



Clustering-based spatial transfer learning for short-term ozone forecasting

Tuo Deng^{a,*}, Astrid Manders^b, Jianbing Jin^c, Hai Xiang Lin^{a,d}

^a Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands

^b TNO, Department of Climate, Air and Sustainability, Utrecht, the Netherlands

^c Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China

^d Institute of Environmental Sciences, Leiden University, The Netherlands



ARTICLE INFO

Keywords:

Air quality
Ozone prediction
Clustering
Transfer learning

ABSTRACT

Ground-level ozone is a critical atmospheric pollutant, and high concentrations of ozone can damage human health, affect plant growth and cause ecological harm. Traditional chemical transport models and popular machine learning models have difficulty in predicting ozone concentrations, especially in times with high concentrations. We propose a clustering-based spatial transfer learning Multilayer Perceptron (SPTL-MLP) to predict ozone concentration at the target observation station for the next three days. We use k-means clustering algorithm to find similar stations and train them together to get a base model for spatial transfer learning. For practical applications, a weighted loss function has been designed with an extra emphasis on reducing prediction errors of high ozone concentrations. Evaluation using historical data of stations in Germany shows that our SPTL-MLP model has a smaller error (reduced by 9.13%) and higher prediction accuracies of ozone exceedances (improved by 8.21% and 16.9%) compared to MLP (without spatial transfer). The results demonstrate the effectiveness of the SPTL-MLP in the short-term ozone forecast. It can be used for timely warning of ozone exceedances and help governments to detect air quality.

1. Introduction

Tropospheric ozone is an air pollutant that is detrimental to human health and plant growth (Fang et al., 2013). Long-term exposure to high ozone concentrations increases the risk of respiratory disease and death (Sicard et al., 2019). For the protection of human health, the European Union's (EU) air quality directives and the World Health Organization (WHO) guideline both set thresholds for maximum daily 8-hour O_3 mean concentrations (EU: $120\mu\text{g}/\text{m}^3$, WHO: $100\mu\text{g}/\text{m}^3$). According to EU, the number of days with an ozone concentration exceeding $120\mu\text{g}/\text{m}^3$ should not be more than 25 days per calendar year. It is necessary to make short-term ozone forecasts such that timely warnings or measures can be issued.

In the presence of sunlight, tropospheric ozone is produced by chemical reactions between oxides of nitrogen (NO_x gases) and volatile organic compounds (VOCs) (Lu et al., 2019). The formation of tropospheric ozone is described through the following reaction. NO_2 is photodissociated by sunlight to generate atomic oxygen (1), which combines with O_2 to create O_3 (2). VOC oxidizes NO emitted by combustion to NO_2 , the precursor of ozone.



Meanwhile, at nighttime and in the vicinity of large emissions of NO (e.g., power plants), ozone will react with NO_2 and form nitrate radical (NO_3) in reaction 3 (Finlayson-Pitts and Pitts Jr, 1993).



Changes in meteorological conditions such as incoming solar radiation, temperature and humidity will affect the production of ozone. Meanwhile, wind speed, wind direction, boundary layer height, etc. will also affect the advection processes related to ozone and change the ozone concentration. Besides, a crucial factor that affects ozone concentration is the emission of ozone precursors. The majority of emissions are caused directly by human activities (e.g., industrial, agricultural, etc.) and can vary dramatically over a short time, which increases the difficulty of short-term ozone forecasting. It should be noticed that ozone depletion in the stratosphere can also affect the tropospheric ozone. Although the ground-level ozone can hardly be affected by the slow exchange of air across the tropospheric-stratospheric border, it can be affected by the radiation and climate changes caused by stratospheric ozone depletion (Williamson et al., 2019).

* Corresponding author.

E-mail address: t.deng@tudelft.nl (T. Deng).

Chemical transport models (CTMs) are often used to predict short-term ozone concentration (Sharma et al., 2017). Meanwhile, data assimilation (DA) methods are used in most of those studies to optimize the initial or boundary conditions. Curier et al. (2012) used LOTOS-EUROS chemical transport model and an Ensemble Kalman Filter (EnKF) to assimilate ground-level ozone concentrations over Europe. It is found that LOTOS-EUROS sometimes underestimates the ozone daily maximum. Ryu et al. (2019) used the WRF-Chem driven by the Rapid Refresh (RAP) forecasting system and the Global Forecast System (GFS) forecasts to simulate ozone for the next day over the Contiguous United States (CONUS). The RAP data assimilation can reduce ozone forecast errors at the first 3 forecast hours due to the improvements in the cloud forecast skill during the initial forecast hours. However, it also turns to overpredict the ozone concentration and gives more frequent false alarms.

Previous studies have also discussed the limitation of CTMs and DA methods. The results are uncertain due to the error in meteorological forecasting (temperature, wind, humidity, etc.) (Hu et al., 2019). Besides, it was found in previous model validation and intercomparison studies that CTMs have uncertainties and errors (uncertainties in emissions, meteorological parameters, etc.) that will affect the prediction of air pollutants (Otero et al., 2018).

Recently, machine learning (ML) methods have been used to predict short-term ozone concentration (Cheng et al., 2021; Wang et al., 2022). Instead of simulating complex atmospheric physical and chemical processes, data-driven machine learning models make predictions of ozone concentrations based on data itself. Zhan et al. (2018) used a random forest (RF) model to predict the daily maximum 8-h average ozone concentration (MDA8) in China. Feng et al. (2019) compared the performance of the traditional atmospheric model (WRT coupled with CMAQ) and multiple machine learning models (Multi-layer Perceptron (MLP), Random Forest, Recurrent Neural Network (RNN), etc.) by predicting the surface MDA8 in Hangzhou. RNN performs the best among all machine learning models considered and has less error than the traditional atmospheric model. Besides, a hybrid model based on convolutional neural networks and long short-term memory (CNN-LSTM) is utilized to estimate MDA8 for the next day in Beijing City (Pak et al., 2018). CNNs can extract the inherent features of huge amounts of air quality and meteorological data, and LSTMs will capture the long-term dependencies of the input time series data.

Unlike traditional chemical transport models, machine learning models are not restricted by resolution and can better solve local problems (e.g., one-station ozone prediction). Meanwhile, traditional models require a pollutant emission inventory to simulate air pollution, which is costly to create and must be updated regularly. Machine learning models do not require an emission inventory, nor do they need the simulation of complex physical and chemical processes (Feng et al., 2019). It will take much less time and resources to train the ML models.

However, problems with machine learning ozone prediction models cannot be ignored. First, the response of the machine learning model to ozone changes is slow (Sayeed et al., 2020). Typically, the model underpredicts when the true concentration suddenly increases and overpredicts when it suddenly declines. Meanwhile, it is difficult for machine learning models to capture extreme situations, such as ozone exceedances. Because high ozone concentration can harm human health, accurate prediction of high ozone concentration is essential for providing early warning. However, due to the relatively small amount of high-concentration ozone observation data, it is difficult to train machine learning models to predict such events accurately (Gong and Ordieres-Meré, 2016).

Transfer learning is a machine learning method that learns the target task through the transfer of the knowledge from a related task that has already been learned (Torrey and Shavlik, 2010). Transfer learning is currently widely used in the field of natural language processing (Weiss et al., 2016), but it is rarely used in air pollution prediction. Ma et al. (2019) used temporal transfer learning to transfer the knowl-

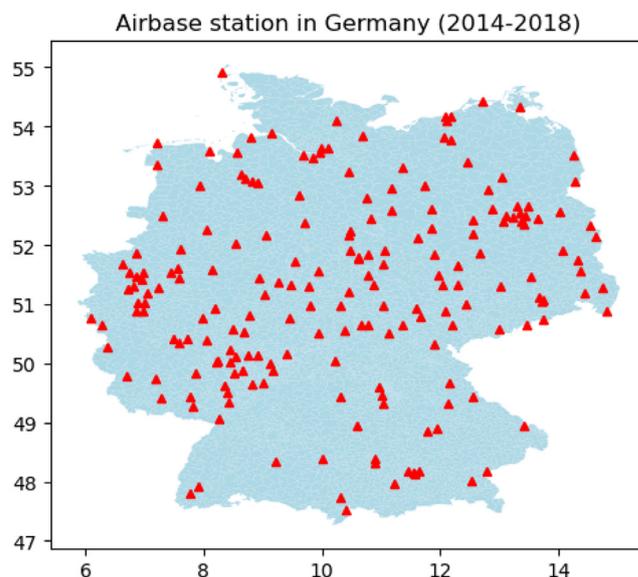


Fig. 1. Distribution of air quality stations in Germany.

edge from smaller temporal resolutions to larger temporal resolutions to predict PM_{2.5} in one day and one week. Fong et al. (2020) used transfer learning to predict PM_{2.5} at the target station based on the pre-trained LSTM model for PM₁₀. For stations that lack data, Ma et al. (2020) also proposed a transfer learning-based stacked bidirectional long short term memory (TLS-BLSTM) network to predict air quality 1 h ahead.

This paper proposes a spatial transfer learning model to predict ozone concentration for the next one and three days in Germany. To reduce the prediction error, we do the clustering for all stations in Germany to find similar stations. The cluster in northern Germany is taken, and one multi-station model is trained based on all stations in the same cluster. Spatial transfer learning ozone forecast models are trained for target stations based on the multi-station model, and the experimental results show that it can reduce the prediction error of ozone. To improve the prediction accuracy of ozone exceedances, a weighted loss function has been designed to ensure smaller errors in high ozone concentration data. With the weighted loss function, our transfer learning model can effectively improve the prediction accuracy of ozone exceedances.

We start in Section 2 by describing the data processing and analysis. Related machine learning methods and evaluation criteria in our research are illustrated in Section 3. Section 4 shows the result of clustering and comparison of the different ozone forecast models. Section 5 shows the conclusion of the paper.

2. Data description and analysis

2.1. Data collection

The data we used contains two parts: Air pollution data and meteorological data. The air pollution data is from German Environment Agency (Umweltbundesamt, UBA). We used the data from 2014 to 2018 for the experiment. 199 observation stations can provide the ozone and nitrogen oxides data we need at the same time. Figure 1 shows the distribution of the stations in Germany.

The meteorological data we used is from German Meteorological Service (Deutscher Wetterdienst, DWD) (Bollmeyer et al., 2015). It provides hourly reanalysis data over Europe with the grids of 6km * 6km resolution. The reanalysis is based on DWD's operational NWP model COSMO (COSMO-REA6). We selected 6 meteorological factors (Rain, Incoming Solar Radiation, Relative Humidity, 10m Wind Direction, 10m Wind Speed and 2m Temperature) and the grid cell value of the grid cell in which station was located is taken. The details of interpolation is de-

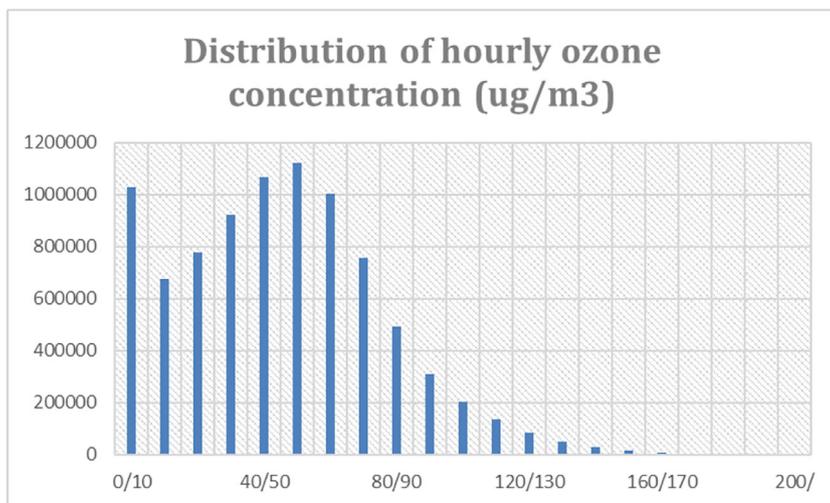


Fig. 2. Amounts of hourly ozone concentration.

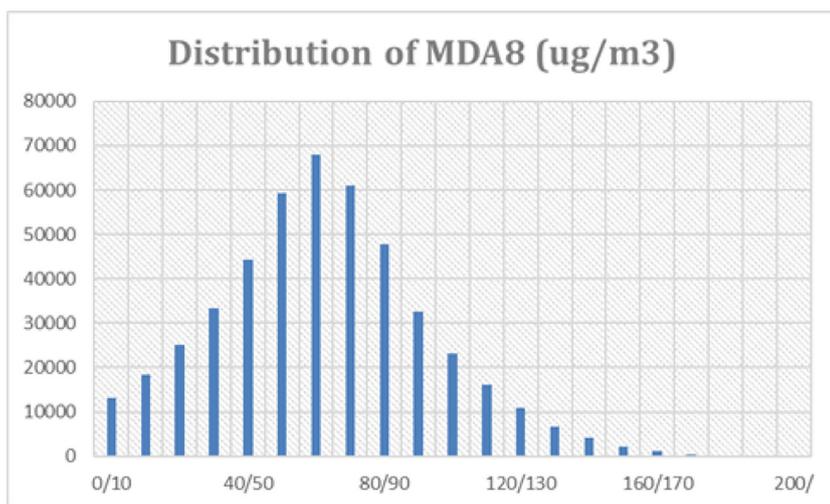


Fig. 3. Amounts of MDA8 data.

scribed in Section 2.2. After that, we selected the data from 2014 to 2018 in the corresponding 199 stations for experiments.

2.2. Data interpolation and processing

We first need to interpolate the missing data in the air pollution data set. For ozone data, the highest missing rate for a single station is 13.19% and the average missing rate across all the stations is 2.72%. For NO_x data, the highest missing rate for a single station is 13.59% and the average missing rate is 2.17%. We use the temporal nearest-neighbor interpolation method to interpolate data. Missing values are filled by the nearest value in the same time series (same station) with python package ‘Scipy’ (Virtanen et al., 2020).

As mentioned in Section 2.1, it is necessary to obtain the meteorological data of each station from the grid cell data. For each station, we first find the nearest grid point and use the data from the grid point as the data for the target station. L1 norm is used to calculate the distance. In addition, since we want to detect the ozone exceedances of MDA8, we need to convert hourly data into daily data. We calculate MDA8 for ozone data and the daily mean value for meteorological data and NO_x .

2.3. Data analysis

We first use bar plots to show the distribution of ozone data from all stations. We use intervals of $10 \mu g/m^3$ in Figs. 2 and 3 plot the dis-

Table 1

Statistical characteristics of centroids of different clusters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	67.69	65.71	66.41	80.05	59.31
Std	30.08	26.30	33.79	26.79	31.72
Min	1.901	2.37	0.99	4.68	0.44
25%	47.10	49.05	42.80	62.75	37.84
50%	66.26	65.29	65.08	76.68	57.03
75%	86.22	80.52	88.23	95.04	77.44
Max	176.59	177.22	183.80	190.30	200.29

tribution of hourly ozone and MDA8. We can see that hourly ozone is concentrated in the range of $0-10 \mu g/m^3$ (nighttime concentrations) and $50-60 \mu g/m^3$. The number of samples decreases with increasing ozone concentration when the ozone concentration is above $60 \mu g/m^3$. There are only a few cases where the ozone concentration is higher than $100 \mu g/m^3$. Similar behavior appears in Fig. 3, with the difference that the peak occurs around $70 \mu g/m^3$. For MDA8, the amount of low concentration ozone data ($<10 \mu g/m^3$) is greatly reduced since the nighttime concentrations will not appear in 8-h max, which always takes place during the day.

We selected a representative station DEBB029 (Schwedt, suburban, see Table 2) to analyze the temporal pattern of ozone concentration. We can first find the periodicity through autocorrelation of ozone concen-

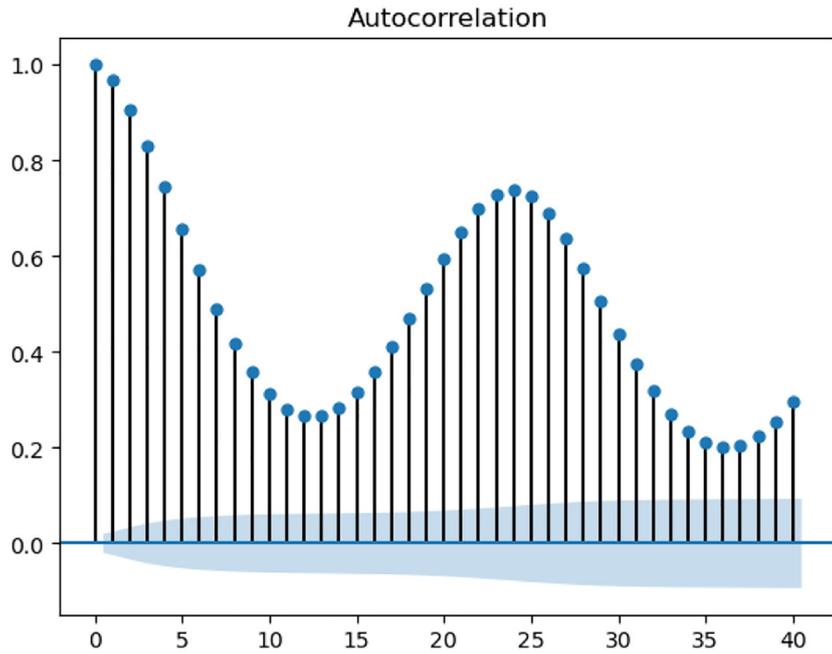


Fig. 4. Autocorrelation of hourly ozone concentration.

Table 2
Station Information.

Station code	Lat	Lon	station name	station type
DEBB007	51.46	13.53	Elsterwerda	suburban
DEBB021	52.40	13.05	Potsdam-Zentrum	urban
DEBB029	53.06	14.29	Schwedt (Oder)	suburban
DEBB032	52.15	14.64	Eisenhüttenstadt	suburban
DEBB053	52.56	14.01	Hasenholz (Buckow)	rural
DEBB055	52.42	12.55	Brandenburg a.d. Havel	suburban
DEBB064	51.75	14.33	Cottbus	urban
DEBB065	52.19	12.56	Luthe(Belzig)	rural
DEBB066	51.90	14.06	Spreewald	rural
DEBB067	52.61	12.89	Nauen	suburban

tration in Figs. 4 and 5. The autocorrelation of hourly ozone reaches a peak every 24-time steps, which matches the daily periodicity. However, no apparent periodicity of MDA8 can be found from autocorrelation analysis. The daily data will not be affected by solar radiation and temperature changes during the day and night. Meanwhile, MDA8 is mainly influenced by the daily maximum temperature, which has no obvious periodic (synoptic scale weather) pattern in the short term.

2.4. Data transformation

In machine learning, we consider ozone prediction as a time series prediction problem, which requires us to reconstruct the data (Bontempi et al., 2012). Without loss of generality, the time series forecast model can be summarized in Eq. (4)

$$Y_{t+d} = f(Y_t, Y_{t-1}, \dots, Y_{t-h+1}) + \epsilon \quad (4)$$

where $\{Y_1, Y_2, \dots, Y_{t+d}\}$ is the time series, h is the number of time-steps used as predictors, d is the time-step ahead we want to predict and f is the forecast model. ϵ is the irreducible error.

In the forecasting setting, the training data should be reconstructed into a $[(t-h+1) * h]$ input matrix in Eq. (5) and a $[(t-h+1) * 1]$ output vector in Eq. (6).

$$\begin{pmatrix} Y_t & Y_{t-1} & \dots & Y_{t-h+1} \\ Y_{t-1} & Y_{t-2} & \dots & Y_{t-h} \\ \vdots & \vdots & \vdots & \vdots \\ Y_h & Y_{h-1} & \dots & Y_1 \end{pmatrix} \quad (5)$$

$$\begin{pmatrix} Y_{t+d} \\ Y_{t+d-1} \\ \vdots \\ Y_{h+d} \end{pmatrix} \quad (6)$$

In our experiments, the input of the model can be divided into two parts. The first part includes historical air pollution data, while the second part contains both historical and feature reanalysis meteorological data. The machine learning model can be represented by the following equation

$$O_{t+d} = f(M_{t+d}, M_{t+d-1}, \dots, M_{t-h_1+d+1}; X_t, X_{t-1}, \dots, X_{t-h_2+1}) + \epsilon \quad (7)$$

where M is the meteorological data, X is the air pollution data and O is MDA8. d is the time-step ahead we want to predict, while h_1 and h_2 are number of time-steps data used as input. f is the forecast model and ϵ is the irreducible error.

3. Methods and evaluation

3.1. Clustering

Clustering is a fundamental task in data mining that is used mostly as an unsupervised learning method (Rokach and Maimon, 2005). Clustering is the process of dividing a data set into different clusters according to a specific criterion (e.g., distance), so that data with similar properties will be grouped together. Unlike classification, the goal of clustering is not to predict the categories of the data but to discover a new set of categories. The structure of clustering can be represented as follows

$$S = \bigcap_1^k C_i \quad (8)$$

$$C_i \cap C_j = \emptyset \text{ for } i \neq j \quad (9)$$

where C_i is a subset of dataset S . Any instance in S belongs to one and only one subset.

In our experiments, we need to divide all stations into different groups and provide a basis for spatial transfer learning. Here we choose the k-means algorithm for clustering (Franklin, 2005). In the k-means algorithm, we first select the centroids of k clusters. Then we calculate

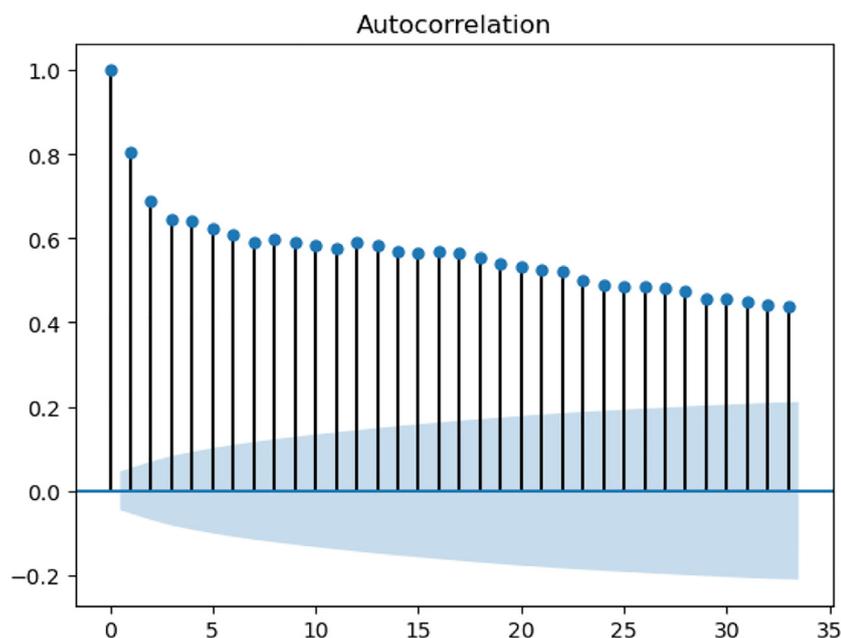


Fig. 5. Autocorrelation of MDA8.

the distance to the k cluster centroids for each sample and assign it to the cluster with the smallest distance. We recalculate the cluster centroids and all distances until a stop condition is reached (e.g., number of iterations). The k -means algorithm has a low time complexity and is also suitable for clusters of different sizes and shapes. Besides, the elbow method is used to determine the k -value. It is based on the idea that one should choose a number of clusters so that adding another cluster does not give much better modeling of the data (Bholowalia and Kumar, 2014). We use Python library scikit-learn (Pedregosa et al., 2011) for the k -means algorithm and elbow method.

3.2. Multilayer perceptron

Multilayer Perceptron (MLP) is a class of feed-forward artificial neural network (ANN) that consists of a system of interconnected neurons (Gardner and Dorling, 1998). It represents a nonlinear mapping between input and output. MLP generally includes three different layers: one input layer, one or more hidden layers, and one output layer. Each layer consists of one or more neurons. The input layer passes the inputs to the network and does not perform calculations. In the hidden layer and the output layer, the nodes have weights and activation functions. Each neural layer receives information from the previous layer and passes it on to the next layer after processing. In MLP, neurons in each layer are fully interconnected with neurons in the next layer, while neurons in the same layer are not connected to each other. The structure of MLP can be seen in Fig. 6.

3.3. Transfer learning

Transfer learning is a machine learning approach where a pre-trained model is reused in another task (Tan et al., 2018). The goal of transfer learning is to improve the model performance in the target task by transferring the information from the source task (Torrey and Shavlik, 2010). It is usually used when the sample size is insufficient or the model is too complex. Transfer learning enables the target model to gain knowledge before training, which gives the model a better initial performance. Compared to training the model from scratch, transfer learning can significantly reduce the time required for model training. Most importantly, when there is insufficient information in the target domain

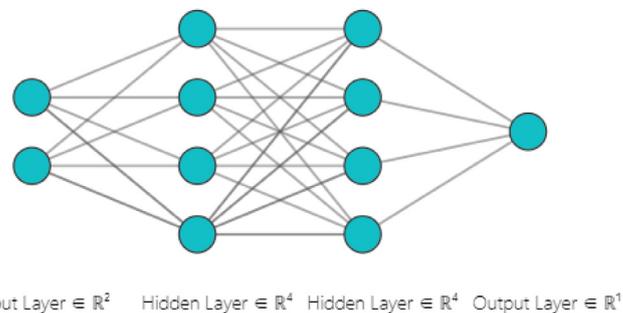


Fig. 6. Multilayer Perceptron.

(insufficient amount of data), it can transfer knowledge from the source domain to the target domain and improve the model performance.

In our research, we use network-based deep transfer learning to optimize the model. We use Python library Keras (Gulli and Pal, 2017) with the TensorFlow backend (Abadi et al., 2015) for all deep learning models. A multi-station MLP model based on the data from all stations in the same cluster is trained and used as a base model. The first few layers of the base model are treated as a feature extractor; the features learned from the multiple similar stations are retained by the structure and parameters of the front layers. The remaining layers are treated as a predictor and need to be retrained with the data from the target station. Our framework will provide additional information from similar stations and a better initial state for the model in the target station. The structure of transfer learning is depicted in Fig. 7.

3.4. Evaluation criterion

We use root mean squared error (RMSE) (Chai and Draxler, 2014) and coefficient of determination (R^2) to evaluate the forecast model. There are two different thresholds ($100\mu\text{g}/\text{m}^3$ and $120\mu\text{g}/\text{m}^3$) of ozone exceedances and we use them both to evaluate the prediction accuracy (PA100 and PA120) of ozone exceedances. The result of the machine learning model will be compared with the result for the free run (no bias correction, no data assimilation) of chemical transport model LOTOS-EUROS (Manders et al., 2017). Ozone concentrations are extracted from an operational simulation of air quality with LOTOS-EUROS model for

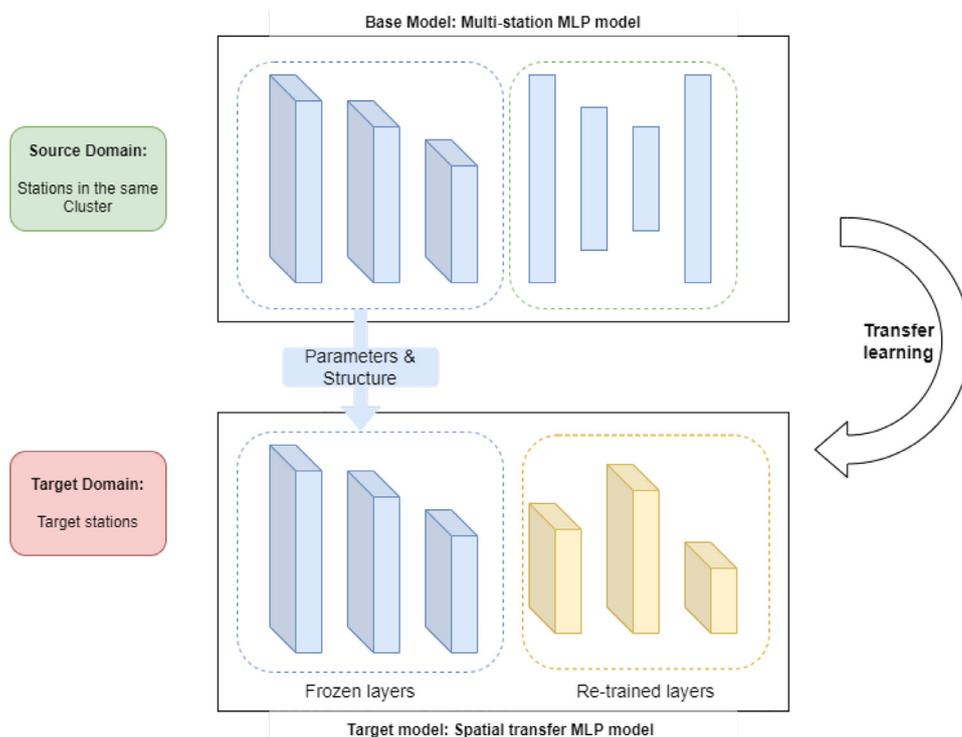


Fig. 7. Network-based transfer learning.

2018 (CAM5) at observation locations. The resolution of the simulation is 0.1×0.1 degrees.

4. Experiment and analysis

4.1. Clustering

In this section, we cluster the stations based on the statistical characteristics of the ozone data and the geographic location of each station. There are 9 input features for each station. For all 199 stations where ozone data exist from 2014 to 2018, we first calculate seven statistical characteristics of MDA8 for each station: mean, maximum, minimum, standard deviation, first quartile, second quartile, and third quartile. At the same time, the longitude and latitude of the stations are used as two additional features. Although stations already have category labels, these labels are mainly based on the geographic location (e.g., urban, suburban, rural, etc.). Our clustering study focuses on the statistical characteristics of the ozone data, which can help us find stations with similar ozone pollution and provide the basis for our transfer learning study. We also use location information as an additional feature so that stations in the same cluster could have similar emission sources and meteorological conditions.

We use k-means algorithm to cluster the stations based on the above features. To reduce the influence of different numerical ranges of the features, we first normalize the data so that the mean of a single feature is 0 and the variance is 1. We also need to set the number of clusters (k), which is selected by elbow methods. When the number of clusters exceeds the actual number of groups (optimal number of clusters), the added information drops dramatically. The position of the bend (knee) in the figure is considered an indicator of the appropriate number of clusters. As shown in Fig. 8, there is an inflection point at $k=5$, so we set the number of clusters to 5.

The results of the clustering are shown in Fig. 9. It can be seen that the clustering result is influenced by the topographic distribution: Clusters 1, 2, and 5 are located in the plains of central, northern, and western Germany, respectively; Cluster 3 is located in the mountainous region of southern Germany. The stations in Cluster 4 are mainly located near

the mountainous regions of Germany's southwestern and southeastern borders.

In order to better compare the difference of ozone between different clusters, we compared the centroids of all 5 clusters. Table 1 shows that the average annual ozone value is significantly higher at the southern and southwestern borders (Cluster 4) and lower in western Germany (Cluster 5). It might be caused by variations in average annual temperatures in different regions. The ozone peak in cluster 5 is much higher than in other clusters due to the dense population and relatively warm summer conditions in western Germany. Owing to the colder winters in southern Germany, the lower quartile and median of ozone in cluster 5 are noticeably lower. Besides, the statistical characteristics of the ozone data are similar for most of the stations in middle and eastern Germany (Cluster 1–3), and the fluctuation of ozone gradually increases from north to south, which also fits the pattern of temperature fluctuations in eastern Germany throughout the year.

4.2. 1-Station model

4.2.1. Model performance

In our experiments, we first investigate the performance of different neural networks to select the base model for spatial transfer learning. MLP, the most frequently used fully connected neural network, is compared to CNN-LSTM (Pak et al., 2018) and S-BLSTM (Ma et al., 2020), two recurrent neural networks that perform well in recent ozone prediction studies. In Cluster 1, ten stations with the starting code "DEBB" were selected for testing (see Table 2). Based on the data of a single station, MLP, CNN-LSTM and S-BLSTM are used respectively to predict MDA8 for the next day. The data from 2014 to 2017 are used as training data, and the data in 2018 are used to test.

We need to determine the input time step length of input data h_1 and h_2 in Eq. (7). To simplify the model, we use the same values for h_1 and h_2 and use h to represent them. From 1 to 7, we gradually increase the length of the input time step and study the effect on the prediction results by cross-validation. We found that when $h > 3$, further increasing the input time step does not have a significant impact on the prediction results. From the autocorrelation analysis in Section 2.3, the autocorre-

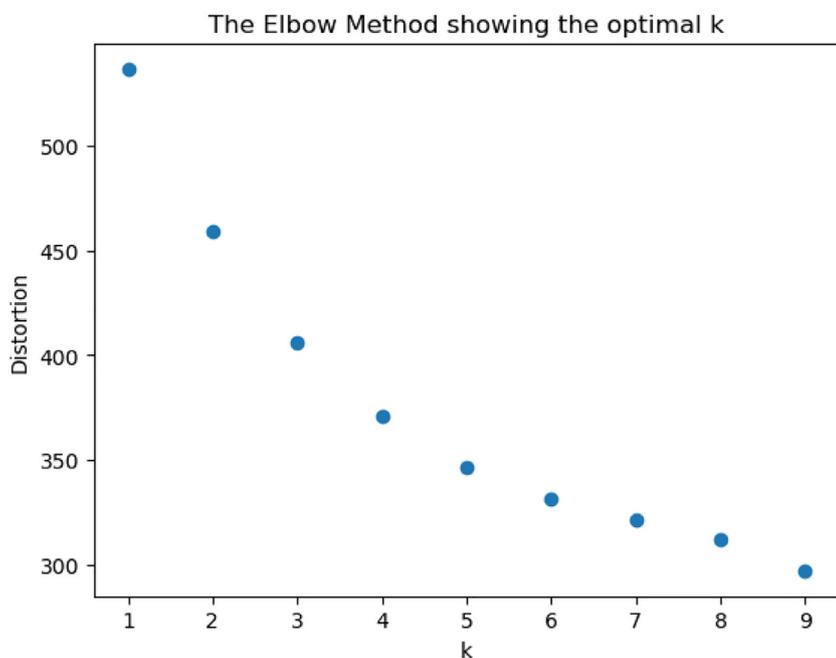


Fig. 8. Graph of Sum of Square Error (SSE) with different number of clusters.

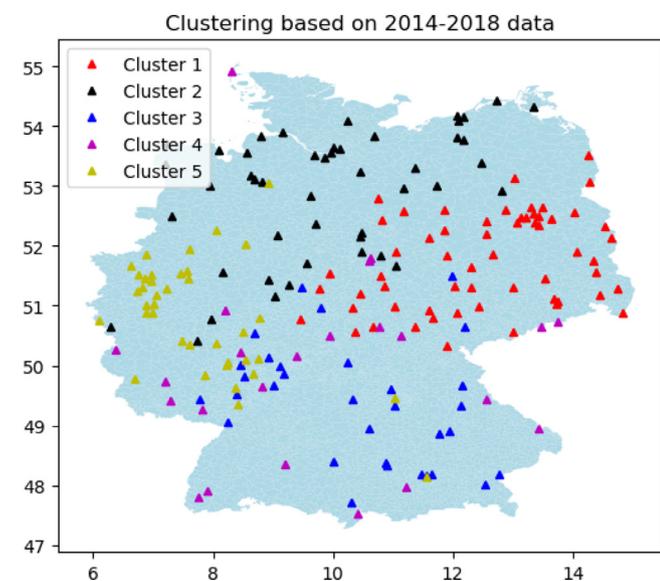


Fig. 9. Different clusters based on k-means algorithm for all German stations.

lation coefficient of the ozone concentration time series does not change significantly after the time step interval exceeds 5. We choose $h = 5$ in the following experiment.

A number of hyperparameters of the neural network model needs to be determined before model fitting. For each set of inputs, we use all 8 elements as features, including 2 air pollution elements and 6 meteorological elements. Relu function is selected as the activation function, and Adam is selected as the optimizer. The structure of the forecast models should be determined through cross-validation. The MLP contains two fully-connected layers. The CNN-LSTM contains two CNN layers and two LSTM layers, and the B-LSTM contains three Bidirectional LSTM layers. Each hidden layer contains 100 neurons. An output layer containing one neuron is added after the hidden layers of each model.

For practical applications, we designed a weighted loss function with an extra emphasis on reducing prediction errors of high ozone concentrations (Jafari et al., 2019). We are more concerned with high ozone

Table 3

Result of MLP with different α values at station DEBB029.

α	RMSE	R^2	PA(> 100)	PA(> 120)
0	11.73	0.82	69%	65%
2	11.96	0.83	77%	69%
4	12.47	0.82	77%	67%
6	13.09	0.81	81%	69%
8	14.18	0.79	85%	79%

concentrations because of their greater impact on human health and plant growth. Therefore, we give more importance and a higher weight to high concentration ozone data during the model training process. The loss function is shown in equation 10 and the hyperparameter α will be determined through cross-validation.

$$L(y) = e^{\alpha * y} * (y - \hat{y})^2 \tag{10}$$

By adjusting the weight term α , we can adjust the weights of different concentrations of ozone data to further optimize the model. The result with different α values can be seen in Table 3.

When α is 0, the loss function is the standard mean square error, and we consider all ozone data equally important. We can find that although the overall error of the model is low (RMSE is 11.52), it tends to underestimate high concentrations of ozone, which is a common problem and has been reported in previous studies (Fong et al., 2020; Ma et al., 2022). As α value becomes larger, the weight of the error of high ozone concentrations becomes higher. When an appropriate α is chosen, we can slightly sacrifice the overall error to improve the prediction accuracy of high ozone concentrations. When α is 2, the prediction accuracy of ozone exceedances is improved to 77% and 69%, despite the increase in RMSE (11.96). As the value of α keeps increasing, the model error improves further, while the model tends to overestimate ozone concentrations.

Table 4 shows the prediction performance of three deep learning models and the LOTOS-EUROS, where station DEBB029 is selected for illustration in this section and the other stations will be further evaluated in Section 4.3.2. The weight term α is decided by cross-validation for different models. In terms of RMSE and R^2 , the performance of the three ML models is similar and is slightly better than LOTOS-EUROS. However, in terms of prediction accuracy of ozone exceedances, espe-

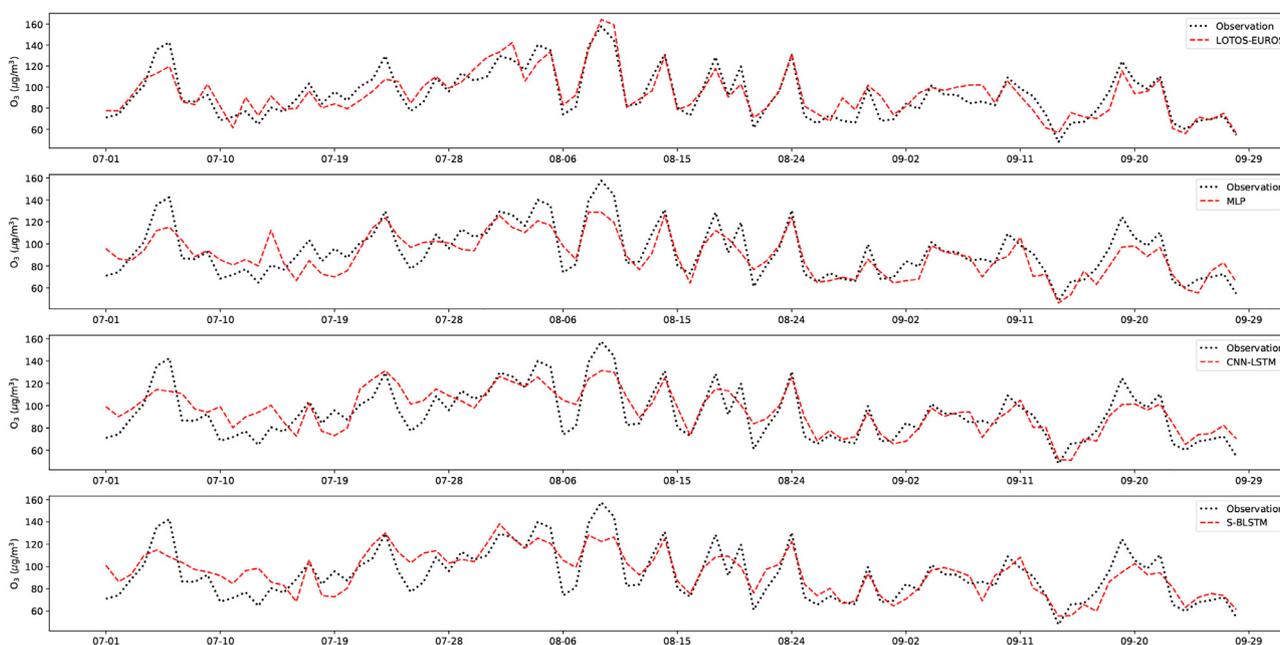


Fig. 10. Prediction of ozone concentrations in the summer of 2018 by different models.

Table 4

Result of different models on station DEBB029.

Model	RMSE	R^2	PA(> 100)	PA(> 120)
MLP	11.96	0.83	77%	69%
CNN-LSTM	11.48	0.84	81%	61%
S-BLSTM	11.23	0.85	75%	65%
LOTOS-EUROS	12.6	0.81	81%	74%

cially for ozone exceedances over $120 \mu\text{g}/\text{m}^3$, all machine learning models are much worse than LOTOS-EUROS. It suggests that 1-station machine learning models can predict overall trends in MDA8 but have difficulty predicting extreme cases, such as ozone exceedances. Figure 10 shows the predicted results for the summer (July-August-September) of station DEBB029. Overall, LOTS-EUROS performs the best in predicting high-level ozone concentration, while all three machine learning models tend to underestimate ozone concentrations in the summer. However, none of the models mentioned above were able to predict the peak ozone occurred in early July. Compared to general conditions (low/medium ozone), the number of high ozone days (especially for exceedances) is much lower (Ma et al., 2022). It makes it challenging to learn the relationship between ozone and meteorological conditions.

In our experiments, MLP and LSTM models performed similarly in ozone prediction. Unlike general time series forecasting problems, the input data in our experiments include future meteorological data, making the air pollution data and meteorological data at each time step not correspond perfectly, which adds difficulty to the training of LSTM. In addition, the data analysis in Section 2.3 shows that there is no obvious temporal pattern in the short-term MDA8, which makes it difficult for the model to find the temporal correlation in the time series. It prevents LSTM from exerting its greatest advantage. Besides, the MLP converges faster and requires less training time. Therefore, we use MLP instead of LSTM in the later experiments.

4.2.2. Feature importance

The feature importance of all input is ranked via permutation importance (Galkin et al., 2018). When a single feature value is randomly shuffled, the importance values are calculated by the decrease in a model score (RMSE in our case). A larger decline in the score indicates that the

feature is more important to the output. In our research, we consider the input features at each time step separately to detect the temporal dependence in the forecast model. Our input features, according to Eq. (7), contain two parts: (1) meteorological data (T,SWR,RH,RAIN,WD,WS) in the future time step ($t+1$); (2) meteorological and air pollution data (O_3, NO_x) in the historical time step ($t, t-1, t-2, \dots$) (Fig. 11).

The profiles of feature importance are similar but feature ranks can differ for the ten stations tested (see Table 2). Temperature $T(t+1)$, shortwave radiation $SWR(t+1)$, and relative humidity $RH(t+1)$ for the future time step, and ozone $O_3(t)$ for the current time step, are rated in the top five for all ten stations. Regarding the sensitivity of data at different time steps, the latest data have the most significant impact on the future ozone concentration. The meteorological features of the coming time steps and the latest ozone data are critical.

In terms of different features, T, SWR, and RH have more impact on ozone concentration (Mao et al., 2020). Wind speed and direction data will influence future ozone in our model (Wang et al., 2017). However, they are significantly less critical than meteorological features such as the temperature since we currently consider only individual stations and do not introduce spatial correlation. Rain data have almost no effect on ozone prediction due to the small fluctuations and the inclusion of relative humidity as input. Although NO_x is a vital ozone precursor, historical NO_x data can hardly affect the model's predictive performance. It could be due to the rapidly changing anthropogenic NO_x emissions (Bae et al., 2020).

4.3. Spatial transfer learning model

4.3.1. Multi-station model

In this section, we trained a multi-station 1-day ozone forecast model using machine learning methods. Training with multiple stations can introduce more data into the model and avoid the problem of underfitting. It also gives the model additional information about the spatial correlation between different stations. To reduce the bias introduced by station differences, we use the data from all stations in Cluster 1 to train a general MLP model first. A total of 59 stations are included in Cluster 1, where each station contains 1826 sets of daily data. We use the data from 2014 to 2017 of all stations as the training and validation set and the 2018 data as test data. To evaluate the applicability of the model, we selected data from different stations as the test set.

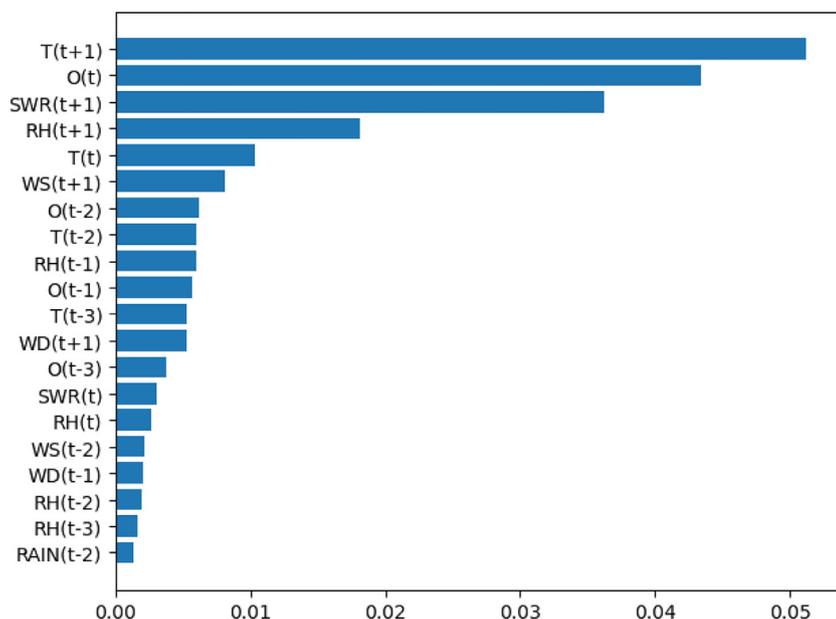


Fig. 11. Top 20 most important features of station DEBB029.

Table 5

Error analysis of results from 1S-MLP, RF, CNN-LSTM, MS-MLP, SPTL-MLP and LOTOS-EUROS. The observations in 2018 are taken as references (True values).

Criteria	Methods	Station Code										Mean
		DEBB007	DEBB021	DEBB029	DEBB032	DEBB053	DEBB055	DEBB064	DEBB065	DEBB066	DEBB067	
RMSE	1S-MLP	13.07	12.51	11.96	12.4	12.9	12.87	13.11	13.44	12.52	12.34	12.71
	RF	12.25	12.49	10.95	11.74	11.84	11.86	12.43	11.89	11.54	11.79	11.88
	CNN-LSTM	13.29	12.51	11.48	12.64	13.25	12.64	12.36	12.39	11.72	12.02	12.43
	MS-MLP	11.87	11.64	10.93	11.39	11.52	11.28	11.76	12.49	11.13	11.65	11.57
	SPTL-MLP	11.59	12.35	11.43	11.58	11.48	11.33	11.86	11.69	10.99	11.21	11.55
	LOTOS-EUROS	14.06	12.73	12.6	12.61	13.52	12.44	13.17	13.23	12.73	13.18	13.03
R ²	1S-MLP	0.85	0.86	0.83	0.85	0.81	0.82	0.83	0.86	0.80	0.85	0.84
	RF	0.87	0.86	0.86	0.86	0.84	0.85	0.84	0.88	0.86	0.86	0.86
	CNN-LSTM	0.84	0.86	0.84	0.84	0.80	0.83	0.84	0.87	0.85	0.86	0.84
	MS-MLP	0.88	0.88	0.86	0.87	0.85	0.86	0.85	0.87	0.85	0.86	0.86
	SPTL-MLP	0.88	0.85	0.85	0.87	0.85	0.86	0.85	0.88	0.87	0.87	0.86
	LOTOS-EUROS	0.83	0.85	0.81	0.84	0.80	0.84	0.82	0.85	0.82	0.83	0.83
PA (100) ^a	1S-MLP	85%	75%	77%	88%	80%	78%	78%	80%	70%	80%	79%
	RF	80%	68%	73%	79%	71%	76%	78%	78%	70%	73%	74%
	CNN-LSTM	86%	70%	81%	88%	81%	82%	73%	80%	75%	85%	80%
	MS-MLP	75%	72%	74%	69%	73%	71%	69%	73%	73%	72%	72%
	SPTL-MLP	88%	85%	83%	88%	85%	86%	85%	86%	85%	86%	86%
	LOTOS-EUROS	84%	86%	81%	81%	78%	79%	82%	81%	84%	80%	82%
PA (120) ^b	1S-MLP	77%	65%	69%	81%	63%	67%	69%	62%	65%	69%	69%
	RF	57%	65%	69%	53%	75%	67%	49%	62%	48%	50%	60%
	CNN-LSTM	79%	50%	65%	86%	75%	66%	63%	62%	61%	70%	68%
	MS-MLP	60%	62%	70%	75%	79%	62%	63%	62%	77%	58%	67%
	SPTL-MLP	80%	83%	82%	84%	83%	76%	83%	76%	81%	75%	80%
	LOTOS-EUROS	65%	81%	74%	63%	87%	76%	69%	57%	71%	73%	72%

^aPrediction accuracy of ozone exceedances with 100 $\mu\text{g}/\text{m}^3$ threshold; ^bPrediction accuracy of ozone exceedances with 120 $\mu\text{g}/\text{m}^3$ threshold

There are some hyperparameters about neural networks that need to be determined before model fitting. We still set the length of the input time window to 5. For each set of inputs, we use all eight elements as features. MDA8 for the next day is used as the output. In different experiments, the number of layers varies from 1 to 5 and the number of neurons in each layer is set to 50, 100, 150, and 200, respectively. Based on the results of the 5-fold cross-validation, we choose a 4-layer neural network with 50 neurons per layer as the configuration of our neural network. Without loss of generality, we set $\alpha=0$, i.e., the regular MSE loss function. We repeat the experiment 10 times and calculate the average of the predicted results. The result performance of different models for the ten stations is shown in Table 5.

We conclude that the trained multi-station model (MS-MLP) is capable of predicting the ozone concentrations for different stations in the same cluster. The MS-MLP has smaller errors compared with the 1-

station model (1S-MLP). The RMSE of the MS-MLP is smaller than that of the 1S-MLP at all tested stations, and it is able to capture the ozone trend better (larger R^2). However, for the prediction of ozone exceedances at individual stations, MS-MLP is not only much inferior to LOTOS-EUROS but also worse than the 1S-MLP model. Such as discussed earlier, the extreme events (i.e., ozone exceedances) are hard to predict. Training the neural network using a larger data set from multiple stations seems to strengthen the behavior of trying to optimize for the average cases since extreme events have an even lower share in the total input data when data from multiple stations are combined.

4.3.2. Spatial transfer learning model

In this section, we use network-based transfer learning for spatial transfer to train models applicable to individual stations. The multi-station model is used as the base model. Ten stations selected in Cluster

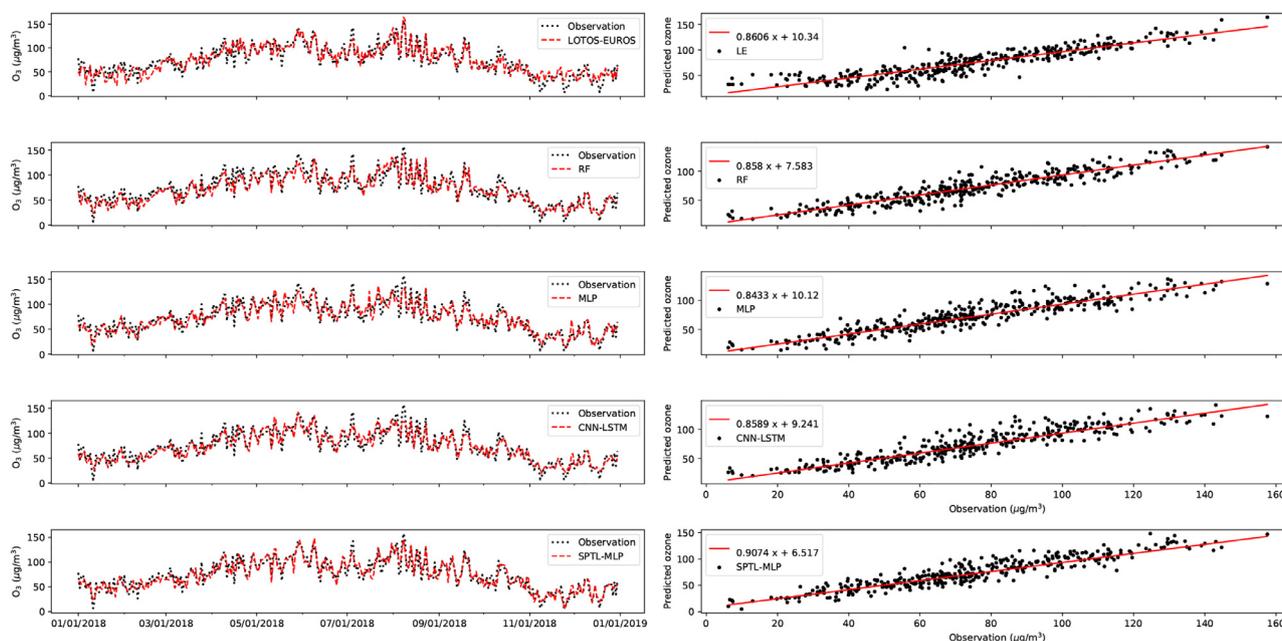


Fig. 12. A comparison of the temporal variation of MDA8 from the SPTL model and the observation in 2018 at station DEBB029.

1 are used to test the effect of spatial transfer. The data from 2014 to 2017 is used to train the model, and the 2018 data is used for the test. We will conduct 1-day and 3-day MDA8 predictions for each station, respectively. To evaluate the prediction performance of our transfer learning model, we compare our model with two different machine learning models that are used in the latest ozone prediction studies: Random Forest (Zhan et al., 2018) and CNN-LSTM (Pak et al., 2018). The prediction performance of S-BLSTM (Ma et al., 2020) is similar to that of CNN-LSTM, so the results are not included. All machine learning models are compared with the 1-day forecast result of LOTOS-EUROS.

Different local models that are applicable to different stations are trained through spatial transfer. In the training process of the Spatial Transfer Multilayer Perceptron (SPTL-MLP) model, one new parameter is the number of frozen layers extracted from the base model. It determines how many layers of the base model need to be retained. In our experiment, the number of frozen layers is set from 1 to 4 and up to 4 MLP layers are added to the new model after frozen layers; each of them contains 100 neurons. The parameters of the model, including the weight term of the loss function, are fine-tuned per station. Table 5 shows the performance of the 1-day forecast in the ten stations by different models. Besides, although the geographic characteristics were not considered as an input in our model, retraining the models with specific meteorological features at the target stations implicitly incorporates the geographical differences of the stations. It ensures that the prediction accuracy will not be significantly affected by geographical characteristics (e.g., altitude). It is noticed that the model performances are not affected by the original station type (Urban, Suburban and Rural, see Table 2).

In terms of RMSE and R^2 , all machine learning models perform better compared to LOTOS-EUROS, and MS-MLP and SPTL-MLP perform the best. Most machine learning models perform poorly for the prediction accuracy of ozone exceedances due to the insufficient high ozone data. Compared to 1S-MLP, SPTL-MLP has a much smaller error (reduced by 9.13%) and higher prediction accuracies (improved by 8.21% and 16.9%) based on the mean values of ten stations. It implies that spatial transfer learning can retain additional information from the multi-station model and transfer it to the target model, which could significantly improve the prediction performance of MDA8. Meanwhile, retraining based on data from the target station ensures that the transfer learning model can improve the prediction skills for the local problem.

Generally, SPTL-MLP outperforms all other models, including the traditional CTM. It indicates that SPTL-MLP can be used for MDA8 prediction in German (see Appendix A) to make timely warnings of ozone exceedances.

Figure 12 shows the detailed predictions for station DEBB029. The temporal variation of MDA8 during the 2018 predictive period is shown on the left, and the scatter plots between observation and prediction of MDA8 are illustrated on the right. It can be seen that the predicted ozone concentrations of all machine learning approaches are in relative good agreement with the observations. Compared to LOTOS-EUROS, the machine learning approaches are more accurate in predicting ozone trends for the beginning and end of 2018, while they tend to underestimate ozone peaks in the middle of the year (especially in the summer). The poor performance of LOTOS-EUROS in the winter may be due to the increased uncertainty of the boundary conditions, which do not affect machine learning models. Among all machine learning models, the prediction result from SPTL-MLP exhibits better agreement with the observation, while it can also capture more ozone peaks during the summer. It should be noticed that although machine learning models tend to underestimate peak ozone concentrations, there are some cases of overestimation of the high ozone concentrations in the summer. Machine learning models have difficulty in responding on time to the sudden drop in ozone concentrations following a peak ozone occurring in summer, resulting in a potential overestimation in the next few days. For all machine learning prediction methods, it can be seen that overestimation occurs after the peak ozone at the end of July.

The same framework can also be used to predict MDA8 for the next three days. First, data of all stations in Cluster 1 are used to train the 3-day forecast multi-station model, which is used as the base model for spatial transfer learning. Then, we retrain the model with spatial transfer for each station and adjust the model parameters by cross-validation. Since we do not have the three-day prediction results for LOTOS-EUROS, we only show the result of machine learning models. The average of all stations is shown in Table 6. We can find that the model performance of SPTL-MLP is better than others, especially for the prediction accuracy of ozone exceedances (81% and 75%). It suggests that the spatial transfer can also be used for 3-day ozone concentration predictions, although the result is not as accurate as the 1-day forecast model.

Table 6
3-day forecast result of different models on ten stations.

Model	RMSE	R ²	PA(> 100)	PA(> 120)
MLP	13.08	0.82	72%	63%
CNN-LSTM	12.95	0.81	70%	55%
RF	12.71	0.83	65%	49%
SPTL-MLP	12.46	0.84	81%	75%

5. Conclusion

This paper proposes a methodology framework with spatial transfer learning to predict daily ozone concentration (MDA8) for the next three days. The data from 2014 to 2018 from stations in Germany are used to test the models. The main contribution of this study can be summarized as follows: (1) the multi-station model based on clustering is used as a base model, which will provide additional information for the target station while reducing the bias caused by differences between stations. (2) the spatial transfer learning models can improve the prediction performance of MDA8 for the next three days. With a proper weighted loss function, SPTL-MLP can get accurate results for the prediction accuracy of ozone exceedances, which is difficult for both CTMs and ML models. (3) Compared to traditional chemical transport models, the transfer learning model is not limited by grid resolution. Given the base model, the model training and prediction require extremely little time (less than one minute), making it simple to update for varied scenarios. Although the effect of substantial emission changes and special events (e.g., traffic reduction due to COVID-19 lockdown) are not considered in our model, regular retraining could be made to bring the model forward to the more recent regime.

To improve the prediction performance of MDA8 for the local problem, we trained the 1-station model by spatial transfer based on the multi-station model, which is trained by all stations in the same cluster. The SPTL-MLP model has smaller errors than LOTOS-EUROS and other machine learning models (including 1-Station MLP). It indicates that spatial transfer learning can retain the information from the multi-station model (information from other similar stations) and transfer it to the 1-station model. With the weighted loss function, the SPTL-MLP is able to predict ozone exceedances accurately. The proposed methodology framework could be used for MDA8 prediction in German, including

newly established air quality monitoring stations with limited data (see Appendix B), to help government and policymakers to detect ozone pollution.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the German Environment Agency UBA for making the air pollution data available which are used in the case study for training the neural network models. The author Hai Xiang Lin thanks the Timman foundation for the partial financial support in doing this work.

Appendix A. A case study of cities in other clusters

To show the generalisability of proposed methods, cities located in different clusters were selected to test the prediction performance. In this section, we selected four cities to evaluate the model: Berlin (Cluster 1), Hamburg (Cluster 2), Munchen (Cluster 3), and Koln (Cluster 5). Based on the results of clustering (Section 4.1), we first trained four different multi-station (MS) models for Cluster 1, 2, 3, and 5, respectively. The parameter settings are the same as in Section 4.3.1. For the four tested cities, four different spatial transfer learning multilayer perceptrons (SPTL-MLP) were trained based on different MS models. 1-station MLPs were trained separately to investigate the improvement of transfer learning (see Section 4.2.1). The data from 2014 to 2017 is used to train the model, and the 2018 data is used for the test.

Figures A.1, A.2, A.3, A.4 shows the prediction performances of transfer learning models for different cities. The predicted ozone concentrations of both machine learning approaches are in relative good agreement with the observations, while the SPTL-MLP can better capture the ozone peaks in summer. It suggests that the proposed clustering-based spatial transfer learning model could be used to improve the prediction performance of MODA8 O₃ in whole German.

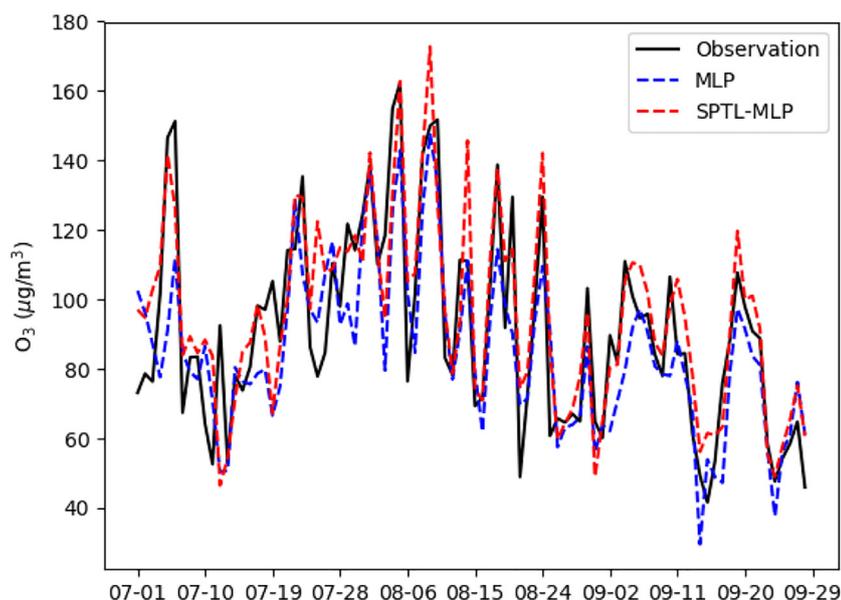


Fig. A.1. Berlin (Cluster 1).

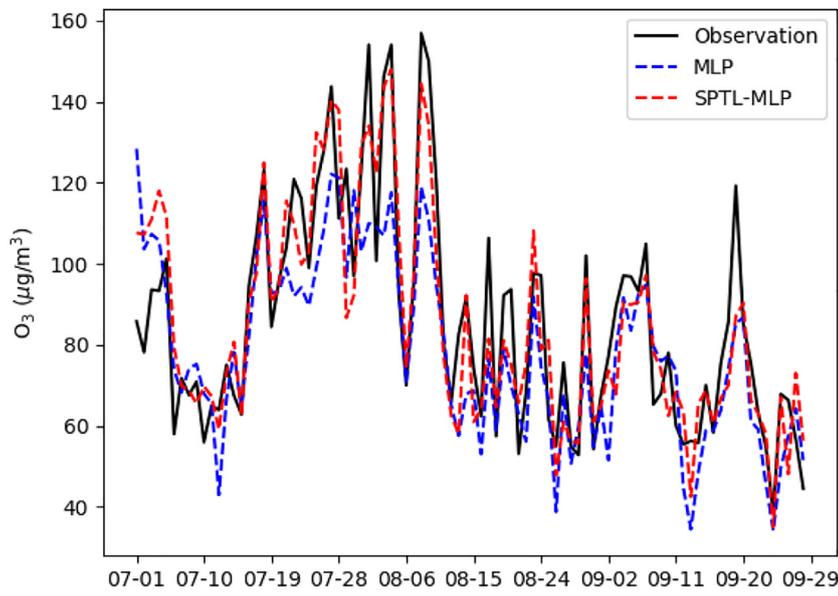


Fig. A.2. Hamburg (Cluster 2).

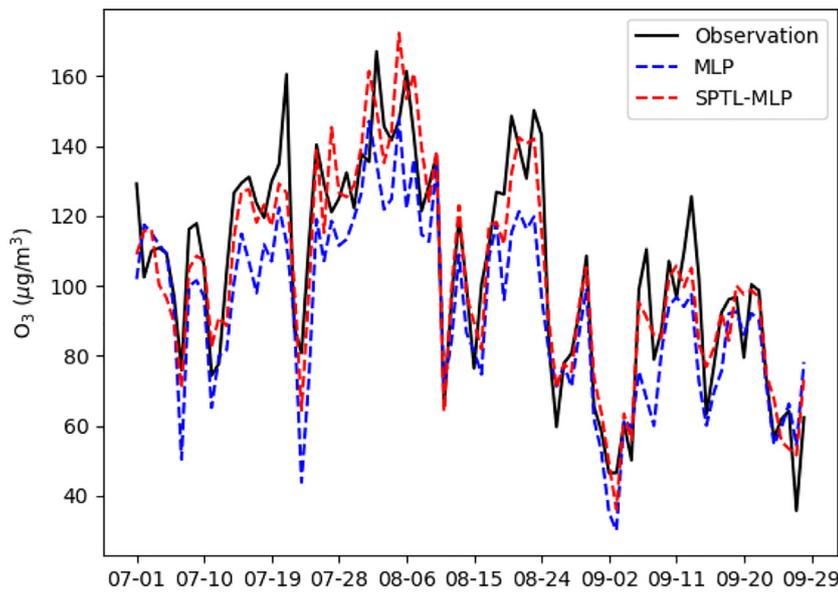


Fig. A.3. Munchen (Cluster 3).

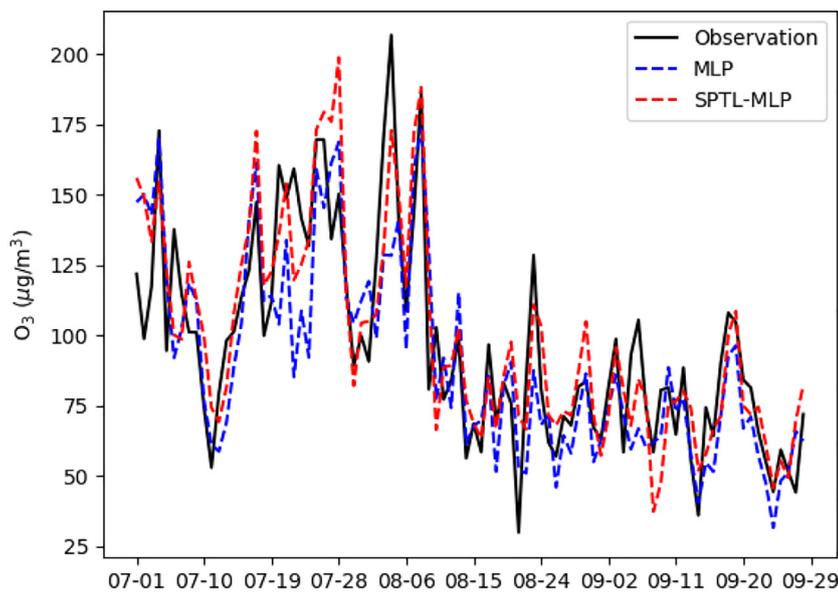


Fig. A.4. Koln (Cluster 5).

Appendix B. Impact of the time length of input data

In previous studies, spatial transfer learning has been used for ozone prediction, which aims primarily to improve model performance for new stations with insufficient data (Ma et al., 2020). The input of their model contains only ozone data and is only used for one-hour ozone forecasts. Due to the differences in model application, input parameters, and forecast length, we did not compare the two transfer learning methods. However, it is mentioned in the study that the time length of the input data has an impact on the effect of transfer learning. For the target station, the influence of transfer learning decreases as the time length of the input data increases. In our research, we also analyze the impact of the size of input data on transfer learning. Table B.1 shows the results of MLP and SPTL-MLP with different amounts of training data, where 2014/1/1 is the starting date, while the different end dates make the lengths of the input data different.

Table B.1
RMSE of different time lengths on transfer learning at station DEBB029.

Cutoff point (start: 2014/1/1)	2014-7-1	2015-1-1	2016-1-1	2018-1-1
MLP	16.88	14.5	13.15	11.96
SPTL-MLP	15.10	13.19	12.04	11.43
Improvement	10.55%	9.03%	5.70%	4.43%

We can find that the improvement of transfer learning for MLP gradually decreases as the length of the input data increases. The improvement of spatial transfer is the largest (10.55%) with six months of input data and the lowest (4.43%) with four years of input data. It suggests that our spatial transfer learning model can be utilized to improve short-term ozone forecast accuracy for newly established air quality monitoring stations with little observed data. However, to ensure the prediction performance of ozone exceedances, we still need a certain length of historical data to train the model due to the insufficient high-level ozone data.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al., 2015. TensorFlow: large-scale machine learning on heterogeneous systems.

Bae, C., Kim, H.C., Kim, B.-U., Kim, S., 2020. Surface ozone response to satellite-constrained NO_x emission adjustments and its implications. *Environ. Pollut.* 258, 113469.

Bholowalia, P., Kumar, A., 2014. EBK-means: a clustering technique based on elbow method and k-means in WSN. *Int. J. Comput. Appl.* 105 (9).

Bollmeyer, C., Keller, J., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., Scheidt, L., et al., 2015. Towards a high-resolution regional reanalysis for the European CORDEX domain. *Q. J. R. Meteorol. Soc.* 141 (686), 1–15.

Bontempi, G., Taieb, S.B., Le Borgne, Y.-A., 2012. Machine learning strategies for time series forecasting. In: *European Business Intelligence Summer School*. Springer, pp. 62–77.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 7 (3), 1247–1250.

Cheng, Y., He, L.-Y., Huang, X.-F., 2021. Development of a high-performance machine learning model to predict ground ozone pollution in typical cities of China. *J. Environ. Manage.* 299, 113670.

Curier, R., Timmermans, R., Calabretta-Jongen, S., Eskes, H., Segers, A., Swart, D., Schaap, M., 2012. Improving ozone forecasts over Europe by synergistic use of the LOTOS-EUROS chemical transport model and in-situ measurements. *Atmos. Environ.* 60, 217–226.

Fang, Y., Naik, V., Horowitz, L., Mauzerall, D.L., 2013. Air pollution and associated human mortality: the role of air pollutant emissions, climate change and methane concentration increases from the preindustrial period to present. *Atmos. Chem. Phys.* 13 (3), 1377–1394.

Feng, R., Zheng, H.-j., Gao, H., Zhang, A.-r., Huang, C., Zhang, J.-x., Luo, K., Fan, J.-r., 2019. Recurrent neural network and random forest for analysis and accurate forecast of atmospheric pollutants: a case study in hangzhou, china. *J. Clean. Prod.* 231, 1005–1015.

Finlayson-Pitts, B., Pitts Jr., J., 1993. Atmospheric chemistry of tropospheric ozone formation: scientific and regulatory implications. *Air Waste* 43 (8), 1091–1100.

Fong, I.H., Li, T., Fong, S., Wong, R.K., Tallon-Ballesteros, A.J., 2020. Predicting concentration levels of air pollutants by transfer learning and recurrent neural network. *Knowl. Based Syst.* 192, 105622.

Franklin, J., 2005. The elements of statistical learning: data mining, inference and prediction. *Math. Intell.* 27 (2), 83–85.

Galkin, F., Aliper, A., Putin, E., Kuznetsov, I., Gladyshev, V.N., Zhavoronkov, A., 2018. Human microbiome aging clocks based on deep learning and tandem of permutation feature importance and accumulated local effects. *BioRxiv* 507780.

Gardner, M.W., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos. Environ.* 32 (14–15), 2627–2636.

Gong, B., Ordieres-Meré, J., 2016. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: case study of hong kong. *Environ. Model. Softw.* 84, 290–303.

Gulli, A., Pal, S., 2017. Deep Learning with Keras. Packt Publishing Ltd.

Hu, X.-M., Xue, M., Kong, F., Zhang, H., 2019. Meteorological conditions during an ozone episode in dallas-fort worth, texas, and impact of their modeling uncertainties on air quality prediction. *J. Geophys. Res. Atmos.* 124 (4), 1941–1961.

Jafari, M., Li, R., Xing, Y., Auer, D., Francis, S., Garibaldi, J., Chen, X., 2019. FU-Net: multi-class image segmentation using feedback weighted U-Net. In: *International Conference on Image and Graphics*. Springer, pp. 529–537.

Lu, H., Lyu, X., Cheng, H., Ling, Z., Guo, H., 2019. Overview on the spatial-temporal characteristics of the ozone formation regime in China. *Environ. Sci. Process. Impacts* 21 (6), 916–929.

Ma, J., Cheng, J.C., Lin, C., Tan, Y., Zhang, J., 2019. Improving air quality prediction accuracy at large temporal resolutions using deep learning and transfer learning techniques. *Atmos. Environ.* 214, 116885.

Ma, J., Li, Z., Cheng, J.C., Ding, Y., Lin, C., Xu, Z., 2020. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Sci. Total Environ.* 705, 135771.

Ma, W., Yuan, Z., Lau, A.K., Wang, L., Liao, C., Zhang, Y., 2022. Optimized neural network for daily-scale ozone prediction based on transfer learning. *Sci. Total Environ.* 827, 154279.

Manders, A.M., Builtjes, P.J., Curier, L., Denier van der Gon, H.A., Hendriks, C., Jonkers, S., Kranenburg, R., Kuenen, J.J., Segers, A.J., Timmermans, R., et al., 2017. Curriculum vitae of the LOTOS-EUROS (v2. 0) chemistry transport model. *Geosci. Model Dev.* 10 (11), 4145–4173.

Mao, J., Wang, L., Lu, C., Liu, J., Li, M., Tang, G., Ji, D., Zhang, N., Wang, Y., 2020. Meteorological mechanism for a large-scale persistent severe ozone pollution event over Eastern China in 2017. *J. Environ. Sci.* 92, 187–199.

Otero, N., Sillmann, J., Mar, K.A., Rust, H.W., Solberg, S., Andersson, C., Engardt, M., Bergström, R., Bessagnet, B., Colette, A., et al., 2018. A multi-model comparison of meteorological drivers of surface ozone over Europe. *Atmos. Chem. Phys.* 18 (16), 12269–12288.

Pak, U., Kim, C., Ryu, U., Sok, K., Pak, S., 2018. A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Qual. Atmos. Health* 11 (8), 883–895.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Rokach, L., Maimon, O., 2005. Clustering methods. In: *Data Mining and Knowledge Discovery handbook*. Springer, pp. 321–352.

Ryu, Y.-H., Hodzic, A., Descombes, G., Hu, M., Barré, J., 2019. Toward a better regional ozone forecast over conus using rapid data assimilation of clouds and meteorology in WRF-chem. *J. Geophys. Res. Atmos.* 124 (23), 13576–13592.

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., Jung, J., 2020. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 h in advance. *Neural Netw.* 121, 396–408.

Sharma, S., Sharma, P., Khare, M., 2017. Photo-chemical transport modelling of tropospheric ozone: a review. *Atmos. Environ.* 159, 34–54.

Sicard, P., Khaniabadi, Y.O., Perez, S., Gualtieri, M., De Marco, A., 2019. Effect of O₃, PM₁₀ and PM_{2.5} on cardiovascular and respiratory diseases in cities of France, Iran and Italy. *Environ. Sci. Pollut. Res.* 26 (31), 32645–32665.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., 2018. A survey on deep transfer learning. In: *International Conference on Artificial Neural Networks*. Springer, pp. 270–279.

Torrey, L., Shavlik, J., 2010. Transfer learning. In: *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI Global, pp. 242–264.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L., 2017. Ozone pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total Environ.* 575, 1582–1596.

Wang, W., Liu, X., Bi, J., Liu, Y., 2022. A machine learning model to estimate ground-level ozone concentrations in california using TROPOMI data and high-resolution meteorology. *Environ. Int.* 158, 106917.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. *J. Big Data* 3 (1), 1–40.

Williamson, C.E., Neale, P.J., Hylander, S., Rose, K.C., Figueroa, F.L., Robinson, S.A., Häder, D.-P., Wängberg, S.-Å., Worrest, R.C., 2019. The interactive effects of stratospheric ozone depletion, UV radiation, and climate change on aquatic ecosystems. *Photochem. Photobiol. Sci.* 18 (3), 717–746.

Zhan, Y., Luo, Y., Deng, X., Grieneisen, M.L., Zhang, M., Di, B., 2018. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environ. Pollut.* 233, 464–473.