irradiance and used it as an additional input feature. This ratio was used in combination with the spectral values of global, direct, and diffuse irradiances, all of which were scaled by the cosine of the solar zenith angle to account for solar geometry. Finally, during model training, we applied a natural logarithmic transformation to both the target variable (PWV) and all irradiance-based input features to enhance learning stability and model accuracy.

Our analyses further revealed that increasing the number of intermediate hidden layers may generally raise model complexity, thereby heightening the risk of overfitting and potentially leading to training instability or failure. Similarly, increasing the number of nodes per layer may further add model complexity and the risk of overfitting without necessarily enhancing predictive performance. From such analyses, we found that a DNN model with three hidden layers (with 32, 16, and 8 nodes) and a single-node output layer performs best on our datasets. Thus, we used model of this configuration in this study.

4.1.1 Loss function evaluation

Figure 3 presents an evaluation of model performance through training and validation loss curves across three spectral bands centered at 940, 820, and 720 nm, and combination of all of them under all-sky conditions (including both clear and cloudy skies). The variation in the number of training epochs among subplots reflects the use of an EarlyStopping criterion, as detailed in the methodology. Each subplot adopts a dual-panel layout: the upper panel displays a broader loss range (0.06–0.09), while the lower panel highlights a finer range (0.002–0.02), allowing for a detailed inspection of convergence behavior. In each subplot, the red line represents the training loss, and the blue line shows the validation loss, both calculated using the MSE loss function using an equation mentioned above. Arrows indicate the minimum validation loss during the training process, representing the point of optimal generalization performance.

Across all three bands and their combination, the loss curves exhibit a rapid initial decline within the first 10–15 epochs, indicating efficient learning during early training. This is followed by a gradual flattening of the curves,

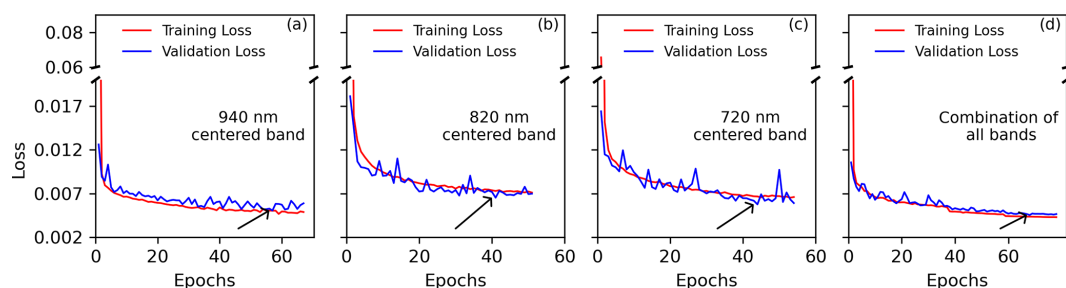


Figure 3. Training and validation losses for spectral direct, diffuse and global irradiance-based input features for band centered at (a) 940 nm, (b) 820 nm, (c) 720 nm, and (d) all of them. The arrows indicate the minimum validation loss.

suggesting convergence is generally achieved within 40–70 epochs depending on the spectral band. The relatively small difference between training and validation losses throughout the training period suggests that the models maintain strong generalization capacity and are not significantly overfitting the training data. Minor fluctuations in the validation loss are normal and are due to the stochastic nature of mini-batch training and variability in the validation data.

For the 940 nm band (Fig. 3a), the model achieves the lowest overall training and validation losses among the three selected bands. This outcome is consistent with the well-known fact that the 940 nm band is the strongest near-infrared water vapor absorption feature (Gueymard, 1995), providing high sensitivity to columnar water vapor content even under cloudy conditions. The model's strong performance across mixed-sky inputs highlights the robustness of this spectral signal and demonstrates the DNN's ability to capture the underlying physical relationship effectively. For the 820 nm band (Fig. 3b), the model also shows good convergence, although the final loss values are slightly higher compared to the 940 nm band. While the 820 nm band exhibits moderate water vapor absorption strength (Gueymard, 1995), it still carries valuable PWV information. Data of this absorption band is particularly useful in scenarios where the 940 nm band is partially obstructed or unavailable. The consistent alignment between training and validation curves suggests that the model trained on this band can still generalize well under diverse atmospheric conditions. The 720 nm band (Fig. 3c), which represents the weakest of the three absorption features analyzed (Gueymard, 1995), also leads to a reasonably converged model. However, it shows relatively higher and slightly more variable validation losses. This is expected due to the weaker absorption of water vapor in the visible range and the stronger influence of confounding factors such as aerosols, clouds, and surface reflectance (Khatri et al., 2016, 2019; Nakajima et al., 2020). Nevertheless, the narrow gap between training and validation losses indicates that the model remains well-regularized and capable of extracting meaningful information, even for such weak absorption band. Crucially, when all three spectral bands were combined as a unified input, the DNN model

achieved exceptionally low and stable losses, comparable to or potentially surpassing the performance of the best individual band (940 nm). This outcome highlights a powerful synergistic effect, where the integration of diverse spectral information leads to enhanced accuracy and robustness. The slightly longer convergence time for the combined band suggests the model requires more iterations to fully exploit the increased complexity and interdependencies within the richer input feature space.

4.1.2 Validation through comparison between predicted and unseen ground truth values

Scatter plot validation

Figure 4 shows the relationship between the predicted and ground-truth PWV values for the three spectral bands centered at (a) 940 nm, (b) 820 nm, (c) 720 nm, and for (d) combination of all of them. These predictions were made using models that achieved the lowest validation losses as indicated by the arrows in Fig. 3. Each panel presents a 2D density scatter plot comparing predicted PWV (y axis) values against PWV (x axis) values observed by MWR, with the red dashed line denoting the 1 : 1 reference line (ideal prediction).

The 940 nm centered band (Fig. 4a) yields the highest performance among three different individual spectral configurations with a Root Mean Square Error (RMSE) of 0.174 cm and correlation coefficient (R^2) of 0.978. Visually, Fig. 4a demonstrates a very tight clustering of data points closely aligned with the 1 : 1 reference line, indicating minimal scatter and high predictive fidelity across the entire range of actual PWV values. This strong performance is consistent with its corresponding loss curves in Fig. 3a, where the model exhibited the lowest training and validation losses and stable convergence among three different individual spectral configurations. This relatively high accuracy is attributed to the strong water vapor absorption feature at 940 nm, which provides high spectral sensitivity to column water vapor under a wide range of atmospheric conditions. The low RMSE signifies that the model's predictions are consistently very close to the true values, while the near-perfect R^2 indicates that al-

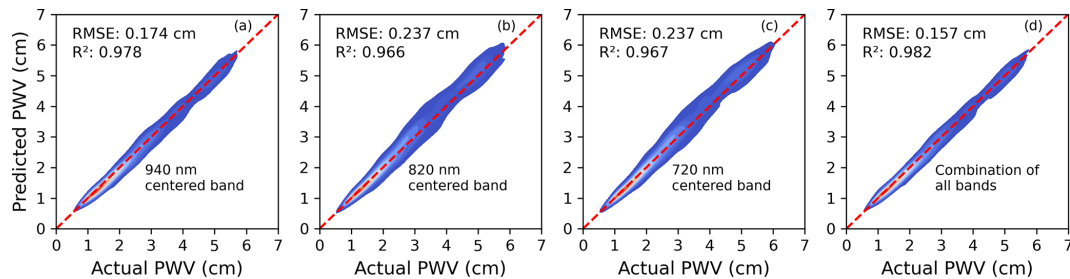


Figure 4. Comparison between predicted and observed PWV values for spectral bands centered at (a) 940 nm, (b) 820 nm, (c) 720 nm, and (d) all of them using DNN models trained on all-sky spectral irradiance data of global, direct, and diffuse. The predictions were made using models that achieved the minimum validation losses (as shown in Fig. 3). The red dashed line indicates the 1 : 1 line.

most all the variability in observed PWV is explained by the model's output.

For the 820 nm band (Fig. 4b), the model also performs commendably, achieving RMSE of 0.237 cm and R^2 of 0.966. While slightly less accurate than the 940 nm band, this result aligns with Fig. 3b, which showed effective model convergence and low loss values, albeit marginally higher than the 940 nm case. Figure 4b shows a good alignment with the 1 : 1 line, though with a slightly wider spread of points compared to the 940 nm band. Although the 820 nm band has weaker absorption, it is still useful for retrieving PWV, especially when 940 nm data are unavailable or less reliable due to instrument issues or very strong absorption that could potentially saturate the signal. Its performance confirms its utility as a valuable alternative or complementary data source, offering a reliable measurement even with a moderately weaker signal.

The 720 nm band (Fig. 4c) achieves a comparable RMSE of 0.237 cm and R^2 of 0.967, similar to the 820 nm result. This reflects the model's robustness in learning from the weaker absorption band, although its performance is slightly more variable, as also observed in the corresponding loss curve (Fig. 3c), which showed higher fluctuation in validation loss. Because this band is less sensitive to water vapor, the predictions are more scattered around the 1 : 1 line – especially at higher PWV levels – likely due to weaker signals and more noise from other factors, such as aerosols, clouds, surface etc. While the DNN demonstrates a remarkable ability to extract meaningful information even from such a challenging signal, the trade-off between signal strength and prediction precision becomes evident, leading to greater uncertainty at higher PWV values where the water vapor signal is less distinct relative to background noise.

The combined spectral band approach (Fig. 4d) visually demonstrated superior performance, appearing as good as or even slightly better than the 940 nm band alone. This is reflected with RMSE of 0.157 cm and an R^2 of 0.982. Figure 4d displays a remarkably tight clustering of predicted values around the 1 : 1 line, indicating exceptional accuracy and linearity across the entire PWV range. This highlights the significant advantage of multi-spectral data fusion. By in-

tegrating information from bands with varying sensitivities, the DNN can leverage complementary data. For instance, while the 940 nm band provides the primary, strong water vapor signal, the 720 nm band, despite its weak water vapor absorption, is more sensitive to confounding factors like aerosols and clouds. The DNN can implicitly use this contextual information to correct for their influence on the stronger absorption bands, effectively performing an “internal atmospheric correction”. This leads to a more robust and accurate PWV retrieval, particularly under complex all-sky conditions where multiple atmospheric components are at play. This synergistic effect means the combined input provides a richer, more comprehensive representation of the atmospheric state, allowing the DNN to achieve optimal accuracy and enhanced resilience against atmospheric variability.

Temporal validation across varying weather and sky conditions

Figure 5 provides a detailed temporal validation of the DNN models' performance by comparing predicted and actual PWV values on a monthly basis. Presented as box plots, the upper panels illustrate the distributions of actual (blue) and predicted (red) PWV values, while the lower panels show the corresponding ratios of actual to predicted PWV (Actual/Predicted) for each month. In the box plots, the box represents the 25th and 75th percentiles, the horizontal line inside the box indicates the median, and the whiskers extend to the most extreme points within 1.5 times the interquartile range. These figures offer a comprehensive view of the central tendency (median), spread (interquartile range), and variability of both actual (blue) and predicted (red) PWV values for each month of the year. This monthly comparison serves as a crucial complement to the overall accuracy metrics (RMSE, R^2) presented in Fig. 4 and the training dynamics observed in Fig. 3, revealing how consistently the models perform across seasonal atmospheric changes.

A general observation across all subplots (upper panel) in Fig. 5 is the clear seasonal cycle of PWV, with higher values typically observed during the summer months (e.g., July–September) and lower values during the winter months (e.g.,

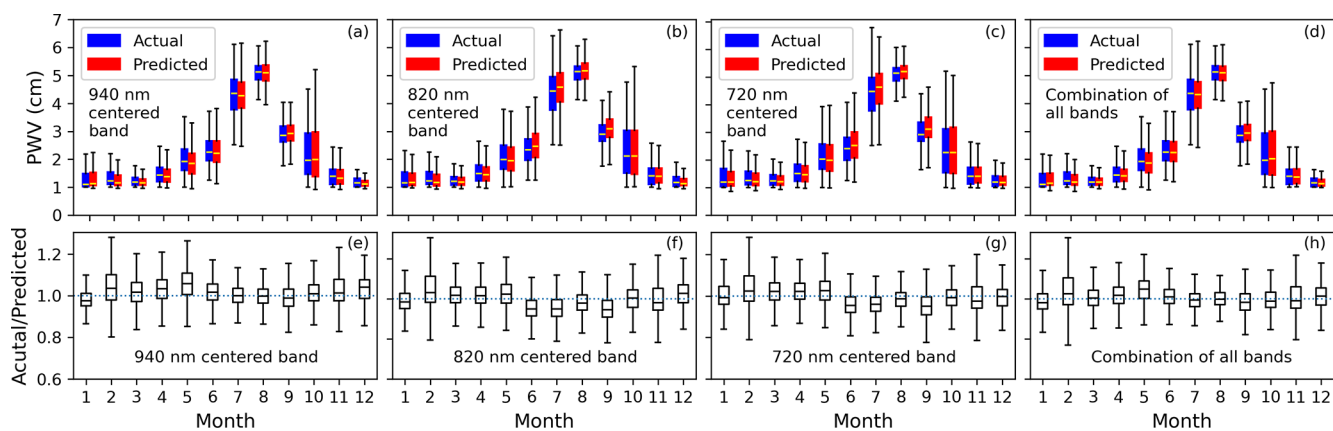


Figure 5. Monthly comparison of DNN-predicted versus actual PWV values for spectral bands centered at (a) 940 nm, (b) 820 nm, (c) 720 nm, and (d) combination of all of them; and ratios of actual to predicted PWVs for spectral bands centered at (e) 940 nm, (f) 820 nm, (g) 720 nm, and (h) combination of all of them. The box represents the 25th and 75th percentiles, the yellow line inside the box indicates the median, and the whiskers extend to the most extreme points within 1.5 times the interquartile range.

January–March). This expected pattern is well captured by the actual PWV data, and the key evaluation lies in how accurately the predicted values track these monthly variations.

For the 940 nm centered band (Fig. 5a), the box plots demonstrate remarkably close agreement between the actual and predicted PWV values across all months. The medians of the predicted (red) boxes consistently align well with those of the actual (blue) boxes, and their interquartile ranges largely overlap. Even the whiskers, representing the full range of data, show strong correspondence. The corresponding ratio plot (Fig. 5e) reinforces this finding. The monthly medians of Actual/Predicted values are nearly constant around 1, and the spreads are narrow, signifying minimal systematic bias and stable retrievals across different atmospheric conditions.

The 820 nm centered band (Fig. 5b) also shows good overall agreement between predicted and actual monthly PWV. The predicted box plots generally follow the seasonal trend of the actual data. However, upon closer inspection, especially during months with higher PWV or greater atmospheric variability, subtle deviations or slightly larger discrepancies between the predicted and actual distributions might be observed compared to the 940 nm band. This visual observation is consistent with the slightly higher RMSE (0.237 cm) and lower R^2 (0.966) for the 820 nm band presented in Fig. 4b, and the marginally higher loss values in Fig. 3b. The Actual/Predicted ratios (Fig. 5f) generally remain close to unity but exhibit slightly broader distributions compared to the 940 nm. Occasional minor departures from 1 indicate small month-specific overestimations or underestimations, likely linked to weaker absorption features at 820 nm.

In the case of the 720 nm centered band (Fig. 5c), the monthly comparison reveals more noticeable discrepancies, particularly in months with higher PWV levels. While the general seasonal trend is captured, the predicted box plots

might show less perfect alignment with the actual data, potentially exhibiting larger differences in medians, wider spreads, or less accurate capture of extreme values. This increased scatter and variability in monthly predictions align directly with the larger scatter observed in Fig. 4c's density plot and the more variable validation losses shown in Fig. 3c. These discrepancies underscore the inherent challenges of retrieving PWV from a very weak absorption band, where confounding factors like aerosols, clouds, and surface reflectance exert a stronger influence, leading to less consistent performance across diverse monthly atmospheric states. In the ratio plot (Fig. 5g), the median values remain roughly centred around 1, but the interquartile ranges and whiskers are, in general, wider than those in the 940 and 820 nm bands. This broader spread indicates relatively greater variability and lesser stable monthly retrievals than those above two wavelength bands.

Finally, the combination of all bands (Fig. 5d) visually presents the most compelling results in terms of monthly tracking of PWV. The alignment between the predicted and actual box plots is remarkably tight, often appearing as good as, if not superior to, the 940 nm band alone. The medians are highly congruent, and the interquartile ranges and overall spread of the predicted values closely mirror the actual data across all months. This outstanding consistency in capturing monthly variations strongly corroborates the best overall RMSE (0.157 cm) and R^2 (0.982) metrics reported for the combined approach in Fig. 4d, as well as the very low and stable loss curves in Fig. 3d. This outcome powerfully demonstrates the synergistic advantage of multi-spectral data fusion. The ratio plot (Fig. 5h) shows that the Actual/Predicted medians are tightly centred at around 1 for all months, and the spread remains very narrow, suggesting minimal bias and higher temporal consistency, consistent with the best

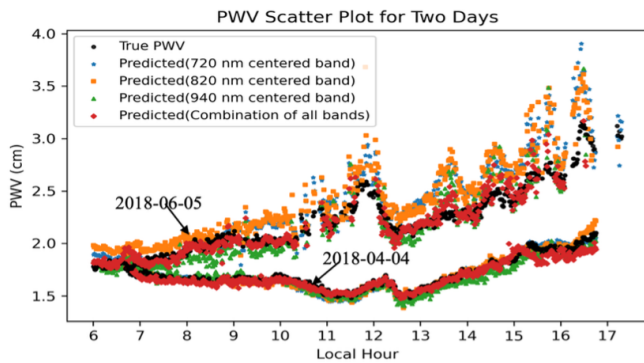


Figure 6. Temporal variations of DNN-predicted and observed PWV values on 4 April 2018 (mostly clear-sky day) and 5 June 2018 (highly cloudy day).

overall RMSE (0.157 cm) and R^2 (0.982) achieved by this model (Fig. 4d).

As the spectral irradiance data used for both training and validation were obtained from the same instrument and processed consistently, instrument-specific systematic effects were effectively learned and compensated by the DNN. As a result, the main source of uncertainty in the predicted PWV is likely associated with the reference MWR measurements. Ground-based MWRs typically retrieve PWV with uncertainties of about ± 0.1 – 0.3 cm (Böck et al., 2025; Elgered and Jarlemark, 1998; Minowa et al., 2024). The RMSE values obtained in this study were below 0.24 cm, well within this range, indicating that the prediction accuracy is consistent with the expected uncertainty of the standard reference measurements – namely, the MWR observations used in this study.

Overall, the upper and lower panels of Fig. 5 provide essential temporal validation across varying weather and sky conditions, confirming that the performance characteristics observed in the overall accuracy metrics (Fig. 4) and training dynamics (Fig. 3) translate consistently across monthly variations.

Further, to highlight the temporal behaviour of the predictions on shorter time scales, Fig. 6 presents example comparisons between predicted and observed PWV for a mostly clear-sky day (4 April 2018) and a highly cloudy day (5 June 2018). These examples provide insight into how the models reproduce temporal fluctuations in PWV under contrasting sky conditions. Overall, the predicted and observed values exhibit similar temporal patterns, although small differences in magnitude occasionally appear, particularly under cloudy conditions or in weaker absorption bands. These results further highlight the ability of models to capture PWC variability on shorter time scales.

Since DOY is included as one of the input features in our DNN models, it is important to examine whether the agreement between predicted and true values shown in Fig. 5 could have been influenced by climatological patterns encoded in

DOY. To assess the relative importance of DOY, together with other input features, we applied the SHAP (SHapley Additive exPlanations) method (Lundberg and Lee, 2017) by computing Shapley values, which represent the average marginal contribution of a feature across all possible feature combinations. This method has been widely used in ML-based atmospheric and remote-sensing studies to evaluate the relative influence of input variables on model predictions (Lundberg et al., 2020; Zhao et al., 2024). Table 2 summarizes the SHAP values (in %) for the input features – global, direct, and diffuse irradiances, the direct-to-diffuse irradiance ratio, and DOY – for individual absorption bands and for their combined dataset. The SHAP values for the radiative components in Table 2 represent averages over all wavelengths, as reporting wavelength-specific SHAP values would be excessively lengthy. The results clearly show that when global, direct, and diffuse irradiances are used jointly, the global irradiance component consistently exhibits the highest relative SHAP importance (32 %–48 %), followed by the direct and diffuse components. The direct-to-diffuse irradiance ratio contributes a moderate but meaningful amount (approximately 14 %–22 %). In contrast, DOY contributes less than 0.3 %, indicating that the seasonal patterns observed in Fig. 5 are dominated by spectral-irradiance-based features rather than by DOY input feature. DOY can have very low importance because it only provides indirect seasonal information, whereas the spectral irradiances directly capture the actual atmospheric state (water-vapor absorption, scattering, SZA effects, etc.).

4.2 Modelling using only spectral global irradiance or spectral direct irradiance

When modelling using either only spectral global irradiance or only spectral direct irradiance as the primary radiation component, we largely followed the same feature engineering approach described in Sect. 4.1. Specifically, for the global-only model, features corresponding to spectral direct and diffuse irradiance were excluded; conversely, for the direct-only model, features related to spectral global and diffuse irradiance were omitted. Additionally, since only a single radiation component was used in each case, the ratio of spectral direct to diffuse irradiance was also excluded from the feature set.

For modelling with spectral direct irradiance as the primary input, we used data corresponding exclusively to clear-sky conditions. These conditions were identified based on clear-sky periods detected by a co-located sky radiometer from the SKYNET network (Hashimoto et al., 2012; Khatri et al., 2014, 2019; Nakajima et al., 2020) installed at the same site of shadow-band spectroradiometer. Cloud screening for the sky radiometer data was performed using the algorithm of Khatri and Takamura (2009). To further reduce the possibility of data affected by very thin clouds, we applied an additional filter to exclude samples with aerosol op-

Table 2. Relative SHAP values (in %) of the input features for the DNN models trained using spectral direct, diffuse, and global irradiances of different absorption bands, together with the direct-to-diffuse irradiance ratio (Ratio) and the day of year (DOY).

Centered band/Features	Direct (%)	Diffuse (%)	Global (%)	Ratio (%)	DOY (%)
940 nm	23.80	17.89	35.67	22.42	0.22
820 nm	18.30	19.20	48.39	13.97	0.14
720 nm	20.03	15.24	44.69	19.95	0.08
Combined (all above)	27.18	19.43	32.92	20.35	0.11

tical thickness (AOT) at 500 nm greater than 0.5. While this additional screening helped to more strictly ensure clear-sky conditions – necessary because clouds near the solar disk can scatter extra light into the sensor and contaminate the direct-beam measurement, making it difficult to accurately assess the contribution of only direct irradiances in PWV – it also significantly reduced the number of available data samples.

Table 3 summarizes the number of data samples used for model training, model validation, and evaluation with unseen data when using only spectral global irradiance or only spectral direct irradiance as the primary radiation component.

4.2.1 Loss function evaluation

Figure 7 provides a direct visual comparison of predictive model performance when input features are restricted to either only spectral global irradiance (Fig. 7a) or only spectral direct irradiance (Fig. 7b). In Fig. 7a, which illustrates models trained exclusively with global irradiance, a consistent pattern of low minimum validation losses and corresponding training losses is observed across all individual wavelength bands (940, 820, 720 nm) and with slightly higher values for their combination. The loss values generally range from approximately 0.010 to 0.017. A key characteristic is the remarkable proximity between the validation loss and its corresponding training loss for each configuration. This suggests that the models have effectively learned the fundamental patterns inherent in global irradiance data, rather than merely memorizing noise or specific examples from the training set.

Conversely, Fig. 7b, depicting models using only clear-sky direct irradiance, presents a different picture. These models generally exhibit higher overall loss values, with training losses ranging approximately from 0.013 to 0.017, and validation losses slightly lower, from 0.010 to 0.013. A more noticeable, and somewhat complex, relationship between training and validation loss is apparent; for instance, at the 820 and 720 nm centered bands, the training loss is notably higher than the validation loss. While typical overfitting is characterized by training loss being significantly lower than validation loss, this observed scenario, where training loss is higher or not consistently lower, suggests that the model might be struggling to adequately learn the complex patterns within the training data itself, potentially indicating a degree of underfitting.

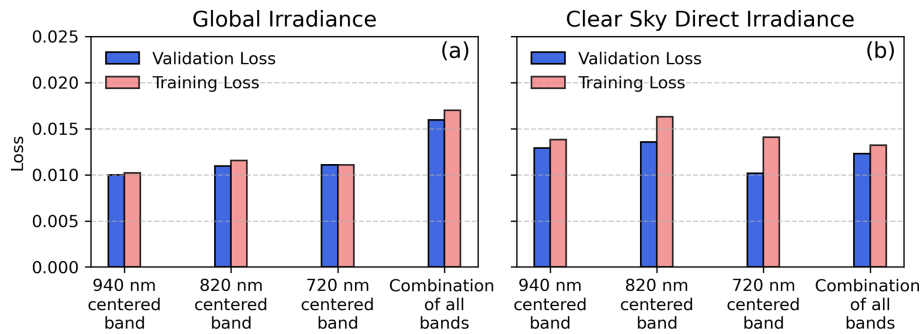
The most significant factor driving this stark difference in performance between the two different radiation data sets is the data volume. As summarized in Table 3, the global irradiance models benefited from a substantially larger dataset for “All-sky conditions”. This ample data provided the models with a rich and diverse set of examples, enabling them to learn more comprehensive and robust patterns. In contrast, the clear-sky direct irradiance models were constrained by a significantly smaller dataset for “Clear-sky conditions”. The reduced data volume makes it challenging for the model to learn robust, generalizable features, increasing the risk of suboptimal learning or underfitting, where the model cannot adequately fit the available data. This highlights that while the choice of specific irradiance components is relevant, the sheer quantity and quality of data for each component are paramount for achieving robust and reliable predictive models.

4.2.2 Validation through comparison between predicted and unseen ground truth values

Figure 8 provides crucial validation of the predictive models’ performance on unseen data, building upon the loss evaluations presented in Fig. 7, by utilizing RMSE in panel (a) and R^2 in panel (b). In Fig. 8a, models using only global irradiance from individual absorption bands consistently exhibit lower RMSE values (e.g., ~ 0.26 cm for 940 nm band), indicating higher predictive precision and smaller average errors on unseen data. This directly aligns with the low training and validation losses observed for global irradiance models in Fig. 7a. Conversely, models using only clear-sky direct irradiance show generally higher RMSE values (e.g., ~ 0.32 cm for 820 nm band) at relatively stronger absorption bands, signifying larger prediction errors, which is consistent with their higher overall loss values in Fig. 7b. These RMSE values are closer to the upper bound of PWV uncertainty corresponding to MWR measurements under favorable conditions (Elgered and Jarlemark, 1998; Minowa et al., 2024; Böck et al., 2025), as discussed in Sect. “Scatter plot validation”. Similarly, in Fig. 8b, models with global irradiance inputs consistently achieve higher R^2 values (ranging from ~ 0.91 to ~ 0.95), demonstrating that a larger proportion of the variance in the data is explained by these models, indicating a better fit and stronger predictive capability. These high R^2 values further validate the robust generalization inferred from the low and

Table 3. Summary of Data Counts Used for Modelling and Validation.

Sky conditions	Modelling			Validation Using Unseen Data
	Training	Validation	Test	
All-sky conditions using global spectral irradiance as input data	57 736	19 245	19 246	97 274
Clear-sky conditions using only direct spectral irradiance as input data	3022	1008	1008	7259

**Figure 7.** Minimum validation loss and the corresponding training loss for models using (a) only global irradiance and (b) only clear-sky direct irradiance as input features, across wavelength bands centered at 940, 820, 720 nm, and their combination.

converging losses in Fig. 7a. In contrast, clear-sky direct irradiance models show consistently lower R^2 values (ranging from ~ 0.90 to ~ 0.91), implying a weaker fit and less explanatory power, which correlates with their less ideal loss profiles in Fig. 7b. This consistent pattern across both Figs. 7 and 8 underscores that the data volume is the primary factor driving model performance; the global irradiance models, benefiting from a substantially larger dataset (over 97 000 total samples for “All-sky conditions”), consistently outperform the clear-sky direct irradiance models, which were constrained by a significantly smaller dataset (only about 7200 total samples for “Clear-sky conditions”) due to stringent filtering. This highlights that ample data can enable models to learn more robust patterns, leading to superior accuracy and generalization on unseen data.

Figure 9 offers a detailed monthly comparison of predicted and observed PWV values specifically for the 940 nm centered spectral band, serving as a representative illustration of the models’ performance. Similar to Fig. 5, panel (a) showcases results for models utilizing only global irradiance as input features, while panel (b) presents results for models based on only clear-sky direct irradiance.

In Fig. 9a, which illustrates the performance of models trained with only global irradiance, a very good alignment is evident between the predicted (red boxes) and actual (blue boxes) PWV values across all months. Both the observed and predicted data clearly capture the expected seasonal cycle, with PWV values generally increasing from winter (e.g., months 11–2, typically around 1–1.5 cm) to summer peaks

(e.g., months 7–8, reaching approximately 4–5 cm) before declining again. Crucially, the medians (yellow lines) of the predicted values closely track those of the actual values for each month. Furthermore, the interquartile ranges (the boxes themselves) and the overall spread indicated by the whiskers demonstrate a very good overlap and consistency. The lower panel Fig. 9c reinforces this performance. The Actual/Predicted ratios remain centered around 1.0 for all months, with relatively narrow boxes and whiskers, suggesting that the global-irradiance model does not exhibit strong seasonal biases and maintains relatively stable accuracy across the annual cycle.

This visual evidence of high fidelity between predictions and observations directly reinforces the quantitative metrics previously discussed in Figs. 7a and 8. The global irradiance models corresponding to stronger absorption bands consistently exhibited lower training and validation losses (Fig. 7a), coupled with lower RMSE and higher R^2 values (Fig. 8a and b), collectively signify superior predictive precision and strong generalization capabilities on unseen data.

Conversely, Fig. 9b, which depicts models based on only clear-sky direct irradiance, reveals relatively less precise alignment between predicted and actual PWV values. While the general seasonal trend is still discernible, noticeable discrepancies emerge, particularly during months characterized by higher PWV values (e.g., July to September). The predicted medians (red yellow lines) do not consistently align as closely with the actual medians (blue yellow lines) as observed in Fig. 9a. Moreover, the spread of the predicted

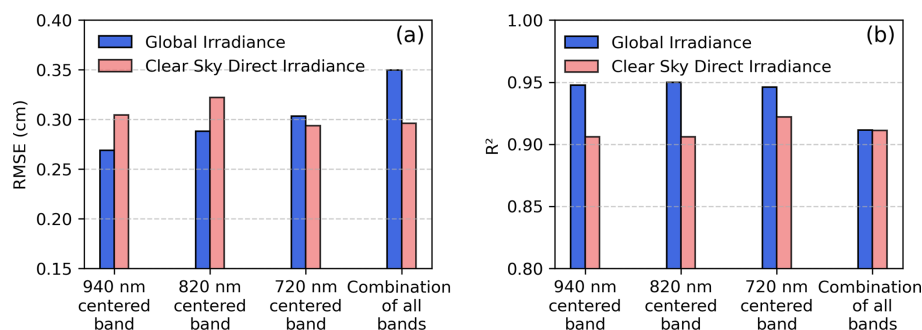


Figure 8. (a) Root Mean Square Error (RMSE) and (b) coefficient of determination (R^2) for models using only global irradiance and only clear-sky direct irradiance as input features, evaluated across wavelength bands centered at 940, 820, 720 nm, and their combination. The predictions were generated using the models that achieved the minimum validation loss, as shown in Fig. 6.

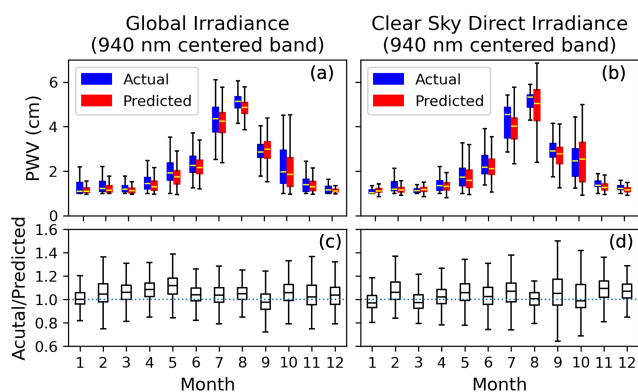


Figure 9. Monthly comparison of predicted and observed PWV values for the spectral band centered at 940 nm, using models based on (a) only global irradiance and (b) only clear-sky direct irradiance as input features; and ratios of actual to predicted values for (c) only global irradiance and (d) only clear-sky direct irradiance as input features. The box represents the 25th and 75th percentiles, the yellow line inside the box indicates the median, and the whiskers extend to the most extreme points within 1.5 times the interquartile range.

values (red boxes and whiskers) sometimes deviates more significantly from the observed spread, indicating that these models struggle to capture the full variability and distribution of PWV as accurately as their global irradiance counterparts. In particular, the discrepancies in Fig. 9b from July to September likely stem from the limited training data during the wet season, when persistent cloud cover greatly reduces the number of usable direct irradiance measurements (Khatri and Takamura, 2009). This behaviour is also reflected in the ratio plot in Fig. 9d. In Fig. 9d, although the median ratio remains close to unity, the interquartile ranges and whiskers are, in general, wider compared with those shown in Fig. 9c, suggesting relatively larger month-to-month deviations and greater uncertainty in the predictions. This visual assessment of reduced precision is entirely consistent with the higher overall loss values observed for clear-sky direct irradiance

models in Fig. 7b, as well as their higher RMSE and lower R^2 values in Fig. 8a and b.

Since DOY is included as one of the input features in our DNN model, we also evaluated SHAP values for the input features – global (or direct) irradiance and DOY – when modelling using spectral global or direct irradiances alone, following the same procedure described for Fig. 5. In both cases, the irradiance-related feature accounts for more than 97%–99% of the total SHAP importance, while the contribution of DOY remains below 2% across all absorption bands. This indicates that the predictive information is overwhelmingly contained in the spectral irradiances themselves, with DOY providing only a very minor contribution.

5 Conclusions

This study highlights the strong potential of surface-observed spectral irradiances for retrieving atmospheric water vapor (PWV) across varying sky conditions, including both clear and cloudy atmospheres. By applying deep learning techniques to spectral irradiance data within three key water vapor absorption bands, we demonstrated that high-fidelity PWV retrieval is achievable even in the presence of atmospheric complexities such as clouds, aerosols, and variable solar geometry. The inclusion of global, direct, and diffuse irradiances provided enhanced sensitivity to radiative transfer characteristics, enabling the model to effectively disentangle water vapor signals from confounding influences, leading to robust performance and accurate seasonal predictions.

Even in observational setups where only global spectral irradiance is available – such as in systems lacking shadow bands – models retained substantial predictive power. These global irradiance-only models consistently exhibited low training and validation losses, along with superior Root Mean Square Error (RMSE) and coefficient of determination (R^2) values, confirming their high predictive precision and strong generalization capabilities. Their monthly predictions of PWV closely aligned with actual observations, accurately capturing seasonal cycles. This finding is particularly valu-

able for expanding the operational viability of PWV monitoring in cost-constrained or logistically limited environments.

In contrast, models trained exclusively on clear-sky direct irradiance generally displayed higher overall losses, larger RMSE values, and lower R^2 values. Their monthly PWV predictions showed less precise alignment with observations, particularly during periods of high water vapor, indicating a struggle to fully capture the data's variability and generalize effectively. This performance disparity underscores a major finding: the volume of training data is a paramount factor influencing model accuracy and generalization. The superior performance of the global irradiance models is directly attributable to the substantially larger and more diverse dataset available for all-sky conditions. Conversely, the clear-sky direct irradiance models were constrained by a significantly smaller dataset, a consequence of stringent filtering to ensure clear-sky purity. This data scarcity limited their capacity to learn sufficient edge cases and diverse scenarios, thereby hindering their generalization and resulting in higher prediction errors.

Across all configurations, the models demonstrated good generalization and stability over seasonal cycles, supporting their applicability in diverse climatic contexts. The results collectively indicate that surface-based spectral observations, when paired with robust machine learning models, offer a scalable and weather-resilient approach to atmospheric water vapor sensing. This capability can complement satellite-based systems, enhance ground-truth validation networks, and contribute to improved understanding of hydrological and radiative processes in Earth's atmosphere. Ultimately, these findings support broader adoption of spectral irradiance-based monitoring strategies in climate science, meteorology, and remote sensing applications, while also emphasizing the critical need for sufficient and diverse training data to unlock a model's full predictive potential.

Data availability. Precipitable water vapor data used in this study are available for download from the SKYNET homepage (<http://atmos3.cr.chiba-u.jp/skyenet/data.html>, last access: 7 January 2026). The spectral irradiance data can be made available upon request to the corresponding author and co-authors.

Author contributions. Conceptualization, PK, TT, and HI; methodology, PK; software, PK; formal analysis, PK; investigation, PK; resources, TT and HI; writing – original draft preparation, PK; writing – review and editing, PK, TT, and HI; funding acquisition, PK. All authors have read and agreed to the published version of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Special issue statement. This article is part of the special issue "SKYNET – the international network for aerosol, clouds, and solar radiation studies and their applications (AMT/ACP inter-journal SI)". It does not belong to a conference.

Acknowledgements. The authors gratefully acknowledge the anonymous reviewers for their valuable suggestions and constructive comments.

Financial support. This research was funded by JSPS KAKENHI Grant Number 24K07129.

Review statement. This paper was edited by Monica Campanelli and reviewed by two anonymous referees.

References

- Alexandrov, M. D., Schmid, B., Turner, D. D., Cairns, B., Oinas, V., Lacis, A. A., Gutman, S. I., Westwater, E. R., Smirnov, A., and Eilers, J.: Columnar water vapor retrievals from multifilter rotating shadowband radiometer data, *Journal of Geophysical Research Atmospheres*, 114, <https://doi.org/10.1029/2008JD010543>, 2009.
- Allan, R. P., Liu, C., Zahn, M., Lavers, D. A., Koukouvas, E., and Bodas-Salcedo, A.: Physically Consistent Responses of the Global Atmospheric Hydrological Cycle in Models and Observations, *Surv. Geophys.*, 35, 533–552, <https://doi.org/10.1007/s10712-012-9213-z>, 2014.
- Anthes, R. A., Bernhardt, P. A., Chen, Y., Cucurull, L., Dymond, K. F., Ector, D., Healy, S. B., Ho, S.-P., Hunt, D. C., Kuo, Y.-H., Liu, H., Manning, K., McCormick, C., Meehan, T. K., Randel, W. J., Rocken, C., Schreiner, W. S., Sokolovskiy, S. V., Snydergaard, S., Thompson, D. C., Trenberth, K. E., Wee, T.-K., Yen, N. L., and Zeng, Z.: THE COSMIC/FORMOSAT-3 MISSION Early Results, *Bull. Am. Meteorol. Soc.*, 89, 313–334, <https://doi.org/10.1175/BAMS-89-3-313>, 2008.
- Araki, K., Murakami, M., Ishimoto, H., and Tajiri, T.: Ground-based microwave radiometer variational analysis during no-rain and rain conditions, *Scientific Online Letters on the Atmosphere*, 11, 108–112, <https://doi.org/10.2151/sola.2015-026>, 2015.
- Bartsevich, M., Rahman, K., Addasi, O., and Ramamurthy, P.: On the Applicability of Ground-Based Microwave Radiometers for Urban Boundary Layer Research, *Sensors*, 24, <https://doi.org/10.3390/s24072101>, 2024.

- Belgiu, M. and Drăguț, L.: Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31, <https://doi.org/10.1016/j.isprsjprs.2016.01.011>, 2016.
- Böck, T., Löffler, M., Marke, T., Pospichal, B., Knist, C., and Löhnert, U.: Instrument uncertainties of network-suitable ground-based microwave radiometers: overview, quantification, and mitigation strategies, *Atmos. Meas. Tech.*, 18, 6251–6270, <https://doi.org/10.5194/amt-18-6251-2025>, 2025.
- Bottou, L.: Large-Scale Machine Learning with Stochastic Gradient Descent, in: *Proceedings of COMPSTAT 2010*, 177–186, https://doi.org/10.1007/978-3-7908-2604-3_16, 2010.
- Campanelli, M., Nakajima, T., Khatri, P., Takamura, T., Uchiyama, A., Estelles, V., Liberti, G. L., and Malvestuto, V.: Retrieval of characteristic parameters for water vapour transmittance in the development of ground-based sun–sky radiometric measurements of columnar water vapour, *Atmos. Meas. Tech.*, 7, 1075–1087, <https://doi.org/10.5194/amt-7-1075-2014>, 2014.
- Chen, D., Guo, H., Gu, X., Wang, J., Liu, Y., Li, Y., and Wu, Y.: Physical-Guided Transfer Deep Neural Network for High-Resolution AOD Retrieval, *Remote Sens.*, 17, <https://doi.org/10.3390/rs17213606>, 2025.
- Chollet, F.: Keras, GitHub, <https://github.com/fchollet/keras> (last access: 7 January 2026), 2015.
- De Mazière, M., Thompson, A. M., Kurylo, M. J., Wild, J. D., Bernhard, G., Blumenstock, T., Braathen, G. O., Hannigan, J. W., Lambert, J.-C., Leblanc, T., McGee, T. J., Nedoluha, G., Petropavlovskikh, I., Seckmeyer, G., Simon, P. C., Steinbrecht, W., and Strahan, S. E.: The Network for the Detection of Atmospheric Composition Change (NDACC): history, status and perspectives, *Atmos. Chem. Phys.*, 18, 4935–4964, <https://doi.org/10.5194/acp-18-4935-2018>, 2018.
- Dessler, A. E.: Observations of Climate Feedbacks over 2000–10 and Comparisons to Climate Models, *J. Clim.*, 1, 333–342, <https://doi.org/10.1175/JCLI-D-11-00640.1>, 2013.
- Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, *Atmos. Meas. Tech.*, 7, 4463–4490, <https://doi.org/10.5194/amt-7-4463-2014>, 2014.
- Dong, S., Li, Y., Zhang, Z., Gou, T., and Xie, M.: A transfer-learning-based windspeed estimation on the ocean surface: implication for the requirements on the spatial-spectral resolution of remote sensors, *Applied Intelligence*, 54, 7603–7620, <https://doi.org/10.1007/s10489-024-05523-w>, 2024.
- Elgered, G. and Jarlemark, P. O. J.: Ground-based microwave radiometry and long-term observations of atmospheric water vapor, *Radio Sci.*, 33, 707–717, <https://doi.org/10.1029/98RS00488>, 1998.
- Gueymard, C.: SMARTS2, A Simple Model of the Atmospheric Radiative Transfer of Sunshine: Algorithms and performance assessment Simple Model for the Atmospheric Radiative Transfer of Sunshine (SMARTS2) Algorithms and performance assessment, Report Number: FSEC-PF-270-95, <https://publications.energyresearch.ucf.edu/wp-content/uploads/2018/06/FSEC-PF-270-95.pdf> (last access: 7 January 2026), 1995.
- Gupta, S., Park, Y., Bi, J., Gupta, S., Züfle, A., Wildani, A., and Liu, Y.: Spatial Transfer Learning for Estimating PM_{2.5} in Data-poor Regions, arXiv [preprint], <https://doi.org/10.48550/arXiv.2404.07308>, 2024.
- Harrison, L., Michalsky, J., and Berndt, J.: Automated multifilter rotating shadow-band radiometer: an instrument for optical depth and radiation measurements, *Appl. Opt.*, 33, 5118–5125, <https://doi.org/10.1364/AO.33.005118>, 1994.
- Hashimoto, M., Nakajima, T., Dubovik, O., Campanelli, M., Che, H., Khatri, P., Takamura, T., and Pandithurai, G.: Development of a new data-processing method for SKYNET sky radiometer observations, *Atmos. Meas. Tech.*, 5, 2723–2737, <https://doi.org/10.5194/amt-5-2723-2012>, 2012.
- Held, I. M. and Soden, B. J.: Water vapor feedback and global warming, *Annual Review of Energy and the Environment*, 25, 441–475, <https://doi.org/10.1146/annurev.energy.25.1.441>, 2000.
- Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A Federated Instrument Network and Data Archive for Aerosol Characterization, *Remote Sens. Environ.*, 66, 1–16, [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5), 1998.
- Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendaal, M.: Eight-component retrievals from ground-based MAX-DOAS observations, *Atmos. Meas. Tech.*, 4, 1027–1044, <https://doi.org/10.5194/amt-4-1027-2011>, 2011.
- Khatri, P. and Takamura, T.: An algorithm to screen cloud-affected data for sky radiometer data analysis, *Journal of the Meteorological Society of Japan*, 87, 189–204, <https://doi.org/10.2151/jmsj.87.189>, 2009.
- Khatri, P., Takamura, T., Yamazaki, A., and Kondo, Y.: Retrieval of key aerosol optical parameters from spectral direct and diffuse: Irradiances observed by a radiometer with nonideal cosine response characteristic, *J. Atmos. Ocean Technol.*, 29, 683–696, <https://doi.org/10.1175/JTECH-D-11-00111.1>, 2012.
- Khatri, P., Takamura, T., Shimizu, A., and Sugimoto, N.: Observation of low single scattering albedo of aerosols in the downwind of the East Asian desert and urban areas during the inflow of dust aerosols, *J. Geophys. Res.*, 119, 787–802, <https://doi.org/10.1002/2013JD019961>, 2014.
- Khatri, P., Irie, H., Takamura, T., and Letu, H.: Optical characteristics of aerosols and clouds studied by using ground-based SKYNET and satellite remote sensing data, *IEEE International Geoscience and Remote Sensing*, 377–380, <https://doi.org/10.1109/IGARSS.2016.7729092>, 2016.
- Khatri, P., Iwabuchi, H., Hayasaka, T., Irie, H., Takamura, T., Yamazaki, A., Damiani, A., Letu, H., and Kai, Q.: Retrieval of cloud properties from spectral zenith radiances observed by sky radiometers, *Atmos. Meas. Tech.*, 12, 6037–6047, <https://doi.org/10.5194/amt-12-6037-2019>, 2019.
- Kiehl, J. T. and Trenberth, K. E.: Earth’s Annual Global Mean Energy Budget, *Bull. Am. Meteorol. Soc.*, 78, 197–208, [https://doi.org/10.1175/1520-0477\(1997\)078<0197:EAGMEB>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2), 1997.
- Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in: *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, arXiv [preprint], <https://doi.org/10.48550/arXiv.1412.6980>, 2015.

- Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, *Commun. ACM*, 60, 84–90, <https://doi.org/10.1145/3065386>, 2017.
- Lecun, Y., Bengio, Y., and Hinton, G.: Deep learning, *Nature*, 521, 436–444, <https://doi.org/10.1038/nature14539>, 2015.
- Li, H., Wang, X., Wu, S., Zhang, K., Chen, X., Qiu, C., Zhang, S., Zhang, J., Xie, M., and Li, L.: Development of an Improved Model for Prediction of Short-Term Heavy Precipitation Based on GNSS-Derived PWV, *Remote Sens.*, 12, 1–22, <https://doi.org/10.3390/rs12244101>, 2020.
- Löhnert, U. and Maier, O.: Operational profiling of temperature using ground-based microwave radiometry at Payerne: prospects and challenges, *Atmos. Meas. Tech.*, 5, 1121–1134, <https://doi.org/10.5194/amt-5-1121-2012>, 2012.
- Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1705.07874>, 2017.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S. I.: From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.*, 2, 56–67, <https://doi.org/10.1038/s42256-019-0138-9>, 2020.
- Minowa, M., Araki, K., and Takashima, Y.: Compact Microwave Radiometer for Water Vapor Estimation with Machine Learning Method, *Scientific Online Letters on the Atmosphere*, 20, 339–346, <https://doi.org/10.2151/sola.2024-045>, 2024.
- Mizobuchi, S., Irie, H., and Shimizu, S.: Long-term continuous observations of the horizontal inhomogeneity in lower-atmospheric water vapor concentration using A-SKY/MAX-DOAS, *Prog. Earth Planet Sci.*, 12, <https://doi.org/10.1186/s40645-025-00724-4>, 2025.
- Mountrakis, G., Im, J., and Ogole, C.: Support vector machines in remote sensing: A review, *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 247–259, <https://doi.org/10.1016/j.isprsjprs.2010.11.001>, 2011.
- Muller, C. J., Back, L. E., O’Gorman, P. A., and Emanuel, K. A.: A model for the relationship between tropical precipitation and column water vapor, *Geophys Res Lett*, 36, <https://doi.org/10.1029/2009GL039667>, 2009.
- Nakajima, T., Campanelli, M., Che, H., Estellés, V., Irie, H., Kim, S.-W., Kim, J., Liu, D., Nishizawa, T., Pandithurai, G., Soni, V. K., Thanu, B., Tugjurn, N.-U., Aoki, K., Go, S., Hashimoto, M., Higurashi, A., Kazadzis, S., Khatri, P., Kouremeti, N., Kudo, R., Marengo, F., Momoi, M., Ningombam, S. S., Ryder, C. L., Uchiyama, A., and Yamazaki, A.: An overview of and issues with sky radiometer technology and SKYNET, *Atmos. Meas. Tech.*, 13, 4195–4218, <https://doi.org/10.5194/amt-13-4195-2020>, 2020.
- Padmanabhan, S., Reising, S. C., Vivekanandan, J., and Iturbide-Sanchez, F.: Retrieval of atmospheric water vapor density with fine spatial resolution using three-dimensional tomographic inversion of microwave brightness temperatures measured by a network of scanning compact radiometers, *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3708–3721, <https://doi.org/10.1109/TGRS.2009.2031107>, 2009.
- Pan, S. J. and Yang, Q.: A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, 22, 1345–1359, <https://doi.org/10.1109/TKDE.2009.191>, 2010.
- Pérez-Ramírez, D., Whiteman, D. N., Smirnov, A., Lyamani, H., Holben, B. N., Pinker, R., Andrade, M., and Alados-Arboledas, L.: Evaluation of AERONET precipitable water vapor versus microwave radiometry, GPS, and radiosondes at ARM sites, *J. Geophys. Res.*, 119, 9596–9613, <https://doi.org/10.1002/2014JD021730>, 2014.
- Qiao, C., Liu, S., Huo, J., Mu, X., Wang, P., Jia, S., Fan, X., and Duan, M.: Retrievals of precipitable water vapor and aerosol optical depth from direct sun measurements with EKO MS711 and MS712 spectroradiometers, *Atmos. Meas. Tech.*, 16, 1539–1549, <https://doi.org/10.5194/amt-16-1539-2023>, 2023.
- Schmidt, G. A., Ruedy, R. A., Miller, R. L., and Lacis, A. A.: Attribution of the present-day total greenhouse effect, *Journal of Geophysical Research Atmospheres*, 115, <https://doi.org/10.1029/2010JD014287>, 2010.
- Schröder, M., Lockhoff, M., Shi, L., August, T., Bennartz, R., Brogniez, H., Calbet, X., Fell, F., Forsythe, J., Gambacorta, A., Ho, S.-P., Kursinski, E. R., Reale, A., Trent, T., and Yang, Q.: The GEWEX Water vapor assessment: Overview and introduction to results and recommendations, *Remote Sens.*, 11, <https://doi.org/10.3390/rs11030251>, 2019.
- Seidel, D. J., Berger, F. H., Diamond, H. J., Dykema, J., Goodrich, D., Immler, F., Murray, W., Peterson, T., Sister-son, D., Sommer, M., Thorne, P., Vömel, H., and Wang, J.: Reference upper-air observations for climate: Rationale, progress, and plans, *Bull. Am. Meteorol. Soc.*, 90, 361–369, <https://doi.org/10.1175/2008BAMS2540.1>, 2009.
- Sengupta, M., Clothiaux, E. E., and Ackerman, T. P.: Climatology of Warm Boundary Layer Clouds at the ARM SGP Site and Their Comparison to Models, *J. Clim.*, 17, 4760–4782, <https://doi.org/10.1175/JCLI-3231.1>, 2004.
- Takamura, T. and Khatri, P.: Uncertainties in Radiation Measurement Using a Rotating Shadow-Band Spectroradiometer, *Journal of the Meteorological Society of Japan. Ser. II*, 99, 1547–1561, <https://doi.org/10.2151/jmsj.2021-075>, 2021.
- Trenberth, K. E., Dai, A., Rasmussen, R. M., and Parsons, D. B.: The changing character of precipitation, 1205–1218, <https://doi.org/10.1175/BAMS-84-9-1205>, 2003.
- Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth’s Global Energy Budget, *Bull. Am. Meteorol. Soc.*, 90, 311–324, <https://doi.org/10.1175/2008BAMS2634.1>, 2009.
- Van Baelen, J., Reverdy, M., Tridon, F., Labbouz, L., Dick, G., Bender, M., and Hagen, M.: On the relationship between water vapour field evolution and the life cycle of precipitation systems, *Quarterly Journal of the Royal Meteorological Society*, 137, 204–223, <https://doi.org/10.1002/qj.785>, 2011.
- Weiss, K., Khoshgoftaar, T. M., and Wang, D. D.: A survey of transfer learning, *J. Big Data*, 3, <https://doi.org/10.1186/s40537-016-0043-6>, 2016.
- Xu, G.: A Review of Remote Sensing of Atmospheric Profiles and Cloud Properties by Ground-Based Microwave Radiometers in Central China, *Remote Sens.*, 16, <https://doi.org/10.3390/rs16060966>, 2024.
- Zhang, H., Yuan, Y., Li, W., and Zhang, B.: A real-time precipitable water vapor monitoring system using the national GNSS network of China: Method and preliminary results, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 12, 1587–1598, <https://doi.org/10.1109/JSTARS.2019.2906950>, 2019.

Zhao, X., Frech, J., Foster, M. J., and Heidinger, A. K.: Studying the Aerosol Effect on Deep Convective Clouds over the Global Oceans by Applying Machine Learning Techniques on Long-Term Satellite Observation, *Remote Sens.*, 16, <https://doi.org/10.3390/rs16132487>, 2024.

Zheng, L., Lin, R., Wang, X., and Chen, W.: The development and application of machine learning in atmospheric environment studies, *Remote Sens.*, 13, <https://doi.org/10.3390/rs13234839>, 2021.