Instantaneous Erythemal UV-B Retrieval with Moderate Resolution Imaging Spectroradiometer (MODIS) based on Machine Learning Methods

Ruixue Zhao[1] and He Tao[1]
[1] School of Remote Sensing and Information Engineering, Wuhan University, No.129 Luoyu Road, Hongshan District, Wuhan, Hubei Province, 430079 China
Email: ruixuezhao@whu.edu.cn; taohes@whu.edu.cn

KEY WORDS: Random Forest; Support Vector Regression; Fully Connected Neural Network

ABSTRACT: Erythemal UV-B has a significant impact on human health and the global ecosystem. Therefore, its monitoring has attracted great attention among researchers in recent years. However, traditional satellite-derived UV-B retrieval products have a low spatial resolution, and algorithms to generate these products rely on experience modeling and parameters provided by other satellite products. This paper attempts to use machine learning methods to retrieve all-sky kilometer-level erythemal UV-B from moderate resolution imaging spectroradiometer (MODIS) data with its multispectral information. Specifically, we design a framework that combines fully connected neural network (FCNN) and random forest (RF) to capture the complex relationship between MODIS TOA and erythemal UV-B. In addition, we establish a dataset based on 7 Surface Radiation Budget Network (SURFRAD) stations between 2007 and 2016, including MODIS TOA reflectance, solar zenith angle (SZA), view zenith angle (VZA), MODIS surface reflectance, altitude, and ozone observations. We conduct extensive experiments on the SURFRAD data set, demonstrating that the FCNN+RF performs significantly better than state-of-the-art machine learning methods such as SVR, RF, and FCNN.

1. INTRODUCTION

UV-B radiation refers to electromagnetic radiation with 280 to 320 nm (Singh *et al.*, 2018). Although UV radiation accounts for only 8.3% of solar energy reaching the top of the atmosphere (Utrillas *et al.*,2018), it plays an essential role in the global environment and human health. Casual UV-B exposure can help the body synthesize vitamin D for building and maintaining bones. However, high UV-B exposure can induce human skin cancer, cataracts, and other diseases (Bernard *et al.*, 2019) by damaging DNA/protein structures and inhibit specific cellular responses (Xiang *et al.*, 2017). The part of the UV spectrum responsible for sunburns on human skin and DNA damage is called the erythemal action spectrum, corresponding to erythemal UV irradiance (McKinlay and Diffey, 1987). Consequently, acquiring erythemal UV-B irradiance is essential for academic research and environmental applications due to its great impact on human beings, flora, and fauna species (Iqbal, 2012).

Multiple factors affect the temporal and spatial changes of the surface erythemal UV-B (Calbó *et al.,* 2005), including cloud, ozone, SZA, altitude, surface albedo, etc. The high SZA and low altitude values indicate a longer length of the radiation transmission path in the atmosphere (Sola *et al.*, 2008). Under the same atmospheric conditions, a longer transmission path leads to a greater degree of radiation attenuation, which thus lowers ground erythemal UV radiation (Wang *et al.,* 2015). In the process of atmospheric transmission, the absorption of UV-B by ozone (Kim *et al.,* 2013) and the scattering and absorption of UV-B by clouds and aerosols (Fountoulakis *et al.*, 2016) will all influence the receive erythemal UV-B radiation on the surface of the earth. Generally, clouds and aerosols attenuate 10-30% of erythemal UV-B (Zempila *et al.*, 2018). Surface reflectance can provide snow cover and terrain information of the surface ground, and when covered with snow, the erythemal UV-B under the clear sky may increase by 15 − 25% due to multiple scattering (Renaud *et al.*. 2000). Therefore, it is also necessary to consider this information when acquiring surface erythemal UV-B estimation over different land cover types.

Ground observation networks, for instance, the Surface Radiation Budget Network (SURFRAD), the UV-B Monitoring and Research Program (UVMRP), and the International Network for the Detection of Atmospheric Composition Change (NDACC) (Driemel *et al.*, 2018), can provide point-based accurate erythemal UV-B data with high accuracy and a high temporal resolution, but they cannot provide erythemal UV-B data with broad spatial coverage. Therefore, using satellite data to estimate spatial-temporal continuous erythemal UV-B radiation has been a promising research direction for a long time. Various models have been proposed in recent decades (Zhang *et al.*,2018), and several satellite-based UV products have been developed based on these models. Among these products, Total Ozone Mapping Spectrometer onboard the Earth Probe Satellite (TOMS/EP), Ozone Monitoring Instrument (OMI) Level 2, and Global Ozone Monitoring Experiment-2 (GOME-2) Level 3 are gridded products (~0.1°–1°) (Bhartia, 2007), while OMI level 3 product can provide swath data with a nadir point resolution of 13 × 24 km^2(Levelt *et al.*, 2006). From the resolution perspective, the coarse spatial resolution

remote sensing data of the sensors used for erythemal UV-B retrieval directly leads to the erythemal UV-B product's low resolution. Due to the high spatial-temporal heterogeneity of atmospheric components (particularly for clouds, aerosols, and water vapor), the existing coarse-resolution erythemal UV-B products cannot depict the temporal and spatial variations of erythemal UV-B. From the spectral band's perspective, most existing products use ultraviolet spectral bands to retrieve ozone and erythemal UV-B products. However, compared with ozone, clouds and aerosols also have a significant impact on erythemal UV. The MODIS TOA spectral bands contain the clouds and aerosols information and have a higher spatial resolution (1km×1km). So far, few pieces of research are focusing on erythemal UV-B estimation from MODIS data, which deserves further investigation and is the primary purpose of our paper.

Remote estimation of UV-B has been developed in the past few decades (Wang *et al.*, 2020). The commonly used algorithms can be classified into the following two categories: the look-up-table (LUT) methods and statistical methods. The LUT methods establish the relationship between albedo or reflectance and atmospheric transmittance under different atmospheric conditions to obtain surface erythemal UV-B (Liang *et al.*, 2006). The erythemal UV-B satellite products mentioned in the previous all use the LUT method (Zhang *et al.*,2018) to generate daily erythemal UV-B data. However, the evaluation study of these satellite-based UV-B products all overestimates the surface UV-B to varying degrees. For example, the OMI and TOMS products overestimate erythemal UV-B for some regions, such as the United States, Europe, and high latitudes (Zempila *et al.*,2017). Furthermore, these methods rely on various atmospheric and surface products. However, errors will be introduced for the algorithm when there is no reliable guarantee for these external products' accuracy (Zhang *et al.*, 2015). Fortunately, statistical/empirical methods provide a potential solution for addressing these issues (Liu *et al.*, 2016). The prominent representation ability of machine learning (ML)-based model can capture the nonlinear and complex relationships between the inputs and targets, presenting a potential solution for radiation parameter retrieval (LeCun *et al.*,2015). Although ML models have been applied in solar radiation retrieval and forecast, to our best knowledge, there are few attempts to estimate erythemal UV-B using satellite-based ML models.

This paper uses three ML algorithms to retrieve the erythemal UV-B from MODIS TOA data. The contribution of this study mainly includes three perspectives:

- We use MODIS data whose spectral bands contains cloud and aerosol information as the raw input to retrieve UV-B, which leads to higher resolution compared with traditional satellite-based methods
- We compare different machine learning methods (FCNN, RF, and SVR) to retrieve erythemal UV-B from MODIS TOA input. Moreover, we design an FCNN-based feature extractor to build representations that combine deep and shallow features from raw MODIS inputs for erythemal UV-B retrieval, where the shallow classifiers can be used to perform the final prediction of erythemal UV-B based on the representations extracted from FCNN-based feature extractor.
- We establish datasets at SURFRAD sites to evaluate the model's performance.

2. METHODOLOGY

2.1 Models

The MODIS TOA reflectance data contain atmosphere and surface information. Based on these data, we can directly establish the relationship between MODIS TOA and erythemal UV-B without additional satellite products or LUTs for retrieving atmospheric parameters. Therefore, we try to use ML algorithms to establish the relationship between MODIS TOA and erythemal UV-B so that UV-B products can be automatically generated for downstream tasks.

2.1.1 Random Forest Regressor (RF): Random Forest Regressor (RFR) is an ensemble regression model composed of multiple decision trees. For each decision tree, the root node represents the entire dataset, and for each node, data are recursively split by the selected feature that maximizes a predefined metric such as the GINI index or the information gain, until the data associated with the leaf nodes are purer than a predetermined threshold. The decision tree has the characteristics of slight bias, significant deviation, and tendency to overfit the training data. Random Forest samples multiple subsets from the original training set to train multiple decision trees, and for each tree, it only uses a subset of the features to determine how the nodes are split. Therefore, limiting the sample space and feature space visible to each decision tree can significantly reduce their correlation degree. Finally, the RF averages the predictions of all the decision trees to obtain a more robust final prediction result.

2.1.2 Support Vector Regression Machine (SVR): When the Support Vector Machines (SVM) are applied to regression problems, they would be called Support Vector Regression machines (SVR). Whether SVM or SVR, the core is to map the data to a new optimal high-dimensional representation and use the transformed variables

for classification or regression. Before the popularity of deep neural networks, SVM/SVRs show state-of-the-art performance for their ability to deal with nonlinear problems (Antonanzas *et al.*, 2016). However, SVM/SVRs prove challenging to scale to large datasets. Simultaneously, because they are shallow ML algorithms, SVMs may need to manually derive useful features (feature engineering), which is tricky and cannot be automated (Chollet, 2018).

2.1.3 Fully Connected Neural Network (FCNN): A naive way to use deep neural networks for erythemal UV-B retrieval is to use a fully connected neural network (FCNN), where every node in the current layer is connected to all nodes in the previous layer. The forward process of the lth layer can be formulated as:

$$\mathbf{v}_l = \mathrm{f}_{\mathrm{acts}}(\mathbf{W}_l \mathbf{v}_{l-1} + \mathbf{b}_l), \tag{1}$$

where $\mathbf{v}_l$ is the activations from the l th layer, $\mathbf{W}_l$ is the weight matrix, $\mathbf{b}_l$ is the bias, and $\mathrm{f}_{\mathrm{acts}}$ is the activation function that introduces non-linearity to map data from one domain to another. A commonly used activation function is ReLU (Rectified Linear Unit). In backpropagation, this paper uses the Adam (Adaptive Moment Estimation) optimization algorithm, which updates the model parameters with a fixed learning rate to minimize the loss function, to obtain optimal or approximate optimal network parameters that affect network training and output (Ronneberger *et al.*, 2015).

2.1.3 Fully Connected Neural Network + Random Forest Regressor (FCNN+RF): The FCNN+RF can be regarded as two parts. The first part is an FCNN, which converts MODIS TOA and other parameters in a data-driven manner into a representation that is more relevant to the surface erythemal UV-B retrieval. The second part is an RF, which establishes the relationship between these deep features and the surface erythemal UV-B and predicts the final erythemal UV-B data. Although each depth feature of the RF input in the proposed method cannot point out the specific physical meaning, they are another manifestation of these parameters with specific physical significance.

2.2 Data Collection and Experiment Development

2.2.1 Study Location: For this study, SURFRAD (Surface Radiation Budget Network) is selected, with a total of 7 stations, to obtain accurate surface erythemal UV-B observation data. The site information is shown in Table 1.

The primary objective of the National Oceanic and Atmospheric Administration's (NOAA's) Surface Radiation budget network (SURFRAD) is to support climate research with accurate, continuous, long-term measurements of the surface radiation budget over the United States. Erythemal UV-B is one of the auxiliary observations. The SURFRAD network was established in 1993 and consisted of seven stations operating in diverse climates zones. The geographic locations of the seven stations cover a variety of climatic environments in the United States, such as temperate climate (BON, SXF), moist subtropical environment (GWN, PSU), cool and dry northern plains (FPK), and semi-arid area (TBL)The types of landcover at these stations range from agricultural land (BON, SXF, PSU), pasture (GWN, PSU), uncultivated land (TBL), to the Gobi (DRA) which have been made of rocks and desert shrubs. The elevations of these 7 SURFRAD stations range from 98 meters below sea level to 1,689 meters above sea level.

Table 1. The geographic information of SURFRAD sites used in this paper.

Station ID	Location	Latitude (°N)	Longitude (°W)	Elevation (m)
BON	Bondville, Illinois	40.05	88.37	213
DRA	Desert Rock, Nevada	36.62	116.02	1007
FPK	Fort Peck, Montana	48.31	105.10	634
GWN	Goodwin Creek, Mississippi	34.25	89.87	98
PSU	Penn. State Univ.,Pennsylvania	40.72	77.93	376
SXF	Sioux Falls, South Dakota	43.73	96.62	473
TBL	Table Mountain, Boulder, Colorado	40.12	105.24	1689

2.2.2 Dataset: We establish a new erythemal UV-B retrieval dataset $\mathrm{D}: \left\{\left(p_1^1,\ldots,p_K^1,m^1\right),\ldots,\left(p_1^N,\ldots,p_K^N,m^N\right)\right\}$, where each element in D has two components, the measured erythemal UV-B irradiance m and the corresponding parameters $\mathbf{p}=\{p_1,\ldots,p_K\}$. In the dataset, we collect the ground measurements of erythemal UV-B for 7 stations in the United States from 2007 to 2017 as the model's actual value (m in the dataset D). The parameters mainly include TOA reflectance (Band1-36) provided by MOD021KM/MYD021KM, SZA and VZA of MOD03/MYD03 products, surface reflectance data of MODIS MCD43A4 (Band1-7), total ozone column data from OMI OMTO3d products, and the altitude of the station. Note that all samples in this dataset did not filter out the sunny or cloudy data,

and therefore the dataset we collect contains all-sky information.

2.2.3 Evaluation Strategy: We divide the dataset into three parts: the training set, the validation set, and the test set. The training and validation set is mainly used for model construction and selection to ensured that the data of the test set is never used in this process. To ensure that the machine learning models (SVR and RF) for comparison have the best performance, choosing a simple but effective combination of parameters is essential. After selecting the optimal set of parameters, all models in the experiment are trained on the SURFRAD data train set and draw compared with on the SURFRAD test set.

2.3 The Forecasting Accuracy Evaluation Metrics

The models are evaluated under several statistical metrics, including MBE, nMBE, RMSE nRMSE, and R^2. The mathematical definitions are provided as follows.

Coefficient of determination (R^2):

$$R^2 = \frac{\sum_{i=1}^{N}\left(E_i - M_{ave}\right)^2}{\sum_{i=1}^{N}\left(M_i - M_{ave}\right)^2} = 1 - \frac{\sum_{i=1}^{N}\left(M_i - E_i\right)^2}{\sum_{i=1}^{N}\left(M_i - M_{ave}\right)^2}。 \quad (2),$$

Mean bias error (MBE) and normalized mean bias error (nMBE):

$$MBE = \frac{\sum_{i=1}^{N}\left(E_i - M_i\right)}{N}$$
$$nMBE(\%) = \frac{100}{M_{ave}}\left(\frac{\sum_{i=1}^{N}\left(E_i - M_i\right)}{N}\right)。 \quad (3)$$

The root-mean-square error (RMSE) and the normalized root-mean-square error (nRMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(E_i - M_i\right)^2}{N}}$$
$$nRMSE(\%) = \frac{100}{M_{ave}}\sqrt{\frac{\sum_{i=1}^{N}\left(E_i - M_i\right)^2}{N}}, \quad (4)$$

where E_i is the estimated erythemal UV-B irradiance (i represents a number), M_i is the ground measurements, M_{ave} is the average of the ground measurements, E_{ave} is the average of the estimated measurements and N is the number of all samples in the test set.

3. EXPERIMENTAL RESULTS

3.1 Parameter sensitivity analysis

To determine the best model input, different parameter combinations are considered in the RF, SVR, and FCNN models. The initial parameters include TOA reflectance, SZA, VZA, and view azimuth angle (VAA), solar azimuth angles (SAA), altitude, latitude, ozone, and surface reflectance. The composition of these parameters and the corresponding results of the models are shown in Table 2.

From Table 2, we can find that the model with inputs composed of TOA reflectance + SZA + VZA + ozone + altitude + surface reflectance performs the best. When SZA, VZA, and altitude already exist in the parameter combination, the accuracy of the model has not been significantly improved if latitude, SAA, and VAA are included in the

combination. The reason could be that parameters SZA, VZA, RAA, SAA, VAA, altitude, and latitude overlap each other, which introduce more trainable parameters in the network but provide less extra information, making the network prone to over-fitting. Combined the prior knowledge of the factors affecting erythemal UV-B introduced in section 1, we finally choose SZA, VZA, and altitude as the input parameters. The ozone parameters can provide information about ozone concentration in the atmosphere to help supplement the UV band information that MODIS lacks. The surface reflectance parameter can distinguish different surface types and is essential for erythemal UV-B retrieval when there is snow cover at high latitudes. The parameter combination of TOA reflectance + SZA + VZA + ozone + altitude + surface reflectance will be the default setting for all subsequent experiments regarding the erythemal UV-B retrieval unless specified otherwise.

Table 2. Performance evaluation of the RF model under different combinations of parameters. TOA refers to the spectral TOA reflectance in MODIS bands.

Model input	R^2	nMBE (%)	nRMSE (%)
TOA	0.8595	-1.50 %	26.17 %
TOA + SZA + VZA	0.8890	-1.09 %	23.50 %
TOA + SZA + VZA + altitude	0.9004	-0.92 %	21.68 %
TOA + SZA + VZA + altitude + VAA + SAA	0.8933	0.80 %	22.71%
TOA + SZA + VZA + altitude + latitude	0.8934	-0.92 %	22.79 %
TOA + SZA + VZA + altitude + ozone	0.9078	-0.80 %	20.71 %
TOA + SZA + VZA + altitude + ozone + surface reflectance	0.9107	0.37 %	19.96 %

3.2 Methods comparison

After training model on the SURFRAD training set, the performance of the four ML models (SVR, RF, FCNN, and FCNN+RF) are evaluated on the SURFRAD test set and Table 3 summarizes their results. The data marked in red is the best result under each statistical metric. Figure 1 shows the scatterplot between ground measurements and estimated erythemal UV-B on the test set of the four models. In order to show the performance of the four models on each site more intuitively, we drew a line graph of the RMSE changes of the seven SURFRAD stations (Figure 2).

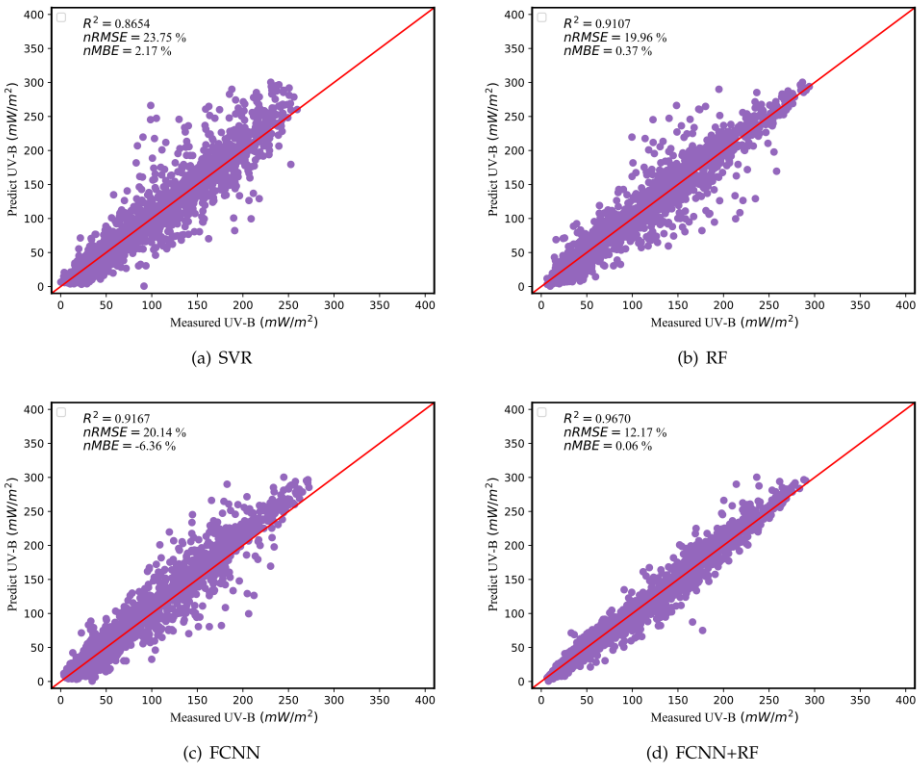

Figure 1. Comparison between measured erythemal UV-B and estimated erythemal UV-B on SURFRAD test data. (a) Estimated erythemal UV-B data by the SVR model. (b) Estimated erythemal UV-B data by RF model. (d) Estimated erythemal UV-B data by the FCNN model. (d) Estimated erythemal UV-B data by the FCNN+RF model.

Table 3. Performance evaluation of the SVR, RF, and FCNN +RF models by R^2 and nRMSE.

All station					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.8654	2.14	2.17 %	23.43	23.75 %
RF	0.9107	0.37	0.37 %	20.04	19.96 %
FCNN	0.9167	-6.41	-6.36 %	20.30	20.14%
FCNN+RF	0.9670	0.06	0.06 %	12.26	12.17 %
BON					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.8706	2.28	2.63 %	21.28	24.47 %
RF	0.8706	-1.10	-1.25 %	20.34	23.10 %
FCNN	0.9034	-3.08	-3.54 %	18.38	21.14 %
FCNN+RF	0.9640	-0.66	-0.76 %	11.19	12.77 %
DRA					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.9102	-6.21	-4.38 %	24.15	17.05 %
RF	0.9557	0.94	0.67 %	16.56	11.77 %
FCNN	0.9441	-9.77	-6.90 %	19.05	13.45 %
FCNN+RF	0.9771	1.49	1.06 %	11.68	8.33 %
FPK					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.9094	1.34	1.61 %	17.47	20.98 %
RF	0.9188	0.12	0.14 %	15.72	18.90 %
FCNN	0.9312	-3.43	-4.12 %	15.23	18.29 %
FCNN+RF	0.9572	-2.55	-2.97 %	11.83	13.78 %
GWN					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.8359	-5.37	-5.23 %	26.97	26.25 %
RF	0.8566	1.43	1.41 %	23.02	22.41%
FCNN	0.8769	-9.55	-9.30 %	23.35	22.73 %
FCNN+RF	0.9475	4.78	4.88 %	13.98	14.28 %
PSU					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.8691	2.58	3.03 %	21.41	25.07 %
RF	0.8596	-2.58	-2.93 %	20.32	23.10 %
FCNN	0.8804	-4.37	-5.12 %	20.47	23.97 %
FCNN+RF	0.9561	-1.73	-1.98 %	12.00	13.77 %
SXF					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.9135	-0.16	-0.21 %	17.29	22.69 %
RF	0.9245	-0.81	-1.05 %	16.05	20.84 %
FCNN	0.9170	-6.14	-8.06 %	16.94	22.23 %
FCNN+RF	0.9626	-0.87	-1.12 %	10.95	14.21 %
TBL					
Model	R^2	MBE (mW/m^2)	nMBE (%)	RMSE (mW/m^2)	nRMSE (%)
SVR	0.8101	-9.01	-7.60 %	32.72	27.60 %
RF	0.8473	4.20	3.67 %	27.64	24.17 %
FCNN	0.8623	-8.72	-7.35 %	27.86	23.50 %
FCNN+RF	0.9593	0.93	0.79 %	14.27	12.13 %

From the synthetic data of seven stations, the optimal model on each site is the FCNN+RF (R^2 = 0.9670, RMSE = 12.17%) evaluation under the three sets of statistical indicators. The performance of RF (R^2 = 0.9107, RMSE = 19.96%) and FCNN (R^2 = 0.9167, RMSE = 20.14%) models are second, while the performance of SVR (R^2 = 0.8654, RMSE = 23.75%) rank the second. It can also be seen from the scatter diagram in Figure 1 and from subgraph 1(a) to 1(d) that the scatter is gradually approaching the ideal position, as is illustrated by the red dotted line. As for subgraph 1(d), almost all points are close to the red line and its vicinity, and only a few points deviate from its ideal position, and the degree of deviation is smaller than that of the previous three subgraphs. This means that FCNN+RF can better capture the relationship between MODIS and UV-B. For RF and FCNN, the RMSE and R^2 of the two are very close. However, from the perspective of the MBE, the retrieval result of FCNN has a negative bias, and the degree of this

bias is higher than the degree of the positive bias of RF. Hence, the performance of the RF is slightly better than the FCNN.

From both Table 3 and Figure 2, we find that all four methods on three stations of DRA, SXF, and FPK perform better than the other four stations. According to the station information that we summarized in section 2.2.1, the DRA station has a hot arid climate, the SXF and FPK stations are in the northern plains. All three stations are in a dry, sunny climate with less extreme weather. Thus, the results on the test set show a high correlation, low MBE, and RMSE. From Figure 2, we can find that compared to SVR, RF, and FCNN, the FCNN+RF model has the smallest change in RMSE at the seven sites. Similarly, the specific data in Table 2 can also show the same conclusion. The difference in RMSE of the FCNN+RF model on the seven sites does not exceed 6%. In contrast, among the SVR, RF, and FCNN, these differences reach 11%, 12%, and 11%, respectively. In other words, the site environment has a more significant impact on the performance of SVR, RF, and FCNN. The FCNN+RF model is most robust towards the impact of the site environment on erythemal UV-B retrieval.

For FCNN + RF, seven stations have R^2 ranging from 0.9814 (TBL) to 0.9925 (DRA), RMSE varying from 4.92% (DRA) to 9.36% (PSU), and MBE staying below ±1%. The accuracy of the estimated erythemal UV-B values obtained by the model at each station of SURFRAD is at the same level. From Figure 1 and Figure 2, we can find that FCNN+RF has the best performance on erythemal UV-B retrieval from MODIS, both in terms of accuracy and robustness. Two reasons could explain such a result. First, the random forest can obtain more reliable regression results, and it is more robust to overfitting due to its multi-decision tree ensembling characteristics and the use of the Bagging algorithm. The second point is that we need to use the cloud and aerosol information contained in MODIS TOA together to help retrieve erythemal UV-B. RF and SVR can fit the relationship between the parameters very well, but they are not good at capturing the complex relationship between the input parameters and the UV-B level, which, on the contrary, is the strength of FCNNs. Therefore, FCNN helps capture implicit aerosol and cloud information from MODIS TOA information and other additional parameters to form new expressions, and RF is used to fit the relationship between these expressions and the surface UV-B.

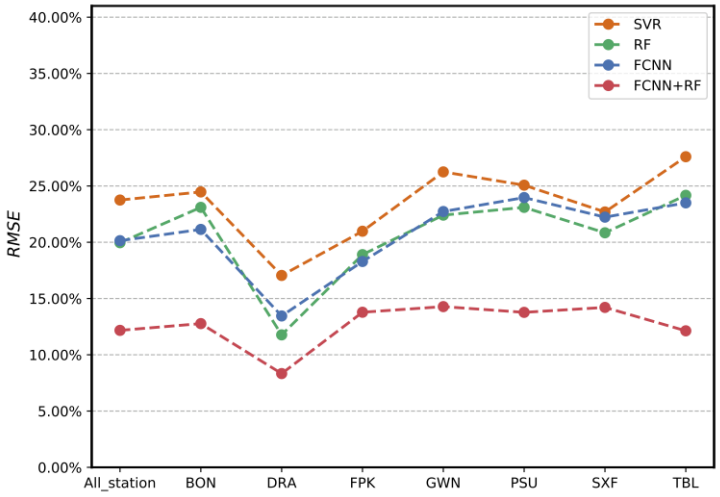

Figure 2. A line chart compares the RMSE generated by SVR, RF, FCNN, and FCNN+RF models.

4. CONCLUSIONS

This article attempts to retrieve all-sky erythemal UV-B from 1 km MODIS TOA data without the ultraviolet band. A dataset is established based on the SURFRAD observation stations and contains seven sets of parameters obtained from MODIS products such as MOD02IKM / MYD021KM, MOD03/ MYD03, MCD43A4, and OMI's ozone products. On this data set, we first use the three most common machine learning algorithms, SVR, RF, and FCNN, to obtain surface erythema UV-B. A framework for combining FCNN and RF is then constructed to model the relationship between MODIS and UV-B on the surface erythema.

To evaluate the model performance, we conduct a sensitivity analysis to determine the optimal combination of parameters. The "TOA reflectance + SZA + VZA + latitude + ozone + surface reflectance" is selected from the nine groups of candidate parameters as the default setting to all subsequent experiments. Furthermore, these four machine learning methods are compared on the dataset. Later, a comparative analysis of the four models on the SURFRAD test set found that the RF (R^2 = 0.9107, RMSE = 19.96%, MBE = 0.37 %) method is superior to the SVR (R^2 = 0.8654,

RMSE = 23.75%, MBE = 2.17 %) and FCNN (R^2 = 0.9167, RMSE = 20.14%, MBE = -6.36 %) methods in fitting the relationship between MODIS and UV-B due to the algorithmic characteristics of its multiple decision trees. Finally, FCNN+RF (R^2 = 0.9670, RMSE = 12.17%, MBE = 0.06%), which combines the representation learning ability of FCNN and classification ability of RF classifier, performs best in retrieving erythema UVB from MODIS TOA. It is mainly due to the combination of FCNN's advantages in capturing MODIS implicit cloud and aerosol information and RF's ability to fit the relationship between parameters well.

In the future, we will further explore the possibility of machine learning algorithms to obtain surface erythema UV-B based on MODIS and continue to expand the dataset to verify its effectiveness on more radiation observation networks.

5. REFERENCES

Antonanzas, J, Osorio N, Escobar R, et al., 2016. Review of photovoltaic power forecasting. Solar Energy, 136(Oct.), pp. 78-111.

Bernard J J, Gallo R L, Krutmann J, et al., 2019. Photoimmunology: how ultraviolet radiation affects the immune system. Nature reviews. Immunology, 19(11), pp. 688-701.

Bhartia P K, Herman J, Mcpeters R D, et al., 2007. Total ozone from backscattered ultraviolet measurements. In Observing Systems for Atmospheric Composition; Springer, pp. 48–63.

Calbó J, Pages D, González J A., 2005. Empirical studies of cloud effects on UV radiation: A review. Reviews of Geophysics, pp. 43.

Chollet, F., 2018. Deep learning with Python.

Driemel, A, Augustine J, Behrens K, et al., 2018. Baseline Surface Radiation Network (BSRN): structure and data description (1992-2017). Earth System Science Data, 10(3), pp. 1491-1501.

Fountoulakis I, Bais A F, Fragkos K, et al., 2016. Short-and long-term variability of spectral solar UV irradiance at Thessaloniki, Greece: effects of changes in aerosols, total ozone and clouds. Atmospheric Chemistry & Physics, 16(4), pp. 2493-2505.

He K, Zhang X, Ren S, Sun J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Iqbal M., 2012. An introduction to solar radiation. Elsevier.

Kim J, Cho H K, Mok J, et al., 2013. Effects of ozone and aerosol on surface UV radiation variability. Journal of [9]. Photochemistry and Photobiology B: Biology, 119, pp. 46–51.

LeCun Y, Bengio Y, Hinton G., 2015. Deep learning. nature, 521, pp. 436–444.

Levelt P F, Van D, Dobber M R, et al., 2006. The ozone monitoring instrument. IEEE Transactions on Geoscience & Remote Sensing, 44(5), pp. 1093-1101.

Liang S, Zheng T, Liu R, et al., 2006. Estimation of incident photosynthetically active radiation from Moderate Resolution Imaging Spectrometer data. Journal of Geophysical Research. Atmospheres, 111.

McKinlay A, Diffey B., 1987. A reference spectrum for ultraviolet induced erythema in human skin.

Renaud A, Staehelin J, Fröhlich C, et al., 2000. Influence of snow and clouds on erythemal UV radiation: Analysis of Swiss measurements and comparison with models. Journal of Geophysical Research: Atmospheres, 105(D4), pp. 4961–4969.

Ronneberger O, Fischer P, Brox T., 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241.

Singh S, Lodhi N K, Mishra A K, et al., 2018. Assessment of satellite-retrieved surface UVA and UVB radiation by comparison with ground-measurements and trends over Mega-city Delhi. Atmospheric Environment, 188, pp. 60–70.

Sola Y, Lorente J, Campmany E, et al., 2008. Altitude effect in UV radiation during the Evaluation of the Effects of Elevation and Aerosols on the Ultraviolet Radiation 2002 (VELETA-2002) field campaign. Journal of Geophysical Research: Atmospheres, 113.

Utrillas M, Marín M, Esteve A, et al., 2018. Relationship between erythemal UV and broadband solar irradiation at high altitude in Northwestern Argentina. Energy, 162(NOV.1), pp. 136-147.

Wang D, Liang S, Zhang Y, et al., 2020. A New Set of MODIS Land Products (MCD18): Downward Shortwave Radiation and Photosynthetically Active Radiation. Remote Sensing, 12(1), pp. 168.

Xiang Y, Laurent B, Hsu C H, et al., 2017. RNA m 6 A methylation regulates the ultraviolet-induced DNA damage response. Nature, 543, pp. 573–576.

Zempila M M, Fountoulakis I, Taylor M, et al., 2018. Validation of OMI erythemal doses with multi-sensor ground-based measurements in Thessaloniki, Greece. Atmospheric environment, 183(jun.), pp. 106-121.

Zempila M M, van Geffen J H, Taylor M, et al., 2017. TEMIS UV product validation using NILU-UV ground-based measurements in Thessaloniki, Greece. Atmospheric Chemistry and Physics, 17(11), pp. 7157-7174.

Zhang Y, He T, Liang S, et al., 2018. Estimation of all-sky instantaneous surface incident shortwave radiation from Moderate Resolution Imaging Spectroradiometer data using optimization method. Remote Sensing of Environment an Interdisciplinary Journal.