Influence of land-sea breeze on long-term PM2.5 prediction in central and southern Taiwan using composite neural network

George William Kibirige
Academia Sinica
Chiao-Cheng Huang (neil1013@iis.sinica.edu.tw)
Academia Sinica
Chao-Lin Liu
National Chengchi University
Meng-Chang Chen
Academia Sinica

Article

Keywords:

Posted Date: September 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1993037/v1

Influence of land-sea breeze on long-term $PM_{2.5}$ prediction in central and southern Taiwan using composite neural network

George William Kibirige[1,2,+], Chiao-Cheng Huang[1,*,+], Chao-Lin Liu[2], and Meng-Chang Chen[1]

[1]Institute of Information Science, Academia Sinica, 128 Academia Road, Sec. 2, Nankang, Taipei, 11529, Taiwan
[2]National Chengchi University, No. 64, Sec. 2, ZhiNan Rd., Wenshan District, Taipei, 11605, Taiwan
[*]neil1013@iis.sinica.edu.tw
[+]these authors contributed equally to this work

ABSTRACT

$PM_{2.5}$ prediction plays an important role for governments in establishing policies to control the emission of excessive atmospheric pollutants to protect the health of citizens. However, traditional machine learning methods that use data collected from ground-level monitoring stations have reached their limit with poor model generalization and insufficient data. We propose a composite neural network trained with aerosol optical depth (AOD) and weather data collected from satellites, as well as interpolated ocean wind features. We investigate the model outputs of different components of the composite neural network, concluding that the proposed composite neural network architecture yields significant improvements in overall performance compared to each component and the ensemble model benchmarks. The monthly analysis also demonstrates the overall superiority of the proposed architecture for the southern and central Taiwan stations in the months when land-sea breeze events frequently occur.

1 Introduction

Particulate matter (PM) is composed of air pollutants emitted into the atmosphere through human activities, urban development, and industrialization. PM with an aerodynamic diameter less than or equal to 2.5 micrometers (µm) ($PM_{2.5}$) has been linked with cerebrovascular, cardiovascular, and pulmonary diseases[1–5]. In the Global Burden of Diseases study, $PM_{2.5}$ was ranked the sixth leading cause of human death[6]. One measure against $PM_{2.5}$ harm is to predict precise $PM_{2.5}$ concentrations; many governments have established ground monitoring stations to record $PM_{2.5}$ concentration to enact policies to control excessive atmospheric pollutants.

Taiwan's Environmental Protection Administration (EPA) has divided Taiwan into seven air quality zones according to geographical and meteorological conditions. Of these air quality zones, the middle and southern air quality zones suffer the most serious air pollution, which is associated with local air pollution emission, unfavorable atmospheric diffusion conditions, and another important reason: land-sea breezes[7,8]. We propose a neural network architecture to improve $PM_{2.5}$ prediction in southern and central Taiwan with wind data merged from monitoring stations on land and sea. The literature shows that the characteristics of weather and air pollution are widely considered and play important roles in $PM_{2.5}$ prediction. However, the low spatial coverage of air pollution monitoring stations presents a challenge that limits the performance of common air quality prediction models trained using known factors[9] (remotely transported $PM_{2.5}$ and other air pollutants).

In addition to $PM_{2.5}$ events caused by local emission, poor atmospheric diffusion conditions, and remote transport, $PM_{2.5}$ concentrations in central and southern Taiwan often reach the national warning threshold due to land-sea breezes[10]. Simulations from the literature have shown that land-sea breeze events occur with a northwest wind onshore formed during the day and east winds offshore at night[8]. However, this land-sea breeze effect is difficult to detect merely by monitoring station data. We further introduce large-coverage wind features which enable our models to detect land-sea breezes by interpolating monitoring data collected from Central Weather Bureau (CWB) stations on land and on sea.

The literature shows that the introduction of machine learning (ML) methods such as feedforward neural networks (FNNs)[11,12], convolutional neural networks (CNNs)[13], convolutional long short-term memory (ConvLSTM)[9,14] and random forests[15–17] improve the performance of $PM_{2.5}$ prediction. Recently, the development of deep neural network (DNN) approaches has overcome the weakness of other ML methods with their ability to capture complex interactions between datasets from different domains. In our case, the introduction of DNN techniques facilitates the learning of spatiotemporal variation and the distribution of air pollutants from massive datasets. The presence of unknown factors also affects $PM_{2.5}$ prediction, especially for long-term prediction[18]. For better prediction, the ensemble models (EMs) produce the softmax-weighted average of several ML model outputs to outperform DNNs[19]. AdaBoost (AD)[20], generalized additive models (GAM)[21,22], random forests (RF)[20,22], and extreme gradient boosting (XGBoost)[20,21,23] are popular EMs for $PM_{2.5}$ prediction. Recently, the composite neural network[24] has outperformed the EM methods in $PM_{2.5}$ prediction[9]. A composite neural network consists of individually pre-trained DNN components, each of which utilizes knowledge from datasets; component outputs are then connected as an acyclic tree. The leaf outputs are weighted by trained variables and collectively taken as an ensemble node, instead of being softmax weighted as in EM. In this work, we build a remotely transported pollutants (RTP) model[9], a composite neural network consisting of two DNN components pre-trained by heterogeneous datasets from multiple sources to improve $PM_{2.5}$ prediction in southern and central Taiwan.

We not only train the $PM_{2.5}$ prediction model using local meteorological and air pollution monitoring data, but we also introduce large-scale satellite images of East Asia and large-scale weather data to aid our model in capturing the spatiotemporal distribution of remotely transported $PM_{2.5}$. Composite neural network models that include multiple factors are also introduced to improve $PM_{2.5}$ prediction in southern and central Taiwan. We present the results of the model performance analysis in different dimensions and scales and interpret the effect of the proposed input features and architecture for $PM_{2.5}$ prediction for 37 monitoring stations in southern and central Taiwan.

2 Materials

2.1 Study region and air quality data

The study region is located in the south and central part of Taiwan between 21°25′ south and longitude 120°18′ and 120°97′ east. We created a grid area of $234 \times 80 = 18720$ km^2 that covers the study area, where each individual grid cell has a spatial resolution of 1 km. The air pollutants from gas include PM_{10} with a diameter of 10 µm, nitrogen dioxide (NO_2), other nitrogen oxides (NOx), ozone (O3), carbon monoxide (CO), and sulfur dioxide (SO2), all of which strongly influence the formation and future status of $PM_{2.5}$[25]. In this work, we used hourly air pollutant data for three years (2014, 2015, 2016) from the Taiwan EPA (`https://opendata.epa.gov.tw`).

2.2 Aerosol optical depth data from MAIAC algorithm

Aerosol optical depth (AOD) products are typically generated by dark target (DT) and deep blue (DB) algorithms at spatial resolutions of 3 to 10 km. However, AOD retrieval is challenging, especially when thick smoke is observed by satellite-based monitoring devices, which view the smoke as clouds. This makes retrieved AOD data unreliable.

Multiangle atmospheric correlation implementation (MAIAC) is an advanced AOD retrieval algorithm based on time series analysis that has been proven reliable for predicting $PM_{2.5}$[26]. The accuracy of MAIAC AOD in China and East Asia has been validated by the AErosol RObotic NETwork (AERONET) ground measurement network[27]. Given MAIAC's strong performance and global coverage, we use these data to capture information on remote $PM_{2.5}$ transported long distances, for example, from one country to another[9].

In this work, we collected three years (2014, 2015, and 2016) of MAIAC AOD data at a 1×1 km^2 spatial resolution from NASA (`https://ladsweb.modaps.eosdis.nasa.gov`). The AOD products cover two tiles of the investigation area. AOD preprocessing is described in Section 3.2.

2.3 Remote meteorological data

$PM_{2.5}$ can float in the air for 4 to 7 days[28] and can be transported from one place to another with the help of meteorological features. Meteorological features are also involved in the formation of $PM_{2.5}$[28].

We used three years (2014, 2015, and 2016) of meteorological data from two different sources available in the remote area to capture more remote pollutants. The first source is data on temperature, pressure, vertical velocity (VVEL), absolute vorticity (ABSV), lifted index (LFTX), wind speed (ws) and wind direction (θ) at pressure levels from 10 mb (millibars) to 1000 mb from the National Center for Environmental Prediction Final (NCEP FNL) Operational Global Analysis data (`https://rda.ucar.edu/datasets/`). NCEP FNL data is provided in grids of 28×28 km^2 at six-hour intervals. We pre-processed the data and converted them to hourly intervals, as explained in Section 3.2.

The second source is buoy monitoring stations that record the hourly wind speed and direction over the oceans. The ocean wind (OW) influences the transportation of remote $PM_{2.5}$ across the East China Sea to Taiwan. Therefore, we created a grid area that covers the remote area where each individual grid cell has a spatial coverage of 1×1 km^2. We constructed feature maps by filling nonobserved grid cells using kriging interpolation, based on windcos, windsin, and wind speed calculated from Center Weather Bureau (CWB) stations and buoy weather monitoring devices. Other preprocessing is described in Section 3.2.

By interpolating non-observed grid cells with CWB stations on land and buoy weather monitoring devices on the ocean, we assemble wind feature maps which are reliable within our research area, which is encircled by buoy monitoring devices.

2.4 Local meteorological data

The dispersion and transportation of $PM_{2.5}$ is strongly influenced by meteorological features (rainfall, pressure, temperature, humidity, wind speed, and wind direction)[26]. In this work we downloaded these features from the CWB website (http://opendata.cwb.gov.tw/index), which included hourly data from 337 monitoring stations. We preprocessed the data as explained in Section 3.2 using spatial interpolation to populate all nonobserved grid cells and vectorize wind speed and wind direction data as described in Section 3.1.

3 Methods

3.1 Wind feature vectorization

Our wind feature maps derive from wind features composed of speed and direction. Wind direction data are usually represented in polar coordinates, which must be converted to vector form. We vectorized the wind feature from wind speed at a particular angle into meridional (v-wind) and zonal (u-wind) components. To isolate the wind speed feature from the direction features, we then normalized u-wind and v-wind by dividing them by the wind speed to yield the u-wind (*windcos*) and v-wind (*windsin*) direction, which represent the meridional and zonal components of the unit wind vector.

3.2 Data preprocessing

Data preprocessing includes conversion from the monitoring station-based areas to a grid, linear interpolation, spatial interpolation to populate empty grid cells, data cleaning, and spatial downscaling. For the hourly prediction task, we vectorized the wind direction into zonal (*windcos*) and meridian (*windsin*) components of the meteorological dataset (NCEP) as described above. We also used linear interpolation to convert the meteorological dataset (NCEP) to hourly intervals from a six-hour interval.

We cleaned the MAIAC AOD data at 550 nm by filtering out poor quality grid values, after which we interpolated using the remaining grid cells. We also downscaled the spatial dimension of each remote tile (h28v06 and h29v06) to 300×300 km^2 from 1200×1200 km^2 using maximum pooling[14] to fit the available memory of the GPU. Then, for the hourly predictions, we repeated the daily reading of each grid cell 24 times to match the hourly interval of other datasets.

To capture the spatiotemporal characteristics of the speed and direction of the ocean wind over the sea, we created a grid area (492×396 = 194,832 km^2) inside the remote area with each grid cell covering 1×1 km^2. We created a feature map by populating the dataset in the grid area according to the latitude and longitude coordinates of the monitoring stations (CWB and buoys). We used kriging interpolation to populate the remaining grid cells that did not match the station coordinates. Shown in Fig. 1 is an example of the results after kriging interpolation on the CWB and buoy dataset. Maximum pooling was applied to the grid area to reduce the spatial dimensions to 246×198 km^2 to match the memory of the computing resource.

Similarly, we converted the study regions to the grid area (234×80 cells) and created the feature map by populating the grid cells with the observed air quality and meteorology data according to the coordinates of the monitoring stations (37 EPA, 174 CWB) and using four nearest neighbors (4-NN) to populate grid cells outside of these coordinates.

3.3 Modeling methods

The proposed composite neural network models —- RTP with DNN components (base, STRI) —- were trained over two years (2014, 2015) of data and tested on one year (2016). All models were constructed using Keras with a TensorFlow backend and trained on an NVIDIA GPU with 11 GB of memory.

3.3.1 STRI component

The spatiotemporal remote information neural network (STRI)[9] is a component of the RTP model that captures remotely transported $PM_{2.5}$ and predicts local $PM_{2.5}$ concentration. We added ML layers (CNN, ConvLSTM) to the STRI model to capture the spatiotemporal characteristics of the new heterogeneous dataset (ocean wind). We included the ocean wind data to capture more spatial-temporal data on long-range $PM_{2.5}$ transported towards Taiwan.

In this work, the large STRI model with multiple ML layers predicts local $PM_{2.5}$ concentration of 37 EPA stations from the next 4 hours (+4hr) to the next 72 hours (+72hr). The model uses large and heterogeneous datasets from remote areas (AOD, meteorology, ocean wind) with local $PM_{2.5}$ as input. The idea is to capture spatiotemporal characteristics of individual features of the remote area, concatenate these, and then merge them with local features ($PM_{2.5}$) to predict local $PM_{2.5}$ concentration.

Furthermore, to fit the large STRI model into the GPU memory, we divided the model into two components, as shown in Fig. 2. STRI_fe, the first component[9], is used for the extraction of remote pollutants (ERP) given the AOD input from two tiles with their meteorology dataset. STRI_p, the second component, is used for prediction given the ERP input, local features, and spatiotemporal ocean wind features (Fig. 2). The detailed configuration of STRI model is described in detail in Supplementary Table 1.

After dividing the model into two components, we borrowed techniques from previous work[9] to fine-tune the individual components with fewer training parameters to improve the final prediction results.

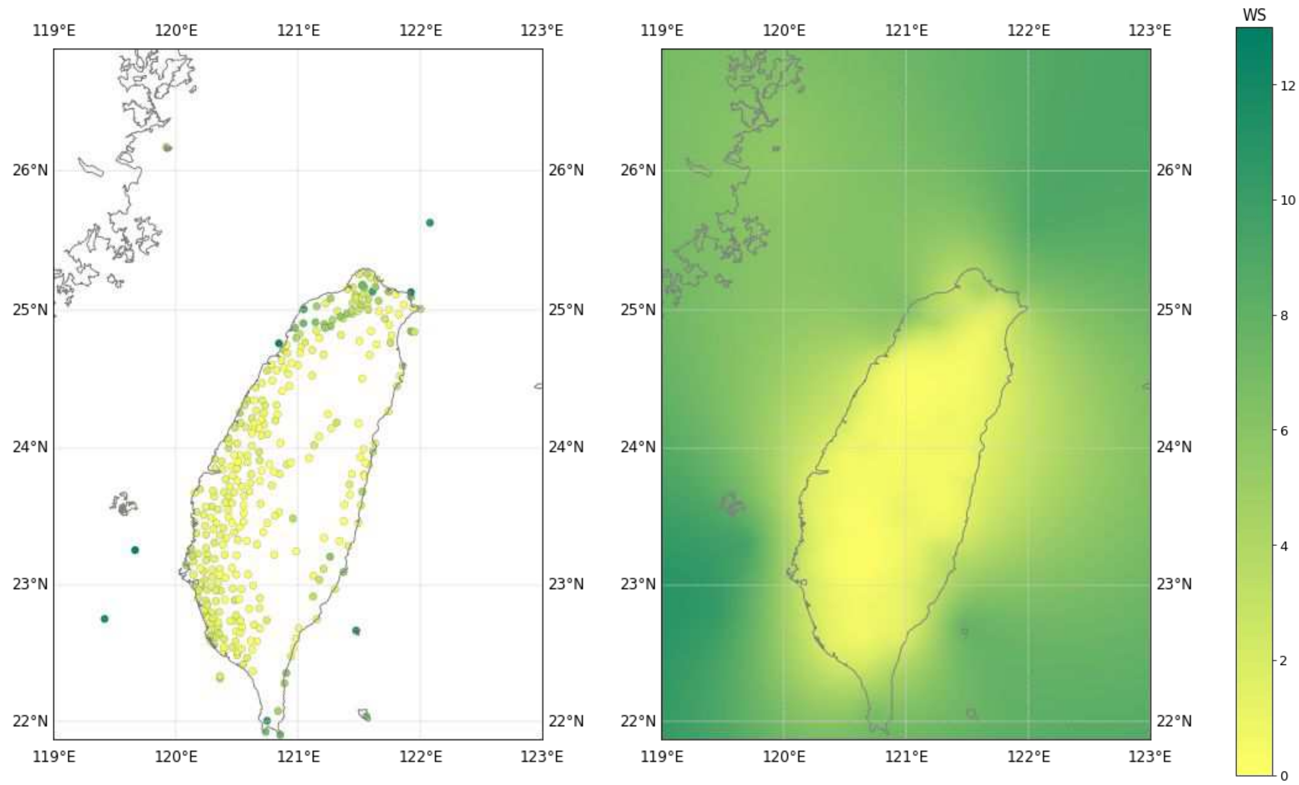

Figure 1. Left side: CWB and buoy monitoring stations. Right side: distribution of ocean wind dataset after kriging interpolation.

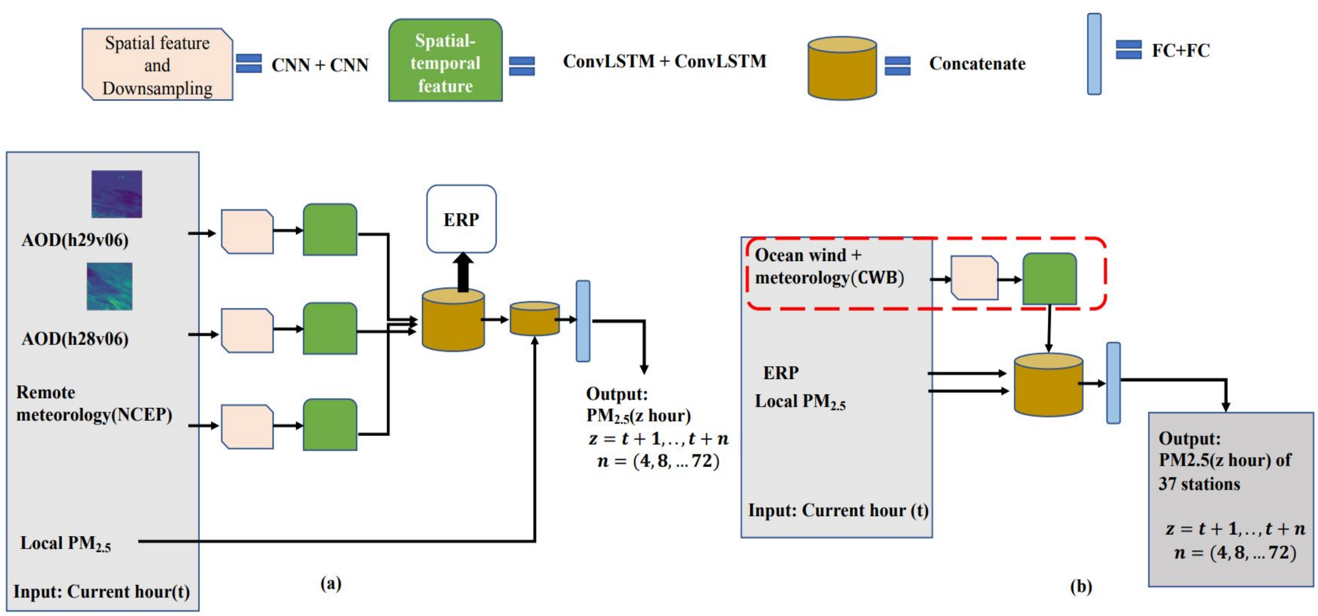

Figure 2. STRI model components STRI_fe (a) and STRI_p (b) with modifications indicated by red dashed line

3.3.2 Base component

The base model[9] is a component of the RTP model that predicts local +4hr to +72hr $PM_{2.5}$ concentration using local features only. The input to the base model is the air quality features (EPA), the weather features (CWB), the weather forecast covering the study area, and the time to predict $PM_{2.5}$ of 37 EPA stations. The base model predicts $PM_{2.5}$ concentration using a single dataset as input. The model is described in detail in Supplementary Fig. 1.

3.3.3 RTP model

Given the prediction output of its pre-trained components (STRI and the base model), the RTP model outputs the final $PM_{2.5}$ predictions for the 37 EPA stations by hour. The RTP model is described in detail in Supplementary Fig. 1.

4 Evaluation

4.1 Metrics

We evaluated the proposed models using the root mean square error (RMSE), which measures the difference between the hourly predicted $PM_{2.5}$ and its true value. In this work, the RMSE is the squared mean of the error between the ground truth and the predicted value at every hour among the monitoring stations of interest:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}(y_{t,i} - \hat{y}_{t,i})^2}, \tag{1}$$

where $y_{t,i}$ and $\hat{y_{t,i}}$ are the true and predicted value of monitoring station i at hour t respectively, T is the length of the prediction sequence and n is the total number of monitoring stations.

4.2 Evaluation of proposed architecture

We conducted experiments to show the $PM_{2.5}$ hourly prediction performance of the proposed architecture by comparing them with benchmarks and also to evaluate the contribution of each input feature to the prediction performance. In our architecture, each model was trained with the corresponding data from 2014 and 2015 and evaluated with the data from 2016.

1. We first compared the prediction performance of RTP_ow with its components (STRI pre-trained with ocean wind and the base model) to evaluate the improvement of the composite neural network architecture with respect to $PM_{2.5}$ prediction.
2. We compared the $PM_{2.5}$ prediction performance of RTP models composed of the STRI pre-trained with ocean wind (RTP_ow) and RTP models composed of the STRI pre-trained without ocean wind (RTP_no_ow) components respectively to evaluate the effect of pre-training with ocean wind data on hourly $PM_{2.5}$ prediction.
3. We further compared the $PM_{2.5}$ prediction performance of the RTP model with its components, after which we compared the RTP model with other ensemble models (ADA, GAM, RF, XGB). RTP and the ensemble models use the same inputs: the prediction output of STRI and the base model. The main objective of these comparisons is to show that RTP outperforms its pre-trained components and other ensemble techniques.
4. Finally, we present the monthly $PM_{2.5}$ prediction performance of models trained under the proposed architecture to compare the model performance during months when land-sea breezes affect southern and central Taiwan to the performance during the rest of the year.

4.3 Grouping of monitoring stations

To accurately analyze the effect of ocean wind on $PM_{2.5}$ prediction with land-sea breezes, we selected the 28 stations most affected by land-sea breezes and annotated these as LS stations, as listed in Supplementary Table 2. The remaining nine stations (Xianxi, Lunbei, Mailiao, Taixi, Xingang, Puzi, Xinying, Annan, and Hengchun) are annotated as normal stations.

5 Results

5.1 RTP and its components

Fig. 3 (top left) shows that RTP and STRI both significantly outperform the base model for short-term prediction (+4hr to +32hr). However, for long-term prediction, STRI is worse than the base model, while RTP exhibits the best $PM_{2.5}$ prediction performance. Fig. 3 (top right) shows the relative improvement in the average RMSE over the base model by prediction hour for both RTP and STRI for the LS and normal stations: the proposed composite neural network architecture greatly improves the LS stations.

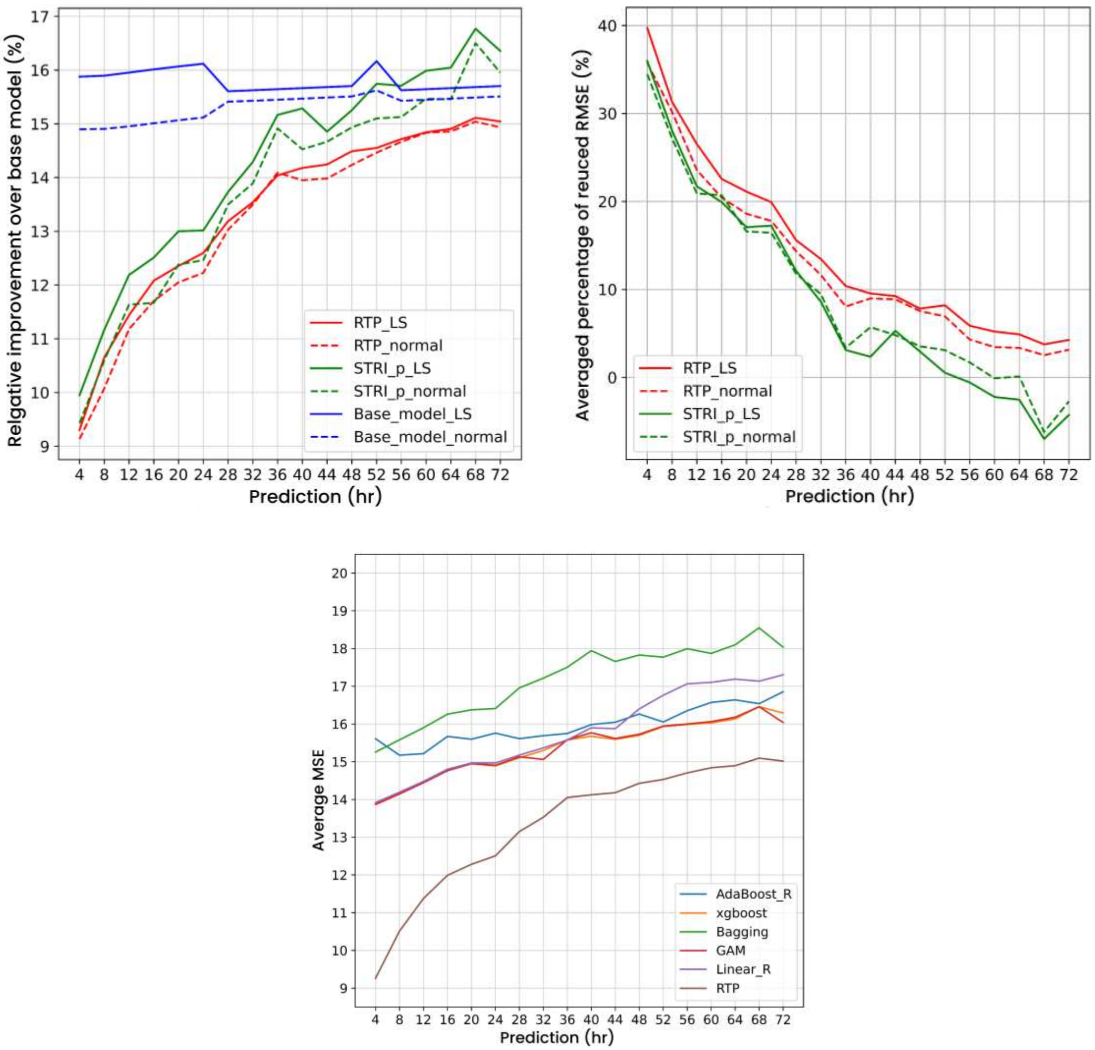

Figure 3. Top left: average RMSE for RTP and components. Top right: relative improvement (%) in average RMSE over base model for RTP and components. Bottom: average RMSE of RTP and other ensemble models.

5.2 Effect of pre-trained components on RTP model

To evaluate the effect of ocean wind data on pre-trained STRI components with respect to hourly $PM_{2.5}$ prediction, we compared RTP composed of two different pre-trained STRI models: STRI pre-trained with $PM_{2.5}$ and AOD data (RTP_no_ow), and STRI pre-trained with $PM_{2.5}$, AOD, and ocean wind data (RTP_ow). In Fig. 4, comparing RTP_ow to RTP_no_ow shows that ocean wind features do not help long-term $PM_{2.5}$ prediction; however, Table 1 shows that in terms of average RMSE over the prediction hours, RTP_ow outperforms RTP_no_ow. This shows that ocean wind helps pre-trained components of composite neural network models to improve the overall $PM_{2.5}$ prediction performance.

5.3 RTP model and other ensemble models

As ensemble models such as AdaBoost, generalized additive models, random forests, and XGBoost have been widely used for $PM_{2.5}$ prediction, we further compared the RTP model trained under the proposed architecture with these models. In this experiment, we input the prediction output from both the STRI and the base model components into RTP and the ensemble models. In Fig. 3 (bottom), the RTP model outperforms the ensemble models (ADA, GAM, RF, XGBoost) at every prediction hour. This shows that the proposed composite neural network architecture has the best overall $PM_{2.5}$ prediction performance in southern and central Taiwan with components pre-trained using large-scale AOD, weather and ocean wind data.

Table 1. Average RMSE for land-sea (LS) and normal stations for RTP pre-trained with (RTP_ow) and without (RTP_no_ow) ocean wind

	RTP_no_ow	**RTP_ow**
LS stations	13.4834	13.4037
Normal stations	13.2319	13.2190
All stations	13.4222	13.3588

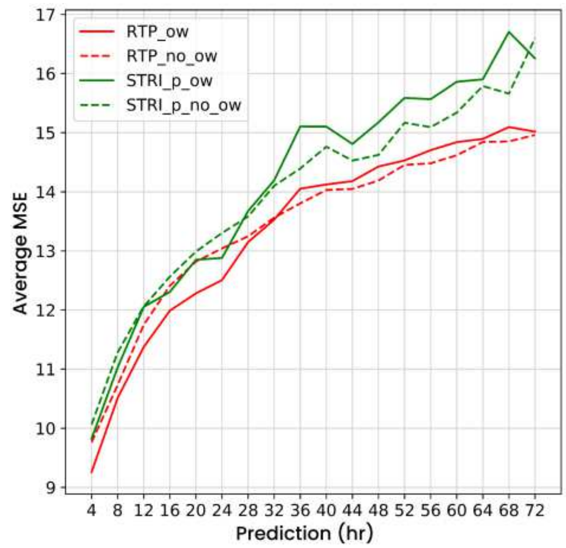

Figure 4. Average RMSE in every prediction hour for RTP and STRI pre-trained with (_ow) or without (_no_ow) ocean wind

5.4 Monthly analysis

Above, we show that RTP_ow significantly improves the $PM_{2.5}$ prediction performance in the overall time sequence. However, land-sea breeze events do not occur throughout the year. Since the purpose of our work is to improve the $PM_{2.5}$ prediction performance for stations in southern and central Taiwan using the proposed model trained with ocean wind features, which we expect to allow the recognition of land-sea breeze events, we separated the $PM_{2.5}$ prediction performance of RTP_ow in all of 2016 into one-month intervals for both LS and normal stations, as shown in Fig. 5. Previous studies show that land-sea breezes occur frequently in spring and early summer. These monthly prediction results show that, in terms of average RMSE, LS stations outperform both normal stations and all stations for every prediction horizon in March, April, and May. In June and August, when $PM_{2.5}$ pollution is the lowest in the year, it is difficult to distinguish RMSEs of LS and normal stations. In Supplementary Fig. 2, we present the monthly prediction results for autumn and winter: clearly, RTP_ow exhibits superior prediction performance for normal stations in September, October, December, and January. In summary, the proposed model produces an improved $PM_{2.5}$ prediction performance for stations in southern and central Taiwan, especially LS stations, in months when land-sea breezes occur frequently.

6 Conclusion

We propose a composite neural network architecture that uses components pre-trained with large-scale weather features and ocean wind to improve $PM_{2.5}$ prediction in southern and central Taiwan. The neural network RTP_ow, which uses STRI, pre-trained with $PM_{2.5}$, AOD, large-scale weather features, and ocean wind features as components, achieved the best overall $PM_{2.5}$ prediction performance compared to its individual components and other ensemble models. Monthly analysis reveals that the proposed model yields improved $PM_{2.5}$ prediction for stations in southern and central Taiwan in months when land-sea breeze events occur frequently.

7 Data availability

Te data that support the fndings of this study are available from the corresponding author upon reasonable request.

Figure 5. Monthly average RMSE

References

1. Lippmann, M. Toxicological and epidemiological studies of cardiovascular effects of ambient air fine particulate matter (PM2.5) and its chemical components: Coherence and public health implications. *Critical Rev. Toxicol.* **44**, 299–347 (2014).
2. Liang, R. *et al.* Effect of exposure to PM2.5 on blood pressure: A systematic review and meta-analysis. *J. Hypertens.* **32**, 2130–2141 (2014).
3. Stafoggia, M. *et al.* Long-term exposure to ambient air pollution and incidence of cerebrovascular events: Results from 11 European cohorts within the ESCAPE project. *Environ. Heal. Perspectives* **122**, 919–925 (2014).
4. Puett, R. C. *et al.* Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. *Environ. Heal. Perspectives* **117**, 1697–1701 (2009).
5. Wu, C.-F. *et al.* Association of short-term exposure to fine particulate matter and nitrogen dioxide with acute cardiovascular effects. *Sci. Total. Environ.* **569**, 300–305 (2016).
6. Gakidou, E. *et al.* Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **390**, 1345–1422 (2017).
7. Tsai, H.-H., Yuan, C.-S., Hung, C.-H., Lin, C. *et al.* Physicochemical properties of PM2.5 and PM2.5–10 at inland and offshore sites over southeastern coastal region of Taiwan Strait. *Aerosol Air Qual. Res.* **11**, 664–678 (2011).
8. Cheng, F.-Y., Chin, S.-C. & Liu, T.-H. The role of boundary layer schemes in meteorological and air quality simulations of the Taiwan area. *Atmospheric Environ.* **54**, 714–727 (2012).
9. Kibirige, G., Yang, M.-C., Liu, C.-L. & Chen, M. C. Using satellite data on remote transportation of air pollutants for PM2.5 prediction in northern Taiwan (2021).
10. Hsu, C.-H., Cheng, F.-Y. *et al.* Synoptic weather patterns and associated air pollution in Taiwan. *Aerosol Air Qual. Res.* **19**, 1139–1151 (2019).

11. Feng, X. *et al.* Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environ.* **107**, 118–128 (2015).

12. Biancofiore, F. *et al.* Recursive neural network model for analysis and forecast of PM10 and PM2.5. *Atmospheric Pollut. Res.* **8**, 652–659 (2017).

13. Di, Q. *et al.* Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. & Technol.* **50**, 4712–4721 (2016).

14. Sønderby, C. K. *et al.* MetNet: A Neural Weather Model for Precipitation Forecasting. *arXiv preprint arXiv:2003.12140* (2020).

15. Huang, K. *et al.* Predicting monthly high-resolution PM2.5 concentrations with random forest model in the North China Plain. *Environ. Pollut.* **242**, 675–683 (2018).

16. Brokamp, C., Jandarov, R., Rao, M., LeMasters, G. & Ryan, P. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environ.* **151**, 1–11 (2017).

17. Wei, J. *et al.* Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. *Remote. Sens. Environ.* **231**, 111221 (2019).

18. Yang, M.-C. & Chen, M. C. PM2.5 forecasting using pre-trained components. In *2018 IEEE International Conference on Big Data (Big Data)*, 4488–4491 (IEEE, 2018).

19. Li, L. *et al.* Ensemble-based deep learning for estimating PM2.5 over California with multisource big data including wildfire smoke. *Environ. Int.* **145**, 106143 (2020).

20. Zhai, B. & Chen, J. Development of a stacked ensemble model for forecasting and analyzing daily average PM2.5 concentrations in Beijing, China. *Sci. The Total. Environ.* **635**, 644–658 (2018).

21. Shtein, A. *et al.* Estimating daily PM2.5 and PM10 over Italy using an ensemble model. *Environ. Sci. & Technol.* **54**, 120–128 (2019).

22. Xiao, Q., Chang, H. H., Geng, G. & Liu, Y. An ensemble machine-learning model to predict historical PM2.5 concentrations in China from satellite data. *Environ. Sci. & Technol.* **52**, 13260–13269 (2018).

23. Li, L. *et al.* Estimation of PM2.5 concentrations at a high spatiotemporal resolution using constrained mixed-effect bagging models with MAIAC aerosol optical depth. *Remote. Sens. Environ.* **217**, 573–586 (2018).

24. Meng, X. & Karniadakis, G. E. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. *J. Comput. Phys.* **401**, 109020 (2020).

25. Yi, X., Zhang, J., Wang, Z., Li, T. & Zheng, Y. Deep distributed fusion network for air quality prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 965–973 (2018).

26. Chu, Y. *et al.* A review on predicting ground PM2.5 concentration using satellite aerosol optical depth. *Atmosphere* **7**, 129 (2016).

27. Su, X., Wang, L., Zhang, M., Qin, W. & Bilal, M. A High-Precision Aerosol Retrieval Algorithm (HiPARA) for Advanced Himawari Imager (AHI) data: Development and verification. *Remote. Sens. Environ.* **253**, 112221 (2021).

28. Cobourn, W. G. An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations. *Atmospheric Environ.* **44**, 3015–3023 (2010).

8 Author contributions

G.K. and C.H. contributed to data collection, the main manuscript text writing and literature survey. G.K. contributed to the main architecture establishment and prepared fig 2, 6 and table 2. C.H. contributed to the the architecture perfomance analysis and interpretation of data, and prepared fig1, 3, 4, 5, 7 and table 1. M.C. and C.L. administrated and supervised this preojects. All authors reviewed the manuscript.

9 Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf