

相関関係を考慮したパラメータ推定に基づく

PM2.5 成分の年平均値推定

久恒 邦裕, 山神 真紀子

Annual Average Value Estimation of PM2.5 Components based on Parameter Estimation with Correlation taken into Account

Kunihiro Hisatsune, Makiko Yamagami

PM2.5 の成分分析は、4 季節ごとに 2 週間行われており、それらの値が成分の年間データを推測する基になっていることがある。しかし、限られた期間のデータでは、それが代表性を十分に担保するとは限らない。2013~2015 年のデータでみると、質量濃度平均の年間データは経年的に低下したが、成分測定期間のデータでは下に凸となった。硫酸イオン濃度も下に凸となったが、成分測定していない期間の質量濃度を用いて年平均値を推定すると、下に凸となる傾向が緩和された。

はじめに

全国で PM2.5 の常時監視局が整備され、多くの地点における PM2.5 の質量濃度のデータが年間を通じて入手できるようになった。一方で、成分分析のデータについても全国的に調査が行われているが、測定期間は 4 季節ごとに 2 週間と限られている。そのため、測定した期間が、当該年度における高濃度期間や低濃度期間だった場合、結果の代表性は乏しくなり、その評価には注意が必要である。特に、成分分析結果の経年変化を論じることは、上記のような測定上での制約があるため、本来的には不可能である。この解決には年間を通じた成分分析が求められるが、分析体制やコストの面から、その実施は困難である。一方、多大な労力をかけて測定されたデータについて、より広い活用が求められる。そこで今回、上記のような限定的な成分分析結果を活用するため、統計的な手法で年平均値の推定を試みた。

実験方法

1. 調査年度

データは 2013~2015 年度のデータを用いた。各年度におけるそれぞれの測定地点に関して、年間を通じた PM2.5 質量濃度は自動測定機の値を、成分分析結果は

4 季節 2 週間の結果を用いた。測定地点は継続して測定を行っている八幡中学校局（名古屋市中川区，一般局：以降，八幡）と元塩公園局（名古屋市中南区，自排局：以降，元塩）のデータを用いた。測定時期などの詳細な条件は報告書¹⁻³⁾の通りである。

2. 推定方法

成分結果の年平均値の推定は、ベイズ統計の手法を用いた。

PM2.5 の質量濃度を PM，例として硫酸イオンの成分濃度を SO4 とし、それぞれの年平均値を μ_{PM} , μ_{SO4} ，標準偏差を σ_{PM} , σ_{SO4} ，PM2.5 質量濃度と成分濃度の共分散を $\nu_{PM,SO4}$ とすると、事後尤度は以下の式 1 で示される。

式 1 の前半は、PM2.5 質量濃度と成分濃度がそろった 4 季節 2 週間分のデータの尤度を計算しており、具体的には 2 変量共分散正規分布に従うモデルを想定し、尤度の計算を行った。後半はそれ以外の期間における質量濃度のみの尤度を正規分布から計算している。この式 1 の尤度が最大化するパラメータを見つけることで、成分の年平均値 μ_{SO4} を推定する。欠測値のあるデータから、全体の分布のパラメータを推定する方法については、相関係数を推定する方法を応用した⁴⁾。

$$\prod_i p(PM_i, SO4_i | \mu_{PM}, \mu_{SO4}, \sigma_{PM}^2, \sigma_{SO4}^2, \nu_{PM,SO4}) \prod_{i'} p(PM_{i'} | \mu_{PM}, \sigma_{PM}^2) \dots \text{式 1}$$

なおパラメータは、測定データが対数正規分布に従うと仮定して設定した。そのため、最後には対数正規分布のパラメータから平均や分散を計算している。

推定には R(3.1.0) とそのパッケージである rstan(2.9.0-3)を用いた。計算は4000回行い、最後の2000回を結果として用いた。chainは3本で、結果の間引きは行わなかった。それぞれのchainでR-hat値を検証し、すべてのパラメータで1.1未満になっていることを確認し、トレースプロットの形状で収束を確認した。

今回は、上記推定方法の制限により、質量濃度と相関関係の高い成分データの推定となる。そのため、本報告では硫酸イオン濃度について検証を行った。

結果と考察

モデルの妥当性を判断するため、質量濃度と成分分析結果がそろっているデータセットが望ましい。そこで、名古屋市環境科学調査センター（名古屋市区）屋上において23.5～24時間のスパンで採取しているサンプルのうち、2013年4月1日から2014年3月31日に採取された281サンプルを用いて、モデルの検証を行った。365日のうち4季節2週間の56サンプルは約15%に相当し、これは281サンプルにおいては42サンプルに相当する。そこで、本検証では281サンプルのうち10×4=40サンプルについては、質量濃度と硫酸イオン濃度のデータセットがあり、残りの241サンプルについては質量濃度のみであるとして検証を行った。40サンプルについては、281のサンプルを大まかに4つに分割しそれぞれから連続して10サンプルをランダムに抽出した。ただし、連続したサンプルにおいても、採取日が必ず連続しているとは限らない。このような抽出を1000回繰り返し、それぞれでの40サンプルの硫酸イオン濃度の平均値とベイズ推定された平均値の中央値を算出し、その変動を比較した。結果を表1に示す。

表1 シミュレーションによる推定結果の比較

(単位: $\mu\text{g}/\text{m}^3$)

	硫酸イオン 平均濃度 (年平均値)	40日間の硫酸イオン濃度平均 値(±標準偏差) :1000回シミュレーションの結果
実測結果	4.41	4.51±0.86
推定結果	—	4.60±0.52

硫酸イオンの実測値の平均値である $4.41\mu\text{g}/\text{m}^3$ を真値(SO4true)と仮定した場合、抽出した40サンプルの硫酸イオンの平均値(SO4calc)の平均は $4.51\mu\text{g}/\text{m}^3$ で、ベイズ推定による平均値(SO4bayes)の平均である $4.60\mu\text{g}/\text{m}^3$ よりも近くなった。ただし、その標準偏差は0.86から0.52とベイズ推定による結果のほうが低くなっていた。データを詳細に検証した場合、40サンプルの実測平均値(SO4calc)が年平均値(SO4true)と近い場合には、ベイズ推定値(SO4bayes)のほうが真値よりも離れている場合があったが、40サンプルが高濃度または低濃度期間を選択していた場合は、実測平均値(SO4calc)は年平均値(SO4true)と大きく異なっており、その場合、ベイズ推定値(SO4bayes)の方が年平均値(SO4true)と近くなった。得られたデータに偏りがあった場合、ベイズ推定を行うことでデータの偏りの影響を少なくすることができることが示唆された。

ただし、SO4calcがどの程度であった場合に、SO4bayesの方がSO4trueに近くなるかについて、定量的な判定方法はまだ解明できておらず、今後のさらなる検証が必要となる。

この結果を踏まえて、実測値のデータの解析を行った。表2に、検討対象である2013～2015年のPM2.5の年平均値と成分分析期間における平均値をそれぞれ示した。年平均値は、八幡および元塩いずれも経年的に減少しており、2013年度と2014年度の変化よりも、2014年度と2015年度の変化のほうが大きくなっている。数字の分布を検証するため、年間データをまとめたバイオリンプロットを図1に示す。バイオリンプロットの中央には箱ひげ図も示し、中央値を線で、白丸は平均値を示す。いずれも、最頻値(バイオリンプロットのふくらの大きなところ)、中央値、平均値が一致しておらず、一般的な算術平均がデータの要約になっていないことが示された。ただし、年を経るごとに測定値が低下している傾向は確認でき、平均値の傾向と一致するのが確認できた。

表2 平均値比較 (単位: $\mu\text{g}/\text{m}^3$)

	年平均値		成分分析期間の平均値			
	八幡	元塩	八幡		元塩	
	PM2.5		PM2.5	SO4	PM2.5	SO4
2013	16.8	21.6	20.1	4.8	21.4	4.7
2014	16.1	19.3	13.2	2.8	15.8	2.8
2015	13.9	15.0	15.6	4.1	16.9	4.4

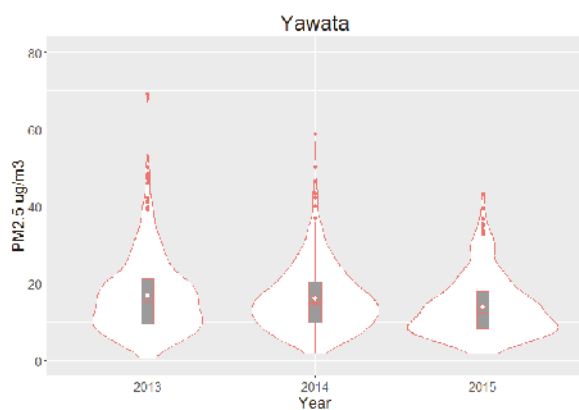


図 1 八幡および元塩の測定データプロット図

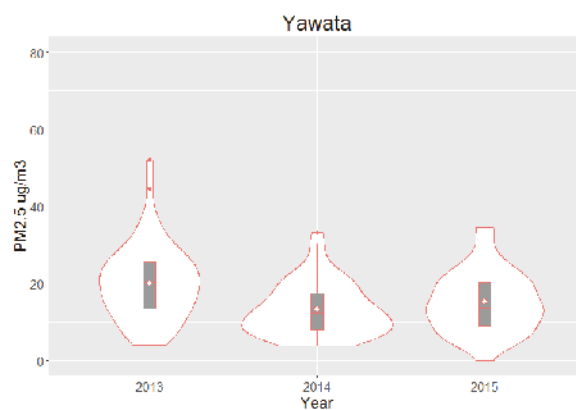


図 2 成分分析期間の質量濃度データ

次に、成分分析期間のデータについて、同様のプロットをしたのが図 2 である。成分分析期間のサンプルが年間データの代表性を持っている場合、同じ形状が示されることが考えられるが、実際には図 1 の形状とは大きく異なることが確認できた。八幡では、年間データ（図 1）は 2015 年度のグラフが高い尖度を持っているが、成分分析期間（図 2）ではその特徴がみられなかった。

元塩においても 2015 年データの尖度が、図 1 と比較して図 2 では小さくなっている。これらの形状から、詳細な検証を重ねることなく、成分分析期間のデータは年間データとはその性質が大きく異なり、代表性は低いと判断できる。そのため、測定された硫酸イオン濃度平均値は、測定期間の質量濃度平均値と同様に 2013 年から 2014 年に減少し、2015 年に上昇に転じている（表 2）が、それが年間データの傾向を反映しているとは言えない。これは、前述のモデルの妥当性確認で示した、年間データのうち極端な時期を抽出した例と見なすことができる。

そのため、ベイズモデル（式 1）を使ったパラメータの補正が有効に機能する可能性がある。そこで、前述の通り $PM_{2.5}$ の質量濃度を用いて、測定していない期間を含めた硫酸イオンの年平均値を推定した。

推定結果(estimated_value : ▲と点線)と、4 季節の平均値(measured_value, 表 1 に記載 : ●と実線)をグラフにしたのが図 3 である。

4 季節の平均値と比較すると、推定値では 2014 年と 2015 年の差が小さくなっている。これは、図 1 および 2 で示した年間データと 4 季節のデータの差が補正されたことを示す。ただし、質量濃度の年平均値が経年的に減っていったのに対して、硫酸イオン濃度は 2015 年度には 2015 年度と比較してまだ上昇する傾向がみられており、この点については、別のデータや推定モデルの改良による検証でさらに検討する必要がある。

(2016)

4) http://www.slideshare.net/hoxo_m/stan-36896297

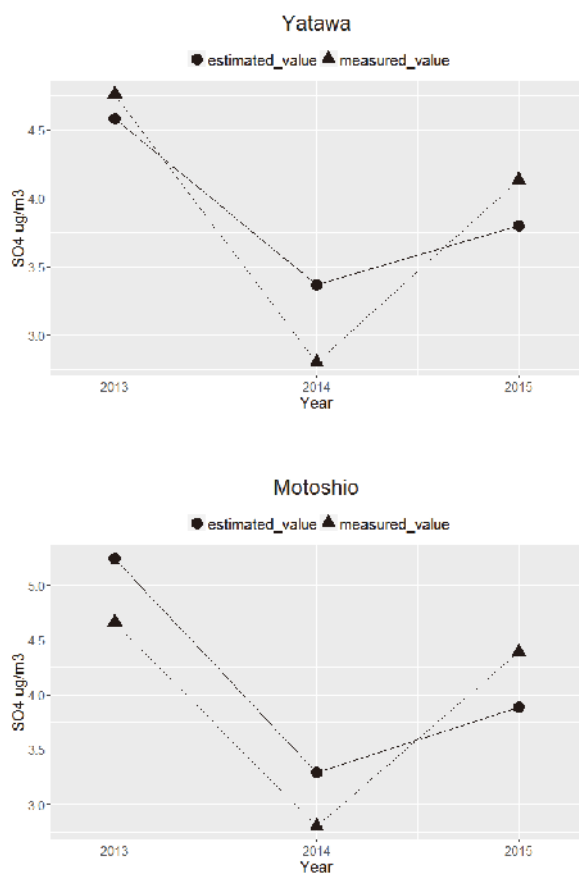


図 3 推定値と 4 季節の平均値

まとめ

不完全なデータについて、その共分散を考慮することでデータの全体像を推定した。特に、極端な期間のみ採取されたデータの場合、平均値などのパラメータを真値に近い状態で推定することができた。また、そのモデルを使って実測値について推定を行い、自然な解釈ができる結果を得た。

ただし、モデルは比較的単純な形をしており、妥当性の検証では必ずしもすべてのケースにおいて真値に近づいたわけではなく、今後の改良などが課題となる。

文献

- 1) 名古屋市，平成 25 年度大気汚染常時監視結果，(2014)
- 2) 名古屋市，平成 26 年度大気汚染常時監視結果，(2015)
- 3) 名古屋市，平成 27 年度大気汚染常時監視結果，