Chula
Chulalongkorn University

CHULA ENGINEERING
Foundation toward Innovation

Investigating the use of machine learning-derived weighted mean temperature for GPS-PWVs estimation

Chaiyut Charoenphon[1], Chokchai Trakolkul[2], and Chalermchon Satirapod[1*]
[1]*Mapping and Positioning from Space (MAPS) Technology Research Center, Department of Survey Engineering, Chulalongkorn University, Bangkok, Thailand*
[2]*Faculty of Engineering and Architecture, Rajamangala University of Technology Suvarnabhumi, Thailand*
Email: chaiyut.c@chula.ac.th; chokchai.t@rmutsb.ac.th; and chalermchon.s@chula.ac.th (corresponding author)*

Abstract: GNSS data has become a widely recognized alternative tool for estimating Precipitable Water Vapor (PWV) used in meteorological applications. It enables all-weather PWV tracking, 24 hours a day, 7 days a week. However, accurate GNSS-PWV determination, especially in tropical climates, relies heavily on the weighted mean temperature (Tm). Unfortunately, the traditional method of obtaining Tm involves expensive meteorological tools like radiosondes or microwave radiometers, which are often unavailable at most GNSS stations in Thailand. An alternative approach utilizes a linear relationship between Tm and surface temperature (Ts). However, this method is susceptible to local seasonal variations, time-dependent factors, and spatial coordinates, all of which can affect the accuracy of the linear regression model. Additionally, traditional linear empirical models have inherent limitations in precision. They cannot capture complex spatial and temporal variations, leading to an inability to depict these changes accurately. In this study, we compared the performance of two methods for calculating the Tm models: a countrywide linear model and a regression Artificial Neural Network (ANN) algorithm with Gated Recurrent Units (GRU) for estimating GNSS-PWVs. Tm values were derived from ERA5 Reanalysis data (ECMWF) between 2016 and 2021. Our primary results showed a mean bias and Root Mean Square Error (RMSE) of -0.95±1.5K and -0.06±1.1K, respectively, for the countrywide and ANN-Tm models compared to the reference ECMWF-Tm values in 2022. The ANN-Tm model demonstrated a noteworthy improvement in performance, achieving a 27% reduction in RMSE compared to the countrywide linear model. This enhanced capability likely stems from the ANN's ability to capture complex spatial and temporal variations in the data, a limitation inherent to the simpler linear model. To evaluate the effectiveness of the ANN-Tm model for GPS-PWV estimation, we investigated and validated GPS-PWVs from 280 GNSS CORS stations around Thailand against ERA5-PWVs from ECMWF. The performance using the countrywide and ANN-Tm models showed mean bias and RMSE of -0.34±2.57 mm and -0.14±2.57 mm, respectively. The ANN-Tm model displayed a statistically significant improvement in the mean bias of GPS-PWV estimation ($p < 0.05$), although the overall performance improvement was modest. Nonetheless, using Tm values from the ANN-Tm model offers potential for improved GNSS meteorology, particularly regarding mean bias reduction. Our findings suggest that estimating GPS-PWVs in tropical regions remains challenging.

1. Introduction

The importance of accurate weighted mean temperature (T_m) is paramount for acquiring precise Precipitable Water Vapor (PWV) values from GNSS data. Following Bevis et al. (1992), researchers have made significant progress in developing and refining T_m models to improve their global and local accuracy. Studies by Mendes et al. (2000), Schueler et al. (2001), Suwantong et al. (2016), and Charoenphon et al. (2022) have established successful linear empirical models relating T_m to air surface temperature (T_s). While these models can be enhanced by incorporating additional factors like spatial coordinates, temporal variations, and seasonal effects, their inherent limitations prevent them from capturing complex temporal and spatial variations. This study aims to address these limitations by employing machine learning techniques. Machine learning allows for incorporating a larger number of parameters, potentially leading to significant improvements in T_m estimation accuracy, thereby reducing the root mean square error (RMS) of residual distributions and mitigating seasonal variations in GPS-PWVs.

2. Methodology

The weighted mean temperature (T_m), defined by equation (1), cannot be directly obtained at GNSS sites because it requires vertical pressure (e) and temperature (T) profiles. Therefore, an alternative method often utilizes a linear function relating T_m to air surface temperature (T_s), as shown in equation (2).

$$T_m = \int_h^\infty \frac{e}{T} dh \Big/ \int_h^\infty \frac{e}{T^2} dh \quad (1)$$

$$T_m = \alpha + \beta \cdot T_s \quad (2)$$

In Thailand, the initial countrywide T_m model was introduced by Suwantong et al. (2016), followed by Chaiyut et al. (2022) who introduced ERA5-Tm and AIRS-Tm based models. This study investigated using a regression artificial neural network (ANN) to improve the accuracy of T_m estimation. Hourly T_m values computed using ERA5 for the period between 2011 and 2020 served as the training data. The best result on the test dataset was achieved by employing a multivariate regression model with eight variables and 7 layers, which included two BiGRU and Attention layers with 8 neurons each. In this model, we utilized seven hours of preceding data to forecast the T_m for the subsequent hour.

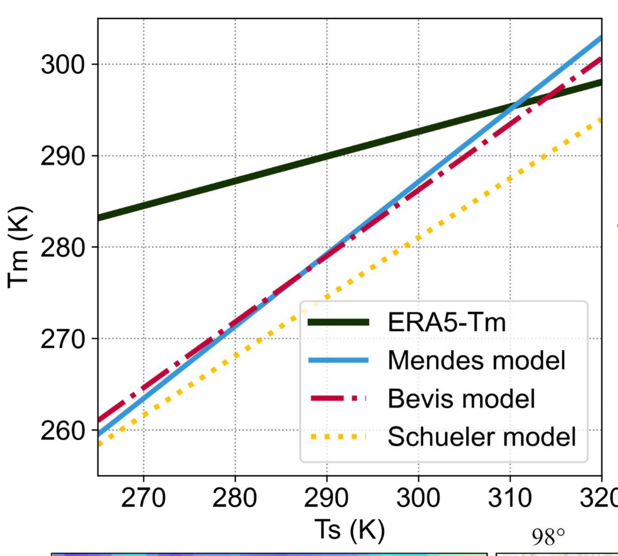
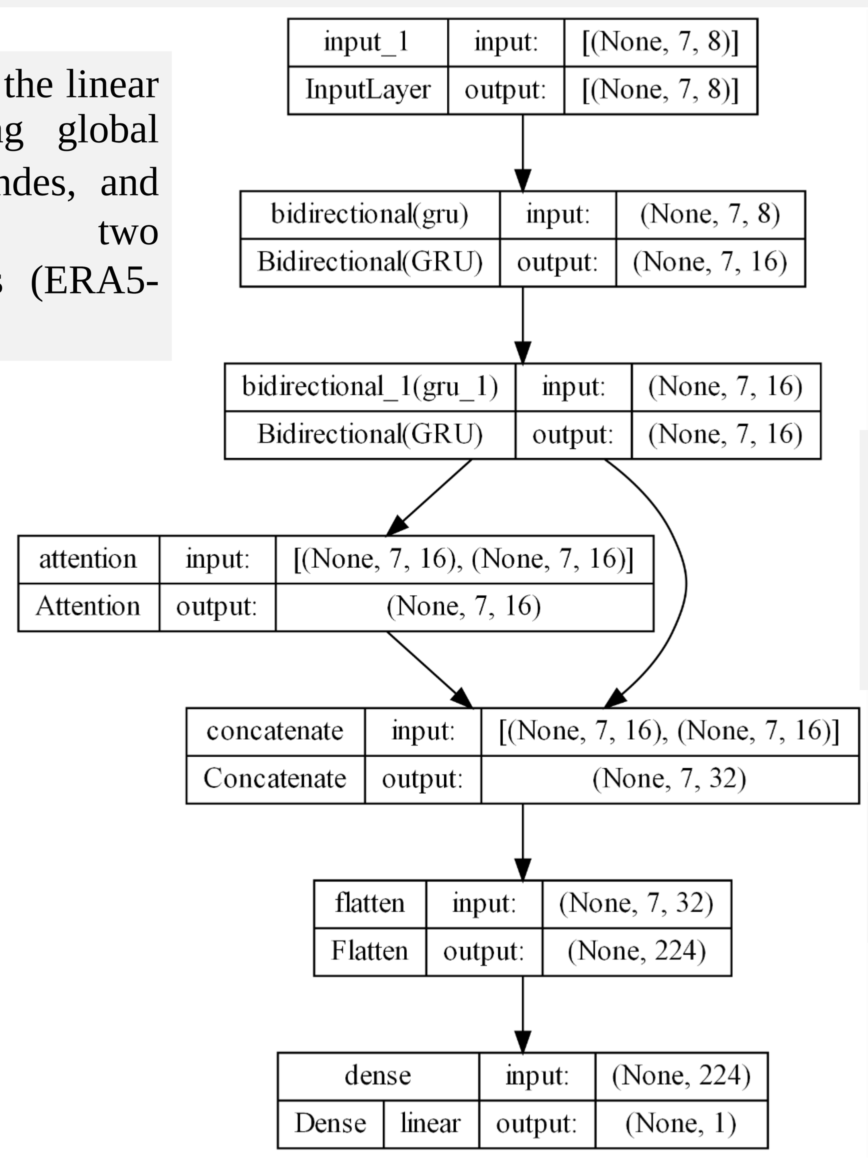


Fig 1. Illustration of the linear Tm models including global models (Bevis, Mendes, and Schueler) and two countrywide models (ERA5-Tm, and Suwantong)



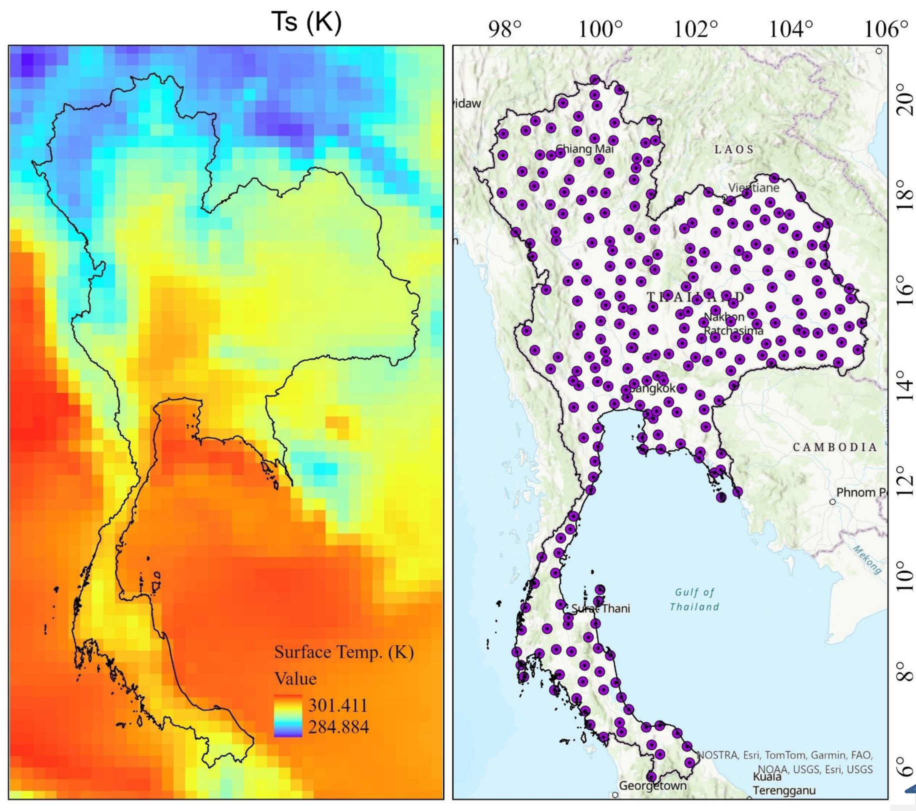


Fig 3. The T_s values are from ERA5 (left). The 280 GNSS stations are utilized to derive GPS-PWVs (right).

Fig 2. The proposed model is composed for 7 layers embedding the input, BiGRU and Attention layer, and the dense layer performs the prediction. The input contains 7 time-steps of previous data (hourly).

3. Results

3.1 ANN-Tm Model

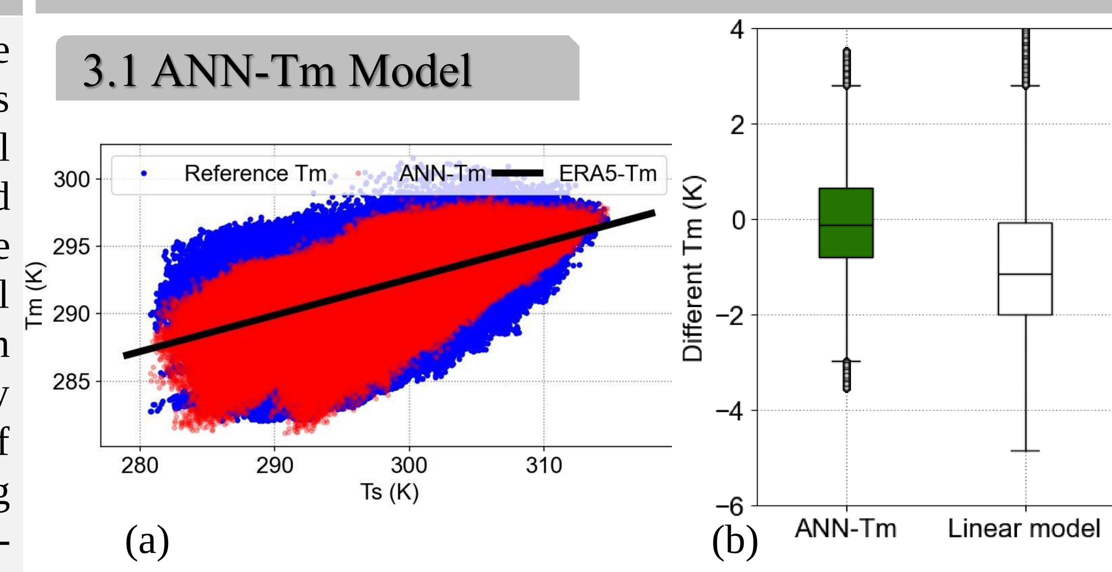


Table 1. The different T_m estimated from ANN-Tm, the countrywide linear model (ERA5-Tm) against ECMWF-Tm as reference.

Statistics	Different Tm (K).	
	Ann-Tm	Linear-Tm
Count	6723593	
Mean	-0.06	-0.95
Median	-0.12	-1.15
STD	1.14	1.52

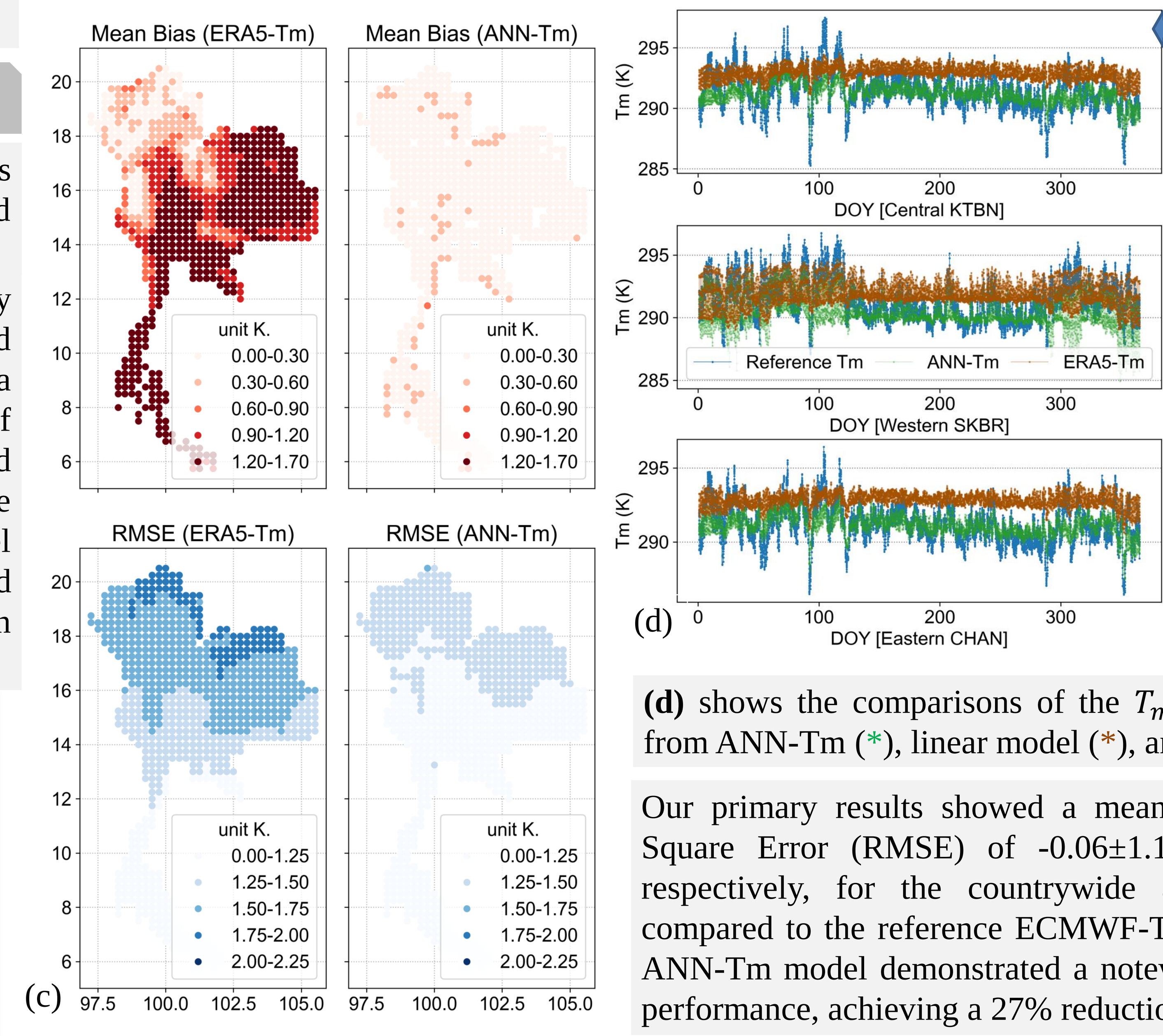


Fig 4. (a) shows the T_m values estimated from ANN-Tm (*), the linear model (**—**), and The true values from ECMWF-Tm (*) (Reference values). **(b)** show the distribution of the different Tm. **(c)** shows the mean bias and RMSE of different T_m calculated from linear (ERA5-Tm) model (left) and ANN-Tm (right) across country. **(d)** shows the comparisons of the T_m time series estimated from ANN-Tm (*), linear model (*), and ECWF-Tm (*).

Our primary results showed a mean bias and Root Mean Square Error (RMSE) of -0.06±1.14K and -0.95±1.52K, respectively, for the countrywide and ANN-Tm models compared to the reference ECMWF-Tm values in 2022. The ANN-Tm model demonstrated a noteworthy improvement in performance, achieving a 27% reduction in RMSE compared to the countrywide linear model. This enhanced capability likely stems from the ANN's ability to capture complex spatial and temporal variations in the data, a limitation inherent to the simpler linear model as shown in Fig4. (d)

3.2 Estimating GPS-PWVs

Table 2. Descriptive Statistics of GPS-PWV Estimates from Linear Model, ANN-Tm Model, and ECMWF-Tm (Reference Tm values)

Statistics	Different PWVs (mm).		
	Ann-Tm	Linear-Tm	Ref-Tm
Count	2209319		
Mean	-0.13	-0.34	-0.14
Median	-0.2	-0.5	-0.2
STD	2.57	2.57	2.56

To evaluate the effectiveness of the ANN-Tm model for GPS-PWV estimation, we investigated and validated GPS-PWVs from 280 GNSS CORS stations around Thailand against ERA5-PWVs from ECMWF. The performance using the Linear-Tm and ANN-Tm models showed mean bias and RMSE of -0.34±2.57 mm and -0.13±2.57 mm, respectively. the overall performance improvement was modest. Our findings suggest that estimating GPS-PWVs in tropical regions remains challenging.

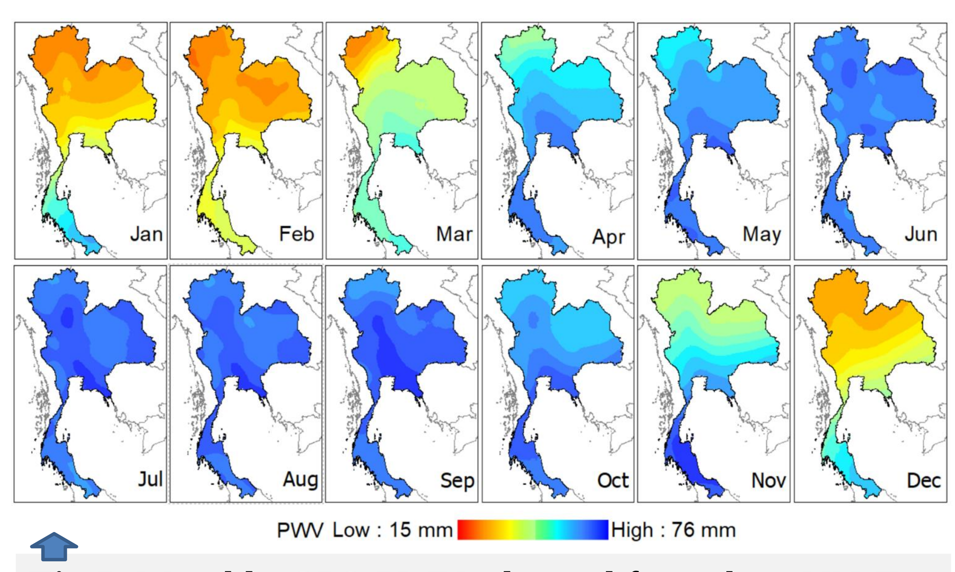


Fig 5. Monthly GPS-PWVs derived from the ANN-Tm model in 2022.

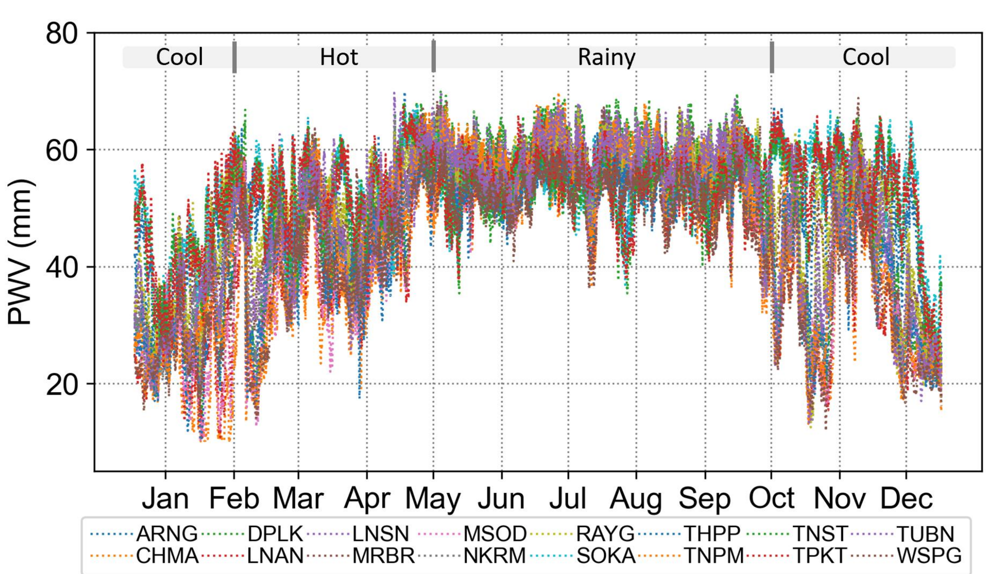


Fig 6. GPS-PWV time series derived from the ANN-Tm model from 16 GNSS stations (2022).

4. Summary and Discussion.

The ANN-Tm model displayed a statistically significant improvement in the mean bias of GPS-PWV estimation ($p < 0.05$), although the overall performance improvement was modest. Nonetheless, using Tm values from the ANN-Tm model offers potential for improved GNSS meteorology, particularly regarding mean bias reduction. Our findings suggest that estimating GPS-PWVs in tropical regions remains challenging.

Acknowledgements

This study was supported by grants from the Mapping and Positioning from Space (MAPS) Technology Research Center, Department of Survey Engineering, Chulalongkorn University.