RETRIEVAL OF PRECIPITABLE WATER FROM HIMAWARI-8 DATA

Hiroki Otsubo (1), Keiji Imaoka (1), Hidenori Shingin (1), Kakuji Ogawara (1)

[1] Yamaguchi University, 16-1 Tokiwadai 2-chome, Ube-shi, Yamaguchi, 755-8611, Japan
Email: b014vd@yamaguchi-u.ac.jp; k.imaoka@yamaguchi-u.ac.jp
shingin@yamaguchi-u.ac.jp; ogawara@yamaguchi-u.ac.jp

KEY WORDS: Himawari-8, AHI, precipitable water, radiance, support vector regression

ABSTRACT: Highly-frequent precipitable water (vertically integrated water vapor) measurements are required for short-term forecasting of severe weather events such as heavy rain. Although there are a number of methods for measuring precipitable water already exist, wide-area observations with high spatial and temporal resolutions are still difficult. Since Himawari-8 began its observation, it became possible to observe the area of Japan every 2.5 minutes with the spatial resolution of 2 km by infrared bands. Although there are restrictions due to clouds, we can estimate the precipitable water by using infrared brightness temperatures (Tbs) observed by the Advanced Himawari Imager (AHI) onboard Himawari-8. In this study, we performed two types of regression analyses, multiple linear regression (MLR) and support vector regression (SVR) to estimate the precipitable water from AHI Tbs by using simulation- and observation-based datasets. Also, we validated the accuracy of the estimates by using precipitable water observed by radiosondes. We prepared several datasets for estimating parameters of regression models and validating the performance of the models. For the parameter estimation, we used two datasets. One dataset consists of simulated Tbs computed by the Radiative Transfer for TOVS (RTTOV) and corresponding precipitable water values (RTTOV dataset). When simulating Tbs, we used temperatures, relative humidity, and air pressure data provided from the Japanese 55-year Reanalysis or JRA-55. Another dataset was created by using collocated data from AHI infrared Tbs and GPS precipitable water estimates (GPS dataset). For the validation, collocated dataset of AHI infrared Tbs and precipitable water estimates from radiosonde data was prepared (RAOB dataset). As explanatory variables of regression models, multiband Tbs and satellite zenith angle were used. As a result of band selection by using SVR models, Tbs at bands 13, 15, and 16 were selected as optimum and minimum explanatory variables. Band 13 and 15 are already known as important bands for estimating precipitable water, such as in the traditional split-window model. The probable reason of selecting band 16, located on the edge of CO_2 absorption lines, is the better correlation with air temperature. As the results of MLR models, the root mean squared error (RMSE) values against RAOB dataset were 8.1 and 7.5 mm for 3-band algorithms optimized by RTTOV and GPS dataset, respectively. For the SVR models, RMSE values against RAOB dataset were 8.6 and 7.3 mm for the optimized by RTTOV and GPS dataset, respectively.

1. INTRODUCTION

For short-term forecasting of severe weather events such as heavy rain, highly-frequent water vapor measurements are required. Although there are a number of existing methods for measuring water vapor content, such as radiosondes, and Global Navigation Satellite Systems (GNSS), wide-area observations with high spatial and temporal resolutions are still difficult. As one of the satellite-based remote sensing observation techniques, visible- and infrared- based observation techniques, such as the split-window technique (Akatsuka et al., 2011), have been intensively investigated. Hereafter, we call the value obtained by integrating the amount of water vapor in the vertical direction precipitable water.

Since Himawari-8 began its observation on July 7, 2015, it became possible to observe the entire Earth as seen from the satellite (hereafter called Full-Disk) every 10 minutes with the spatial resolution of 2 km, and the area of Japan every 2.5 minutes with the same spatial resolution that of

Full-Disk. The number of bands increased from 5 to 16 and the horizontal resolution was improved from 4 to 2 km in infrared bands compared to the predecessor imager on the Multi-functional Transport Satellite (MTSAT) series. Since atmospheric conditions change in a short period of time, observed precipitable water with short time interval like by Himawari-8 are useful for forecasting weather changes. Also, the increased spatial resolution may help improving the accuracy. In this paper, we will describe and discuss the regression models to estimate the precipitable water from Himawari-8 infrared bands, and the validation results against collocated in-situ measurements.

2. INSTRUMENT AND DATA

2.1 Himawari-8 and AHI

The Himawari-8 satellite was developed and is being operated by the Japan Meteorological Agency (JMA). The Advanced Himawari Imager (AHI) on Himawari-8 consists of 3 visible bands, 3 near infrared bands, and 10 infrared bands. We used only infrared bands in this study. Table 1 shows the specification of infrared bands of Himawari-8 (Bessho et al., 2016).

We downloaded the Himawari L1 Gridded data from the P-Tree system maintained by the Japan Exploration Agency (JAXA). The data cover AHI Full-Disk area with 2 km spatial resolution every 10 minutes.

Table 1. Specification of infrared bands of Himawari-8.

Channel Number	Spatial Resolution at Sub Satellite Point (km)	Central wave length (μm)
7	2	3.8853
8	2	6.2429
9	2	6.9410
10	2	7.3467
11	2	8.5926
12	2	9.6372
13	2	10.4073
14	2	11.2395
15	2	12.3806
16	2	13.2807

2.2 JRA-55 and MGDSST

In simulating AHI Tbs, we used the Japanese 55-year Reanalysis (JRA-55) (Kobayashi et al., 2015) and the Merged Satellite and In-situ Data Global Daily Sea Surface Temperatures (MGDSST) (Sakurai et al., 2005), both provided by JMA. The JRA-55 data contain analyzed atmospheric profiles such as air temperature, pressure, and relative humidity with 1.25-degree resolution and every 6 hours. Table 2 shows the specification of JRA-55 data used in our analysis.

Table 2. Specification of JRA-55 data used in our analysis.

Date	Time	Area	Number of profiles
January 10, 2016 July 10, 2016 October 10, 2016 April 10, 2017	12UTC	AHI Full-Disk Area	9409 for each

2.3 GPS AND RADIOSONDE PRECIPITABLE WATER DATA

The Global Positioning System (GPS) precipitable water dataset was created by JAXA to validate

the accuracy of satellite precipitable water. The dataset contains precipitable water values estimated from the GPS signals over 123 points, which were selected from the GPS points registered by the International GNSS service (IGS). We used the data over 22 stations located within AHI Full-Disk area. Figure 1 (a) shows the location of GPS stations used in this study.

The radiosonde dataset was also created by JAXA. This dataset contains the atmospheric vertical profiles for temperature, air pressure, relative humidity, and calculated precipitable water. We used the locations within the AHI Full-Disk area. Figure 1 (b) shows the location of radiosondes used in this study.

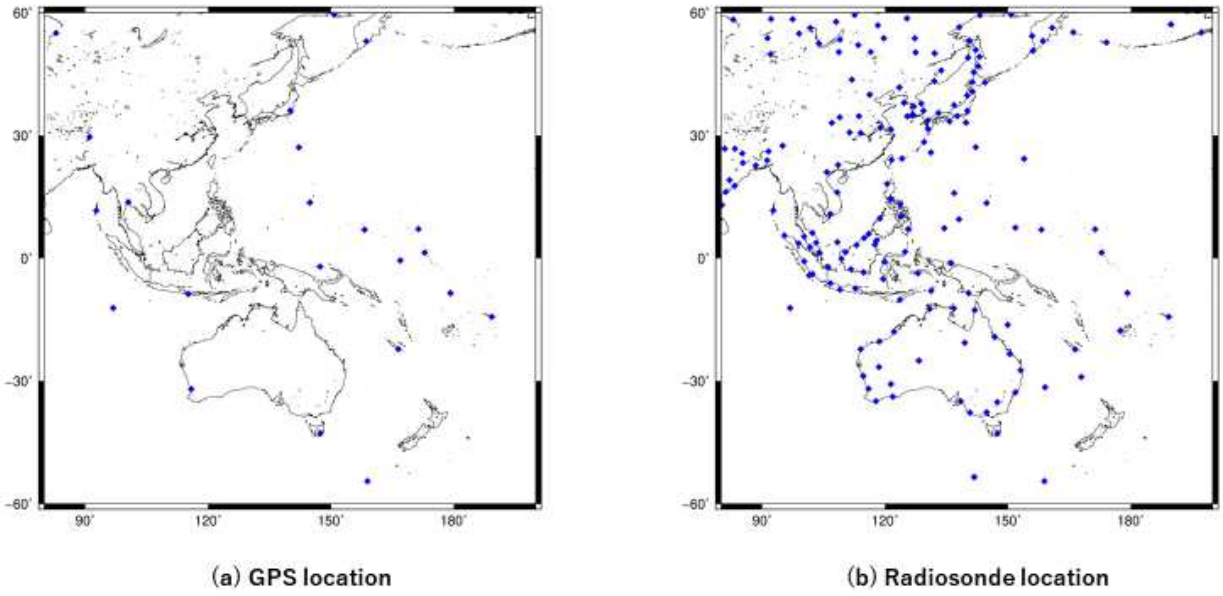

Figure 1. Locations of (a) GPS and (b) radiosonde stations used in the present study. Blue points represent the station positions.

3. METHOD

For retrieving precipitable water, we tested two types of regression models including multiple linear regression (MLR) and support vector regression (SVR). To determine parameters of these models and validate the performance, we prepared several datasets. For the parameter determination, we used simulation-based and observation-based datasets. The simulation-based dataset consists of simulated Tbs computed by the Radiative Transfer for TOVS (RTTOV) (Saunders et al., 2018) and corresponding precipitable water values (RTTOV dataset). The observation-based dataset consists of collocated data of AHI infrared Tbs and GPS precipitable water estimates (GPS dataset). For the validation, collocated dataset of AHI infrared Tbs and precipitable water estimates from radiosonde data was prepared (RAOB dataset).

3.1 Dataset preparation

To prepare the RTTOV dataset, we simulated AHI Tbs by using RTTOV with atmospheric profiles from JRA-55 including temperature, air pressure, and relative humidity. For all the atmospheric profiles, we set the satellite zenith angle (SZA) to 5, 43, and 60 degrees. As the surface temperatures, we used MGDSST values over ocean. Due to the lack of good data source, we used air temperatures at 1-m height over land instead of land surface temperatures. Corresponding precipitable water amounts were available in JRA-55 data. The available number of data is 112,908 in total. Because of the inability to estimate the precipitable water over cloud-covered areas by using infrared Tbs, we need to remove the cloud-covered data in creating the GPS and RAOB datasets. In this study, we adopted simple and empirical method using Tbs decreases due to clouds at Band 13, which is the

most transparent band against water vapor absorption. First, we picked up the highest hourly Tb value of Band 13 in each month over each pixel. We assumed this value as a cloud-free Tb of the hour, month, and pixel. When a difference between a Tb at targeted pixel and the corresponding cloud-free Tb exceeds the threshold value, we judged the pixel as cloudy. By the visible inspection of images, we adopted the value of 4 K as the appropriate threshold. After this cloud masking process over the AHI Tbs, temporally and spatially collocated Tbs against GPS and RAOB measurements were extracted. In total, the available numbers of data are 248115 and 10704 for GPS and RAOB dataset, respectively.

3.2 Regression models

3.2.1 MLR

Based on the existing investigations on split-window method (e.g., Akatsuka et al., 2011), MLR models will work to some extent, although they may not be perfect due to the nonlinear nature of the precipitable water retrieval and multi-collinearity problems. To investigate the advantage of newly-installed infrared bands for AHI, we tested MLR models to retrieve precipitable water. Precipitable water PW can be expressed as follows.

$$PW = a_0 + \sum_{i=1}^{9} a_i * T_{B(i+7)} + a_{10} * \cos(\theta_{sza}) \quad (1)$$

Here, a_x and T_{Bx} are the regression coefficients and Tbs for band x, respectively, and θ_{sza} represents SZA.

3.2.2 SVR

Because of the ability to handle nonlinear problems and the robust characteristics for noise and multi-collinearity issues, we wanted to test Epsilon-SVR for comparing with the results by MLR. As is the case in MLR analysis, explanatory variables are multi-band infrared Tbs and SZA. To train the models, we used LIBSVM software (Chang and Lin, 2011) implemented in e1071 packages on R. Although we need to decide hyper-parameters to obtain ideal model, we used default values in this analysis. In creating the models, we randomly divided the dataset into equal-sized training and validation datasets to check the over-training issue.

3.3 Band selection

We evaluated the appropriate bands for SVR models by using the GPS dataset, and the same bands were used for MLR models. For the band selection, we adopted the step-down approach. Starting from 10 bands, we created the model by using the training dataset, and sequentially excluded a band that provided the least contribution (in other words, did not make the performance worse) for the validation dataset at each stage.

We used root mean squared error (RMSE) for the validation dataset as a performance indicator. The equation of RMSE is as follows,

$$RMSE = \sqrt{\frac{1}{N}(y_i - \hat{y}_i)^2} \quad (2)$$

In Equation (2), N, y_i, and $\hat{y}_i$ indicate the number of data, predict value, and actual value, respectively.

4. RESULTS AND DISCUSSION

4.1 Band selection

Figure 2 summarizes the result of band selection for SVR. We can see the noticeable improvement of RMSE by 3-band algorithm compared to that by 2-band algorithm. After that, the improvement was not so significant, although increasing the number of bands gradually decreased the RMSE. The selected bands for 3-band algorithm were Bands 13, 15, and 16. Bands 13 and 15 are the typical bands used in the split-window algorithm. As indicated in Fig. 2, Band 16 is located in the CO_2 absorption band. It is well known that there is a good correlation between observed Tbs at CO_2 absorption bands and air temperature, and they are utilized as air temperature sounding. Since the typical split-window algorithm actually needs the mean-radiation temperature of the atmospheric layer, it is reasonable to think that Band 16 provided the information of air temperature and improved the performance. Based on this result, we will describe the results only for 2- and 3-band algorithms in the next section.

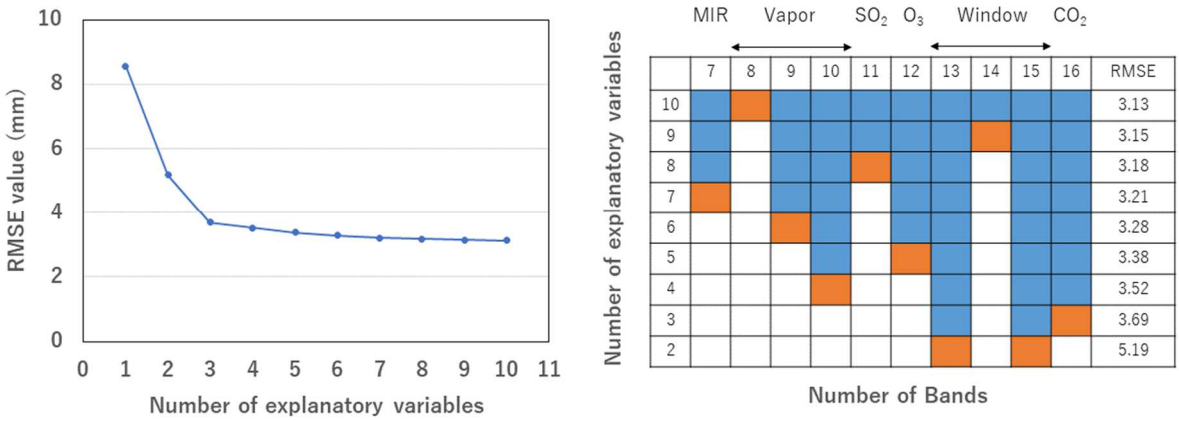

Figure 2. Summary of band selection for SVR. Left panel indicates the number of bands used and corresponding RMSE. Right panel shows the bands used in each stage.

4.2 Algorithm performance

Table 3 shows the RMSE values for different algorithm. For example, MLR-SIM-2 stands for the algorithm with 2-band MLR model trained by RTTOV dataset, and MLR-OBS-3 stands for the case with 3-band MLR model trained by GPS dataset. Figures 3 and 4 show the scatter plots for different algorithms. We should note that, for MLR-OBS-2 and MLR-OBS-3, RMSE values for GPS dataset are just for reference, since they were actually trained by the GPS dataset.

Table 3. Performance in RMSE for different MLR algorithms.

Algorithm	RMSE [mm]	
	GPS	RAOB
MLR-SIM-2	9.7	9.7
MLR-SIM-3	7.9	8.1
MLR-OBS-2	(8.1)	9.0
MLR-OBS-3	(6.3)	7.5

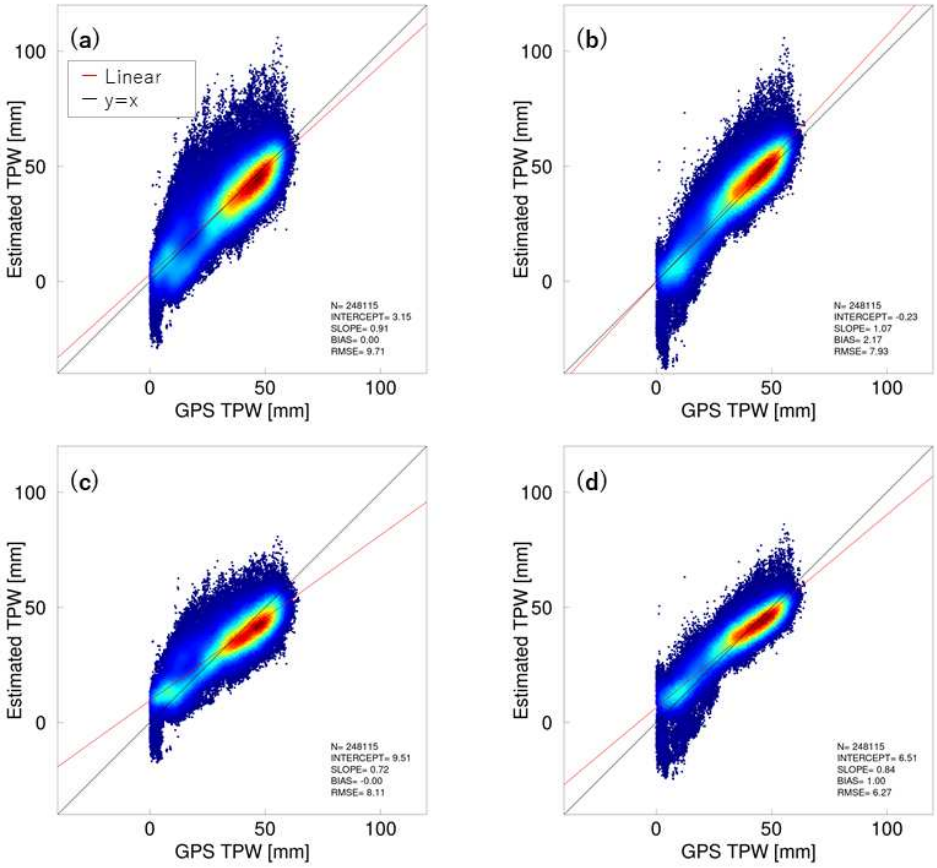

Figure 3. Scatter plots of (a) MLR-SIM-2, (b) MLR-SIM-3, (c) MLR-OBS-2, and (d) MLR-OBS-3 against precipitable water observed by GPS. Horizontal and vertical axes are GPS precipitable water and estimated precipitable water, respectively.

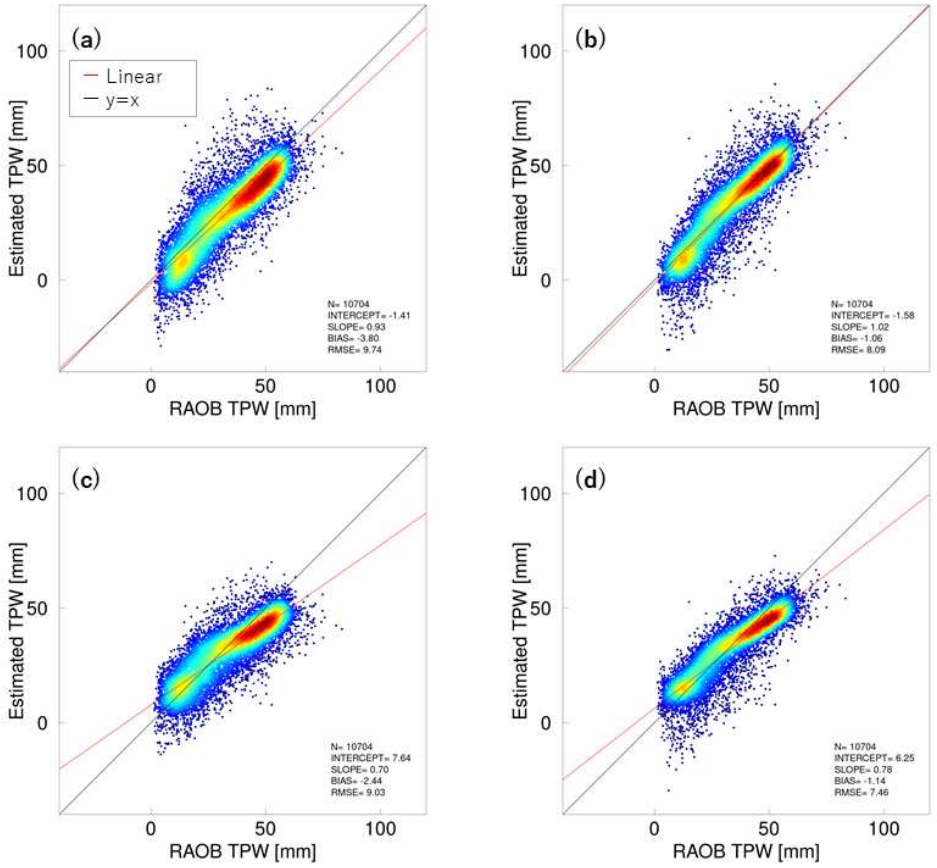

Figure 4. Same as Fig. 3, except against precipitable water observed by radiosondes.

As we can see from the results, for against both GPS and RAOB datasets, the 3-band algorithms with Band 16 performed better than the traditional 2-band algorithms. From the training dataset point of view, the algorithms trained by the GPS dataset performed better than those trained by the RTTOV dataset.

Table 4 shows the RMSE values for different SVR algorithms. The naming convention of the algorithms is similar to that for MLR algorithms. Figures 5 and 6 are the scatter plots for different SVR algorithms. Also similar to the MLR algorithms, for SVR-OBS-2 and SVR-OBS-3, RMSE values for GPS dataset are just for reference.

Table 4. Performance in RMSE for different SVR algorithms.

Algorithm	RMSE [mm]	
	GPS	RAOB
SVR-SIM-2	7.2	8.3
SVR-SIM-3	7.6	8.6
SVR-OBS-2	(5.2)	9.5
SVR-OBS-3	(3.7)	7.3

For the observation-based algorithms, SVR-OBS-2 and SVR-OBS-3, the 3-band algorithm performed better than the 2-band algorithm similar to the MLR algorithms. However, this is not the case for simulation-based algorithm, SVR-SIM-2 and SVR-SIM-3. The 3-band algorithm showed slightly larger RMSE values. This is probably due to the tendency of overestimation shown by larger slopes than unity in Fig. 5(b) and Fig. 6(b). One potential reason is the difference between simulated and observed Tbs due to errors in both radiative transfer model and sensor calibration. The SVR algorithms may be more sensitive to this issue due to the better expressive ability.

When we compare the performances between MLR and SVR algorithms, at least in the current research phase, we should state that the advantages in SVR algorithms are not clear, particularly for RAOB dataset validation. In this paper, we did not perform any hyper-parameter optimizations for the SVR algorithms. We need to optimize the algorithms without any overfitting issues in future. Regarding the training dataset, we originally expected that the RTTOV dataset could be better because of the wide coverage of geophysical conditions. However, this was also not clear in this study. In the next step, we need to confirm the consistency between simulated and observed Tbs. In addition, several assumptions and configurations in the Tb simulation should be re-considered. We arbitrarily set the SZAs for all atmospheric profiles in this study. However, for example, it is unlikely to couple a tropical atmospheric profile and large SZA. We need to select realistic conditions. As surface temperatures over land, we used 1-m height air temperature instead of land surface temperatures. We need to confirm the validity of this assumption. Additionally, to create ideal datasets for regression analysis, we need to confirm the performance of the cloud mask.

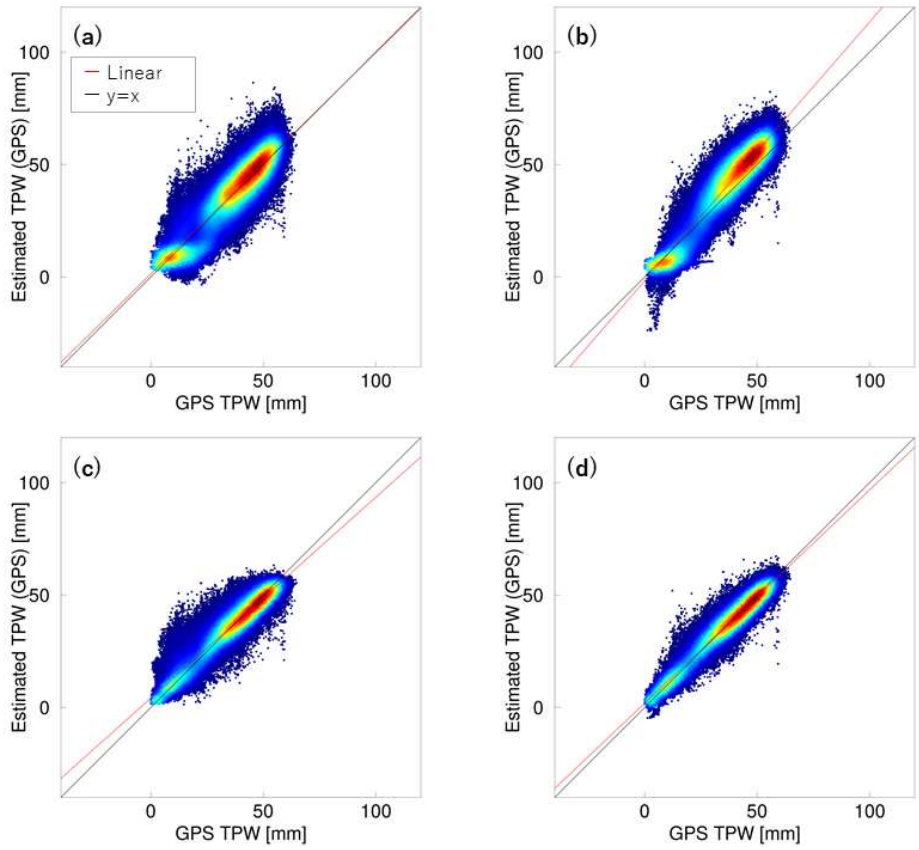

Figure 5. Scatter plots of (a) SVR-SIM-2, (b) SVR-SIM-3, (c) SVR-OBS-2, and (d) SVR-OBS-3 against precipitable water observed by GPS. Horizontal and vertical axes are GPS precipitable water and estimated precipitable water, respectively.

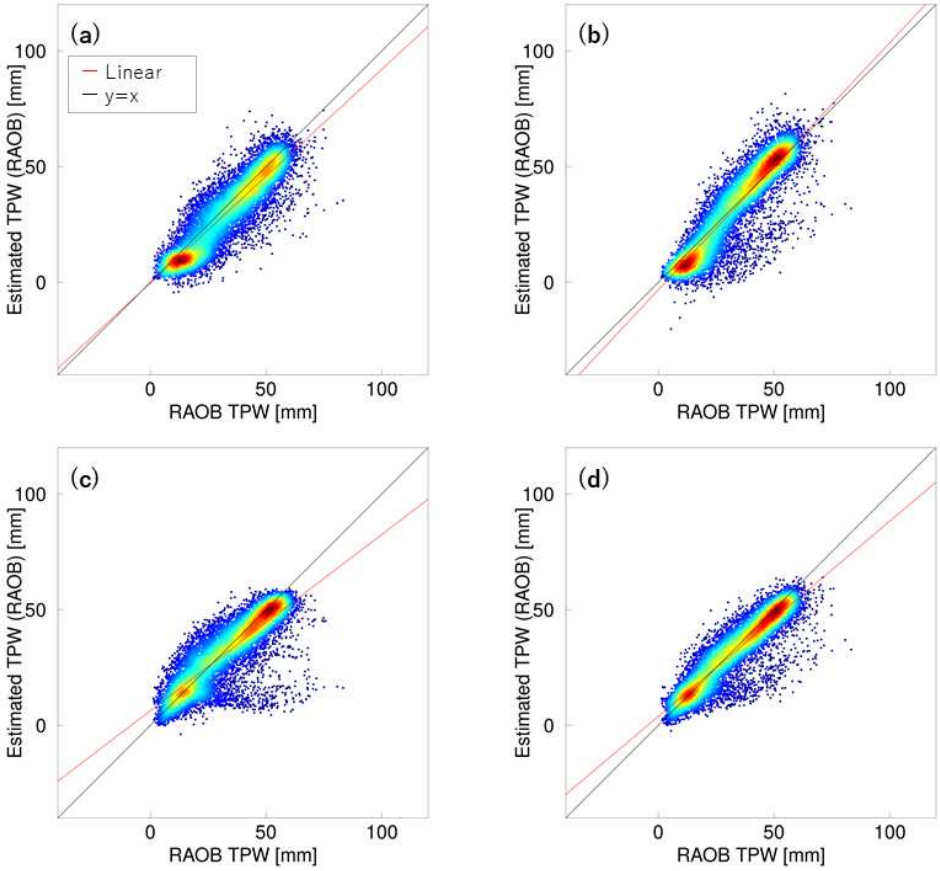

Figure 6. Same as in Fig. 5, except against precipitable water observed by radiosondes.

5.CONCLUSION

We performed two types of regression analyses, MLR and SVR, to estimate the precipitable water amount from AHI Tbs by using simulation-based (RTTOV) and observation-based (GPS) dataset for training, and validated the accuracy of the estimates by using RAOB dataset. Through the band selection by using SVR models, Bands 13, 15, and 16 were selected as optimum and minimum explanatory variables. The probable reason of selecting band 16, located on the edge of CO_2 absorption lines, is the better correlation with air temperature. As the results of MLR models, the RMSE values against RAOB dataset were 8.1 and 7.5 mm for the 3-band algorithms trained by RTTOV and GPS dataset, respectively. For the SVR models, RMSE values against RAOB dataset were 8.6 and 7.3 mm for the 3-band algorithms trained by RTTOV and GPS dataset, respectively.

REFERENCES

Akatsuka, S., Oyoshi, K., and Takeuchi, W., 2011. Development of a precipitable water estimation method using MTSAT measurement. Journal of Remote Sensing Society of Japan, 31 (5), pp.481-489.

Bessho, K., Date, K., and Hayashi, M., 2016. An introduction to Himawari-8/9 – Japan's New-Generation Geostationary Meteorological Satellites. Journal of the Meteorological Society of Japan, 94 (2), pp.151-183.

Chang, C-C., and Lin, C-J., 2011. LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology, 2:27:1—27:27.

Kleespies, J. T., and McMillin, L. M., 1990. Retrieval of precipitable water from observations in the Split Window over varying surface temperatures. Journal of Applied Meteorology, 29, pp.851-862.

Kobayashi, S., Ota, Y., and Harada, Y., 2015. The JRA-55 Reanalysis – General specification and basic characteristics. Journal of the Meteorological Society of Japan, 93 (1), pp.5-48.

Sakurai, T., Kurihara, Y., and Kuragano, T., 2005. Merged satellite and in-situ data global daily SST. Proc. IEEE Int. Geoscience and Remote Sensing Symp., pp. 2606-2608.

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., Bormann, N., and Lupu, C., 2018. An update on the RTTOV fast radiative transfer model (currently at version 12). Geosci. Model Dev., 11, pp.2717-2737.

ACKNOWLEDGEMENT

The Himawari-8 L1 gridded data for Full-Disk area were downloaded by the P-Tree System maintained by JAXA. Precipitable water amount data observed by GPS and radiosonde were provided by JAXA. This work is conducted and supported partly under the collaboration research with JAXA (PI No. ER2GWF201).