remote sensing

MDPI

Article

Machine Learning-Based Estimation of Hourly GNSS Precipitable Water Vapour

Zohreh Adavi [1,*], Babak Ghassemi [2], Robert Weber [1] and Natalia Hanna [1]

[1] Department of Geodesy and Geoinformation, TU Wien, Wiedner Hauptstraße 8, 1040 Vienna, Austria; robert.weber@tuwien.ac.at (R.W.); natalia.hanna@tuwien.ac.at (N.H.)
[2] Institute of Geomatics, University of Natural Resources and Life Sciences, BOKU, Peter-Jordan-Straße 82, 1190 Vienna, Austria; babak.ghassemi@boku.ac.at
* Correspondence: zohreh.adavi@tuwien.ac.at

Abstract: Water vapour plays a key role in long-term climate studies and short-term weather forecasting. Therefore, to understand atmospheric variations, it is crucial to observe water vapour and its spatial distribution. In the current era, Global Navigation Satellite Systems (GNSS) are widely used to monitor this critical atmospheric component because GNSS signals pass through the atmosphere, allowing us to estimate water vapour at various locations and times. The amount of precipitable water vapour (*PWV*) is one of the most fascinating quantities, which provides meteorologists and climate scientists with valuable information. However, calculating *PWV* accurately from processing GNSS observations usually requires the input of further observed meteorological parameters with adequate quality and latency. To bypass this problem, hourly *PWV*s without meteorological parameters are computed using the Random Forest and Artificial Neural Network algorithms in this research. The first step towards this objective is establishing a regional weighted mean temperature model for Austria. To achieve this, measurements of radiosondes launched from different locations in Austria are employed. The results indicate that Random Forest is the most accurate method compared to regression (linear and polynomial), Artificial Neural Network, and empirical methods. *PWV* models are then developed using data from 39 GNSS stations that cover Austria's entire territory. The models are afterwards tested under different atmospheric conditions with four radiosonde stations. Based on the obtained results, the Artificial Neural Network model with a single hidden layer slightly outperforms other investigated models, with only a 5% difference in mean absolute error. As a result, the hourly *PWV* can be estimated without relying on measured meteorological parameters with an average mean absolute error of less than 2.5 mm in Austria.

Keywords: weighted mean temperature; precipitable water vapour; GNSS; machine learning

Citation: Adavi, Z.; Ghassemi, B.; Weber, R.; Hanna, N. Machine Learning-Based Estimation of Hourly GNSS Precipitable Water Vapour. *Remote Sens.* **2023**, *15*, 4551. https://doi.org/10.3390/rs15184551

Academic Editors: Olivier Bock, Tong Ning, Galina Dick and Florian Zus

Received: 28 July 2023
Revised: 3 September 2023
Accepted: 5 September 2023
Published: 15 September 2023

1. Introduction

Water vapour is a crucial component of the earth's atmosphere, and it plays a significant role in climate and weather systems. As a result, monitoring this variable atmospheric greenhouse gas has a direct impact on short-term and long-term studies of the weather, including climate change and weather forecast. Precipitable Water Vapour (*PWV*) is a valuable water vapour product defined as the vertically integrated atmospheric water vapour in a column of a unit area [1,2] and can be estimated with conventional techniques, such as radiosonde measurements, water vapour radiometer, or data derived from numerical weather models. There are still some drawbacks to these techniques, such as their high cost and poor spatial-temporal resolution, even though they have many advantages. These limitations can be mitigated by GNSS due to its continuous scanning of the troposphere at a low cost with a higher spatial-temporal resolution in all-weather conditions [1,3–5]. Thereby, GNSS observations can be used to estimate *PWV* with long-term stability, high accuracy, and high spatiotemporal resolution due to the development of GNSS technology and the increasing number of permanently tracking GNSS receivers [1,6–10].

Calculating GNSS PWV requires estimation of Zenith Wet Delay (ZWD). To do so, GNSS measurements are first processed with the GNSS software (here, Bernese version 5.2 [11]) to estimate Zenith Tropospheric Delay (ZTD). Then, the Zenith Hydrostatic Delay (ZHD) can be calculated by utilizing well-known hydrostatic models such as Saastamoinen by introducing the required meteorological parameters [12]. In the next step, ZWD is obtained by subtracting ZHD from ZTD, and then ZWD is converted to PWV using a conversion factor (Π) [1,13]. Since this factor strongly depends on weighted mean temperature (T_m), PWV estimation is affected directly by T_m, whose value differs depending on the geographical location (φ, λ, h), season, and weather condition [7,14,15].

One of the earliest and most influential works on GPS meteorology was by Bevis et al. [16] which paved the way for further research on GNSS-based PWV estimation and modelling. The authors found that GPS data provided accurate and continuous measurements of PWV that could potentially be used to enhance the Numerical Weather Model accuracy, predicting severe rainstorms and studying the climate [8,17–24]. Based on this exemplary study, some other researchers have done valuable work in the past three decades to model PWV with different methods at various spatial scales and temporal resolutions. In 2019, Zhang et al. [25] developed a real-time GNSS PWV monitoring system utilizing the Chinese national GNSS network. Real-time GNSS $ZTDs$ were generated with high accuracy and performance using the precise point positioning technique. A validation of the GPS PWV using radiosonde data found the system to be highly accurate, with a mean bias of -0.1 mm and root mean square of 1.7 mm. A method for analysing PWV using a low-cost GNSS receiver installed onboard the ISABU vessel, operated by the Korea Institute of Ocean Science and Technology, was developed by Sohn et al., (2020) [26]. Radiosonde data and GNSS PWV were found to be in close agreement in their validation. In spite of their results, shipborne GNSS offers the potential for accurate PWV derivation via kinematic precise point positioning. In 2022, Li et al. [27] used multiple linear regression to establish multi-factor PWV conversion models, focusing on the correlations between PWV, ZTD, surface temperature, and atmospheric pressure. Their results indicated that the multi-factor models significantly enhanced the models' robustness and applicability across the China region.

Some researchers have also explored the use of machine learning methods to model PWV. For example, a generalized regression neural network was used by Zhang and Yao in 2021 [28] to fuse PWV from GNSS, Moderate-Resolution Imaging Spectroradiometer (MODIS), and European Centre for Medium-Range Weather Forecasts Reanalysis 5 (ERA5) in North America. In the proposed method, sparse high-quality data could be used to enrich low-quality data to enhance their quality and utility. Consequently, a homogeneous PWV product can be derived with reduced temporal and spatial variations in accuracy, thus raising data usability for weather and climate monitoring. A stacked machine learning model was proposed in 2022 by Zheng et al. [29] to map ZTD into PWV without meteorological parameters at the global scale. To validate the PWV values, ERA5 and radiosondes measurements were utilized, which showed a consistency within 2.5 mm. According to the study results, time-critical meteorological applications can benefit from this model, and other methods that are capable of sensing ZTD can take advantage of it for determining water vapour in real -time.

As already stated, T_m is a crucial parameter in PWV retrieval. In general, T_m modelling methods can be categorized into surface meteorological factor models and non-meteorological factor models based on the use of surface meteorological data [30,31]. In the first method, a linear or nonlinear model of T_m is developed based on the relationships between this parameter and surface meteorological variables, such as surface temperature (T_s), pressure (P_s), and water vapour pressure (e_s) [31]. The seminal study by Bevis et al. (1992) [1] pioneered establishing the linear model between T_m and T_s, which has been the basis for water vapour modelling studies since then, see for instance [7,32–37]. In addition, this technique has also been used in numerous studies to obtain a more accurate global or local model of T_m [19,37–41]. For the second method, instead of using in situ meteorological

data, other variables such as geographical location and time are taken into account to model T_m [30,31,42]. There have been a number of empirical models developed in recent years, including different models of Global Pressure and Temperature (GPT) [43–45], Global Weighted Mean Temperature [46,47], Global Tropospheric Model [48], Hourly Global Pressure and Temperature model [49], GTm_R [34] model, and the Global Weighted Mean Temperature model [50].

In the following, we summarize some recent studies on modelling T_m on local and global scales. In 2017, Manandhar et al. [35] proposed a simplified model for the conversion factor for retrieval of PWV from GPS signals based on latitude and day of year. The proposed model was compared with temperature-dependent models using data from 174 stations. PWV values derived from the proposed model were validated with temperature-dependent models using three databases. Based on the obtained results, the simplified model can be applied universally, since it is computationally efficient, and has high accuracy for a wide range of geodetic applications. In 2020, Wan et al. [51] utilized radiosonde profiles from 12 Antarctic and 58 Arctic stations between 2008 and 2015 to establish a model relating T_m and T_s in polar regions. Two regional T_m models were developed, one based on linear regression and the other based on the quadratic function. These models were found to be more accurate than the global GPT2w (GPT 2 wet) [44] model. Even though the quadratic function T_m model had slightly a higher accuracy, both could be used for retrieving PWVs. In 2021, Long, Hu et al. [42] established T_m models over China using the neural network technique with three different schemes, namely, a non-meteorological-factor T_m model, a single-meteorological-factor T_m model and a multi-meteorological-factor T_m model. In this study, the three-layer feedforward neural network method was used with the help of an ensemble learning in order to combine multiple models and consequently improve the accuracy and robustness of the predictions. The new models were found to be very capable of capturing regional spatiotemporal variations in T_m and simulating the interactions between T_m and a variety of surface meteorological parameters. Zhao et al. [52] developed a global conversion factor model of non-meteorological parameters using gridded data taken from the 2006 to 2014 Global Geodetic Observing System. In addition to the geographical location, the changes in atmospheric water vapour on an annual basis, semiannually, and day by day are taken into account in this model. Based on European Centre for Medium-Range Weather Forecasts Reanalysis (ERA5) reanalysis data from 1990 to 2018, Zhang et al. [31] proposed a principal component analysis method for modelling T_m in Greenland where in situ meteorological data are limited. Their model validated using 11 radiosonde stations (2015–2019) showed a bias of -0.110 K and root mean square error of 4.447 K. Additionally, the principal component analysis model performed better than GPT3 and Global Tropospheric Model, required fewer parameters, and improved computation efficiency.

Since Artificial Neural Networks (ANNs) and Random Forest (RF) are well-established machine learning algorithms that have the potential to learn complex relationships between input and output variables and can handle high-dimensional data well, this study is motivated to explore their effectiveness for T_m and PWV models. Therefore, based on three radiosonde stations in this study, the T_m model for Austria is developed using RF, ANN, and linear regression models. We then analyse the model performance against the radiosonde control station along with the Bevis and GPT3 models. Next, the hourly PWV for 39 GNSS stations located in the EPOSA (Echtzeit Positionierung Austria) network is modelled using RF and ANN algorithms. In the next step, the retrieved PWV is validated by the test period and four radiosonde stations in different period (August 2022 to April 2023).

This paper is organized as follows. Section 2 describes the data sources that were used for this study, including the GNSS ZTD, radiosonde observations, and the ERA5 model. Section 3 discusses the techniques that are used for modelling T_m and PWV. We investigate both the T_m and PWV data series in Section 4. Finally, we present our outlook and conclusions gained from this study in Section 5.

2. Data

In this section, we first describe the GNSS ZTD time series employed in this study. Then, radiosonde observations are defined, which are used for both T_m modelling and PWV validation. In the end, the pre-processing step is described, which removes outliers in the data. Detailed information on all the datasets used in this study is provided in Tables 1 and 2.

Table 1. Information of the used data to model T_m and PWV parameters.

Dataset	Period	Spatial Resolution	Temporal Resolution	Usage
GNSS	Train: Jan 2018–Jun 2021 Test: Jul 2021–Jul 2022	39 Stations	1 h	PWV modelling
RS *	Train (T_m): Jan 2010–Jun 2021 Test (T_m): Jul 2021–Jul 2022 Predict (T_m): Aug 2022–April 2023 Predict (PWV): Aug 2022–April 2023	4 Stations	3, 6, 12 h (see Table 2)	T_m modelling and PWV external validation
ERA5	Jul 2021, April 2023	0.25° × 0.25°	1 h	PWV assimilation (External validation)

* Radiosonde.

Table 2. Information on the used radiosonde stations in this study.

Station	Lat (deg)	Lon (deg)	H (metre)	Period	No. Launch 00:00	No. Launch 06:00	No. Launch 12:00	No. Launch 18:00
11035	48.25	16.36	196	Jan 2010–April 2023	5203	253	5203	139
11240	46.99	15.45	330	Jan 2010–April 2023	6006	66	39	1
11120	47.26	11.35	579	Jan 2010–April 2023	5816	111	65	40
11010	48.23	14.18	313	August 2022–April 2023	450	4	2	0

2.1. GNSS ZTDs

A total of 39 permanent GNSS stations of the EPOSA network are considered in this study for the period between June 2018 and July 2022. GNSS sites in this network range in height from 172 m to 2215 m, extending from eastern to western Austria. The hourly GNSS ZTD data are used in this study over two different time periods: training (January 2018–June 2021) and testing (July 2021–July 2022). To estimate ZTD from GNSS measurements, we have used the Bernese software version 5.2 employing double-difference mode (please see [11,53] for more details). In Figure 1, the distribution of the various multi-GNSS (GPS + GLONASS) stations is visualized along with the location of the four control radiosonde stations (RS11035, RS11120, RS11010, and RS11240).

2.2. Radiosonde Observations

Radiosondes are highly accurate standard sensors in meteorology that measure a variety of meteorological parameters, including temperature ($\sigma_T = \pm 0.5$ °C), pressure ($\sigma_P = [\pm 1, \pm 2]$ hpa), relative humidity ($\sigma_{RH} = \pm 5\%$), and geopotential height [54–56]. Figure 2 depicts the distribution of radiosonde stations on the Austrian territory used in this study. To train and test the regional T_m model, we take into account the period of January 2010–June 2022. In addition, it should be highlighted that radiosonde profiles associated with the radiosonde stations have been downloaded free of charge via the website of the University of Wyoming. It is important to note that the radiosonde stations shown here correspond to the radiosonde stations in Figure 1. However, three of them are used for training while one is used to test the T_m model.

Table 2 shows the location of the radiosonde stations in the case study, as well as the number of launches at different times (00, 06, 12, and 18) over the experimental period.

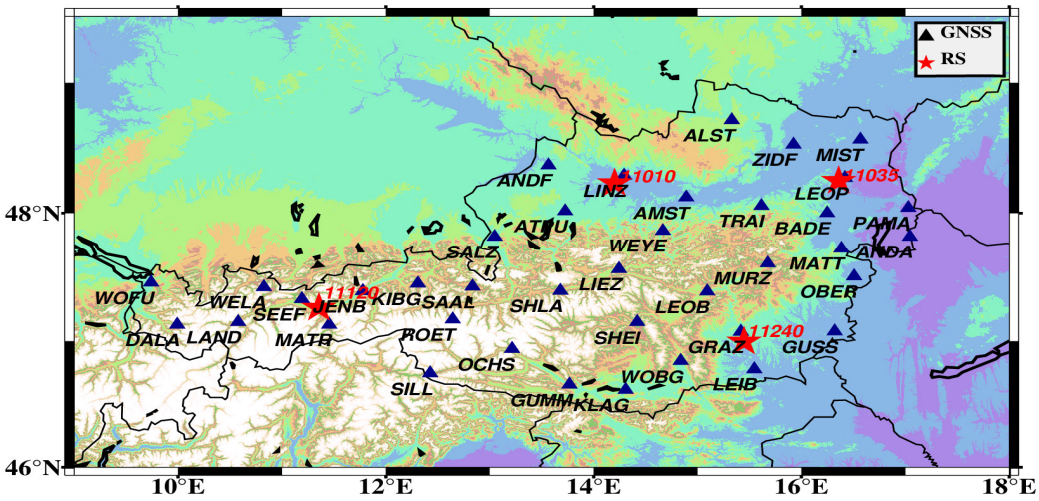

Figure 1. Distribution of GNSS stations and four radiosonde control stations in Austria.

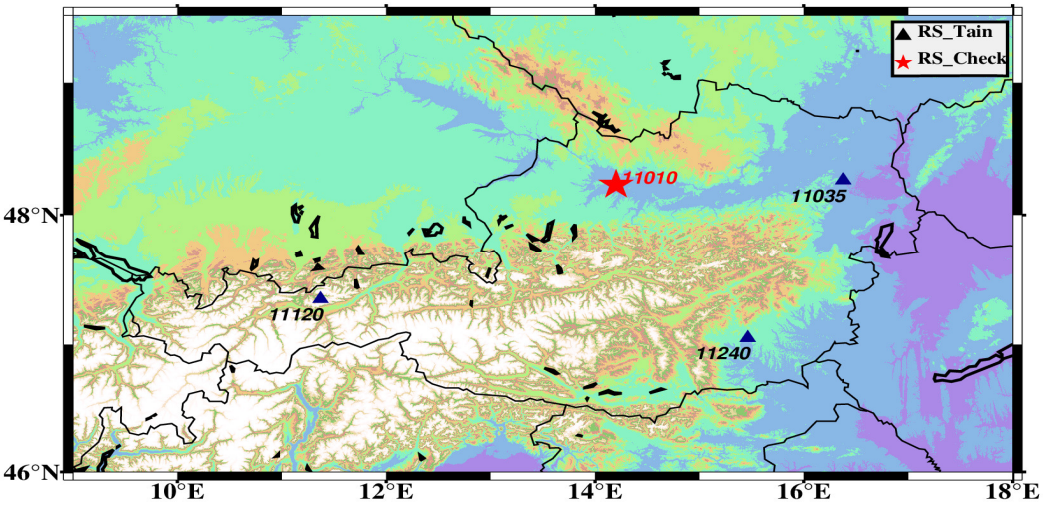

Figure 2. Distribution of three radiosonde stations to train the T_m model and one radiosonde control station to test the T_m model in Austria.

2.3. Data Pre-Processing

The presence of outliers in the measurements is inevitable and can negatively affect the performance and accuracy of a model if they are not addressed appropriately [30,57–59]. In data pre-processing, the Interquartile Range (IQR) method is widely used to detect outliers. It is based on the idea of quartiles, namely the first quartile (Q_1), the second quartile (Q_2), and the third quartile (Q_3). Based on this, the IQR is defined as the difference between Q_1 and Q_3, as follows:

$$IQR = Q_3 - Q_1 \tag{1}$$

In the next step, using Equations (2) and (3), the lower and upper bounds for the data, Lb and Ub, are determined as follows:

$$Lb = Q_1 - (const \times IQR) \tag{2}$$

$$Ub = Q_3 + (const \times IQR) \tag{3}$$

where *const* is usually equal to 1.5. All data outside these bands are considered outliers and removed from the dataset. As a result, model performance will be enhanced.

3. Methodology

First, we describe Tm modelling using different techniques and data sources. Next, we discuss modelling with GNSS and radiosonde data sources. The main characteristics and features we employ in the RF and ANN models for T_m and PWV parameters are also described in those sections since we will apply these methods to model the T_m and PWV parameters later in this study. The main strategy of this research is demonstrated in Figure 3.

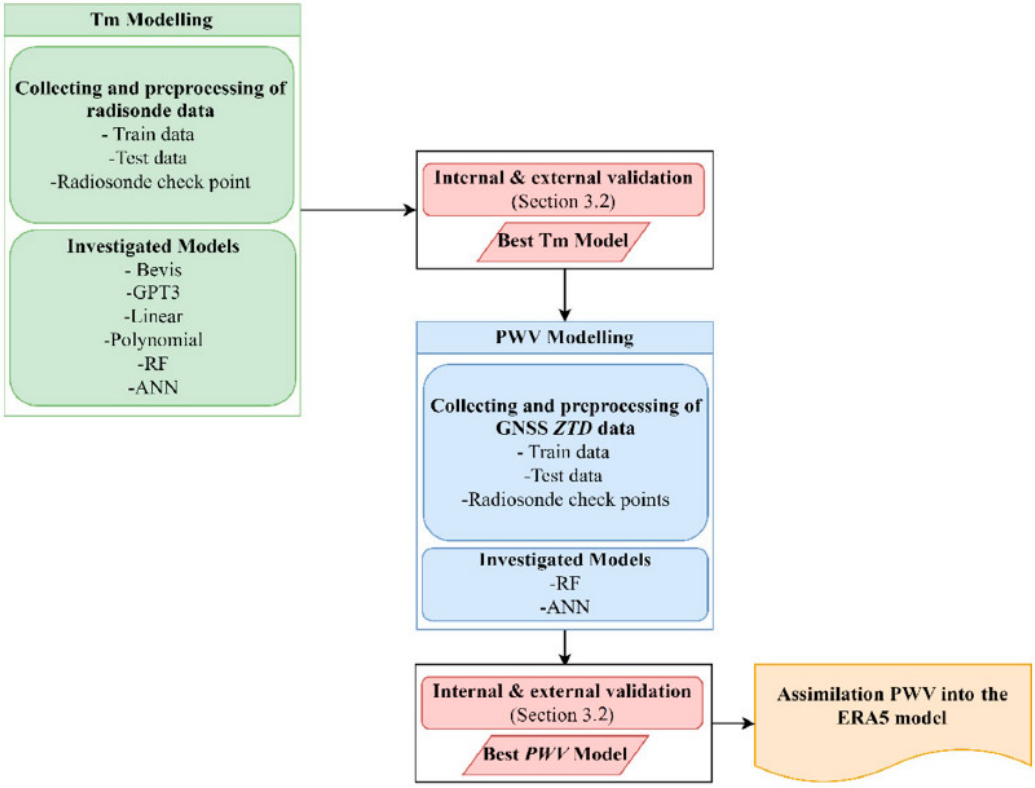

Figure 3. Main procedure for estimating T_m and PWV in this research.

3.1. T_m Modelling

Here, Bevis and GPT3, two of the most commonly used models, are described with their main properties and formulas. Next, we present four different regional models based on regression (linear and polynomial) and machine learning for Austria.

In general, by considering the meteorological data of a radiosonde profile in different layers, T_m (K) can be calculated as follows [42,60]:

$$T_m = \frac{\sum_{i=1}^{n}\left(\frac{e}{T}\Delta h\right)}{\sum_{i=1}^{n}\left(\frac{e}{T^2}\Delta h\right)} \quad (4)$$

here, e (hpa) and T (K) are water vapor pressure and temperature derived from the radiosonde profile, respectively; Δh (m) and n refer to the height difference between the adjacent data points, and the total number of layers, respectively.

3.1.1. Empirical Models

As part of this study, we examine and compare the accuracy of two widely-used T_m models, namely the Bevis and GPT3 models. The primary features of the Bevis and GPT3 models are outlined in Table 3.

Table 3. Characteristics of the two selected models, Bevis and GPT3.

Model	Inputs	Scale	Trained Data
Bevis	T_s	Global/Regional	RS
GPT3	$\varphi, \lambda, h, mjd^*$	Global	ECMWF, VLBI

* Modified Julian date.

- Bevis model

The Bevis model is an empirical model that estimates T_m using the linear relationship between T_m and T_s as follows [1]:

$$T_m[K] = 0.72\ T_s[K] + 70.2 \tag{5}$$

To determine this model, Bevis et al. [1] utilized data from 8718 radiosonde stations in the United States over the years 1990–1991. In spite of the fact that this model has been widely used for many years in GNSS meteorology, it has some limitations, such as its simplicity as well as time and location independence.

- GPT3 model

GPT3 is the latest version of the empirical model GPT provided on a global grid of $5° \times 5°$ and $1° \times 1°$ [61,62]. Various meteorological parameters can be obtained through this model, including pressure, temperature, and weighted mean temperature. Due to its high accuracy and simplicity [61], GPT3 is widely used in geodetic and meteorological fields. In Equation (6), the formula for estimating T_m for a given doy (day of the year) is provided [45,61]:

$$T_m[k] = A_0 + A_1\cos\left(\frac{doy}{365.25}\,2\pi\right) + B_1\sin\left(\frac{doy}{365.25}\,2\pi\right) + A_2\cos\left(\frac{doy}{365.25}\,4\pi\right) + B_2\sin\left(\frac{doy}{365.25}\,4\pi\right) \tag{6}$$

A_0 represents the mean T_m value, while (A_1, B_1) and (A_2, B_2) represent the annual and semi-annual variations. These coefficients were estimated by means of an least-square adjustment using ERA-Interim data from ECMWF covering the years 2001–2010 [62]. The GPT3 model can then be used to interpolate all different meteorological parameters, including T_m, for a specific location and time. In spite of this, it is still an empirical model, and thus it may not accurately reflect the actual state of the atmosphere in different regions or weather conditions [34,41,56].

3.1.2. Developed T_m Models

To develop a more accurate and consistent model for Austria, we established four models. Two of these models use traditional regression techniques, linear and polynomial, and the other two models use machine learning techniques, RF and ANN. The main characteristics of these models are presented in Table 4.

Table 4. Characteristics of the developed models based on traditional regression and machine learning techniques.

Model	Features
Linear	T_s
Polynomial	φ, h, T_s
RF	$\varphi, h, T_s, doy, time$
ANN	$\varphi, h, T_s, doy, time$

- Linear Model

As shown below, the general form of this model is based solely on surface temperature (T_s):

$$T_m[K] = a\ T_s[K] + b \tag{7}$$

For the calculation of the a and b coefficients, we used data obtained from three radiosonde stations (RS11035, RS11240, RS1120) located in different parts of Austria between January 2010 and June 2021 (see Figure 2). By deriving linear regression coefficients, Equation (7) can be fitted for the Austrian region as shown below:

$$T_m = 0.76\ T_s + 59.94 \tag{8}$$

- Polynomial Model

As follows, this model is defined by the geographic location of the desired point (φ, h) plus T_s:

$$T_m[K] = a_0\ \varphi[deg] + a_1 h[m] + a_2\ T_s[k] + a_3 \tag{9}$$

Using the same data (RS11035, RS11240, RS1120) from January 2010 to June 2021, the unknown coefficients (a_0, a_1, a_2, a_3) were derived using the least-square method. Introducing these coefficients, the regional model for Austria states as follows:

$$T_m = -2.65\ \varphi + 0.03\ h + 0.78\ T_s + 180.20 \tag{10}$$

- Machine Learning Models (RF and ANN)

In the 1950s, artificial intelligence pioneer Arthur Samuel defined machine learning as a subset of artificial intelligence [63–65]. According to Samuel's definition, it is the field of study that enables computers to acquire knowledge without explicit programming [63]. Plenty of complex problems can be modelled with this method, such as image classification, object detection, crop monitoring, wind speed prediction, groundwater modelling, and climate modelling [66–69]. A number of scholars have used machine learning methods to model T_m on global and regional scales (e.g., [56,70,71]). To enhance the quality of T_m for Austria, we also apply two well-known machine learning techniques, namely RF and ANN.

Accordingly, the following model is considered for both RF and ANN algorithms:

$$T_m = f(T_s, \varphi, h, doy, time) \tag{11}$$

here, $f()$ refers to the machine-learning based estimation model for T_m. In order to obtain the machine learning models, again we used measurements at three different radiosonde stations (RS11035, RS11240, RS1120) located across Austria between January 2010 and June 2021 (see Figure 2).

- RF Method

Beriman proposed the RF method as a supervised machine learning algorithm in 2001 [72], which can be used in classification and regression problems (for e.g., [71,73–76]). To perform this method, a set of decision trees is created using random subsets of data and features [72,77]. By creating such an uncorrelated set of trees, the risk of bias and overfitting is reduced [72]. Once multiple decision trees have been calculated, the outputs are then summed for regression and voted for classification to come up with a single result [72,78]. A schematic diagram of how this method is used to estimate the final output is shown in Figure 4.

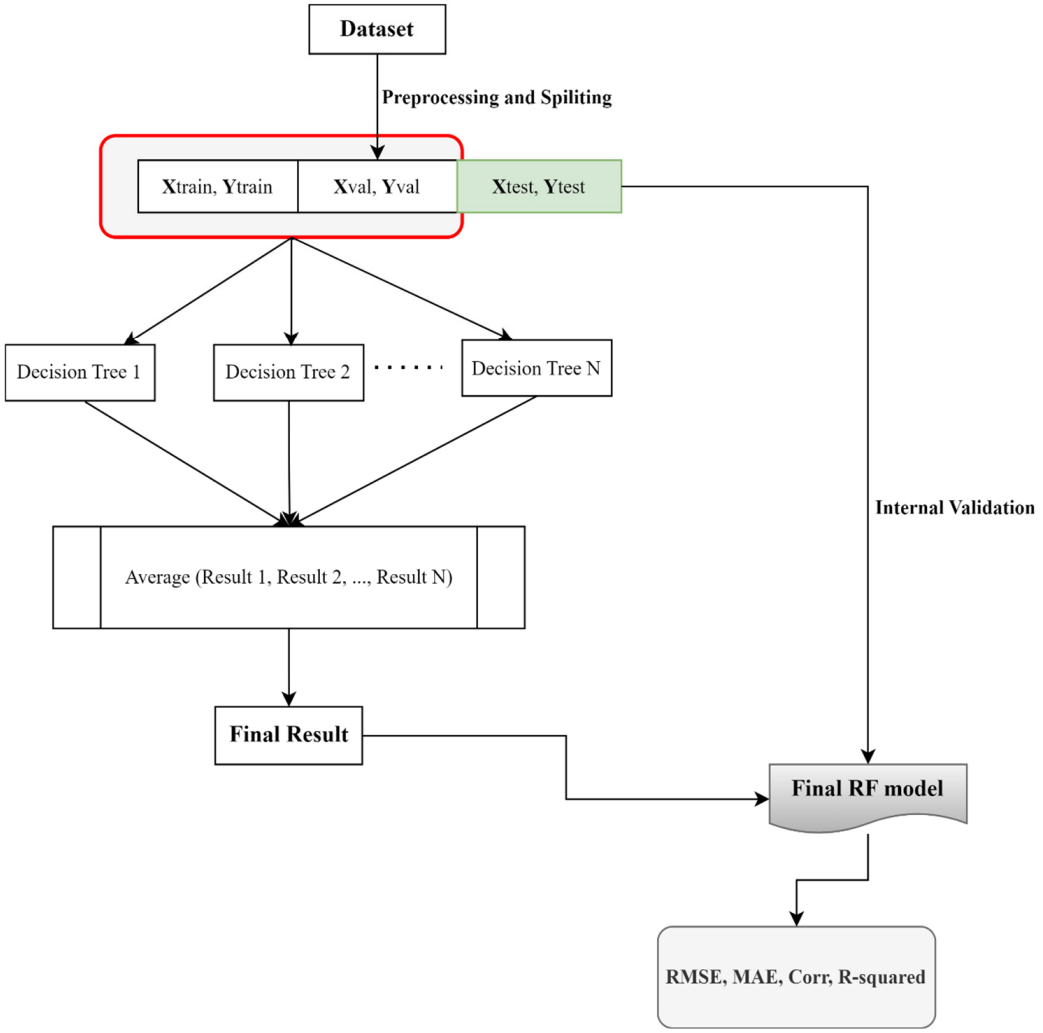

Figure 4. The flowchart of the RF method in regression mode.

Equation (10) is implemented with the Scikit-learn machine learning library [79], module randomforest.regressor. The main parameters of this method have been tuned using the grid search method (see [80,81] for more details) by considering five cross-folds. Table A1 in Appendix A summarizes the hyper-parameters of the RF technique model.

- ANN Method

The idea behind neural networks was that they would simulate the human brain in some way [82]. An ANN model consists of interconnected artificial neurons that can determine a general relationship between inputs and outputs without having a prior knowledge of the variables [82,83]. As shown in Figure 5, the ANN architecture consists of an input layer, one or more hidden layers, and an output layer, with several neurons connected between them and running simultaneously [84]. A variety of tasks can be accomplished using this method, such as environmental remote sensing [85], prediction of tropospheric wet delay and rainfall [84,86], and classification of crop yield [87].

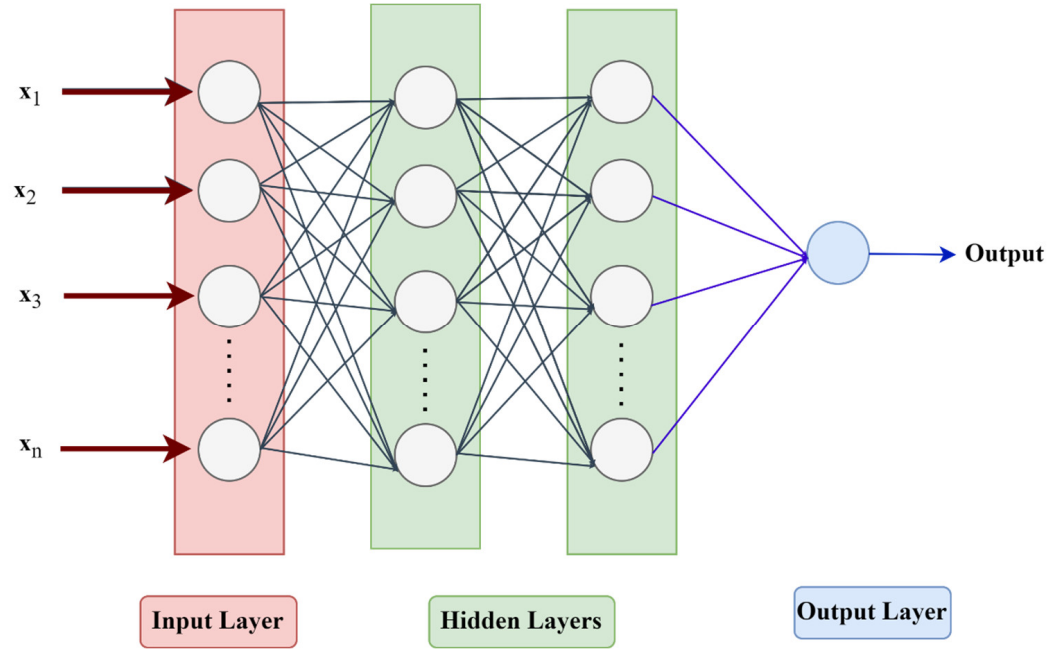

Figure 5. Schematic diagram of ANN architecture for regression.

There are several hyperparameters that are critical for ANN, including the number of hidden layers, the number of neurons in each layer, the activation function, and the dropout rate. One and two hidden layers have been examined in the present study in order to determine how they affect the model's accuracy. Moreover, the Bayesian optimization method from the Python keras-tuner toolkit [88] has been used to determine the optimal set of hyperparameters. The method consists of predicting the optimum hyperparameter sets for a complex computational function from a probabilistic model [89,90]. By using surrogate models, like Gaussian processes, the Bayesian optimization technique approximates a true objective function by using a distribution over the objective function [89–91]. With Bayesian optimization, the model's performance can be evaluated computationally inexpensively even when the search space is large, as it smartly selects which hyperparameters to examine next, thus reducing the number of tests [92,93]. Additionally, the validation mean square error (val-MSE) loss has been used to assess the performance of the model on unseen data in order to prevent overfitting. To do this, 20 per cent of the training dataset was used to monitor val-MSE's behaviour during training.

In Appendix A Table A2, we detail the best parameter settings for an ANN with one hidden layer to model T_m. Furthermore, another hidden layer was added in order to compare the performance of the double hidden layer neural network with the single hidden layer neural network. Table A3, shows the best hyperparameter sets for the double hidden layer neural network.

In the numerical results section, we evaluate the results obtained with the different methods described here in order to find the best model of T_m for the Austrian region. This model will then determine PWV in this area in the next step.

3.2. PWV Modelling

We first define the PWV based on GNSS data and radiosonde measurements. Following that, four different regional models of PWV are detailed adapted for the Austrian region based on RF regressions and ANNs techniques.

3.2.1. PWV Derived from GNSS Data

In order to derive PWV from GNSS data, we have first estimated ZTD using Bernese GNSS software in baseline mode [11]. Tropospheric zenith delay parameters were estimated on top of the apriori tropospheric delay model as piecewise linear functions. Loose

constraints of 5 m for the initial parameter and 1 m for subsequent parameters were applied. A summary of the primary inputs and configurations used to estimate ZTD in this study can be found in Table 5.

Table 5. Bernese GNSS software processing settings in this study.

Parameters	Bernese Processing
Reference Frame	ITRF2014
Satellite Orbit and Clock	IGS Final Products (CDDIS)/30 s
A priori Troposphere Model	Dry GMF
Gradient Model	CHENHER
Mapping Function	VMF1
Ionospheric Model	Global Ionospheric Models 'GIMs' (CODE)
Ocean tidal loading	FES2004 model (Chalmers)
Ambiguity Fixing Strategy	Quasi Ionosphere-Free (QIF)
Cut-off angle	5 degrees
Observation sampling rate	30 s

In the next step, ZHD has been computed using the Saastamoinen model as follows [12]:

$$ZHD = \frac{0.002277\ P_s}{1 - 0.00266\cos(2\varphi) - 0.00028\ H} \tag{12}$$

where, H (km), φ (rad), and P_s (hpa) are the ellipsoid height, latitude, and surface pressure of a GNSS station, respectively. Using ZWD as the result of subtracting ZHD from ZTD, the PWV can be derived by multiplication of ZWD with the conversion factor (Π) as noted below:

$$PWV = \Pi \times ZWD \tag{13}$$

With

$$\Pi = \frac{10^6}{\left(k_2' + k_3/T_m\right) \times (\rho_w\ R_v)} \tag{14}$$

whereby, ρ_w (1000 kg/m^3) and R_v (461.51 J/K×kg) denote the density of liquid water and the specific gas constant of water vapour, respectively. Moreover, the atmospheric refraction constants k_2' and k_3 have respective values of 22.97 K/hpa and 375,463 K^2/hpa [94]. The T_m used in Equation (14) has been calculated using the best model in Section 3.1, which is investigated in the numerical results.

3.2.2. PWV Derived from Radiosonde Data

As already stated in Section 2.2, radiosonde provides accurate and reliable atmospheric parameter measurements. This allows us to calculate PWV by using Equation (15) [25]:

$$PWV = \sum_{i=1}^{n}\left(\frac{q_i \times \Delta p}{g}\right) \tag{15}$$

In this equation, Δp (hpa) and g (m/s^2) represent the pressure difference between adjacent layers and gravity acceleration, respectively. q_i (kg) refers to specific humidity and can be obtained as below [54]:

$$q = \frac{0.622 \times P_w}{P - 0.378\ P_w} \tag{16}$$

here P_w (hpa) is the water vapour pressure and can be computed as follows [95]:

$$P_w = \left(\frac{RH}{100}\right) \times \left(6.112 \times \exp\left(\frac{17.67(T[K] - 273.15)}{(T[K] - 273.15) + 243.5}\right)\right) \tag{17}$$

whereby T (K) is the temperature at each pressure level. It is important to note that, for the PWV machine learning method developed from ZTDs of the GNSS stations, Equation (15) serves as an external validation reference.

3.2.3. Developed PWV Models

As for T_m, we have used RF and ANN techniques in order to determine an hourly PWV in conjunction with GNSS ZTD for the Austrian region. This is mainly aimed at improving weather forecasting and climate monitoring by delivering timely and accurate PWV. Equations (18) and (19) show the general structure of the machine learning method, which describes the relationship between input parameters and output PWV:

$$PWV = g_1(ZTD, \varphi, h, doy, time) \tag{18}$$

$$PWV = g_2(ZTD, \varphi, \lambda, h, doy, time) \tag{19}$$

where $g_1()$ and $g_2()$ refer to the machine-learning based estimation model for timely PWV in presence of four and five input variables, respectively. From now on, we will refer to the model with five parameters as Scheme#1 and the model with six parameters as Scheme#2. Further, to develop the machine learning model, we used data from 39 GNSS stations situated across Austria between January 2018 and June 2021 (see Table 1 and Figure 1).

- RF method

To develop the RF models in Scheme#1 and Scheme#2, we relied on Scikit-learn machine learning library [79], module randomforest.regressor. The grid search method has also been used for finding the best set of hyperparameters. Using grid search with five cross-folds, the best hyperparameters for Scheme#1 have been determined. Table A4 in Appendix B shows the obtained result. Additionally, Table A5 details the best set of hyperparameters for Scheme#2 obtained through the grid search with five cross-folds. In the numerical result section, the results gained using the RF method for Scheme#1 and Scheme#2 are discussed.

- ANN method

ANN parameters have been tuned using Bayesian optimization techniques; for more information, please refer to Section 3.1.2. Accordingly, we have implemented ANN with one hidden layer for Scheme#1 and Scheme#2. The hyperparameters for the ANN method in Scheme #1 are listed in Table A6 in Appendix B. Further, a set of best hyperparameters for estimating hourly PWV using the ANN method of Scheme#2 is presented in Table A7.

The double hidden layer model has also been implemented for Scheme#1 to assess the impact of the extra hidden layer on ANN performance. Due to the fact that Scheme#2 did not differ significantly from Scheme#1 in terms of numerical results (please see Section 4.3), we did not implement a double hidden layer for Scheme#2. Table A8 details the best hyperparameter sets for the two-layer hidden neural network in Scheme#1.

Using the various approaches discussed here, we assess the numerical outcomes in order to determine the strengths and limitations of each, leading to the selection of the PWV model that performs best for Austria.

3.3. Statistical Metrics

To measure model fitness, the R-squared indicator is used, which indicates how well the model fits the data [51,96]. Generally, the closer the R-squared value is to 1, the better the model fitted and, consequently, the more credible the fit. A formula for determining this statistical metric is as follows [96]:

$$R\text{-squared} = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n}(Y_{o_i} - Y_{p_i})}{\sum_{i=1}^{n}(Y_{o_i} - \overline{Y}_o)} \tag{20}$$

In this formula, SSR stands for Sum of Squared Residuals, and SST stands for Sum of Squares Total. Moreover, Y_{o_i}, Y_{p_i}, and $\overline{Y}_o$ represent the observed value (PWV or T_m), the predicted value, and the average of the observed value. According to the Equation (20), R-squared reflects the proportion of the variance explained by the model to its total variance.

Additionally, we have employed a set of statistical indicators to measure the model's performance quantitatively. The following indicators are included in the study: Mean Absolute Error (MAE), Root Mean Square Error ($RMSE$), and Pearson correlation coefficient ($corr$) [51,56,97,98]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_{o_i} - Y_{p_i}\right| \tag{21}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_{o_i} - Y_{p_i}\right)^2} \tag{22}$$

$$corr = \frac{\sum_{i=1}^{n}\left[\left(Y_{o_i} - \overline{Y}_o\right)\left(Y_{p_i} - \overline{Y}_p\right)\right]}{\left(\sqrt{\sum_{i=1}^{n}\left(Y_{o_i} - \overline{Y}_o\right)^2}\right)\left(\sqrt{\sum_{i=1}^{n}\left(Y_{p_i} - \overline{Y}_p\right)^2}\right)} \tag{23}$$

4. Numerical Results

The performance of both, T_m and PWV models is examined in this section. The developed model for T_m is evaluated internally using the radiosonde test data and externally using data from a check station (RS11010). A PWV model is also validated internally using GNSS test data and externally with radiosonde stations (RS11035, RS11120, RS11120, and RS11240).

4.1. Accuracy Evaluation of T_m

Different statistical metrics like $RMSE$ and MAE are used here to evaluate the performance of the T_m models described in Section 3.1.2. We start by assessing the model's reliability and precision on the test dataset. In the next step, we will validate the efficiency of different models using RS11010 data. Based on the results, we will select the best T_m model to be applied to PWV calculation.

4.1.1. Internal Model Testing

Internal validation is a crucial step in the machine learning procedure in order to validate the performance and robustness of the trained model. The goal of this step is to demonstrate how well the developed model can generalize to unseen data and meet users' expectations regarding accuracy and reliability. Here, we have used the test dataset, which was separated from the whole dataset (see Table 1), covering the period July 2021–July 2022. A summary of the statistical indicators for the different T_m models is shown in Table 6. According to the reported results, the GPT3 model (MAE of 3.58 K and R-squared of 50%) delivers slightly less accurate results in comparison with other models. Furthermore, the RF model provides better results with a MAE of 2.38 K, which is up to 30% better than for other methods but still comparable to ANN. A further analysis shows no significant difference between ANNs in both single-hidden (MAE of 2.45 K, R-squared of 77%) and double-hidden layers (MAE of 2.46 K, R-squared of 77%). This demonstrates that the T_m problem may not require more than one hidden layer because there is a nearly linear relationship between the defined features and the input data have relatively simple patterns. According to the training speed of the RF model, it would be more practical to use the RF model since it can achieve almost the same results as ANN within a shorter period of time.

Table 6. Statistical results of *MAE*, *RMSE*, *corr*, and *R*-squared of different T_m models for test dataset over the period of July 2021–July 2022.

Models	*MAE* (K)	*RMSE* (K)	*corr* (%)	*R*-Squared (%)
Polynomial	2.82	3.46	84	70
Linear	3.14	3.85	80	62
GPT3	3.58	4.46	79	50
Bevis	3.10	3.84	80	63
RF	2.38	2.97	89	79
ANN (1 Hidden layer)	2.45	3.02	88	77
ANN (2 Hidden layer)	2.46	3.04	88	77

4.1.2. External Model Testing

The purpose of external model testing is to verify that the model can adapt to changes in conditions and maintain its quality and reliability over time. In order to accomplish this, data from RS11010 have been used during August 2022–April 2023 to evaluate the performance of the discussed T_m models.

In Figure 6, the results of the empirical (Bevis and GPT3) and regression models (Polynomial and Linear) are shown. According to this figure, polynomial regression outperformed other models with an *MAE* of 2.70 K and a *RMSE* of 3.25 K. Moreover, GPT3 has the weakest performance among the other models, degraded by about 17% in terms of *MAE*.

Figure 7 illustrates the developed models using RF and ANN methods. It can be seen from this figure that the RF model shows a slight improvement in *MAE* compared to the two other models, namely ANN with a single hidden layer (1 HL) and ANN with double hidden layers (2 HL). Moreover, the ANN model with one and two hidden layers produces almost identical results. Therefore, the results confirm once again that using two hidden layers for modelling T_m in this study has no significant advantage over one hidden layer.

Based on the performance of different strategies to model Tm, RF appears to be the superior model in this case study with *MAE* improvements of approximately 2–19%. As a result, we employ the RF model in the next step to model *PWV* in Austria.

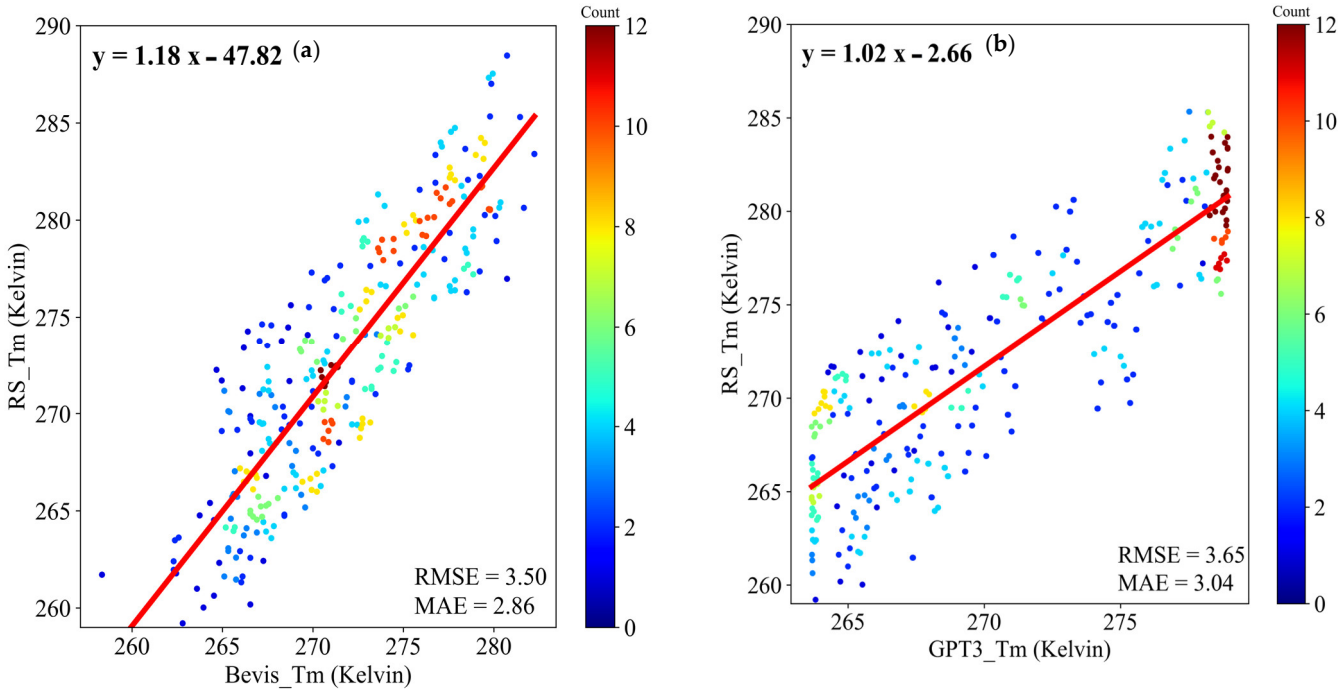

Figure 6. *Cont.*

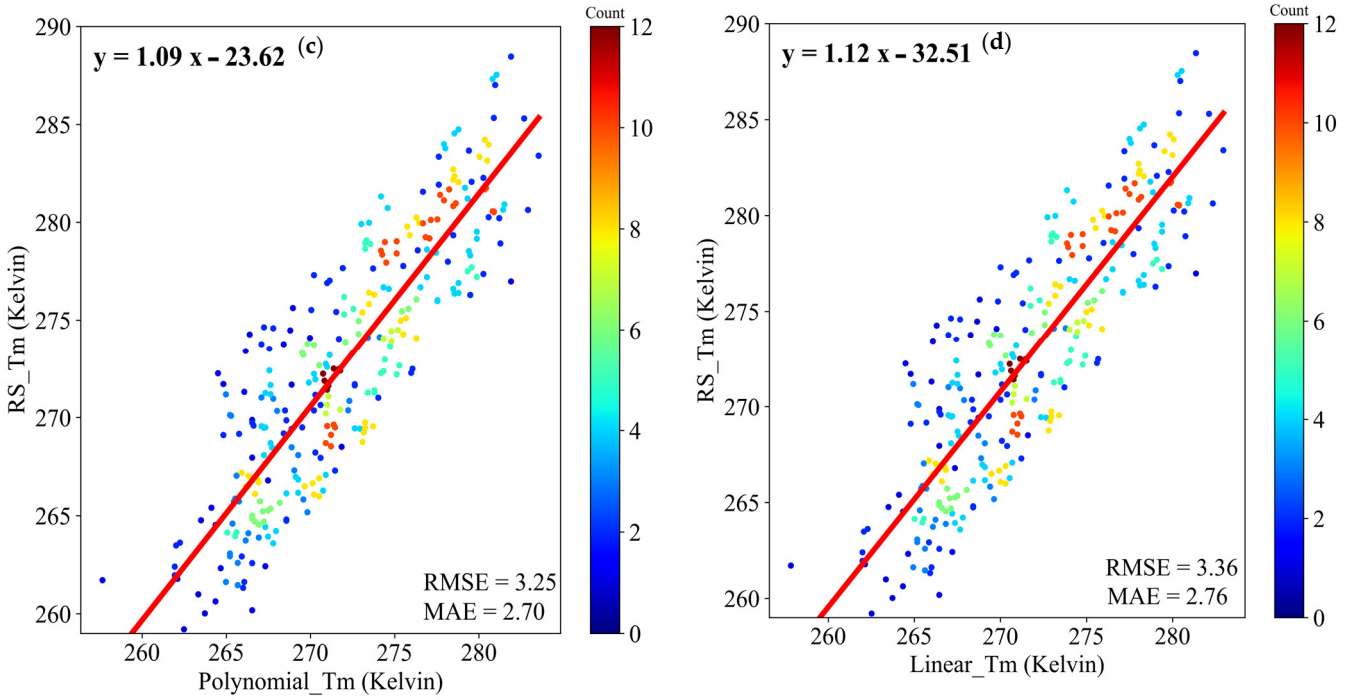

Figure 6. Scatter plots of observed T_m against estimated T_m using different models: (**a**) Bevis, (**b**) GPT3, (**c**) Polynomial, and (**d**) Linear.

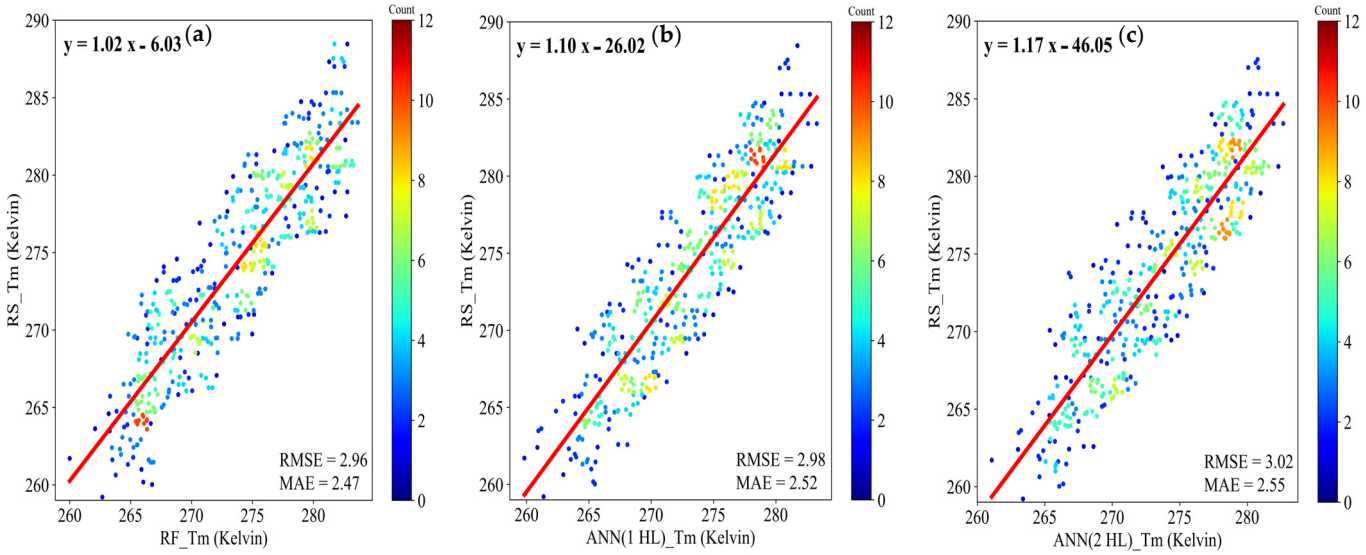

Figure 7. Scatter plots of observed T_m against predicted T_m using different models: (**a**) RF, (**b**) ANN with one hidden layer, and (**c**) ANN with two hidden layer.

4.2. Accuracy Evaluation of PWV

We first assess the reliability of the model using the test dataset in this section. Afterwards, four radiosonde stations in Austria were analyzed to determine the accuracy of the model for predicting hourly *PWV*.

4.3. Internal Model Testing

The GNSS *PWV* dataset covering July 2021–2022 was used to ensure the robustness of the *PWV* estimation using RF and ANN methods. Based on *MAE*, *RMSE*, *corr*, and *R*-squared (see Section 3.3 for more details), we have calculated the differences between *PWV* derived directly from GNSS and *PWV* estimated from RF and ANN models. As

shown in Table 7, there is no significant difference between Scheme#1 and Scheme#2, which means adding longitude as an extra feature could only affect the model accuracy (*MAE*, and *RMSE*) by less than 2%. This indicates that the significant variations in both water vapour and pressure, and accordingly troposphere, tend to be more dependent on latitude and height than longitude. Furthermore, the double hidden layer was relatively ineffective in improving ANN model performance, and even one hidden layer was better than the double hidden layer. Accordingly, Scheme#1, using the ANN model with a single hidden layer, outperformed other models with *MAE* improvements by 1–15%.

Table 7. Statistical results of *MAE*, *RMSE*, *corr*, and *R*-squared of different *PWV* models for test dataset over the period of July 2021–July 2022.

Models	*MAE* (mm)	*RMSE* (mm)	*corr* (%)	*R*-Squared (%)
RF Scheme#1	2.14	2.72	94	87
RF Scheme#2	2.11	2.69	94	88
ANN 1 HL Scheme#1	1.83	2.37	95	90
ANN 2 HL Scheme#1	1.85	2.40	95	90
ANN 1 HL Scheme#2	1.85	2.41	95	90

4.4. External Model Testing

For the purpose of comparing the effectiveness of the RF and ANN models when applied to new, unseen data, we used data from four radiosonde stations (see Figure 1) from August 2022 to April 2023. This allows us to determine the accuracy and reliability of both models in real-world applications. Figure 8 illustrates the relationship between the predicted *PWV* using the RF method and radiosonde for RS11035 for Scheme#1 and Scheme#2. It can be seen from this figure that both methods have nearly the same overall pattern of points, although Scheme#2 has a slightly lower *MAE* and *RMSE* than Scheme#1. However, in the case of RS11120 (see Appendix C, Table A9), there were some improvements in Scheme#1 over Scheme#2, which may be due to the fact that it is located in a more mountainous area than other radiosonde stations which might play a significant role in influencing *PWV* values. Nevertheless, similar to RS11035, two other radiosonde stations (RS11010 and RS11240) confirmed that Scheme#2 performs generally slightly better than Scheme#1.

Figure 9 depicts the distribution of predicted *PWV* using ANN models compared to RS11035-derived *PWV*. This figure shows that the estimated *PWV* using Scheme#1 with one hidden layer overcomes two other models with a *MAE* and *RMSE* of 2.15 mm and 2.74 mm, respectively. Other radiosonde stations also support the same inference, as shown in Appendix C.

To acquire a better understanding of the models' performance, the *MAE*, *RMSE*, and *corr* of all radiosonde stations included in the external validation process were averaged. Based on the presented results in Table 8, Scheme#1 using the ANN with a single hidden layer outperformed the other models by almost 5% in *MAE* and 1–5% in *RMSE*.

Table 8. Results of average *MAE*, *RMSE*, and *corr* of different *PWV* models over all radiosonde stations (RS11035, RS11120, RS11010, RS11240) during August 2022 to April 2023.

Models	*MAE* (mm)	*RMSE* (mm)	*corr* (%)
RF Scheme#1	2.17	2.78	94
RF Scheme#2	2.14	2.68	94
ANN 1 HL Scheme#1	2.05	2.64	95
ANN 2 HL Scheme#1	2.13	2.71	95
ANN 1 HL Scheme#2	2.13	2.70	95

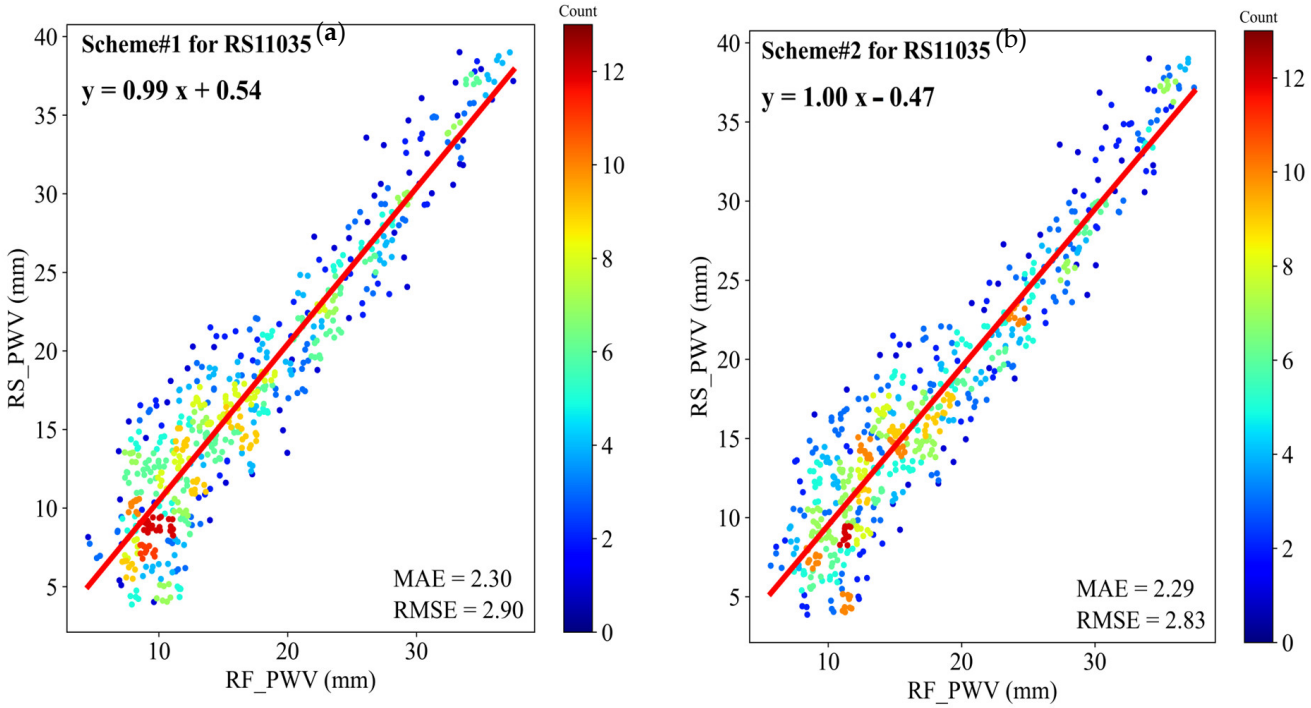

Figure 8. Scatter plots of RS11035 *PWV* against predicted *PWV* using RF model for: (**a**) Scheme#1, and (**b**) Scheme#2.

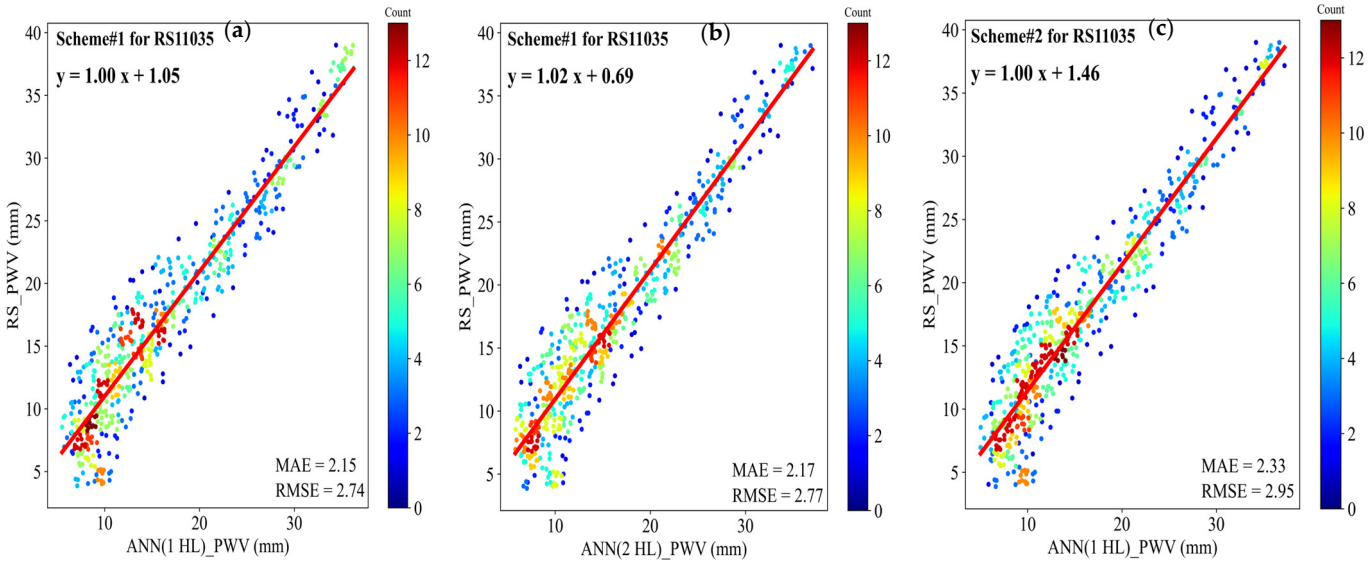

Figure 9. Scatter plots of RS11035 *PWV* against predicted *PWV* using ANN model with: (**a**) single hidden layer for Scheme#1, (**b**) double hidden layer for Scheme#1, and (**c**) single hidden layer for Scheme#2.

This result suggests that the developed ANN model (Scheme#1) can predict the *PWV* amounts in Austria with an average *MAE* of less than 2.5 mm. Moreover, as there was no significant difference between the RF and ANN models, the RF model may also be a good replacement for the ANN model in this case study.

5. Conclusions

In this study, a non-meteorological model was developed for estimating *PWV* in the Austrian region using a neural network technique. In order to accomplish this, first, we

developed a regional model based on the location of the intended station, the surface temperature, the time of interest, and the day of the year. For the best T_m model, we compared the empirical (GPT3 and Bevis) models to the regression (linear and polynomial) as well as RF and ANN models. Based on the data from three radiosonde stations in different parts of Austria (RS11035, RS11120, and RS11240) from January 2010 to June 2021, regression and RF/ANN models were developed. A comparison of the RF model and the other models at the location of RS10101 revealed that the RF model provided superior results with MAE improvements of approximately 2–19% in the prediction phase from August 2022 to April 2023. Following the consideration of RF T_m, three models were constructed, namely RF, ANN with a single hidden layer, and ANN with a double hidden layer, in two separate schemes: Scheme#1 ($ZTD, \varphi, h, doy, time$), and Scheme#2 ($ZTD, \varphi, \lambda, h, doy, time$). These models were developed using ZTD data from 39 GNSS stations covering the entire Austria territory between January 2018 and June 2021. A two-stage validation process was then performed on the developed models: an internal validation and an external validation. As a result, the performance of ANNs with one or two hidden layers was not significantly different. Additionally, longitude did not appear to be an important factor in improving PWV, as tropospheric variation is influenced by latitude and height. According to the statistical results, it is possible to predict the hourly PWV in the Austria region without the need for sensed meteorological parameters with an average MAE of less than 2.5 mm.

Author Contributions: Conceptualization Z.A. and B.G.; methodology Z.A. and B.G.; software Z.A.; validation Z.A. and B.G.; investigation Z.A., R.W. and N.H.; visualization Z.A.; supervision R.W.; writing original draft Z.A.; review and editing all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Radiosonde data were also obtained from the University of Wyoming (http://weather.uwyo.edu/upperair/sounding.html; accessed on 11 July 2023). The corresponding author can provide the ZTD upon reasonable request.

Acknowledgments: Special thanks go out to EPOSA (Echtzeit Positioning Austria) for providing GNSS observation data, as well as to the University of Wyoming for supplying radiosonde measurements. Furthermore, CDDIS (Crustal Dynamics Data Information System) and IGS (International GNSS Service) provided valuable satellite products and ionospheric maps, which are very much appreciated. The authors also wish to acknowledge the Open Access Funding provided by TU Wien. This manuscript has been greatly improved by the three anonymous reviewers' valuable feedback and constructive suggestions. Moreover, we appreciate the guidance and support provided by the editors throughout the review process.

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

Appendix A

The best set of hyperparameters for RF and ANN techniques for modelling T_m.

Table A1. Hyperparameters tuning process of the final RF regressor for T_m modelling.

Parameter	Value	Description
n_estimators	1000	The number of trees in the forest (100:100:1000)
max_features	sqrt	The number of features to consider when looking for the best split ('sqrt', 'log2')
max_depth	None	The maximum depth of the tree (None, (1:1:5))
min_samples_leaf	5	The minimum number of samples required to be at a leaf node (1:1:5)
bootstrap	True	Whether bootstrap samples are used when building trees. If False, the whole dataset is used to build each tree [False, True]

Table A2. Final tuned parameters for the single hidden layer ANN for T_m modelling.

Parameter	Description
Input layer	1000 neurons, 'selu' activation function
Dropout layer	0.0 dropout rate
Hidden layer	1008 neurons, 'selu' activation function
Dropout layer	0.0 dropout rate
Output layer	1 neuron, 'linear' activation function
Optimizer	'rmsprop'

Table A3. Final tuned parameters for the double hidden layer ANN for T_m modelling.

Parameter	Description
Input layer	1000 neurons, 'selu' activation function
Dropout layer	0.0 dropout rate
First hidden layer	1008 neurons, 'sigmoid' activation function
Dropout layer	0.5 dropout rate
Second hidden layer	1008 neurons, 'sigmoid' activation function
Dropout layer	0.0 dropout rate
Output layer	1 neuron, 'linear' activation function
Optimizer	'rmsprop'

Appendix B

The best set of hyperparameters for RF and ANN techniques for modelling PWV.

Table A4. Hyperparameters tuning process of the final RF regressor for PWV modelling in Scheme#1.

Parameter	Value	Description
max_features	sqrt	The number of trees in the forest (100:100:1000)
n_estimators	800	The number of features to consider when looking for the best split ['sqrt', 'log2']
max_depth	None	The maximum depth of the tree (None, (1:1:5))
min_samples_leaf	5	The minimum number of samples required to be at a leaf node (1:1:5)
bootstrap	True	Whether bootstrap samples are used when building trees. If False, the whole dataset is used to build each tree [False, True]

Table A5. Hyperparameters tuning process of the final RF regressor for PWV modelling in Scheme#2.

Parameter	Value	Description
n_estimators	900	The number of trees in the forest (100:100:1000)
max_features	sqrt	The number of features to consider when looking for the best split ['sqrt', 'log2']
max_depth	None	The maximum depth of the tree (None, (1:1:5))
min_samples_leaf	5	The minimum number of samples required to be at a leaf node (1:1:5)
bootstrap	True	Whether bootstrap samples are used when building trees. If False, the whole dataset is used to build each tree [False, True]

Table A6. Final tuned parameters for the single hidden layer ANN for *PWV* modelling in Scheme#1.

Parameter	Description
Input layer	360 neurons, 'sigmoid' activation function
Dropout layer	0.3 dropout rate
Hidden layer	80 neurons, 'sigmoid' activation function
Dropout layer	0.2 dropout rate
Output layer	1 neuron, 'linear' activation function
Optimizer	'Adam'

Table A7. Final tuned parameters for the single hidden layer ANN for *PWV* modelling in Scheme#2.

Parameter	Description
Input layer	360 neurons, 'sigmoid' activation function
Dropout layer	0.3 dropout rate
Hidden layer	752 neurons, 'sigmoid' activation function
Dropout layer	0.1 dropout rate
Output layer	1 neuron, 'linear' activation function
Optimizer	'Adam'

Table A8. Final tuned parameters for the double hidden layer ANN for *PWV* modelling in Scheme#1.

Parameter	Description
Input layer	328 neurons, 'sigmoid' activation function
Dropout layer	0.4 dropout rate
First hidden layer	144 neurons, 'sigmoid' activation function
Dropout layer	0.2 dropout rate
Second hidden layer	848 neurons, 'LeakyReLU' activation function
Dropout layer	0.1 dropout rate
Output layer	1 neuron, 'linear' activation function
Optimizer	'Adam'

Appendix C

Statistical results for different *PWV* models at RS11120, RS11010, and RS11240.

Table A9. Statistical results of *MAE*, *RMSE*, and *corr* of different *PWV* models for RS11120 over the period of August 2022–April 2023.

Models	*MAE* (mm)	*RMSE* (mm)	*corr* (%)
RF Scheme#1	1.95	2.48	95
RF Scheme#2	2.11	2.66	95
ANN 1 HL Scheme #1	1.96	2.50	96
ANN 2 HL Scheme #1	2.02	2.57	96
ANN 1 HL Scheme #2	2.04	2.56	96

Table A10. Statistical results of *MAE*, *RMSE*, and *corr* of different *PWV* models for RS11010 over the period of August 2022–April 2023.

Models	*MAE* (mm)	*RMSE* (mm)	*corr* (%)
RF Scheme#1	2.32	2.92	94
RF Scheme#2	2.17	2.69	94
ANN 1 HL Scheme #1	2.09	2.68	95
ANN 2 HL Scheme #1	2.16	2.74	95
ANN 1 HL Scheme #2	2.22	2.81	95

Table A11. Statistical results of *MAE*, *RMSE*, and *corr* of different *PWV* models for RS11240 over the period of August 2022–April 2023.

Models	*MAE* (mm)	*RMSE* (mm)	*corr*
RF Scheme#1	2.09	2.72	95
RF Scheme#2	2.00	2.53	95
ANN 1 HL Scheme #1	2.02	2.64	95
ANN 2 HL Scheme #1	2.15	2.74	95
ANN 1 HL Scheme #2	2.43	3.02	95

References

1. Bevis, M.; Businger, S.; Herring, T.; 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]
2. Li, H.; Wang, X.; Wu, S.; Zhang, K.; Chen, X.; Qiu, C.; Zhang, S.; Zhang, J.; Xie, M.; Li, L. Development of an Improved Model for Prediction of Short-Term Heavy Precipitation Based on GNSS-Derived PWV. *Remote Sens.* **2020**, *12*, 4101. [CrossRef]
3. Li, H.; Wang, X.; Choy, S.; Wu, S.; Jiang, C.; Zhang, J.; Qiu, C.; Li, L.; Zhang, K. A New Cumulative Anomaly-based Model for the Detection of Heavy Precipitation Using GNSS-derived Tropospheric Products. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4105718. [CrossRef]
4. Rocken, C.; Hove, T.V.; Johnson, J.; Solheim, F.; Ware, R.; Bevis, M.; Chiswell, S.; Businger, S. GPS/STORM: GPS sensing of atmospheric water vapor for meteorology. *J. Atmos. Ocean. Technol.* **1995**, *12*, 468–478. [CrossRef]
5. Adavi, Z.; Rohm, W.; Weber, R. Analyzing Different Parameterization Methods in GNSS Tomography Using the COST Benchmark Dataset. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6155–6163. [CrossRef]
6. Emardson, T.R.; Elgered, G.; Johansson, J.M. Three months of continuous monitoring of atmospheric water vapor with a network of global positioning system receivers. *J. Geophys. Res. Atmos.* **1998**, *103*, 1807–1820. [CrossRef]
7. Huang, L.; Liu, L.; Chen, H.; Jiang, W. An improved atmospheric weighted mean temperature model and its impact on GNSS precipitable water vapor estimates for China. *GPS Solut.* **2019**, *23*, 51. [CrossRef]
8. Liu, Z.; Zhang, T.; Li, J.; Xu, J.; Huang, L.; Zhou, L.; Liu, L. Analysis of Precipitable Water Vapor Change Based on GNSS Observations During Typhoon Mangkhut Event. In Proceedings of the China Satellite Navigation Conference (CSNC), Beijing, China, 22–25 May 2022.
9. Rohm, W.; Yuan, Y.; Biadeglgne, B.; Zhang, K.; Le Marshall, J. Ground-based GNSS ZTD/IWV estimation system for numerical weather prediction in challenging weather conditions. *Atmos. Res.* **2014**, *138*, 414–426. [CrossRef]
10. Zhao, Q.; Zhang, X.; Wu, K.; Liu, Y.; Li, Z.; Shi, Y. Comprehensive Precipitable Water Vapor Retrieval and Application Platform Based on Various Water Vapor Detection Techniques. *Remote Sens.* **2022**, *14*, 2507. [CrossRef]
11. Dach, R.; Lutz, S.; Walser, P.; Fridez, P. *Bernese GNSS Software, Version 5.2*; User manual; Astronomical Institute, University of Bern: Bern, Switzerland, 2015.
12. Saastamoinen, J. Contributions to the theory of atmospheric refraction. Part II: Refraction corrections in satellite geodesy. *Bull. Géodésique* **1973**, *107*, 13–34. [CrossRef]
13. Vázquez, B.G.E.; Grejner-Brzezinska, D.A. GPS-PWV estimation and validation with radiosonde data and numerical weather prediction model in Antarctica. *GPS Solut.* **2013**, *17*, 29–39. [CrossRef]
14. Jade, S.; Vijayan, M.S.M.; Gaur, V.K.; Prabhu, T.P.; Sahu, S.C. Estimates of precipitable water vapour from GPS data over the Indian subcontinent. *J. Atmos. Sol. Terr. Phys.* **2005**, *67*, 623–635. [CrossRef]
15. Putri, N.S.E. *Tropospheric Delay Models for GNSS in Indonesia*; Technische Universität Wien: Vienna, Austria, 2021.
16. Bevis, M.; Businger, S.; Chiswell, S.R.; Herring, T.A.; Anthes, R.A.; Rocken, C.; Ware, R. GPS meteorology: Mapping zenith wet delays onto precipitable water. *J. Appl. Meteorol. Climatol.* **1994**, *33*, 359–386. [CrossRef]
17. Huang, L.; Wang, X.; Xiong, S.; Li, J.; Liu, L.; Mo, Z.; Fu, B.; He, H. High-precision GNSS PWV retrieval using dense GNSS sites and in-situ meteorological observations for the evaluation of MERRA-2 and ERA5 reanalysis products over China. *Atmos. Res.* **2022**, *276*, 106247. [CrossRef]
18. Liang, H.; Cao, Y.; Wan, X.; Xu, Z.; Wang, H.; Hu, H. Meteorological applications of precipitable water vapor measurements retrieved by the national GNSS network of China. *Geod. Geodyn.* **2015**, *6*, 135–142. [CrossRef]
19. Huang, L.; Mo, Z.; Xie, S. Spatiotemporal characteristics of GNSS-derived precipitable water vapor during heavy rainfall events in Guilin, China. *Satell. Navig.* **2021**, *2*, 13. [CrossRef]
20. Guo, M.; Zhang, H.; Xia, P. Exploration and analysis of the factors influencing GNSS PWV for nowcasting applications. *Adv. Space Res.* **2021**, *67*, 3960–3978. [CrossRef]
21. Sam Khaniani, A.; Motieyan, H.; Mohammadi, A. Rainfall forecast based on GPS PWV together with meteorological parameters using neural network models. *J. Atmos. Sol.-Terr. Phys.* **2021**, *214*, 105533. [CrossRef]
22. Zhang, Z.; Lou, Y.; Zhang, W.; Liang, H.; Bai, J.; Song, W. Correlation Analysis between Precipitation and Precipitable Water Vapor over China Based on 1999–2015 Ground-Based GPS Observations. *J. Appl. Meteorol. Climatol.* **2022**, *61*, 1669–1683. [CrossRef]
23. Ejigu, Y.G.; Teferle, F.N.; Klos, A.; Bogusz, J.; Hunegnaw, A. *Tracking Hurricanes Using GPS Atmospheric Precipitable Water Vapor Field*; Springer: Cham, Switzerland, 2020; Volume 152.

24. Gong, Y.; Liu, Z.; Chan, P.K.; Hon, K.K. Assimilating GNSS PWV and radiosonde meteorological profiles to improve the PWV and rainfall forecasting performance from the Weather Research and Forecasting (WRF) model over the South China. *Atmos. Res.* **2023**, *286*, 106677. [CrossRef]
25. Zhang, H.; Yuan, Y.; Li, W.; 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.* **2019**, *12*, 1587–1598. [CrossRef]
26. Sohn, D.-H.; Choi, B.-K.; Park, Y.; Kim, Y.C.; Ku, B. Precipitable Water Vapor Retrieval from Shipborne GNSS Observations on the Korean Research Vessel ISABU. *Sensors* **2020**, *20*, 4261. [CrossRef]
27. Li, L.; Wang, X.; Wei, Y.; Wang, H. The New PWV Conversion Models Based on GNSS and Meteorological Elements in the China Region. *Atmosphere* **2022**, *13*, 1810. [CrossRef]
28. Zhang, B.; Yao, Y. Precipitable water vapor fusion based on a generalized regression neural network. *J. Geod.* **2021**, *95*, 36. [CrossRef]
29. Zheng, Y.; Lu, C.; Wu, Z.; Liao, J.; Zhang, Y.; Wang, Q. Machine Learning-Based Model for Real-Time GNSS Precipitable Water Vapor Sensing. *Geophys. Res. Lett.* **2022**, *49*, e2021GL096408. [CrossRef]
30. Ma, Y.; Zhao, Q.; Wu, K.; Yao, W.; Liu, Y.; Li, Z.; Shi, Y. Comprehensive Analysis and Validation of the Atmospheric Weighted Mean Temperature Models in China. *Remote Sens.* **2022**, *14*, 3435. [CrossRef]
31. Zhang, S.; Gong, L.; Gao, W.; Zeng, Q.; Xiao, F.; Liu, Z.; Lei, J. A weighted mean temperature model using principal component analysis for Greenland. *GPS Solut.* **2023**, *27*, 57. [CrossRef]
32. Balidakis, K.; Nilsson, T.; Zus, F.; Glaser, S.; Heinkelmann, R.; Deng, Z.; Schuh, H. Estimating integrated water vapor trends from VLBI, GPS, and numerical weather models: Sensitivity to tropospheric parameterization. *J. Geophys. Res. Atmos.* **2018**, *123*, 6356–6372. [CrossRef]
33. He, Q.; Shen, Z.; Wan, M.; Li, L. Precipitable Water Vapor Converted from GNSS-ZTD and ERA5 Datasets for the Monitoring of Tropical Cyclones. *IEEE Access* **2020**, *8*, 87275–87290. [CrossRef]
34. Li, Q.; Yuan, L.; Chen, P.; Jiang, Z. Global grid-based Tm model with vertical adjustment for GNSS precipitable water retrieval. *GPS Solut.* **2020**, *24*, 73. [CrossRef]
35. Manandhar, S.; Lee, Y.H.; Meng, Y.S.; Ong, J.T. A Simplified Model for the Retrieval of Precipitable Water Vapor from GPS Signal. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6245–6253. [CrossRef]
36. Wang, Z.; Zhou, X.; Liu, Y.; Zhou, D.; Zhang, H.; Sun, W. Precipitable water vapor characterization in the coastal regions of China based on ground-based GPS. *Adv. Space Res.* **2017**, *60*, 2368–2378. [CrossRef]
37. Zhang, F.; Barriot, J.-P.; Xu, G.; Yeh, T.-K. Metrology Assessment of the Accuracy of Precipitable Water Vapor Estimates from GPS Data Acquisition in Tropical Areas: The Tahiti Case. *Remote Sens.* **2018**, *10*, 758. [CrossRef]
38. Emardson, T.R.; Derks, H.J.P. On the relation between the wet delay and the integrated precipitable water vapour in the European atmosphere. *Meteorol. Appl.* **2000**, *7*, 61–68. [CrossRef]
39. Mendes, V.B.; Prates, G.; Santos, L.; Langley, R.B. An Evaluation of the Accuracy of Models for the Determination of the Weighted Mean Temperature of the Atmosphere. In Proceedings of the ION 2000, National Technical Meeting, Anaheim, CA, USA, 26–28 January 2000; pp. 433–438.
40. Suresh Raju, C.; Saha, K.; Thampi, B.V.; Parameswaran, K. Empirical model for mean temperature for Indian zone and estimation of precipitable water vapor from ground based GPS measurements. *Ann. Geophys.* **2007**, *25*, 1935–1948. [CrossRef]
41. Wu, M.; Jin, S.; Li, Z.; Cao, Y.; Ping, F.; Tang, X. High-Precision GNSS PWV and Its Variation Characteristics in China Based on Individual Station Meteorological Data. *Remote Sens.* **2021**, *13*, 1296. [CrossRef]
42. Long, F.; Hu, W.; Dong, Y.; Wang, J. Neural Network-Based Models for Estimating Weighted Mean Temperature in China and Adjacent Areas. *Atmosphere* **2021**, *12*, 169. [CrossRef]
43. Böhm, J.; Heinkelmann, R.; Schuh, H. Short Note: A global model of pressure and temperature for geodetic applications. *J. Geod.* **2007**, *81*, 679–683. [CrossRef]
44. Böhm, J.; Möller, G.; Schindelegger, M.; Pain, G.; Weber, R. Development of an improved empirical model for slant delays in the troposphere (GPT2w). *GPS Solut.* **2015**, *19*, 433–441. [CrossRef]
45. Lagler, K.; Schindelegger, M.; Böhm, J.; Krásná, H.; Nilsson, T. GPT2: Empirical slant delay model for radio space geodetic techniques. *Geophys. Res. Lett.* **2013**, *40*, 1069–1073. [CrossRef]
46. Yao, Y.; Zhang, B.; Xu, C.; Yan, F. Improved one/multi-parameter models that consider seasonal and geographic variations for estimating weighted mean temperature in ground-based GPS meteorology. *J. Geod.* **2014**, *88*, 273–282. [CrossRef]
47. Yao, Y.B.; Zhang, B.; Yue, S.Q.; Xu, C.Q.; Peng, W.F. Global empirical model for mapping zenith wet delays onto precipitable water. *J. Geod.* **2013**, *87*, 439–448. [CrossRef]
48. Sun, Z.; Zhang, B.; Yao, Y. A global model for estimating tropospheric delay and weighted mean temperature developed with atmospheric reanalysis data from 1979 to 2017. *Remote Sens.* **2019**, *11*, 1893. [CrossRef]
49. Mateus, P.; Catalão, J.; Mendes, V.B.; Nico, G. An ERA5-Based Hourly Global Pressure and Temperature (HGPT) Model. *Remote Sens.* **2020**, *12*, 1098. [CrossRef]
50. Yao, Y.; Zhu, S.; Yue, S.A. A globally applicable, season-specific model for estimating the weighted mean temperature of the atmosphere. *J. Geod.* **2012**, *86*, 1125–1135. [CrossRef]
51. Wang, S.; Xu, T.; Nie, W.; Wang, J.; Xu, G. Establishment of atmospheric weighted mean temperature model in the polar regions. *Adv. Space Res.* **2020**, *65*, 518–528. [CrossRef]

52. Zhao, Q.; Liu, K.; Zhang, T.; He, L.; Shen, Z.; Xiong, S.; Shi, Y.; Chen, L.; Liao, W. A Global Conversion Factor Model for Mapping Zenith Total Delay onto Precipitable Water. *Remote Sens.* **2022**, *14*, 1086. [CrossRef]
53. Adavi, Z. *Assessment of Various Processing Schemes and Solution Strategies to Improve the Performance of GNSS Tropospheric Tomography*; Technische Universität Wien: Vienna, Austria, 2022.
54. Li, L.; Zhang, K.; Wu, S.; Li, H.; Wang, X.; Hu, A.; Li, W.; Fu, E.; Zhang, M.; Shen, Z. An Improved Method for Rainfall Forecast Based on GNSS-PWV. *Remote Sens.* **2022**, *14*, 4280. [CrossRef]
55. Sá, A. *Tomographic Determination of the Spatial Distribution of Water Vapour Using GNSS Observations for Real-Time Applications*; Wroclaw University of Environmental and Life Sciences: Wrocław, Poland, 2018.
56. Sun, Z.; Zhang, B.; Yao, Y. Improving the Estimation of Weighted Mean Temperature in China Using Machine Learning Methods. *Remote Sens.* **2021**, *13*, 1016. [CrossRef]
57. Vinutha, H.P.; Poornima, B.; Sager, B.M. Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. *Adv. Intell. Syst. Comput.* **2018**, *701*, 511–518. [CrossRef]
58. Xie, W.; Huang, G.; Fu, W.; Shu, B.; Cui, B.; Li, M.; Yue, F. A quality control method based on improved IQR for estimating multi-GNSS real-time satellite clock offset. *Measurement* **2022**, *201*, 111695. [CrossRef]
59. Yin, A.; Liu, H.; Yang, L.; Li, L. Analyzing the Impact of Climate Factors on GNSS-Derived Displacements by Combining the Extended Helmert Transformation and XGboost Machine Learning Algorithm. *J. Sens.* **2021**, *2021*, 9926442. [CrossRef]
60. Davis, J.L.; Herring, T.A.; Shapiro, I.I.; Rogers, A.E.E.; Elgered, G. Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio Sci.* **1985**, *20*, 1593–1607. [CrossRef]
61. Landskron, D. *Modeling Tropospheric Delays for Space Geodetic Techniques*; Technische Universität Wien: Vienna, Austria, 2017.
62. Landskron, D.; Böhm, J. VMF3/GPT3: Refined discrete and empirical troposphere mapping functions. *J. Geod.* **2018**, *92*, 349–360. [CrossRef] [PubMed]
63. Brown, S. *Machine Learning, Explained*; MIT Sloan School of Management: Cambridge, MA, USA, 2021.
64. Lee, J.A.N. *Computer Pioneer*; IEEE Computer Society, History Committee: Piscataway, NJ, USA, 2021.
65. Tajnai, C.E. Samuel Was Artificial Intelligence Pioneer. In *Computing Research News*; Computing Research Association: Washington, DC, USA, 1991.
66. Kelleher, J.D.; Mac Namee, B.; D'Arcy, A. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*; The MIT Press: Cambridge, MA, USA, 2015.
67. Shinde, P.P.; Shah, S. A Review of Machine Learning and Deep Learning Applications. In Proceedings of the Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–6.
68. Siemuri, A.; Kuusniemi, H.; Elmusrati, M.S.; Välisuo, P.; Shamsuzzoha, A. Machine Learning Utilization in GNSS—Use Cases, Challenges and Future Applications. In Proceedings of the International Conference on Localization and GNSS (ICL-GNSS), Tampere, Finland, 1–3 June 2021; pp. 1–6.
69. Osman, A.I.A.; Ahmed, A.N.; Huang, Y.F.; Kumar, P.; Birmia, A.H.; Sherif, M.; Sefelnasr, A.; Ebraheemand, A.A.; El-Shafie, A. Past, Present and Perspective Methodology for Groundwater Modeling-Based Machine Learning Approaches. *Arch. Computat. Methods Eng.* **2022**, *29*, 3843–3859. [CrossRef]
70. Ding, M. A neural network model for predicting weighted mean temperature. *J. Geod.* **2018**, *92*, 1187–1198. [CrossRef]
71. Li, H.; Li, J.; Liu, L.; Huang, L.; Zhao, Q.; Zhou, L. Random Forest-Based Model for Estimating Weighted Mean Temperature in Mainland China. *Atmosphere* **2022**, *13*, 1368. [CrossRef]
72. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
73. Lazri, M.; Soltane, A. Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVIRI data. *Atmos. Res.* **2018**, *203*, 118–129. [CrossRef]
74. Łoś, M.; Smolak, K.; Guerova, G.; Rohm, W. GNSS-Based Machine Learning Storm Nowcasting. *Remote Sens.* **2020**, *12*, 2536. [CrossRef]
75. Nakagawa, Y.; Miyauchi, T.; Higashino, T.; Okada, M. Application of random forest to classify weather observation into rainfall using GNSS receiver. In Proceedings of the IEEE VTS 17th Asia Pacific Wireless Communications Symposium (APWCS), Osaka, Japan, 30–31 August 2021; pp. 1–5.
76. Ghassemi, B.; Dujakovic, A.; Żółtak, M.; Immitzer, M.; Atzberger, C.; Vuolo, F. Designing a European-Wide Crop Type Mapping Approach Based on Machine Learning Algorithms Using LUCAS Field Survey and Sentinel-2 Data. *Remote Sens.* **2022**, *14*, 541. [CrossRef]
77. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2001**, *2*, 18–22.
78. Liu, Y.; Wang, Y.; Zhang, J. New Machine Learning Algorithm: Random Forest. In *Information Computing and Applications*; ICICA, 2012; Lecture Notes in Computer Science; Liu, B., Ma, M., Chang, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7473.
79. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
80. Brownlee, J. Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python. In *Machine Learning Mastery*; Jason Brownlee, 2020.
81. Liashchynskyi, P.; Liashchynskyi, P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. *arXiv* **2019**, arXiv:1912.06059.

82. Rebala, G.; Ravi, A.; Churiwala, S. *An Introduction to Machine Learning*; Springer: Cham, Switzerland, 2019.
83. Suparta, W.; Alhasa, K.M. Modeling of zenith path delay over Antarctica using an adaptive neuro fuzzy inference system technique. *Expert Syst. Appl.* **2015**, *42*, 1050–1064. [CrossRef]
84. Benevides, P.; Catalao, J.; Nico, G. Neural Network Approach to Forecast Hourly Intense Rainfall Using GNSS Precipitable Water Vapor and Meteorological Sensors. *Remote Sens.* **2019**, *11*, 966. [CrossRef]
85. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
86. Selbesoglu, M.O. Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and GNSS data. *Eng. Sci. Technol. Int. J.* **2020**, *23*, 967–972. [CrossRef]
87. Gupta, A.; Nahar, P. Classification and yield prediction in smart agriculture system using IoT. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *14*, 10235–10244. [CrossRef]
88. Keras. BayesianOptimization Tuner. Available online: https://keras.io/api/keras_tuner/tuners/bayesian/ (accessed on 24 April 2023).
89. Masum, M.; Shahriar, H.; Haddad, H.; Faruk, M.J.H.; Valero, M.; Khan, M.A.; Rahman, M.A.; Adnan, M.I.; Cuzzocrea, A.; Wu, F. Bayesian Hyperparameter Optimization for Deep Neural Network-Based Network Intrusion Detection. In Proceedings of the IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 5413–5419.
90. Nguyen, V. Bayesian Optimization for Accelerating Hyper-Parameter Tuning. In Proceedings of the IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, Italy, 3–5 June 2019; pp. 302–305.
91. Snoek, J.; Rippel, O.; Swersky, K.; Kiros, R.; Satish, N.; Sundaram, N.; Patwary, M.M.; Adams, R.P. Scalable Bayesian Optimization Using Deep Neural Networks. *arXiv* **2015**. [CrossRef]
92. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *arXiv* **2012**. [CrossRef]
93. Victoria, A.H.; Maragatham, G. Automatic tuning of hyperparameters using Bayesian optimization. *Evol. Syst.* **2019**, *12*, 217–223. [CrossRef]
94. Rüeger, J. Refractive index formulae for electronic distance measurements with radio and millimetre waves. *UniServ Rep.* **2002**, *109*, 758–766.
95. Bolton, D. The Computation of Equivalent Potential Temperature. *Mon. Weather. Rev.* **1980**, *108*, 1046–1053. [CrossRef]
96. Witt, L.D.; McKean, J.W.; Naranjo, J.D. Robust measures of associatiosn in the correlation model. *Stat. Probab. Lett.* **1994**, *20*, 295–306. [CrossRef]
97. Osah, S.; Acheampong, A.A.; Fosu, C.; Dadzie, I. Deep learning model for predicting daily IGS zenith tropospheric delays in West Africa using TensorFlow and Keras. *Adv. Space Res.* **2021**, *68*, 1243–1262. [CrossRef]
98. Ibrahim, K.; Huang, Y.F.; Ahmed, A.N.; Koo, C.H.; El-Shafie, A. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alex. Eng. J.* **2022**, *61*, 279–303. [CrossRef]