RETRIEVAL OF REAL-TIME PM$_{2.5}$, TEMPERATURE AND HUMIDITY PROFILES FROM SATELLITE AND GROUND-BASED REMOTE SENSING DATA USING ADVANCED DEEP LEARNING MODELS

Xing Yan (1), Zhou Zang (1), Nana Luo (1) (2), Dan Li (1), Yushan Guo (1)

[1] State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China
[2] Department of Geography, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182-4493, USA
Email: yanxing@bnu.edu.cn; 201921490029@mail.bnu.edu.cn; nluo@sdsu.edu; 202021490031@mail.bnu.edu.cn; 202021490028@mail.bnu.edu.cn

KEY WORDS: Deep learning; Himawari-8; microwave radiometer

ABSTRACT: Monitoring the real-time atmospheric PM$_{2.5}$, temperature and humidity profiles is highly valuable for human health and climate research. To achieve this goal, we applied two new advanced deep learning models. One is EntityDenseNet to estimate real-time ground-level PM$_{2.5}$ from Himawari-8 satellite data, and the other is called batch normalization and robust neural network (BRNN) to retrieve temperature and humidity profiles using data from a ground-based microwave radiometer (MWR). Many features and technologies have been introduced in these two models, in particular: (1) a dropout layer for each hidden layer in the deep learning model has been introduced to reduce the overfitting problem; (2) the problems of saturation and vanishing gradients are overcome by using the ReLU as the activation function; and (3) the data between the inputs in model are normalized by batch normalization technology, which fixes the mean and variance of the inputs to accelerate the training process. A detailed comparison with various traditional machine learning models (backpropagation neural network, extreme gradient boosting, light gradient boosting machine, and random forest) has been conducted in this research, using the same training and test data sets. From the comparison, the new models reduce overfitting and has a greater capacity to describe nonlinear relationships. In addition, the EntityDenseNet can “peek inside the black box” to extract the spatio-temporal features of data. The EntityDenseNet is able to map variables that are close to each other to an embedding space. It enables us to calculate the distance between different variables in this space. The smaller the distance, the higher the correlation between two features could be found. This ability greatly improves interpretability of the deep learning model inversion result. This work reveals that these two advanced deep learning models significantly improves retrieval accuracy, and demonstrates strong potential in the application of EntityDenseNet and BRNN for additional earth-observation datasets and scenarios. We have created an EntityDenseNet Cloud Platform (http://49.233.1.40:8888/), it is free to access and researchers can use it for their own data modelling.

1. INTRODUCTION

Monitoring the real-time atmospheric PM$_{2.5}$, temperature and humidity profiles is highly valuable for understanding earth environmental changing. In recent years, machine learning methods have increased in popularity as a method for estimating PM$_{2.5}$, temperature and humidity profiles using remote sensing data, such as satellite data and microwave radiometer (MWR).

Machine learning methods can offer the best performance for the solution of nonlinear relationships in the model. Even now, the three-layer backpropagation neural network (BPNN) is a very popular method for satellite and MWR data, to retrieve PM$_{2.5}$ and atmospheric vertical profiles (Che et al., 2019; Mao et al., 2017). However, these BPNN models have only a single hidden layer and their capacity to model highly varying functions defining nonlinear structures is much less than using multiple hidden layers (deep learning model). In addition, the BPNN approach cannot directly use categorical data (Guo and Berkhahn, 2016). The neural network requires all input variables and output variables to be numeric (Yan et al., 2015), thus, the One-Hot encoding is typically used for converting the categorical variable which is then input into neural network training and prediction (Chren, 1998; Liu et al., 2002). Therefore, as deep learning is usually a neural network-based model, one of the key issues to overcome was determining how to handle and learn the information using categorical variables. Furthermore, interpreting the prediction from deep learning neural networks also remains challenging. Although Reichstein et al. (2019) indicated that deep machine learning models may act as a promising tool to extract spatial–temporal features from the data, the processes required to open and interpret this "black box" model are difficult.

In this study, we applied two advanced deep learning models to retrieve real-time $PM_{2.5}$ and atmospheric vertical profiles from Himawari-8 satellite data and MWR data. In contrast to the traditional machine learning methods, these advanced deep learning models reduce overfitting and has a greater capacity to describe nonlinear relationships for these remote sensing data.

2. DATA AND METHODS

2.1 MWR and radiosonde data

In this article, the data were measured by an MWR located in the Beijing Nanjiao Meteorological Observatory, China. The MWR used in this research was the Humidity And Temperature PROfiler (HATPRO; Radiometer Physics GmbH, Germany). The radiosonde data were measured by an L-band GTS1 digital radiosonde at the same location; the radiosonde was launched twice a day, at 11:15 and 23:15 UTC during the research period. The collected MWR and radiosonde data from 2017 to 2018 were used in the training process and the data from January to May 2019 were applied for validation.

2.2 Himawari-8 satellite data and ground-based $PM_{2.5}$

The Himarwari-8 reflectance (Bands 1 to 6) and brightness temperature (Bands 7 to 16) data of a spatial resolution of 5 km were extracted from the L1 Gridded Data at 10 min intervals during daytime (UTC 1:00–6:00) from January 2016 to June 2019. Hourly $PM_{2.5}$ concentrations were collected for the same period from 1434 monitoring stations across mainland China.

2.3 Batch normalization and robust neural network (BRNN)

In this research, we applied BRNN (Yan et al., 2020a) to retrieve temperature and humidity profiles using data from a ground-based MWR. The BRNN consists of four layers (see Figure 1): one input layer, two hidden layers, and one output layer. The input layer receives the collected data including the brightness temperature data from the RPG-HATPRO's 14 channels and three features are the surface pressure, temperature, and RH measured by the RPG-HATPRO. In BRNN, each of the two hidden layers includes one fully connected layer, one rectified linear unit (ReLU) layer, one BN layer, and one dropout layer. In the fully connected layer, the number of neurons is 256. In the output layer, the input data will first be processed by a fully connected layer and a sigmoid layer. We introduce the sigmoid layer to scale the output to a reasonable range in both training and prediction processes. For example, the normal range of the RH is 0%–100%. The collected MWR and radiosonde data from 2017 to 2018 were used in the training process and the data from January to May 2019 were applied for validation.

2.4 EntityDenseNet

EntityDenseNet (Yan et al., 2020b) was used to estimate real-time ground-level $PM_{2.5}$ from Himawari-8 satellite data. The schematic diagram of the EntityDenseNet is shown in Figure 1, which is the same as BRNN when the input variables are only continuous type. However, the EntityDenseNet can directly process categorical variables and continuous variables. To process the categorical variables, we used the Entity Embeddings method (Guo and Berkhahn, 2016). In Figure 2, we use the categorical variable “January” as an example to show how the embedding layer works in this EntityDenseNet model. The input data is first separated into two data types: categorical variables and continuous variables. The categorical variables include year, month, date, hour, China administrative divisions and day type (weekday or holiday). The continuous variables consist of TOA reflectance from Himawari-8 bands 1 to 6, brightness temperature data from Himawari-8 bands 7 to 16, satellite zenith angle (SEZ), solar zenith angle (SOZ), satellite azimuthal angle (SEA), solar azimuthal angle (SOA), relative azimuth angle, scattering angle, longitude, latitude, light density, Digital Elevation Model (DEM), and Normalized Vegetation Index (NDVI). The main purpose of EntityDenseNet is to establish the nonlinear relationship between satellite spectral measurements and $PM_{2.5}$ concentration. The integrated training data collected from ground-based $PM_{2.5}$ and satellite data are for the period from 2016 to 2018, which contains 4,490,474 samples. The collected data from 2019 (437,351 samples) was used to test the trained network system.

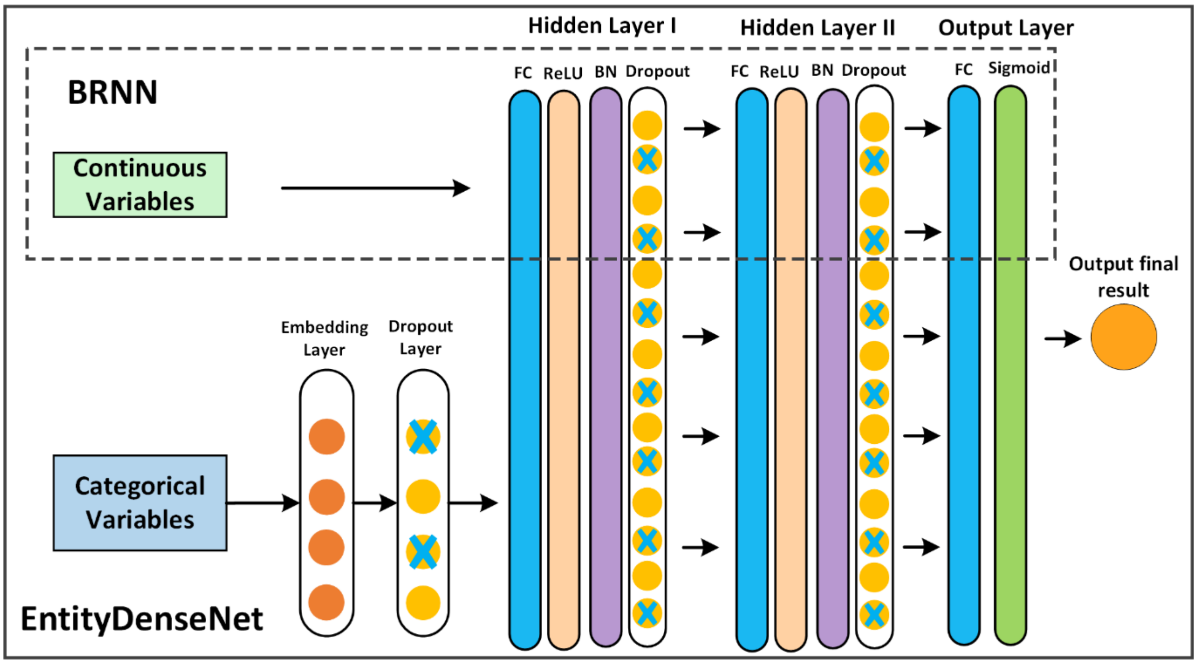

Figure 1. Schematic diagram of the BRNN and EntityDenseNet

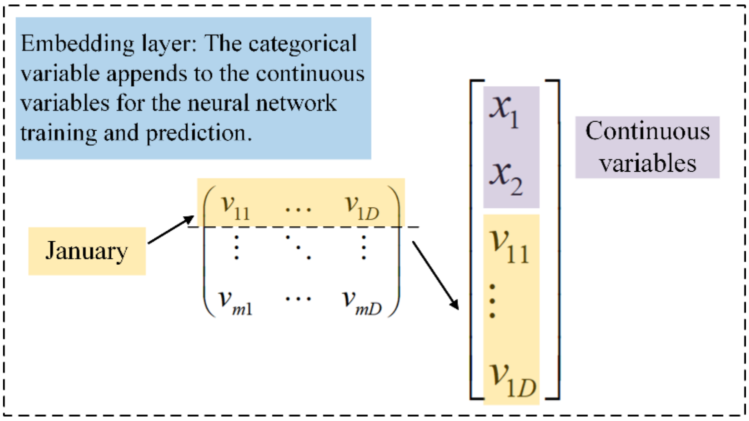
Embedding layer: The categorical variable appends to the continuous variables for the neural network training and prediction.
January
v_{11} … v_{1D}
v_{m1} … v_{mD}
x_1
x_2
v_{11}
v_{1D}
Continuous variables

Figure 2. The diagram of the categorical variable "January" appended to the continuous variables for the EntityDenseNet training and prediction.

3. RESULTS

3.1 Validation of temperature and humidity profiles with Radiosonde

Figure 3(a)–(c) shows the BRNN temperature, water vapor density (WVD), and RH as a function of the radiosonde measurements from all 47 atmospheric vertical layers up to 10 km. High kernel density values with red color show where most of the data lie. As shown in Figure 3(a), the linear regression relation between the BRNN temperature and radiosonde temperature has a slope of 0.98 and a y-intercept of 6.5, with a coefficient of determination (R^2) of 0.99 and a root-mean-square error (RMSE) of 1.70. For WVD [see Figure 3(b)], the R^2 is 0.90 and the slope is 0.9 with an RMSE of 0.26. From Figure 3(a) and (b), we observe that the temperature and WVD from BRNN agree well with the radiosonde measurements. In contrast to the temperature and WVD validation, the result of RH is more scattered. Figure 3(c) shows that the R^2 is 0.44 and the RMSE is 11.72.

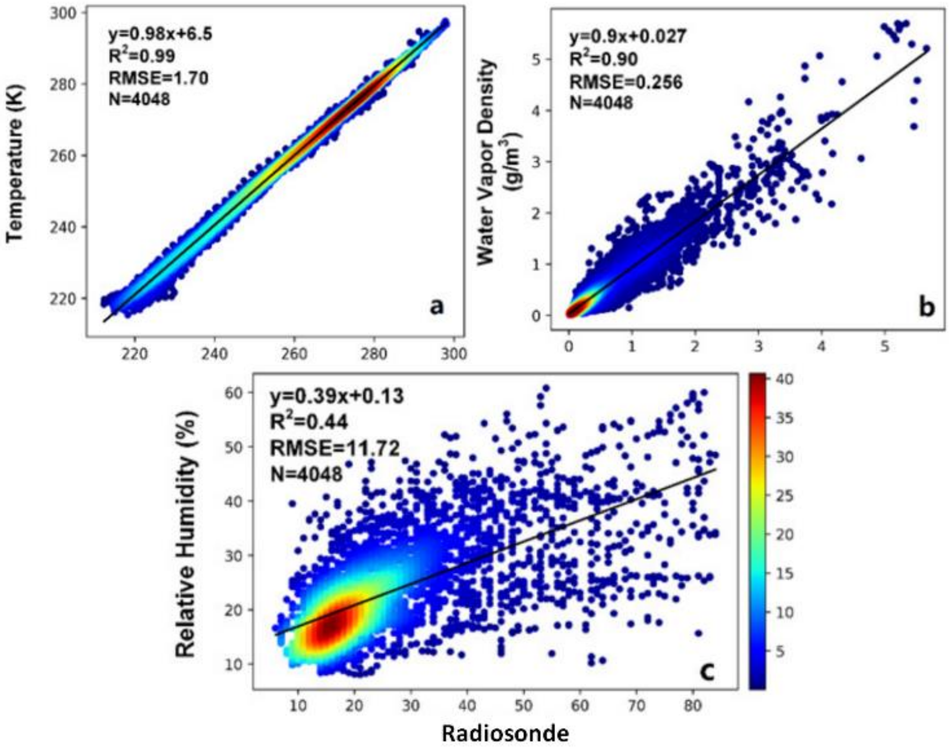

Figure 3. (a)–(c) BRNN retrievals for temperature, WVD, and RH, respectively, as a function of radiosonde data

3.2 Comparison of BRNN with Other Retrieval Techniques

We plot the RMSE of the temperature, RH, and WVD profiles resulting from BRNN, XGBoost, BPNN, and SVM at different heights, using the radiosonde observation as the reference. For the temperature, using the BRNN method, the RMSE below 2 km is around 1 K and remains less than 2 K below 8 km; this is apparently less than the RMSE of XGBoost, BPNN, and SVM. The XGBoost method has a better accuracy than SVM and BPNN below 2 km, but has a large RMSE in the upper atmosphere (above 3.5 km). For RH, the RMSE of BRNN can be controlled at about 5% below 1 km. The XGBoost method also performs well in the lower atmosphere, but its RMSE is significantly larger than the others from 3 to 7 km. In terms of WVD, the situation is exactly the opposite: the RMSE of the lower layers is larger and that of the higher layers is smaller. The superiority of the BRNN method is that its RMSE value does not exceed 0.4 g/m3 from the surface to 10 km; in particular, its RMSE is about 0.35 g/m3 from the surface up to 3.5 km and then decreases to below 0.1 g/m3 at 8 km. Among the remaining three methods, the performance shows little variation, except for the SVM method, which has the largest error in the layer near the ground.

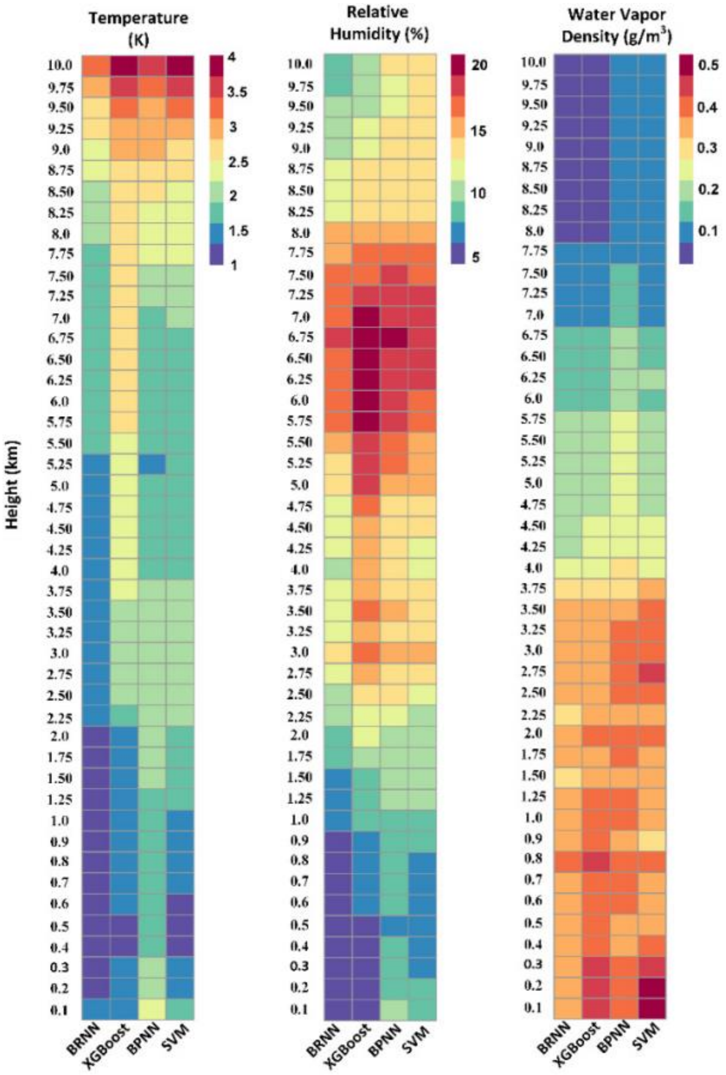

Figure 4. Profile retrieval RMSEs for (Left) temperature, (Center) RH, and (Right) WVD, with respect to radiosonde, for BRNN, XGBoost, BPNN, and SVM techniques

3.3 EntityDenseNet performance

Figure 4 consist of scatterplots depicting the relationship between measured and estimated $PM_{2.5}$ for both training and test data sets. It should be noted that extremely high values were not eliminated from all the datasets to enable us to test the performance of the EntityDenseNet model in the presence of abnormal values. In our validation of the hourly test data, the linear regression between the EntityDenseNet $PM_{2.5}$ and ground-based $PM_{2.5}$ resulted in a slope of 0.56, a y-intercept of 19.45, a coefficient of determination (R^2) of 0.63, and an RMSE of 26.85 μg/m^3. As shown in Figure 4, a small change in R^2/RMSE between training (R^2=0.68, RMSE=24.05 μg/m^3) and hourly test data sets indicates a slight over-fitting in the EntityDenseNet model.

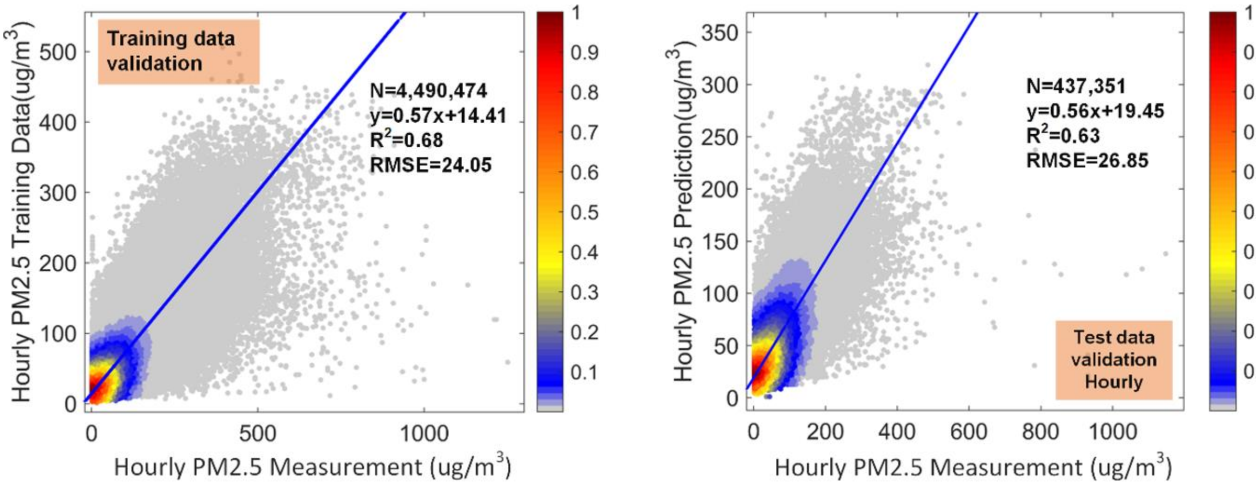

Figure 5. Scatterplot showing the performance of EntityDenseNet on trained (2016–2018) and test (2019) data sets

Figure 12 presents three cases of the application of EntityDenseNet, including true color maps (left), real local time (13:00) (middle), and daily averaged (right) $PM_{2.5}$ distributions in mainland China on April 15–17, 2019. The true color maps show the satellite image in this period including the real cloud covering over China. In cloud-free regions, the $PM_{2.5}$ spatial distributions as well as their temporal variations are well-retrieved. Northern China also showed high concentrations of $PM_{2.5}$ during the same period. Another hotspot of high $PM_{2.5}$ concentrations is Yunnan, where $PM_{2.5}$ concentrations can reach up to 45~55 μg/m^3. Figure 12 confirms that EntityDenseNet can retrieve $PM_{2.5}$ data for a specific local time or daily temporal resolution covering a large spatial scale.

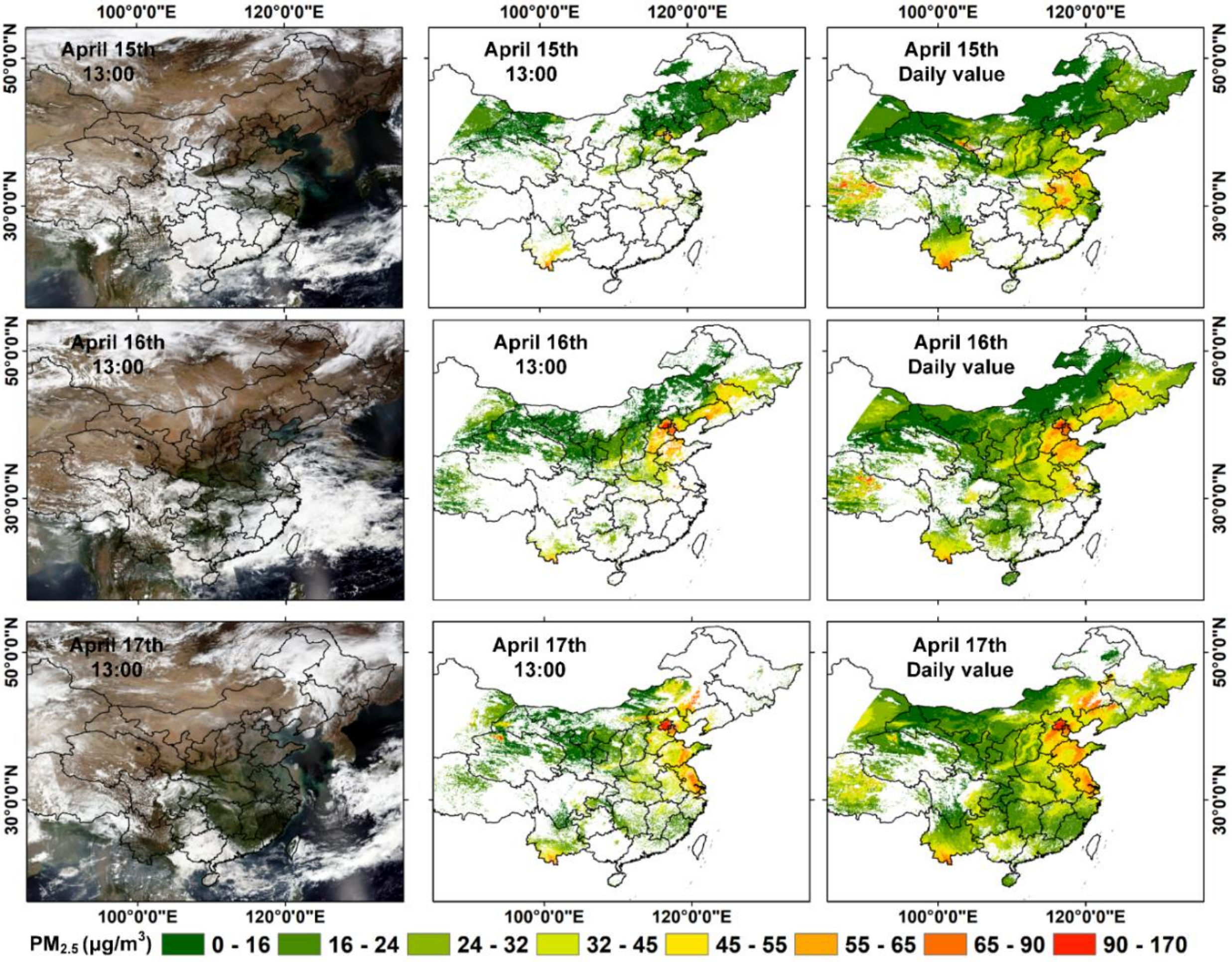

Figure 6. EntityDenseNet $PM_{2.5}$ over mainland China on April 15 –17, 2019. The left column is the true color satellite image. The middle column is the $PM_{2.5}$ concentrations at 13:00 (local time). The right column is the daily averaged $PM_{2.5}$ concentrations.

3.4 Spatial features extraction by EntityDenseNet

The spatial characteristics of $PM_{2.5}$ between different provinces of China extracted by EntityDenseNet is displayed in Figure 7. The province features matrix from the trained EntityDenseNet embedding layer is mapped to 3D (Figure 7a). The Cosine Distance between different provinces was calculated based upon Figure 7a. In the Beijing-Tianjin-Hebei region, Tianjin had a closer Cosine Distance with Hebei (0.26) than with Beijing (0.45). This result illustrates that the $PM_{2.5}$ in Tianjin was influenced more by the $PM_{2.5}$ from Hebei than Beijing (Figure 7a,b). This result is consistent with the WRF-Chem modeling outcome which demonstrated that $PM_{2.5}$ from Hebei had a greater contribution to Tianjin's $PM_{2.5}$ than Beijing's (Meng et al., 2020). Xue et al. (2014) also indicated that the contribution of Hebei to the $PM_{2.5}$ of Tianjin is 26%, while its contribution to Beijing is 24%. In Jiangxi, China, a closer correlation with the $PM_{2.5}$ from Hunan (Cosine Distance=0.51) than the Hubei's $PM_{2.5}$ (Cosine Distance=0.54) or other neighboring provinces (Cosine distance >0.6) was identified. According to source apportionment of the $PM_{2.5}$ in Jiangxi, other provinces contributed approximately 48% of $PM_{2.5}$ annually (Xue et al., 2014). In the present study, EntityDenseNet further indicated that the Jiangxi $PM_{2.5}$ was most closely associated with Hunan. Using Figure 7, we can then extract the spatial information from the training data. Overall, EntityDenseNet improves the understanding of the impact of $PM_{2.5}$ pollution on the provincial scale.

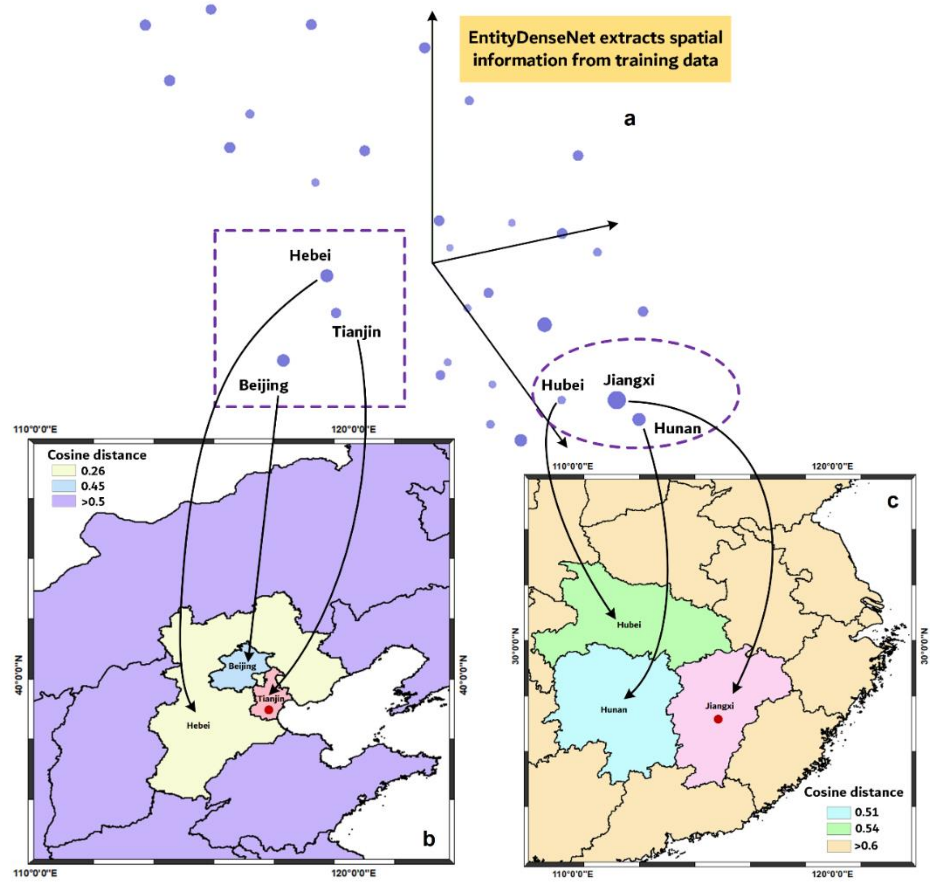

Figure 7. Spatial analysis from EntityDenseNet. a) The province features matrix from the trained EntityDenseNet embedding layer is mapped to 3D. b) The Cosine Distance to Tianjin, China. c) The Cosine Distance to Jiangxi, China.

4. DISCUSSION AND CONCLUSION

BRNN and EntityDenseNet, these two advanced deep learning models have introduced many features and technologies that improve the capacity to describe the nonlinear relationship for remote sensing data. In particular: (1) a dropout layer (Srivastava et al., 2014) for each hidden layer in the EntityDenseNet has been introduced to reduce the overfitting problem; (2) the problems of saturation and vanishing gradients are overcome by using the ReLU as the activation function (Nair and Hinton, 2010); and (3) the data between the inputs are normalized by BN technology (Ioffe and Szegedy, 2015), which fixes the mean and variance of the inputs to accelerate the training process.

In the application, BRNN showed a good retrieval capability with an RMSE of 1.70 K for temperature, 11.72% for RH, and 0.256 g/m3 for WVD. And the results obtained by EntityDenseNet reflected a good retrieval capability at hourly time scales with RMSE values of 26.85 $\mu g/m^3$. The spatial features of $PM_{2.5}$ interpreted by EntityDenseNet demonstrates that in the Beijing-TianjinHebei area, the $PM_{2.5}$ in Tianjin is more subject to impacts from Hebei than Beijing, which is consistent with previous studies (Meng et al., 2020; Xue et al., 2014). This relationship was further explored using the deep learning approach in the present study. This work reveals that these two advanced deep learning models significantly improves retrieval accuracy, and demonstrates strong potential in the application of EntityDenseNet and BRNN for additional earth-observation datasets and scenarios. We have created an EntityDenseNet Cloud Platform (http://49.233.1.40:8888/), it is free to access and researchers can use it for their own data modelling.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 41801329, in part by the National Key Research and Development Plan of China (2017YFC1501702), the Open Fund of State Key Laboratory of Remote Sensing Science (OFSLRSS201915) and the Fundamental Research Funds for the Central Universities.

References

Che, Y., Ma, S., Xing, F., Li, S., & Dai, Y. 2019. An improvement of the retrieval of temperature and relative humidity profiles from a combination of active and passive remote sensing. Meteorology and Atmospheric Physics, 131(3), 681-695.

Chren, W.A., 1998. One-hot residue coding for low delay-power product CMOS design. IEEE T CIRCUITS-II 45 (3), 303–313.

Guo, C., Berkhahn, F., 2016. Entity embeddings of categorical variables. arXiv preprint arXiv:1604.06737.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv:1502.03167.

Liu, H., Hussain, F., Tan, C.L., Dash, M., 2002. Discretization: An Enabling Technique. Data Min. Knowl. Discov. 6, 393–423.

Mao, X., Shen, T., & Feng, X. 2017. Prediction of hourly ground-level PM2. 5 concentrations 3 days in advance using neural networks with satellite data in eastern China. Atmospheric Pollution Research, 8(6), 1005-1015.

Meng, L., Cai, Z., Li, Y., et al., 2020. Spatial and temporal distributions and source simulation during heavy pollution of PM2.5 in Tianjin. Res. Environ. Sci. 1, 9–17.

Nair, V., Hinton, G. E., 2010. Rectified linear units improve restricted boltzmann ma- chines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807–814.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Xue, W., Fu, F., Wang, J., Tang, G., Lei, Y., Yang, J., Wang, Y., 2014. Numerical study on the characteristics of regional transport of PM2.5 in China. China Environ. Sci. 6, 1361–1368.

Yan, X., Shi, W., Zhao, W., & Luo, N. 2015. Mapping dustfall distribution in urban areas using remote sensing and ground spectral data. Science of the Total Environment, 506, 604-612.

Yan, X., Liang, C., Jiang, Y., Luo, N., Zang, Z., Li, Z., 2020a. A Deep Learning Approach to Improve the Retrieval of Temperature and Humidity Profiles From a Ground-Based Microwave Radiometer. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/ TGRS.2020.2987896.

Yan, X., Zang, Z., Luo, N., Jiang, Y. and Li, Z., 2020b. New interpretable deep learning model to monitor real-time PM2. 5 concentrations from satellite data. Environment International, 144, p.106060.