

# 深層学習によるPM<sub>2.5</sub>濃度の予測モデル構築と精度評価

Construction and Accuracy Evaluation of a Deep Learning Model  
for PM<sub>2.5</sub> Concentration Prediction

桐山 悠祐\*、飯野 翔太\*\*、柴崎 みはる\*\*、平野 瑞歩\*\*、吉村 有史\*\*

Yusuke Kiriyama\*, Iino Shota\*\*, Miharu Shibasaki\*\*, Mizuho Hirano\*\*, Yuji Yoshimura\*\*

## 【要約】

わが国での予報手法の拡充と高精度化の第一歩として本研究ではACSA-14により得られた全国10地点のPM<sub>2.5</sub>とその構成成分の連続観測データを用いた深層学習モデルを構築し、PM<sub>2.5</sub>濃度の予測とその精度評価を行った。深層学習モデルのベースとして、ニューラルネットワークの一種のLong Short-Term Memory (LSTM)を採用して基本構造を構築し、データ処理として単独の高濃度と統計的な外れ値を除くパターン、単独高濃度のみを除くパターンの2通りのデータ処理を通じてデータセットを作成した。その結果、予測開始直後は比較的良くPM<sub>2.5</sub>濃度を予測するものの、時間が経つにつれその精度が下がる傾向が見られた。また、一定濃度より高い濃度を予測しない挙動が見られ、改良を要した。この問題に対し、気象データおよびタイムラグを持たせたデータを追加し、モデル中で受け渡すデータを正規化するためのBatch normalization層をモデルに追加することでモデルが改良される結果を得た。

キーワード：深層学習、濃度予測、PM<sub>2.5</sub>

## 1. はじめに

2013年冬季に中国北京での微小粒子状物質(PM<sub>2.5</sub>)による重大な大気汚染が生じた際、越境輸送によるわが国への影響が全国的に大きな関心事として取り上げられた。健康に悪影響を及ぼすほどの大気汚染現象が生じうる場合、屋外での活動を控えるなどの回避策が必要になり、そのための情報として、大気汚染物質濃度の予報は重要である。

わが国ではこの問題に対してPM<sub>2.5</sub>についての注

意喚起の暫定指針が示された。この手法は非常に簡易であり、誰でも実施できるという大きなメリットがある一方で、誤った判定になる場合が多い。また、国立環境研究所が運用するVisual atmospheric Environment Utility System(VENUS)と呼ばれる大気質予報モデル(菅田ら, 2011<sup>1</sup>)や九州大学等により開発、運用されているSPRINTARS(竹村, 2009<sup>2</sup>)をはじめ、シミュレーションモデルを用いた、天気予報と類似した予報システムがある。大気汚染の科学的知見の蓄積や現象解明、成分間の関連性の解析や科学的知見に基づいた情報提供など、シミュレーションモデルが担う役割は大気汚染研究を進める上で非常に重要である。しかし、このような予報システムは計算負荷が非常に大きく、大型計算機を用いた場合でも予報計算に長時間(VENUSでは全工程で7.5時間)かかることや基礎となるモデルに改善すべき点が未だ多いなど、課題や問題が多い。

以上のようにわが国ではいくつかの予報手段が運用されているが、多くの課題が残されている。一方で、わが国では2017年4月から全国10地点で

\* 一般財団法人 日本環境衛生センター アジア大気汚染研究センター 情報管理部、大気圏研究部

Data Management Dept. and Atmospheric Research Dept., Asia Center for Air Pollution Research (ACAP), Japan Environmental Sanitation Center (JESC)

\*\* 一般財団法人 日本環境衛生センター 東日本支局 環境事業本部 環境事業第二部 Environmental Sciences Dept., Environmental Business Group, East Branch Office, Japan Environmental Sanitation Center (JESC)

PM<sub>2.5</sub>の成分自動測定機が稼働しており、高時間分解能でPM<sub>2.5</sub>の成分濃度を測定している。このようなデータを有効に活用する方法の一つとして、深層学習(ここでは多層のニューラルネットワークで構成されたモデルの学習、構築とする)という、統計的な手法が挙げられ、国外では大気汚染研究への応用事例も近年多く発表されている(例えばFeng et al., 2015<sup>3</sup>, Catalano et al., 2016<sup>4</sup>, Kim et al., 2019<sup>5</sup>)。このような手法は、わが国の常時監視網のデータを用いても同様の効果が期待できる。しかし、PM<sub>2.5</sub>の構成成分を用いてモデル構築した例は見られず、わが国のように多地点で高時間分解能のPM<sub>2.5</sub>成分測定を長期的に実施している例もみられない。深層学習を用いたモデル(以下、深層学習モデル)の学習には多くの計算リソースが必要になるが、学習済みの深層学習モデルの運用は比較的簡易であることもメリットである。

本研究では、わが国での予報手法の拡充と高精度化を目的に、その第一歩としてPM<sub>2.5</sub>の構成成分の連続観測データを用いた深層学習モデルを構築し、PM<sub>2.5</sub>濃度の予測とその精度評価を行った。

## 2. 方法

### 2.1 観測データおよびデータ処理

本研究では、環境省が実施している全国10地点でのPM<sub>2.5</sub>成分自動測定機による観測データの2017年度、2018年度および2019年度の確定値([http://www.env.go.jp/air/%20osen/pm\\_result\\_monitoring/post\\_25.html](http://www.env.go.jp/air/%20osen/pm_result_monitoring/post_25.html), last access 2019/12/20)の内、大気エアロゾル化学成分連続自動分析装置(ACSA-14, KIMOTO ELECTRIC Co., Ltd.)の全10地点のデータを使用してモデルの学習を行った。環境省が公開しているACSA-14のデータにはE、M、C、H、H1のフラグと呼ばれる記号が適宜付与されている(環境省、2019<sup>6</sup>)。このうち、E、M、Cはそれぞれ測定エラー、機器メンテナンス、機器校正を表し、事前にデータが削除されている。残るH、H1、は高濃度(各年度の平均値+5σを超過)、および単独高濃度(Hフラグ付与の対象、かつ前後のデ

ータより10倍以上濃度が高い)を示す。本研究では、H1フラグが付与されたデータには、突発的な機器的エラーや特異的なイベントによる影響が含まれているとみなし、除外する方針とした。また、本研究では、データ処理による予測精度の変化を検証するために、1. H1と外れ値を削除(以下、処理①)、2. H1のみ削除(以下、処理②)の2パターンを検証した。ここで、外れ値はデータの四分位幅の1.5倍を超えるデータを指す。

入力データはPM<sub>2.5</sub>質量濃度および微小粒子として分級された硫酸塩、硝酸塩、水溶性有機炭素および黒色炭素の各濃度を用いた。毎日23時に自動的に行われる濃度校正による欠測は、22時と24時のデータが存在する場合は線形内挿する、という条件で連続データを作成した。また、モデルの学習及び検証計算にデータを投入する際、(1)式で示す正規化処理を成分ごとに施した。

$$z_{a,t,i} = \frac{x_{a,t,i} - \mu_{a,i}}{\sigma_{a,i}} \quad (1)$$

ここで、aは地点、tは時刻、iはPM<sub>2.5</sub>および成分濃度、 $z_{a,t,i}$ は地点a、成分iの時刻tにおける正規化後データ、 $\mu_{a,i}$ は地点a、成分iの平均値、 $\sigma_{a,i}$ は地点a、成分iの標準偏差を示す。

以上の処理を施したデータからモデル構築に用いるデータセットを作成した。本研究では、学習データ72時間分、正解データ72時間分を一組とするデータセットを作成した。データセット一組に欠測がある場合、その組は除外した。処理①を施した場合、計7502セット、処理②を施した場合、計80058セットがモデル構築に利用可能であった。データセットは、学習用と検証用を9:1に分割した。

さらに、上記のデータとは別にACSA-14による観測データの2020年度4月から12月までの速報値を用いて、モデル構築後の実用性を評価した。その際、新潟巻局の2020/5/3 9:00および大阪局の2020/8/4 10:00のPM<sub>2.5</sub>濃度を異常高濃度と判断して削除したうえで、上記と同様の処理を施した。

### 2.2 モデル構築環境およびデータ処理環境

本研究では、モデル構築およびデータ処理にはPython version 3.7.12を利用した。また、主要な

モジュール類を表 1 に示す。

表 1 本研究で用いた Python モジュール

モジュール名, バージョン	利用目的
Pandas version 1.1.5	入力データの処理
Numpy version 1.19.5	
Tensorflow Version 2.7.0	深層学習モデルの構築 (TensorflowはKerasをバック エンドとして用いた)
Scikit-learn Version 1.0.2	

### 2. 3 深層学習モデル

以下に、本研究で用いたモデルについて述べる。

本研究で採用する深層学習は、人工ニューロンと呼ばれる、統計処理を行うためのユニットを複数接続してネットワーク化し(ニューラルネットワーク)、それらを多数接続することで複雑なデータの特徴や関連性を抽出し、処理することを可能にする手法である。本研究では、再帰的ニューラルネットワークと呼ばれる、連続データを処理することに長けたニューラルネットワークのうち、Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997<sup>7</sup>) を基本構造として採用した。構築したモデルの概要を図 1 に示す。投入する学習データを時間軸として順方向に認識させる (Forward LSTM) 以外に、同時に逆方向からのデータも認識させる (Backward LSTM) ことで、過去、未来の両面からのデータの特徴抽出を試みた。ForwardとBackwardの両者へ学習データを投入し、時系列データの特徴量を抽出して、それぞれから出力された特徴量を結合した。その後、結合した特徴量と各層の重みを次段のLSTMへ渡して予測結果を出力する構成とした。

深層学習では、モデルを学習する過程で重要なパラメータは自動的に調整されるが、いくつかのパラメータ(ハイパーパラメータ)については、基本的にはユーザーが設定する必要がある、これら

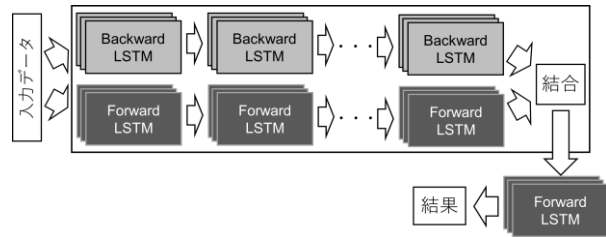


図 1 本研究で用いたモデルの概要

がモデルの精度を左右する場合がある。本研究では入力データの条件ごとに、Optuna (Akiba et al., 2019) を用いてモデルのハイパーパラメータの最適化を行った。最適化の試行回数を30回とした。また、1試行あたりの最大のエポック数を30回とした。これは早期に過学習に陥るパターンが多く、これより多く設定しても効果が見られなかったためである。表 2 に、調整の対象としたハイパーパラメータと、最終的に用いた値を示す。

表 2 調整対象のハイパーパラメータ

パラメータ	処理①	処理②
バッチサイズ	56	32
LSTM層一層あたりの ユニット数	128	84
LSTM層の総数	12	4
Forward LSTMの ドロップアウト率	0.02696	0.12010
Backward LSTMの ドロップアウト率	0.00113	0.09123
結合後LSTMの ドロップアウト率	0.70568	0.01383
学習率	0.00015	0.000025

また、モデルの損失関数をMean Absolute Error (MAE、(2)式)とし、これを指標として判断した。

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

ここで、(2)式において $\hat{y}_i$ は観測値を、 $y_i$ はモデルによる予測値を示す。これより、MAEは値が小さい方が良い結果であることがわかる。評価用データを用いた予測結果の評価指標には Root Mean

Squared Error (RMSE, (3)式)およびMAEを用いた。(3)式における $\hat{y}_i, y_i$ は(2)式と同様であり、RMSEも値が小さいほど良好な結果である。

### 3. 結果と考察

#### 3. 1 データ処理による差異の検証

図 2に、処理①、処理②のそれぞれで、検証用のデータセットで予測した結果の、各時点及び期間全体におけるMAEおよびRMSEを示す。また、図 3に各時点での予測値と正解値の散布図を示す。

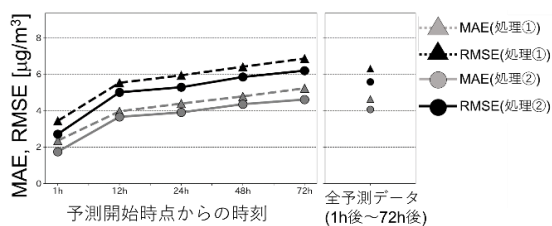


図 2 データ処理①、②によるモデルの精度評価  
 三角+破線：処理①、丸+実線：処理②  
 黒：MAE、灰色：RMSE

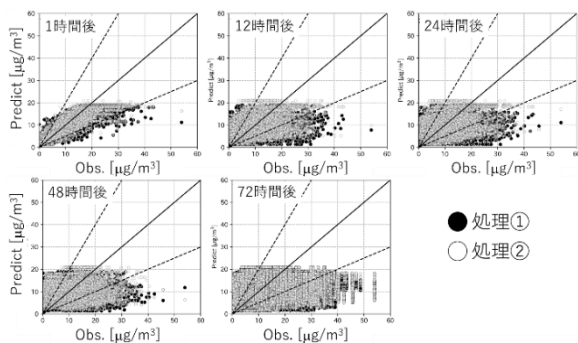


図 3 各時点における散布図  
 黒:処理①、グレー：処理②  
 予測開始時点から  
 上:(左)1時間後、(中)12時間後、(右)24時間後  
 下:(左)48時間後、(中)72時間後  
 の予測結果と対応する実測値の比較  
 正解と予測が黒実線:1対1、破線:1対2、2対1

図 2より、予測開始時点から遠い時点程精度が落ちる傾向が見られた。処理①では全予測データのMAEが4.63、RMSEが6.30である一方、処理②では全期間のMAEが4.06、RMSEが5.57と、両指標値から処理②で構築したモデルの方が、精度が良いことがわかる。一方、図 3から予測値と観測値の分布は両モデルで類似しているが、両モデルに共通して、どの時点でもモデルで予測された値に20  $\mu\text{g}/\text{m}^3$ 付

近を上限とするような傾向がみられた。これにより、高濃度の再現に困難が生じると予測される。

この問題について、モデル構成の改良とデータの追加による解決を試みた。

#### 3. 2 追加的検討

3.1節で示した問題について対応するため、以下の事項を、処理②に基づくモデルを基に検討した。

##### i. データの追加

本研究で対象としているACSA-14のデータセットのみでは、深層学習モデルの学習に対してデータの種類、量が不足であると考えられる。そこで、データの種類を増やすため、気象データを処理②のデータセットに加えた(以下、気象)。データは気象庁のメソ数値予報モデルのGPVデータを京大生存圏研究所の生存圏データベース (<http://database.rish.kyoto-u.ac.jp>) からダウンロードして用いた。海面更正気圧、地上2 m気温、湿度、地上10 mと850 hPa面の風の南北、東西成分を、ACSA-14が設置されている測定局の位置に対応するグリッドから抽出し、3時間間隔データを1時間間隔に線形に内挿した。さらに、PM<sub>2.5</sub>濃度および成分濃度に正時、正時-1時間、正時-3時間のデータが横並びになるようにタイムラグを持ったデータセットを作成した(以下、Lag)。ただし、気象データにはこのラグは適用していない。

##### ii. Batch normalizationの追加

モデルを構成する各層の間でデータを受け渡す際、前の層のパラメータの変化に伴って次層への入力データの分布が変化することで、学習の不安定化や学習速度の低下を引き起こす可能性がある。これを回避するために、Ioffe and Szegedy (2015)が提唱したBatch normalizationと呼ばれる手法を導入し、学習の安定化と速度向上を図った。図

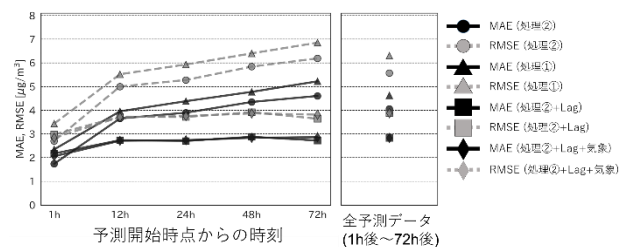


図 4 モデル改良による統計指標の変化

丸:処理②、三角:処理①、四角:処理②+Lag  
 ひし形:処理②+Lag+気象  
 黒:MAE、灰色:RMSE

4および図 5に、図 2および図 3で示した結果に加えて、データとモデル層を追加した結果を示す。

図4より、MAE、RMSEの両指標で、処理②+Lag+気

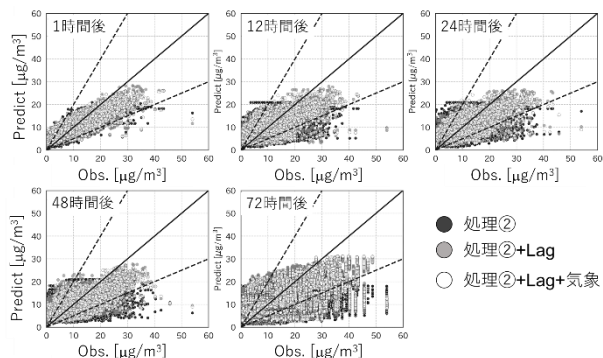


図 5 モデル改良後の各時点における散布図

黒: 処理②, 灰色: 処理②+Lag, 白: 処理②+Lag+気象  
予測開始時点からそれぞれ

上: (左) 1 時間後、(中) 12 時間後、(右) 24 時間後

下: (左) 48 時間後、(中) 72 時間後

の予測結果と対応する実測値の比較

黒実線: 1対1、破線: 1対2、2対1

象と処理②+Lagの精度が大きく改善される傾向にあった。これはデータの変動を表現する特徴量を適切に増やしたことによる効果であると考えられる。また、図5から、処理②で見られた $20 \mu\text{g}/\text{m}^3$ 付近を上限とする振る舞いが、Lagと気象データを加えた場合に解消される傾向にあった。しかし、どの予測時点でも予測が正解値を過小に評価する傾向が処理②+Lag、処理②+Lag+気象に見られ、実際の濃度ピークを十分に予測できていないことを示唆している。図 6に予測事例として、福岡局 2020/6/10 3:00~6/16 2:00と、同局2020/6/16 16:00~6/21 15:00の観測データと処理②、処理②+Lag、処理②+Lag+気象の、それぞれの予測結果を示す。図 6(上)より、処理②、処理②+Lagおよび処理②+Lag+気象のモデルに共通して予測期間の初めに濃度増加がみられ、現実とは異なる予測になっていた。処理②ではその後増加を続けるが、処理②+Lagおよび処理②+Lag+気象では6/14頃に減少を示し、予測期間後半にかけての増加傾向を再現していた。ただし、6/15にみられる急激なピークを予測しない等の問題が残っている。また、図 6(下)から、処理②+Lag+気象は予測期間初めの濃度減少、低濃度を示すような挙動をしているが、その他2つのモデルではその様子が見られな

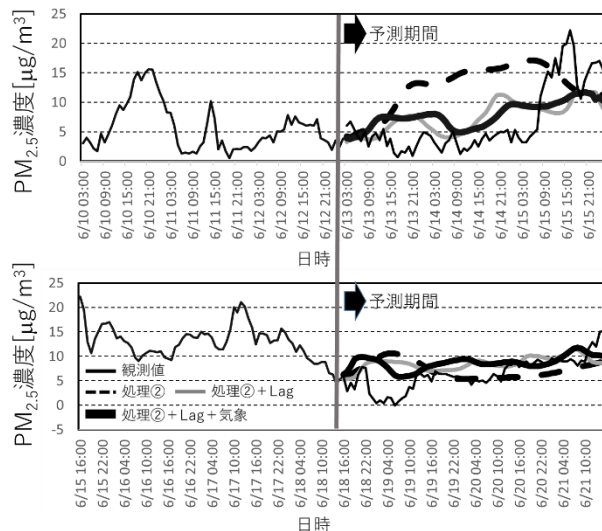


図 6 予測結果の事例検証

上段: 福岡局 2020/6/10 3:00~6/16 2:00

下段: 福岡局2020/6/16 5:00~6/22 4:00

黒細線:  $\text{PM}_{2.5}$ 観測値

黒太破線: 処理②による予測結果

灰色線: 処理②+Lagによる予測結果

黒太線: 処理②+Lag+気象による予測結果

い。一方、期間後半へかけての濃度の増加傾向は、細かな変動を除いて3つのモデルで表現していた。処理②+Lagや処理②+Lag+気象のようにデータを追加するだけでも統計指標は改善される傾向にあるが、日内変動や濃度増大イベントを再現できない等、課題が残る。これらの改良にはモデル構造の大幅な改良やさらに踏み込んだ検討が必要であると考えられる。

#### 4. まとめと課題

本研究では $\text{PM}_{2.5}$ 濃度の予測手段の拡充における第一歩として、ACSA-14による $\text{PM}_{2.5}$ 濃度及び構成成分を入力データとし、LSTMをベースとした深層学習モデルを構築し、将来予測を実施した。学習に用いるデータ処理により、予測結果の精度が明確に変化することが確認された一方、 $\text{PM}_{2.5}$ 成分濃度を単純に使用するのみでは予測値に問題が生じた。これを解決するため、 $\text{PM}_{2.5}$ 濃度に重要な影響を与える気象データと、 $\text{PM}_{2.5}$ 濃度及びその成分濃度にタイムラグを持たせたデータを加え、さらにモデルの学習を安定させるためのモデル構造とすることで、上記のような問題が解決されるとともに、大幅に再現性の指標が改善される結果を得た。ただし、事例検証から、改善されたモデルであって

も、大まかな傾向は再現するが、細かな時間変動や急激なピークを見逃すなど、未だ課題が残る。

#### 謝辞

本研究で用いたACSA-14のデータは環境省の事業により測定、整理されたものである。また、気象データは京都大学生存圏研究所が運営する生存圏データベースによって収集・配布されたものである。

#### 参考文献

- 1) 菅田 誠治, 大原 利眞, 黒川 純一, 早崎 将光 (2011): 大気汚染予測システム (VENUS) の構築と検証, 大気環境学会誌, 46(1), pp. 49-59
- 2) 竹村 俊彦 (2009): 大気エアロゾル予測システムの開発, 天気, 56, pp. 455-461
- 3) Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L. and Wang, J. (2015): Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation., Atmospheric Environment, 107, pp. 118-128
- 4) Catalano, M., Galatioto, F., Bell, M., Namdeo, A. and Bergantino, A. S. (2016): Improving the prediction of air pollution peak episodes generated by urban transport networks. Environmental Science & Policy, 60, pp. 69-83,
- 5) Kim, H. S., Park, I., Song, C. H., Lee, K., Yun, J. W., Kim, H. K., Jeon, M., Lee, J., and Han, K. M. (2019): Development of a daily PM10 and PM2.5 prediction system using a deep long short-term memory neural network model., Atmospheric Chemistry and Physics, 19, pp. 12935-12951
- 6) 環境省: 微小粒子状物質 (PM2.5) 成分自動測定機 データ取扱要領書 第二版, 2019 [http://www.env.go.jp/air/%20osen/pm\\_resultmonitoring/post\\_25.html](http://www.env.go.jp/air/%20osen/pm_resultmonitoring/post_25.html), last access 2022/1/5
- 7) Hochreiter, S. and Schmidhuber, J. (1997): Long short-term memory., Neural Computation, 9 (8), pp. 1735-1780
- 8) Laurent, C., Pereyra, G., Brakel, P., Zhang, Y., and Bengio, Y. (2015): Batch Normalized Recurrent Neural Networks, arXiv preprint,

arXiv:1510.01378

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

#### Summary

As a first step toward expanding and improving the accuracy of forecasting methods in Japan, this study developed a deep learning model that predicted PM2.5 concentrations and evaluated its accuracy using continuous observation data of PM2.5 and its components obtained by ACSA-14 at 10 locations nationwide. As the base of the deep learning model, long short-term memory (LSTM), a type of neural network, was used to construct the basic structure. A dataset was created through two types of data processing: a pattern that excludes single high concentrations and statistical outliers and a pattern that excludes only single high concentrations. The results showed that there was a tendency for the accuracy to decline further away from the starting point of the forecast. Furthermore, it did not predict concentrations higher than a certain level, a behavior that requires improvement. To address this problem, we added meteorological data and data with a time lag as well as a batch normalization layer to the model to normalize the data passed through the model, which eliminated the behavior of not predicting concentrations above a certain level and improved the model.