

# 機械学習による大気汚染物質濃度の予測

## Prediction of air pollutant concentrations by machine learning

太田 耕右  
Kohsuke OHTA

微小粒子状物質（以下「PM<sub>2.5</sub>」という。）による呼吸器・循環器系への影響に関する懸念が、近年高まっている。高濃度警報等の早急な発令のため、濃度予測手法の確立が必要であると考えられるが、ある測定局における 7 日後以降の濃度を予測する手法は確立されていない。本研究では、AI（人工知能）技術の一種であるニューラルネットワーク（以下「NN」という。）を用いて、岩沼一般大気環境測定局（以下「岩沼局」という。）における 1 週間後の PM<sub>2.5</sub> 濃度（1 時間値）を予測した結果、NN に学習させるデータ（以下「教師データ」という。）の種類によって異なる予測精度が得られた。特に、過去の PM<sub>2.5</sub> 濃度及び粒子同士の衝突反応速度定数を学習させた場合に最も高い精度が得られ、PM<sub>2.5</sub> 濃度（1 時間値）の実測値と予測値の相関係数  $r$  は 0.83 となった。この結果から、岩沼一般大気環境測定局における PM<sub>2.5</sub> の生成機構には、粒子同士の衝突反応等が関連している可能性が示唆された。

キーワード：大気汚染常時監視；PM<sub>2.5</sub> 予測；機械学習；再帰的ニューラルネットワーク

*Key words* : Air pollution monitoring; Predictoin of PM<sub>2.5</sub>; Machine learning ;Recurrent neural network

### 1 はじめに

PM<sub>2.5</sub> は代表的な大気汚染物質であり、呼吸器・循環器系疾患の発生率や死亡率を有意に増加させることが明らかになっている。現在、世界人口の約 9 割が汚染された大気のもとで暮らしており、ドイツのマックス・プランク研究所の発表によると、PM<sub>2.5</sub> などの大気汚染を原因とする世界全体の死者数は年間約 880 万人であると考えられている<sup>1)</sup>。宮城県においては、令和元年度の PM<sub>2.5</sub> の環境基準達成率は 100%であったものの、バイオマスを燃料とする発電所等の新規稼働が複数件予定されていることなどから、局所的な汚染も懸念されている。したがって、早急な警報発令体制や大気汚染の未然防止に向けて有用な情報を得るためにも、PM<sub>2.5</sub> 濃度の予測は重要であると考えられる。

現在、大気汚染物質濃度の予測は、輸送や放射等の物理的・化学的過程を考慮して計算する手法が主流となっているが、計算に膨大な費用を要するほか、特定の工場などといった局所的な影響に起因する濃度変化を予測できないという課題がある。このため、局所的な影響を考慮した予測を行うことは本県にとっても喫緊の課題であるものの、現在のところ十分な検討は行なわれていない。

一方、近年 AI に係る技術の進歩によって、これらの課題を解決する手掛かりが見いだされつつある。環境分野における AI の活用は数十年前から試みられてきたが、情報工学分野以外の研究者が容易に AI を利用できる環境ではなく、また予測精度も十分ではなかったため、実用化は進まなかった。しかし最近では、家庭用の一般的なコンピュータに無料で高性能な AI を搭載できる環境が整ってきており、一部の研究機関においても、機械学

習を用いて局所的な影響を考慮した予測を行おうとする動きが、ここ数年で始まりつつある。

以上のことから、本研究では、機械学習を用いた PM<sub>2.5</sub> 濃度の予測を目的とする。この目的の達成により、早急な警報発令体制及び大気汚染の未然防止に資する知見が得られるとともに、大気汚染常時監視測定局の置かれていない地点における濃度をも予測することで常時監視体制の質向上と効率化を両立でき、大気環境保全施策の向上に役立てることができると考えられる。

### 2 予測地点の選定

本研究においては、PM<sub>2.5</sub> 濃度を予測する地点の選定基準を以下のように定めた。

(1) 非メタン炭化水素（以下「NMHC」という。）、光化学オキシダント（以下「Ox」という。）、窒素酸化物（以下「NO<sub>x</sub>」という。）、二酸化硫黄（以下「SO<sub>2</sub>」という。）、及び PM<sub>2.5</sub> をいずれも常時監視している測定局

(2) 人間が居住し活動する地域に設置された測定局

(3) 周囲を建物や樹木に覆われているといった、特殊な地理的要因の無い測定局

(4) 測定データの長期にわたる欠測の無い測定局

まず(1)について説明する。PM<sub>2.5</sub> 濃度の予測地点は、教師データとなり得るデータ、すなわち PM<sub>2.5</sub> の生成に関わる物質のデータをより多く蓄積している測定局が望ましい。PM<sub>2.5</sub> 二次粒子の主な生成経路<sup>2)</sup>を図 1 に示す。図 1 より、PM<sub>2.5</sub> の生成反応には有機化合物すなわち NMHC、Ox、NO<sub>x</sub>、及び SO<sub>2</sub> 等が関連していると考え

られる。このため、一般環境大気測定局のうち、PM<sub>2.5</sub>の生成に関連している NMHC, O<sub>x</sub>, NO<sub>x</sub>, SO<sub>2</sub> 及び PM<sub>2.5</sub>をいずれも常時監視している測定局を、予測地点の候補とする。この観点から、本県における候補としては、塩釜局（塩竈市旭町 1-1 塩竈市役所屋上）、岩沼局（岩沼市中央 3-1-18 仙台保健福祉事務所岩沼支所屋上）、榴岡局（仙台市宮城野区五輪 1-2-3 榴ヶ岡公園内）、中野局（仙台市宮城野区白鳥 1-32-1 仙台市立高砂中学校）及び国設箕岳局（涌谷町小塚字桜清水 2-1-5 箕岳山中腹）の 5 局が候補となる。

次に選定基準（2）として、本研究においては人体への影響を未然防止するために大気汚染の予測を行うという位置づけであるため、人間が居住し活動する地域に設置された測定局を予測地点の候補とする。選定基準（1）で絞り込まれた 5 局のうち、国設箕岳局は周囲に住居がほとんど見られない山の中腹に位置し、大気汚染物質濃度のバックグラウンド測定としての役割があるため、本研究における予測の候補地点からは除外した。

選定基準（3）を設定した理由は、機械学習を用いた予測は本県においては初めての試みであり、予測に着手する第一段階としては特殊な地理的要因のある局舎よりも、より一般的な（局舎固有の特性が少ない）局舎を選定する必要があると判断したためである。

選定基準（1）及び（2）により絞り込まれた 4 測定局について、航空写真を図 2 に示す。図 2 より、塩釜局の北側には丘が存在するため、北側を通過する物質の移動が阻害される可能性が高い。また、榴岡局は、周囲を樹木に囲まれているため、物質の移動及び日射に起因する反応が阻害される可能性が否定できない。そのため、岩沼局及び中野局を予測地点の候補とする。

最後に、選定基準（4）であるが、機械学習においては、蓄積されたデータが多い方が精度が高まるため、欠測期間が短い方が有利である。本研究では、濃度予測を行う期間を 2019 年 1 月 1 日 1 時～12 月 31 日 24 時までとし、この期間内に長期にわたる欠測が無いことが望ましい。ここで、前述の 2 地点のうち、中野局は台風被害のために 2019 年 10 月 13 日～12 月 31 日のデータが欠測となっている。

以上の理由から、本研究では、（1）から（4）の選定基準をもとに判断し、岩沼局を予測地点とした。

### 3 予測方法

教師データによる機械学習モデルの構築には、NN の中でも時系列予測に長けた手法である再帰的ニューラルネットワーク（以下「RNN」という。）を用いた。図 3 に RNN の模式図を示す。通常の NN では、ある層の出力は、次の層の入力に利用されるのみである。一方で、RNN では出力の一部が「再帰的に」同じセルに戻る（図 3 中、隠れ層における赤色の矢印）。この結果、ある時点での入力が、それ以降の出力に影響を及ぼし、結果とし

て時系列データの処理を行うことができる。

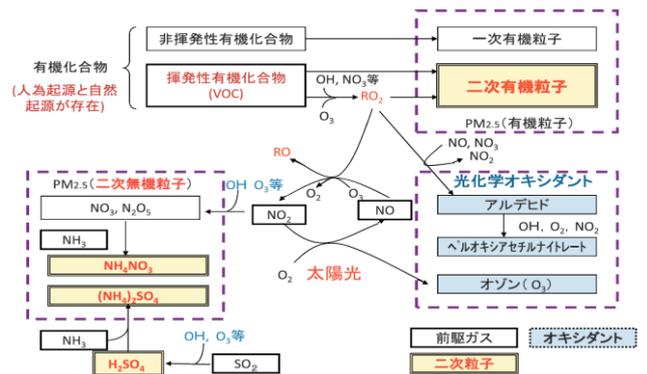
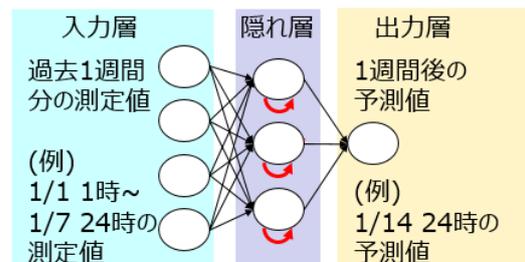


図 1 PM<sub>2.5</sub> 二次粒子の生成経路<sup>2)</sup>

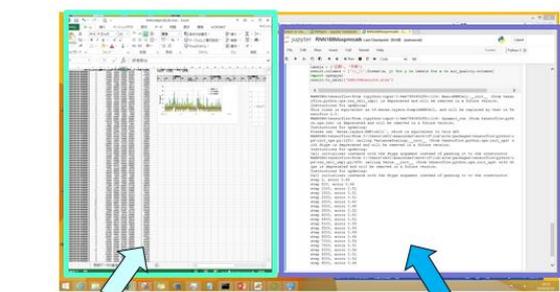


図 2 塩釜局、岩沼局、榴岡局、中野局の航空写真（国土地理院撮影の写真に測定局位置を追記して掲載）



※データの入出力はいずれもExcel形式

図 3 RNN の模式図



出力された予測結果（Excel） 予測コードの記述（Jupyter notebook）

図 4 PC画面における作業イメージ

開発環境についてはJupyter notebook(Python3.5)を使用し、RNNの実装においてはTensorFlow 1.2.1を使用した。PC構成はプロセッサにIntel Core i5、CPUに3.30GHz×2、メモリ4GBを搭載したものを利用した。PC画面上における作業イメージを図4に示すが、後述する期間及び種類のデータを使用した場合、予測に要する時間は20～30分程度である。

予測精度の評価には、2019年1月1日1時～12月31日24時までのデータを使用した。予測対象は1週間(168時間)後のPM<sub>2.5</sub>濃度とし、168時間前～現在までの連続データを用いて予測した。

教師データの期間は、2016年1月1日1時～2018年12月31日24時とした。教師データの種類は、図1及び化学反応論<sup>3)</sup>を参考に、以下を候補とした。

なお、気温、日射量、相対湿度、風速は仙台管区気象台におけるデータを使用した。

- ・NMHC [ppmC], Ox [ppb], NOx [ppb], SO<sub>2</sub> [ppb], 及びPM<sub>2.5</sub> [μg/m<sup>3</sup>]
- ・気温 [°C] (以下「T」という。)
- ・日射量 [MJ/m<sup>2</sup>] (以下「SR」という。)
- ・相対湿度 [%] (以下「HUM」という。)
- ・風速 [m/s] (以下「WV」という。)
- ・遷移状態が生じない場合における気相中の分子同士の衝突による化学反応論<sup>3)</sup>における、反応速度定数の温度関数部分である  $T^{0.5} \times \exp(-E/RT)$  (以下「K」とよぶ。)

参考に、衝突反応の模式図及び反応速度定数の式を図5に示す。ここで、Rは気体定数、Eは活性化エネルギーを表す。なお、PM<sub>2.5</sub>の生成反応に係るEの値については一般的に知られていないため、本研究ではNOxの酸化還元反応に係るEの値<sup>4)</sup>を参考に、E=200 [kJ/mol]と仮定して各時間におけるTの値からKを計算した。

以上のうち1～7種類のデータを選定し、必要に応じてこれらの積を求め、教師データとした。教師データの種類と予測精度の関係については後述する。

#### 4 結果

教師データの種類と、予測精度の指標である実測値と予測値との相関係数rの関係を表1に示す。教師データとしてSO<sub>2</sub>, KもしくはOxとNOxの積を加えた場合、相関係数rが0.8以上となり、高い相関が認められた。一方、表1のNo.20に示すように、特に関係が深いと考えられるPM<sub>2.5</sub>, NMHC, Ox, NOx, SO<sub>2</sub>, T, Kの7種類のデータを使用した場合、rは0.48となった。

次に、最も高い相関となった、KとPM<sub>2.5</sub>を用いた場合及び7種類のデータを使用した場合の予測値を実測値(正解データ)と並べて時系列表示した結果を図6に示す。

また、KとPM<sub>2.5</sub>を用いた予測誤差の平均値は±2.5[μg/m<sup>3</sup>]であり、誤差の95%信頼区間は±0.16～7.2[μg/m<sup>3</sup>]であった。



$$\text{反応速度定数} = \underbrace{\sigma(8k/\pi\mu)^{1/2}}_{\text{定数}} \times \underbrace{(T)^{1/2} \exp(-E/RT)}_{\text{Tの関数: 以下Kと定義}}$$

σ: 反応断面積 k: ボルツマン定数  
 μ: 実効質量  
 T: 気温 R: 気体定数  
 E: 反応の活性化エネルギー  
 (ここでは200 [kJ/mol]と仮定)

図5 気相分子同士の衝突反応の模式図

表1 教師データの種類とrとの関係

No.	教師データの種類	相関係数r
1	PM <sub>2.5</sub>	0.71
2	PM <sub>2.5</sub> , NMHC	0.68
3	PM <sub>2.5</sub> , Ox	0.72
4	PM <sub>2.5</sub> , NOx	0.69
5	PM <sub>2.5</sub> , SO <sub>2</sub>	0.80
6	PM <sub>2.5</sub> , T	0.74
7	PM <sub>2.5</sub> , K	0.83
8	PM <sub>2.5</sub> , SR	0.68
9	PM <sub>2.5</sub> , HUM	0.77
10	PM <sub>2.5</sub> , WV	0.68
11	PM <sub>2.5</sub> , NOx * SR	0.71
12	PM <sub>2.5</sub> , Ox * NOx	0.81
13	PM <sub>2.5</sub> , Ox * NOx * SR	0.74
14	PM <sub>2.5</sub> , NMHC * SR	0.55
15	PM <sub>2.5</sub> , NMHC * Ox	0.63
16	PM <sub>2.5</sub> , NMHC * Ox * SR	0.66
17	PM <sub>2.5</sub> , Ox * SR	0.67
18	PM <sub>2.5</sub> , NMHC * NOx	0.66
19	PM <sub>2.5</sub> , NMHC * Ox * SO <sub>2</sub>	0.68
20	PM <sub>2.5</sub> , NMHC, Ox, NOx, SO <sub>2</sub> , T, K	0.48

表中「\*」は積記号「×」と同義

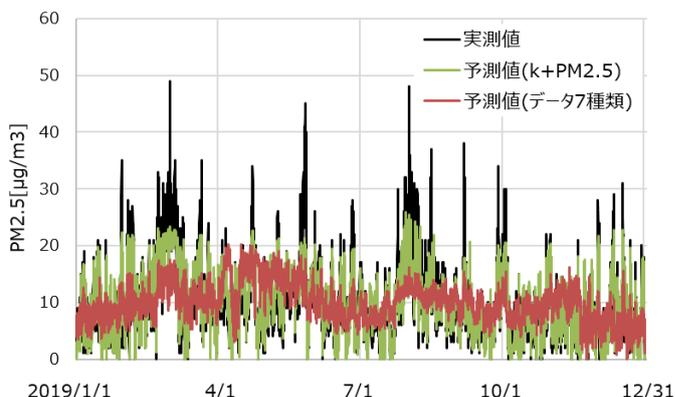


図6 予測値と実測値の時系列グラフ

## 5 考察

K を教師データとした場合に予測値が高精度 ( $r=0.83$ ) となった理由は、PM<sub>2.5</sub> の生成機構には衝突反応が関連しているためであると考えられる。また、SO<sub>2</sub> を教師データとした場合にも高い  $r$  の値 ( $r=0.80$ ) となった理由は、宮城県における PM<sub>2.5</sub> 成分のうち主要なものは硫酸塩である傾向がある<sup>5)</sup>ためと推察される。NO<sub>x</sub> と O<sub>x</sub> の積を用いた場合に  $r=0.81$  となった理由については、図 1 に示されているように PM<sub>2.5</sub> の成分である硝酸アンモニウムや硫酸アンモニウムが生成する際に、NO<sub>x</sub> と O<sub>x</sub> の反応が関連しており、この反応が PM<sub>2.5</sub> の生成に影響を及ぼしている可能性があるためと考えられる。

なお、相関係数  $r$  については、PM<sub>2.5</sub> の 1 時間値を用いた既報<sup>6)</sup>では 0.53、1 か月平均値を用いた既報<sup>7)</sup>は 0.87 であった。予測手法、期間及び使用データが異なるため単純比較は難しいものの、1 か月平均値を用いた方が突発的な濃度上昇が少なくなり、高精度となりやすいと推察されるが、本研究における予測では、1 時間値を用いたものでも 1 か月平均値を用いた既報に近い精度を得ることができた。

## 6 まとめ

本研究では、RNN を用いて岩沼局における 1 週間後の PM<sub>2.5</sub> 濃度の 1 時間値を予測した。結果、教師データの種類によって相関係数  $r$  の値は変化し、気相分子の衝突反応における化学反応速度論を参考に反応速度定数の温度関数部分 K を教師データとした結果、 $r=0.83$  が得られ、既往の研究と比較しても遜色ない精度となった。

また、教師データに O<sub>x</sub> 濃度と NO<sub>x</sub> 濃度の積もしくは SO<sub>2</sub> 濃度を学習させた場合も  $r$  が 0.80 以上となった。これらの結果から、岩沼一般大気環境測定局における PM<sub>2.5</sub> の生成機構には、粒子同士の衝突反応及び O<sub>x</sub>、NO<sub>x</sub>、SO<sub>2</sub> が関連している可能性が示唆された。

今後、高濃度時における予測精度の向上及び O<sub>x</sub> 濃度の予測について検討する。

## 参考文献

- 1) マックス・プランク研究所ホームページ：  
<https://www.mpg.de/12823232/polluted-air-shortens-the-lifespan-of-europeans-by-about-two-years>
- 2) 坂本和彦：表面科学, **36**, 141 (2015)
- 3) アトキンス物理化学, p.935.
- 4) J.Nicholas, "Chemical kinetics", Harper & Row, New York(1976).
- 5) 佐藤ら,宮城県保健環境センター年報, 2014, p.45.
- 6) 井上ら,2019 年大気環境学会年会 ポスター発表 P-144#.
- 7) 荒木ら,2019 年大気環境学会年会 口頭発表 2D0900.