

アンサンブルカルマンフィルタにおける変数局所化を利用した気象場と大気濃度場の同時データ同化

関山 剛[†]・梶野 瑞王[†]

(受付 2022 年 1 月 4 日；改訂 7 月 11 日；採択 7 月 11 日)

要 旨

アンサンブルカルマンフィルタ(EnKF)は背景誤差共分散を陽に導出し、カルマンゲインの計算に使う。そのためデータ同化の途中で導出された背景誤差共分散に修正を加えることもできる。例えば、共分散を恣意的に大きくしてアンサンブル摂動を拡げること(共分散膨張)、あるいは状態変数間の物理的距離に応じて共分散を減衰させること(共分散局所化)も可能である。気象学における EnKF が系の自由度よりも遙かに小さなサンプル数で実用的に動作する主な理由はこの共分散局所化のおかげである。共分散の局所化は物理的距離とは関係なく、相関の小さな状態変数間でも可能である(変数局所化)。本研究では気象観測データ(風速・気温・気圧等)と濃度観測データを EnKF により同時に同化しつつ、変数局所化によって相関の小さな変数の組み合わせで共分散を強制的に 0 にする実験を試みた。それによりサンプリングエラーの影響が減少し、風速観測の情報から濃度分布の解析精度向上を得ることに成功した。一方で、本研究の条件下では濃度観測の情報から風速分布の解析精度向上を得ることは成功しなかった。

キーワード：データ同化，アンサンブルカルマンフィルタ，変数局所化，気象シミュレーション，大気化学シミュレーション。

1. はじめに

気象学においてデータ同化とは、時間発展する複雑な現実大気の状態を観測データと数値モデルを使って最大限確からしく推定するための数理的技法である。時々刻々推定された大気の状態変数(格子点上の気温・気圧・風向風速・密度・水蒸気量など)は気象場解析値と呼ばれ、数値予報の初期値として使われる。しかし、地球大気の状態を精緻に表現するのに必要な自由度の巨大さに比して観測データの数あるいは時空間密度は極めて小さいため、圧倒的な劣決定問題を解く必要に迫られる。その上、数値モデルが計算する大気の状態変数は一般に気象変数と呼ばれるが、その直接観測は必ずしも容易ではない。例えば、海洋や僻地の上空あるいは高度数千mの高層を直接かつ頻繁に観測することは極めて困難である。気象変数の観測データが足りない場合、気象変数からなんらかの物理法則によって導かれる間接的な物理量、例えば大気黒体放射・電波の大気遅延・電波/音波のドップラーシフトなどの観測値をデータ同化に利用する必要に迫られる。

気象学において、より高精度な気象場解析値が得られることは極めて重要である。それは正

[†]気象庁気象研究所：〒305-0052 茨城県つくば市長峰 1-1

確な天気予報に繋がるのみならず、基礎研究においても大気の状態を広範囲かつ正確に把握する手段となる。そのため様々なデータ同化アルゴリズムの開発が進められてきた。観測データが圧倒的に足りない現状では、空間的に遠く離れた情報を精度良く利用でき、気象変数以外の物理量からの情報も漏れなく利用できるデータ同化アルゴリズムが求められる。それらの要件を満たすものとして現在盛んに利用されている手法が4次元変分法(4D-Var)とアンサンブルカルマンフィルタ(EnKF)である。これらの手法では背景誤差共分散が陰に、あるいは陽に動的に推定され、また観測演算子を用いることで間接的な物理量のデータ同化も可能となっている。背景誤差共分散は限られた地点・限られた物理量の観測データからでも大気全体の状態推定を可能にする鍵である。

気象データ同化において、気象変数とは異なる観測データではあるが気象変数とはなんらかの物理法則によって結びついている物理量を利用したい場合、観測演算子を使う以外にもう一つ方法がある。その物理量を数値モデルの中の状態ベクトルに追加してしまい、気象変数と一緒に推定するのである。この場合、実用上は数値モデルの物理化学過程を拡張する必要があり、追加された物理量も誤差共分散の要素に組み込まれることとなる。もし大気汚染物質濃度などのようにその物理量の予報自体にも社会的ニーズがあるならば、この方法は特に有用である。ところで大気汚染物質のような大気微量成分の空間分布は、化学反応や地表面の影響が小さい場合、風速場(風向風速の3次元分布)の影響のみを受けて時間発展していると近似できる。すなわち大気の流れを追跡する“大気トレーサー”と見なすことができる(図1)。

実際、先行研究において大気微量成分を数値モデルで取り扱えるようにして、大気微量成分濃度の観測データを同化することにより、気象変数(特に風速場)の解析精度向上を試みた例が幾つかある(Semane et al., 2009; Han and McNally, 2010; Allen et al., 2018)。これらの例では成層圏下部および対流圏上部のオゾンに着目し、4D-Varを使ってデータ同化が行われた。大気微量成分であるオゾンは、成層圏下部および対流圏上部において(成層圏上部あるいは地上付近に比べて)化学的な生成消滅が少なく、人工衛星による観測も充実しているため大気トレーサーとして選ばれたと考えられる。しかしこれらの先行研究では、オゾンの観測データを加えたことによる気象変数の解析精度向上はほぼ無いか極めて小さかったことが報告されている。Han and McNally (2010)はその原因を数値モデルのオゾン予測精度の低さに帰しているが、データ同化のアルゴリズム改良によって事態を打開できないだろうか。

EnKFの利点の一つは、4D-Varと異なり背景誤差共分散が陽に導出され、カルマンゲインの計算に使われることにある。また、通常は大気全体の巨大な背景誤差共分散を一度に作るこ

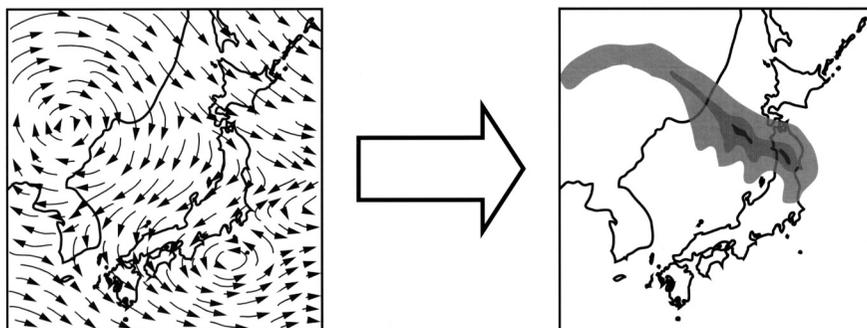


図1. (左)風速場の例、(右)その風速場によって形成された物質濃度分布すなわち濃度場の例。風速場は濃度場に大きな影響を与えるが、濃度場は風速場に直接的な影響を与えない。

とはせず、限定的な範囲の誤差共分散を用いることで小さなアンサンブルサイズでも 4D-Var に匹敵するデータ同化の精度を実現している(時空間的な局所化)。局所化については 2 章にて詳述する。4D-Var と EnKF の性能比較は Kalnay et al. (2007) に詳しい。さらに、データ同化サイクルにおいて数値モデルが過度に信用されることを防ぐため、背景誤差共分散の値を恣意的に膨張させることも行われる(共分散膨張: 2 章にて詳述)。このように EnKF では陽に背景誤差共分散が導出されることを利用して、その値の補正を試みるのが一般的である。その共分散補正は制御変数間においても可能であり、変数局所化と呼ばれる(Kang et al., 2011)。変数局所化では、相関が小さいと事前に予想できる制御変数の組ではサンプリングノイズを避けるために共分散を恣意的にゼロにしてしまう操作も可能である。これについても 2 章にて詳述する。

先行研究において大気トレーサーの濃度観測データ同化が気象場の解析精度向上に繋がらなかった原因の一つには、大気トレーサー濃度を計算する数値モデルの不完全性に伴う誤差に比較して大気濃度場と風速以外の気象場の相関が小さかったことが考えられる。それならば、変数局所化によって相関の小さな制御変数間の共分散は事前にゼロにしてしまえば(すなわち濃度観測の情報は気温や気圧の解析には使わない)、濃度データ同化が風速場の解析精度だけを向上させられるのではないだろうか。実際、Kang et al. (2011) は現実大気を用いない理想化実験(観測システムシミュレーション実験 OSSE)において EnKF の変数局所化を使えば、二酸化炭素濃度データの同化が風速場を改善することを示した。しかし残念なことに Kang et al. (2011) が用いた数値モデルは現業気象予報モデルに比べかなり単純化されたものであり、さらに疑似観測値の作成とデータ同化に同一の数値モデルを使った「双子実験」であった。そのため数値モデルにバイアスが一切含まれておらず、観測誤差には既知のランダム誤差のみを含むという極端な理想化実験であった。

そこで本研究では、現実大気 of EnKF データ同化において変数局所化を導入し、大気微量成分濃度データが風速場の精度を大きく向上させることができるかについて検証する。ただし、Kang et al. (2011) が大気トレーサーとして用いた二酸化炭素は対流圏において化学的に不活性であるという利点を持ち数値モデルの性能低下の主要因である化学反応計算が必要ないが、現実大気では広範囲・高頻度で精度良く濃度を観測することが極めて困難である。一方、4D-Var を使った先行研究において使われた高層オゾンは二酸化炭素に比べて高精度観測データは手に入りやすいが、化学反応が複雑であるため数値モデルの不完全性が顕著化しやすい。そこで我々は化学的に不活性で高精度測定が容易に可能な大気トレーサーとして放射性同位体(原子力発電所事故由来のセシウム-137)を利用し、EnKF データ同化における変数局所化の検証を実施することにした。

もし本研究で放射性同位体濃度のデータ同化による風速場の解析精度向上が確認できれば、そのノウハウを非放射性同位体濃度のデータ同化に活かすことができる。すなわち非放射性的同位体を人為的に環境中へ定期的に放出し、その濃度を各地で精密測定することによって、天気予報の精度向上も期待できる。本研究はその一里塚としての重要性を有する。以下、2 章で背景誤差共分散の修正手法について説明した後、3 章で本研究におけるデータ同化実験の設定を記述し、4 章で実験結果の紹介と考察を行う。

2. データ同化手法

2.1 気象データ同化で行われる誤差共分散の修正

気象学や地球物理学での利用における EnKF の解説は Evensen (1994) や三好 (2008) などに譲る。EnKF によるデータ同化サイクルでは解析値 x_t^a を得るだけでなく、次のサイクルにお

いてアンサンブル予報シミュレーションを実行するために初期値(すなわち解析値)のアンサンブル摂動 $\delta x_i^{a(k)}$ ($1 \leq k \leq m$) が必要である。ここで下添字は時刻(i)を表し、上添字は予報(f)あるいは解析(a)を表す。 m はアンサンブルメンバー数である。このアンサンブル摂動を得るための作業をアンサンブル・アップデートと呼ぶ。しかしアンサンブル・アップデートの式を $\delta x_i^{a(k)}$ について解くことは技術的に煩雑で、計算コストも大きい。そのため様々な近似解法や計算コストを減らすアルゴリズムが開発されている。本稿ではその技術的な詳細には立ち入らないが、現業気象機関や研究機関で現在最も採用されているのは平方根フィルタ法であり、本研究では平方根フィルタの一種である局所アンサンブル変換カルマンフィルタ (LETKF; Hunt et al., 2007) と呼ばれる手法を使用した。

また、EnKFでは数値モデルの接線形演算子やアジョイント演算子を実装上必要としない。気象学において数値モデルは非線形であることが大前提であるから、接線形化による近似を避けることができ非線形モデルを直接使うことのできる EnKF のメリットは大きい (Hunt et al., 2007)。さらに、EnKFは背景誤差共分散 P_i^f をデータ同化サイクル毎に陽に計算しており、その値の修正も以下に述べるとおり逐次可能である。

気象予報では数値モデルのランダム誤差の評価は極めて難しいため、通常はモデルの完全性を仮定する。しかし現実にはモデルの不完全性に起因する誤差は必ず発生するはずなので、モデルを信用しすぎた状態で推定した背景誤差共分散 P_i^f は実際よりも若干小さく見積もられてしまう。また、たとえモデルが完全だったとしても、アンサンブルメンバー数が十分でなければアンサンブル摂動は期待されるほどに大きくなることができず、やはり背景誤差共分散 P_i^f は実際よりも小さく見積もられる。このように EnKF がデータ同化サイクルにおいて背景誤差共分散を繰り返し小さく見積もると、生成されるアンサンブル摂動は徐々に小さくなり、最終的にはカルマンゲイン K_i が極端に小さくなることによって観測情報は解析値にほとんど取り込まれなくなる。この不具合を避けるためにデータ同化サイクル毎に誤差共分散を恣意的に大きくする操作を実施するのが気象学においては通例である。この操作を共分散膨張 (covariance inflation) と呼ぶ。

共分散膨張のやり方には幾つかあるが、最も単純な方法は解析誤差共分散を計算する際に $\tilde{P}_i^a = P_i^a(1 + \delta)$ と恣意的に修正してしまうことである。ここで δ は任意の正数である。これを multiplicative inflation と呼ぶ。これにより、次のサイクルのアンサンブル摂動は全体的に少し大きくなる。あるいは、アンサンブル摂動を $\delta x_i^{a(k)} = \delta x_i^{a(k)} + \varepsilon$ と修正する方法もある。ここで微小摂動ベクトル ε は、例えば $\langle \varepsilon \varepsilon^T \rangle = \alpha \mathbf{A}$ を満たすようなランダムサンプルで選ぶ。記号 $\langle \rangle$ は期待値を表す。 α はチューニング可能な定数であり、 \mathbf{A} は統計的に計算した解析誤差共分散行列などを利用する。この方法を additive inflation と呼ぶ。いずれにせよ、これらの操作は EnKF の実装において自由に選択可能であり、気象変数や地域ごとに異なるチューニングを施すこともできる。

一方、アンサンブル予報による背景誤差共分散 P_i^f の推定はアンサンブルメンバー数 m が十分に大きければ精度良く成り立つと仮定できるが、その妥当性の確保は容易ではない。状態ベクトルの次元数 N に比べてアンサンブルメンバー数 m が圧倒的に小さくならざるをえない気象学においては、背景誤差共分散 P_i^f の推定にはかなりのサンプリングエラーが入り込むと考えるべきである。ところで地球大気は、天気予報の対象となる数時間から数日の時間スケールで見ると、すぐ近くの場所や風上風下に当たる場所との相関が大きい一方で、遠く離れた場所との相関は極めて小さい。例えば地球の裏側の大気との間で相関は微小存在するのかもしれないが、そんな離れた場所との相関はサンプリングエラーに比べたら遙かに小さいであろうことは容易に想像できる。そこで、物理的距離が離れるにつれて共分散を強制的に減衰させることでサンプリングエラーの影響を可能な限り回避する手法が EnKF では一般に採用されている。

離れた場所の相関は初めから無視するのである。これを共分散局所化(covariance localization)と呼ぶ。

共分散局所化を適用する範囲すなわち局所化距離は当てはめる減衰関数の形も含めて任意であるが、アンサンブルメンバー数に大きく依存するチューニングパラメータである。気象学的に大きな相関が発生しうる距離や時間にも依存するので、気象予報モデルの種類によって適切な局所化距離が異なる。例えばアンサンブルメンバー数 m が $O(10 \sim 10^2)$ 程度、すなわち現業天気予報システムの標準的アンサンブルサイズを用いる場合、全球モデルならば水平方向に数千km離れれば共分散をゼロにするのが一般的であるが、計算領域が数百km程度の都市大気汚染モデルならば数十km離れれば共分散がゼロになっても構わない。一方で、このような $m \approx O(10 \sim 10^2)$ 程度の気象データ同化システムでは、共分散局所化を行わないと確実にアンサンブルメンバーの発散(filter divergence)が起こり、データ同化に失敗する。

全球モデルを使った理想化実験の場合、Kondo and Miyoshi (2016)によると、10,240個のアンサンブルメンバーを使えばEnKFに共分散局所化が必要なかったと報告されている。実際、地球大気の構造はランダムではなく、状態変数間にも物理法則が成立している。したがって地球大気の実質的な自由度は次元数 N に比べて極めて小さいと考えるべきである。さらにEnKFで使われる数値モデル(気象予報モデル)はそれら物理法則に基づいて状態変数を予報し、誤差共分散を推定するためのサンプルを提供しているため、小さなサンプル数(=アンサンブルメンバー数 m)で実用的な統計推定精度を実現しているのだと考えられる。

状態ベクトルの次元数 N が $O(10^7)$ 程度である現在の気象予報モデルを使ったEnKFにおいては、誤差共分散行列の特異ベクトルの数も本来ならば $O(10^7)$ 程度と考えるべきであろう。ところが共分散局所化を使えば、上記の自由度の制限も相まって、 $m \approx O(10)$ のアンサンブルメンバー数でEnKFは十分に動作する。これは共分散局所化の適用範囲内の空間において、おそらく 10^3 個程度の特異ベクトルを10個程度の特異ベクトルで近似していることになる。さらにKondo and Miyoshi (2016)の報告では、たかだか $m \approx O(10^4)$ で共分散局所化すら不要になる。これは、気象学におけるEnKFが状態変数のランダムサンプリングによって誤差共分散を推定してのではないことを示唆している。

以上のような共分散局所化の操作は空間方向(あるいは時間方向)でのサンプリング制限を想定している。一方で、この局所化操作は空間的距離に応じて実施できるだけでなく、状態変数間でも実施可能である。そのことについて次節で説明する。

2.2 変数局所化

前述の共分散局所化では空間的距離が離れるにつれて共分散を強制的に減衰させることでサンプリングエラーの影響を可能な限り回避した。しかし真のシグナルよりもサンプリングエラーの方が大きくなってしまふ危険性は空間的に離れた場所に限らない。たとえ同じ場所の大気(空気塊)だろうと、物理法則上関連が低い物理量との間では相関が極めて小さくなるのが当然である。それならば、空間的距離に関係なく大気微量成分濃度と気温・気圧・湿度との間の共分散を最初から無視してしまえばサンプリングエラーの影響を除去できるはずである。このようなEnKF実装上のアイデアを変数局所化と呼ぶ。

詳述すると、化学的に不活性な大気微量成分の濃度を体積混合比で表した場合、理想気体の状態方程式が仮定できれば気温や気圧が変化しても大気との相対的な体積は変化しないため、濃度は気温や気圧との直接的な相関を持たない。濃度を単位大気体積あたりの質量や粒子数で表した場合も、理想気体の状態方程式が近似できている限り、数時間程度の気温変化量や気圧変化量では密度はあまり変化しないため、濃度変化は極めて小さい。一方で大気微量成分の濃度変化は移流によってほぼ決まり、移流は風向風速で完全に決まるため、濃度と風向風速の相

関は大きいはずである．ところが風向風速は気圧差を外力とするニュートン力学第2法則によって決定され，気圧差は気体の状態方程式に基づいて気温差から生まれるため，風向風速は気温や気圧との相関を有する．よって濃度と気温・気圧との間には風向風速を通じて間接的な相関が僅かにだけ存在する．湿度については，水蒸気分子を一種の大気微量成分と考えれば移流によって濃度が変化するため他の大気微量成分濃度と大きな相関を持つはずである．しかし実際には水の相変化によって雲が生成し，雨が降り，水面から水蒸気が供給されるため，著しく水蒸気濃度すなわち湿度は大気中で変化する．したがって湿度と他の大気微量成分濃度との間の相関は必ずしも大きくはない．一方で湿度自体は移流によって風向風速と大きな相関を持つ可能性があり，水の相変化や飽和水蒸気量を決定する気温や気圧とも大きな相関を持つ．まとめると，もし小さな相関を無視したいのならば，大気微量成分濃度は風向風速に限定して共分散を有し，しかしながら風向風速は他の気象変数とも共分散を有すると仮定することが妥当である．

変数局所化はアンサンブルメンバー数の少なさに起因する変数間の共分散誤差にかかるサンプリングエラーの除去に役立つだけではない．加えて，特定の状態変数のシミュレーション精度が低いことに起因する不確かな共分散推定を除去することにも役立つ．数値モデルはすべての状態変数についてその典型値(あるいは気候値)と比べた場合にバイアス誤差が常に同程度となるわけではない．幸い天気予報に関連する気象変数は比較的どれも極端なバイアスを含むことなく予測可能である．例えば気温 300 K に対してバイアス誤差はせいぜい 1~2 K であり，地上気圧 1000 hPa に対しては数 hPa 程度である．一方で大気微量成分濃度は気象変数に比べてバイアス誤差が極めて大きい．例えばエアロゾル濃度のバイアス誤差がその濃度観測値の平均値(気候値)よりも大きいことは頻繁に見られる．なぜなら生成・消滅に関わる化学反応計算は流体力学計算に比べて大きく精度が劣り，地表面からの放出過程(あるいは地表面への沈着過程)や乱流拡散過程も大きな不確定要因となるからである．バイアス誤差が大きければ共分散の正確な推定に支障をきたす．

したがって大気微量成分濃度との相関が大きい風向風速に限れば相関のシグナルが共分散の推定誤差を上回るが，他の気象要素との間では逆に共分散の推定誤差がシグナルを上回る可能性も十分に考えられる．このような場合，変数局所化を使えば風向風速以外の気象要素との共分散を無視することが可能である．ただし同化サイクルを繰り返す中で濃度場が風速場に変化を与え，その変化に影響されて他の気象場も若干変化することが考えられる．したがって変数局所化によって変数同士が完全に独立するわけではない．

本研究では図 2(c)で模式的に示されたような背景誤差共分散行列を設定することで変数局

	(a) w/o concentration DA (STD)					(b) Independent DA (TEST1)					(c) Variable-localized DA (TEST2)					(d) All-simultaneous DA (TEST3)						
	濃度	風速	気温	気圧	湿度	濃度	風速	気温	気圧	湿度	濃度	風速	気温	気圧	湿度	濃度	風速	気温	気圧	湿度		
濃度	0	0	0	0	0	濃度	0	0	0	0	濃度			0	0	0	濃度					
風速	0					風速	0				風速						風速					
気温	0					気温	0				気温	0					気温					
気圧	0					気圧	0				気圧	0					気圧					
湿度	0					湿度	0				湿度	0					湿度					

図 2. 各実験における背景誤差共分散行列の模式図．左から(a)濃度同化無し実験(w/o concentration DA; STD), (b)独立同化実験(Independent DA; TEST1), (c)変数局所化実験(Variable-localized DA; TEST2), および(d)完全同時同化実験(All-simultaneous DA; TEST3)．灰色のセル部分には EnKF が推定した値が入る．

所化を実行した。ここで大気微量成分濃度は濃度自身と風速のみに対して共分散を共有し、他の気象変数との共分散は強制的に0としている。一方、風速は濃度以外にも気温や気圧や湿度との間の共分散も共有している。図中に示した他の3種の実験の詳細については後述する。同様の操作によって、Kang et al. (2011)は変数局所化によるEnKFの性能向上を示した。ただしKang et al. (2011)は二酸化炭素を大気トレーサーとして使い、濃度分布と共にその地表面放出量分布もデータ同化によって逆推定した。一方本研究では放出情報がほぼ既知の放射性セシウムを大気トレーサーとして使うため、解析値を作成するのは放射性セシウム濃度および気象変数に限られる。そのためKang et al. (2011)とは変数局所化の設定パターンが若干異なっている。

真に濃度のみが風向風速と強い相関を持ち、他の気象変数との相関は無視できるのならば、図2(c)で示した変数局所化によってサンプリングエラーの影響を縮小できるはずである。それによって他の条件下の実験(図2(a), (b), (d))よりも風向風速の解析精度が向上することを期待して、本研究を実施した。このような変数局所化は共分散膨張や空間的距離の共分散局所化と同じく、EnKFにおいて誤差共分散が陽に計算され自由に変更できるために効率的に実施可能となる。

3. 実験設定

3.1 変数局所化の設定

本研究における背景誤差共分散行列の変数局所化設定についてまず説明する。実験は4種類実施した。まず標準実験(STD)として、放射性セシウム濃度(以下、濃度と記す)が他のいかなる制御変数とも相関を持たないように設定した(図2(a))。これは濃度観測データを一切同化しないことに相当し、気象観測データのみを同化する通常的气象データ同化と同一である。この場合、濃度分布(濃度場)は気象変数の予報値から物理法則に従い受動的に決定される。風向風速分布(風速場)は局所化距離内に存在する全ての気象観測データの情報を利用して解析値が作成される。

次に、濃度は濃度とのみ相関を持つよう設定した実験(TEST1)を用意した(図2(b))。この場合、濃度場は気象変数の予報値に従ってシミュレーションされつつ、局所化距離内の濃度観測データの情報を利用して解析値が作成される。風速場はSTDと同様に濃度観測データとは無関係に解析値が作成される。2つのデータ同化は完全に独立である。実はこの実験設定は通常の大気化学データ同化(大気汚染物質の予報を目的にした大気微量成分データ同化)と同じである。一般的に大気汚染物質の予報モデルは誰かが気象データ同化で作成した気象場解析値を天賦の情報として受け取り、その気象場に合わせて濃度場を計算する。そのため大気化学データ同化は大気汚染物質濃度観測データと大気汚染物質予報モデルの情報だけを使って実施され、気象場解析値とは別に濃度場解析値を計算するのが一般的である。TEST1はこれと実質的に同じことをしている。

TEST2では前述のとおり濃度場は濃度場自身と風速場に限定して相関を有し、その他の気象場(気温、気圧、湿度)との相関は強制的に削除されている(図2(c))。一方、風速場は濃度場とも他の気象場とも同時に相関を有している。風速場以外の気象場は濃度場との相関は有していないが風速場とは相関している。この場合、濃度場と風速場は相互に影響を与えながらデータ同化が行われるが、風速場以外の気象場は濃度場の情報を与えられずにデータ同化される。ただし風速場以外の気象場がTEST2においてSTDやTEST1と同一の解析値になることはない。なぜなら風速場が濃度場の観測データ情報に影響されて変化し、その風速場の変化に影響されて他の気象場も若干変化するためである。この実験(TEST2)が本研究の目標である「気象

場と濃度場の同時データ同化における変数局所化の影響評価」の要となる．さらに比較のため、変数局所化を一切設定しない完全同時データ同化の実験(TEST3)も実施した(図2(d))．

3.2 観測データと予報モデル

先述したとおり Kang et al. (2011)が利用した二酸化炭素は広範囲に精度良く濃度観測を継続することが極めて困難であるため、我々の知る限り、現実大気によって変数局所化を検証した例はない．一方、高層オゾンは化学反応系が複雑なため予報モデル精度の向上が難しい．そこで我々が着目したのは福島原発事故由来の放射性セシウムである．放射性同位体の壊変速度や壊変生成物は厳密に既知であり、気温・気圧・太陽光や化学反応によっても変化しない．したがって予報モデルは化学反応計算を行う必要がない．放射性セシウムの場合、降雨さえなければ地表面沈着量はほぼ無視でき、放出場所は福島第一原子力発電所ただ一ヶ所に限定された緯度・経度・高度が完全既知の点源放出である．その放出量は時間変化も含めてすでに推定されており (Katata et al., 2015), その推定値を使ったモデル計算によって各地の濃度観測値の再現が報告されている (例えば Sato et al., 2018)．そのため放射性セシウムをトレーサーとして用いれば化学反応過程および生成消滅過程のモデル不完全性に起因するバイアスを減らすことが可能であり、共分散推定における大きなメリットとなる．さらに同位体は試料さえ捕集できれば、極めて微量でも精度良く濃度測定することが可能である．福島原発事故の場合、図3(a)に示された約100地点において事故直後から1時間間隔で地上付近の放射性セシウムが捕集され、その濃度が計測された (Tsuruta et al., 2014; Oura et al., 2015)．本研究ではこの観測値を大気微量成分濃度データとして利用した．

気象変数の観測データについては、気象庁現業メソ予報システム(JNoVA)のために収集されたデータセットおよび気象庁地域気象観測システム(AMeDAS)の地上風観測データセットを利用した．地上風は原則的に高度10mでの測風である．高層の風速データと異なり、AMeDAS風速データは図3(b)に示すとおり観測地点が広範囲に分布しており(観測頻度も10秒単位)、風速場の解析精度検証に最適である．詳細は Sekiyama et al. (2017), Sekiyama and Kajino

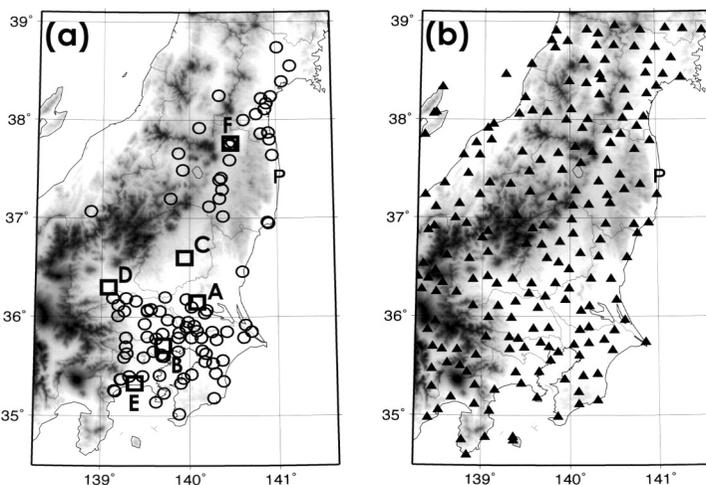


図3. (a) 本研究で用いた放射性セシウム濃度データ(○)および空間線量率データ(□)の地上観測地点(A~Fの記号は表1を参照)．(b) 本研究で用いたAMeDAS地上風データの観測地点(▲)．Pは福島第一原子力発電所の位置を示す．

(2020),あるいは関山 他(2022)を参照願いたい。

予報モデルには福島原発事故当時(2011年3月)気象庁が現業メソ予報に使っていた完全圧縮系非静力学メソモデルのJMA-NHMを水平解像度3kmに設定して使った。大気境界層の乱流スキームにはMellor-Yamada Nakanishi-Niino Level-3乱流クロージャーモデル(MYNN3; Nakanishi and Niino, 2006, 2009)を用いた。積雲対流にパラメタリゼーションは用いていない。気象変数はJMA-NHMで予報できるが、オリジナルのJMA-NHMには大気微量成分の移流拡散シミュレーション機能はないので、気象研究所の開発したエアロゾル化学輸送モジュールをJMA-NHMに組み込んだNHM-Chemと呼ばれるバージョン(Kajino et al., 2019)を本研究では利用した。移流スキームにはWalcek and Aleksic (1998)のMPMAアルゴリズム(Mass-conservative, Peak-preserving, Mixing-ratio-bounded Advection Algorithm)を用いた。これにより放射性セシウム濃度の予報が気象予報と同時に可能となった。予報モデルの詳細な環境設定に関してはSekiyama et al. (2017)およびSekiyama and Kajino (2020)を参照願いたい。

JMA-NHM(NHM-Chemバージョン)にEnKF(本研究ではLETKF)を組み合わせたデータ同化システムのプログラムソースコードはKunii (2014)で使われたものをベースにした。Sekiyama et al. (2017)と同じくアンサンブルメンバー数は20個、同化間隔は3時間、共分散局所化には正規分布関数を用いて $1-\sigma$ に相当する距離すなわち $e^{-0.5}$ に減衰する距離を水平50kmに設定した。この50kmという値は試行錯誤によって決定した。実験期間は放射性セシウムが内陸に流れ込んでいた2011年3月14日21時から15日21時に限定した。それ以外の期間の多くは太平洋側へ流れ出していたためである。予報モデルの初期時刻は2011年3月14日9時とし、データ同化サイクルもこの時点から開始した。したがって最初の解析結果は3月14日12時に得られたが、9時から21時までの12時間はスピンアップ期間とした。気象場の初期値と境界値は気象庁全球予報から取り入れた。また、その初期値と境界値には気象庁全球アンサンブル予報の摂動を加えた。濃度場の初期値と境界値はゼロに設定した。濃度値の観測範囲は 0.1Bq/m^3 程度から 1000Bq/m^3 程度にまで及ぶため、データ同化の際には常用対数を使って値を変換し、値の変動幅を小さくした。その際、濃度の観測誤差は1(すなわち1桁)に設定した。これは濃度勾配の空間代表性誤差を加味すれば妥当な大きさである。それ以外のデータ同化に関する諸設定についてはSekiyama et al. (2017)およびSekiyama and Kajino (2020)を参照願いたい。

加えて、放射性セシウム濃度場の精度検証のために、つくば市・新宿区・宇都宮市・高崎市・茅ヶ崎市・福島市における空間線量率(γ 線エネルギー量)の観測値をそれぞれ大学共同利用機関法人高エネルギー加速器研究機構・東京都庁・栃木県庁・群馬県庁・神奈川県庁・福島県庁から頂いた。

4. 結果と考察

濃度観測が3km以内で実施されていたAMeDAS地上風観測地点を選び出し、地上風速(東西風と南北風それぞれ)の観測と解析の比較を行った(図4)。濃度観測には欠測が多いため常に同じAMeDAS地点が選ばれ続けるわけではないことに注意が必要である。濃度観測地点の近傍のみを選んだのは濃度データ同化の影響を明瞭にするためである。残念ながら散布図を肉眼で眺める限りでは実験ごとの明瞭な違いはない。以下、二乗平均平方根誤差(RMSE)で比較する。STDとTEST1は風速場にとって同じ実験条件であるため、当然ながら同じ解析誤差(1.29m/s)である。本研究では風速の観測誤差は気象庁の現業システムの設定に倣って一律 2m/s と設定しているが、もしその値がEnKFの推定する背景誤差の大きさから乖離していれば

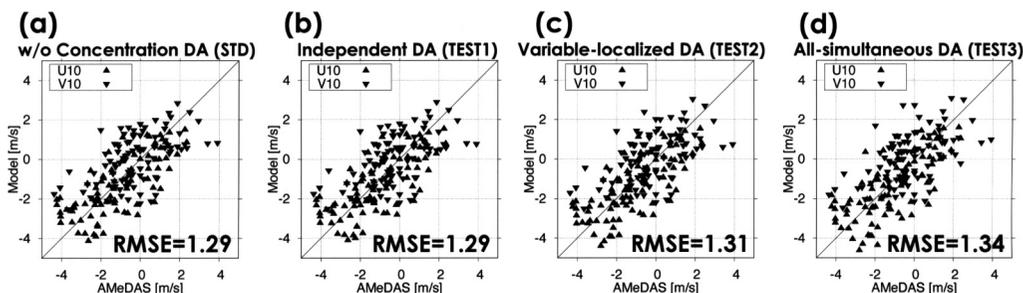


図4. 地上10 m 風速(東西風 U10 および南北風 V10)の AMeDAS 観測値(横軸)と解析値(縦軸)を比較した散布図。サンプルは日本時刻 2011 年 3 月 14 日 21 時から 15 日 21 時までの 3 の倍数時毎の正時。(a)濃度同化無し実験 STD, (b)独立同化実験 TEST1, (c)変数局所化実験 TEST2, (d)完全同時同化実験 TEST3 の結果。二乗平均平方根誤差(RMSE)は東西風と南北風を合わせて計算した。

ば観測の情報あるいはモデルの情報のどちらか一方がデータ同化においてほぼ使われないことになる。どちらが使われないにしてもデータ同化サイクルが最適化されていないことを示すので、観測誤差と背景誤差は大きく乖離していないことが理想的である。そして観測誤差と背景誤差の大きさが同程度の場合、解析誤差は観測誤差よりも若干小さな値になるはずである。したがって上記の RMSE が観測誤差の設定値と乖離しておらず、それでいて 2/3 程度の大きさになっていることは良い傾向である。

期待に反して、変数局所化を行った TEST2 は STD よりも若干誤差が大きくなっている。一方、変数局所化を一切行わない TEST3 は想定どおりサンプリングエラーの影響が大きいと考えられるため最も誤差が大きい。ただしそれらの STD との差は 0.02 m/s あるいは 0.05 m/s と大きくはない。とはいえ、Sekiyama et al. (2017)によると本研究の STD と同一条件でデータ同化を実施した場合、AMeDAS 風観測データを同化に使う実験と使わない実験での地上風解析誤差の変化幅は約 0.1 m/s であるので、 0.05 m/s の悪化は無視できる値ではない。変数局所化実験の TEST2 で STD よりも解析精度が悪化したのは、濃度場と風速場の相関シグナルを共分散推定の不確かさが上回ってしまったためであろう。これは高層オゾンを使った先行研究と同様の結論になってしまった。

では、濃度場の解析誤差は濃度場と風速場の同時データ同化によって改善したのだろうか。図5に濃度場の一例を示す。福島原発事故において関東一円に放射性セシウムが流れ込んだのは2回だけで(3月15日と21日)、うち1回は降雨による激しい沈着が起きたため、明瞭な放射性セシウムのプルームが観測できたのはこの例(3月15日)だけである。3月15日にも北関東山岳域や福島県の一部で弱い降水が見られたが、それらの降水域およびその風下には放射性セシウム観測地点が存在しないため本研究には影響を与えない。これら濃度場に関しては4つの実験間で明瞭な差が見られ、STD だけで北関東の高濃度帯が大きく広がり、神奈川県方面への侵入(penetration)が TEST1, 2, 3 に比べて弱い。濃度観測データ(図中○印)を見ると、現実大気では神奈川県方面へのプルーム侵入は STD の計算結果よりも強く起こっていたようである。

濃度場の再現性を空間線量率との比較で示した結果が表1である。空間線量率は放射性セシウム濃度の推定には直接使えないが、プルームの到達時刻の判別には使える。なぜなら放射性セシウム濃度のピークと空間線量率のピークは一致するためである。関東及び福島における6地点でのプルーム到達時刻の絶対値誤差の合計は STD 実験 > TEST1 実験 > TEST2 実験の順で小さくなる。STD は濃度のデータ同化をしておらず、TEST1 では濃度単独で同化し、

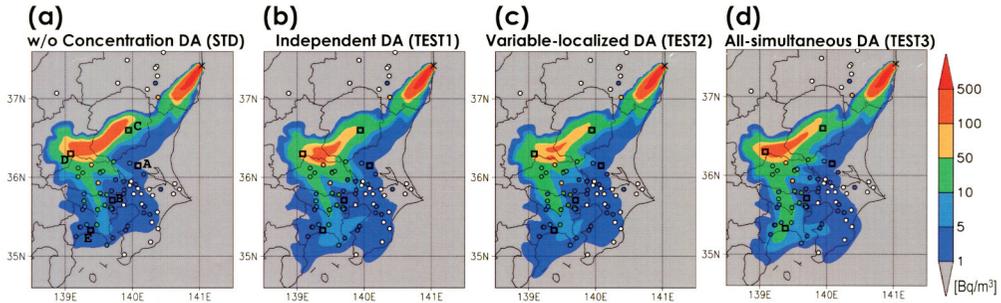


図 5. 放射性セシウム地上濃度の解析値(等値線)と観測値(丸印)の比較例。日本時刻 2011 年 3 月 15 日 11 時から 12 時までの平均値。白抜きの丸印は測定下限値以下あるいは欠測であった地点。□印は空間線量率観測地点(表 1 参照)。(a) 濃度同化無し実験 STD, (b) 独立同化実験 TEST1, (c) 変数局所化実験 TEST2, (d) 完全同時同化実験 TEST3 の結果。

表 1. 関東・東北 6 地点で比較した放射性セシウム濃度のモデル計算ピーク時刻と空間線量率観測値ピーク時刻の相違(日本時刻 2011 年 3 月 15 日に観測されたブルームによる検証)。

観測場所	観測日時	STD 実験	TEST1 実験	TEST2 実験	TEST3 実験
つくば市 (A)	3/15 08-09	-1 時間	-2 時間	-1 時間	-2 時間
新宿区 (B)	3/15 10-11	-1 時間	-2 時間	-1 時間	-2 時間
宇都宮市 (C)	3/15 10-11	-1 時間	±0 時間	±0 時間	-1 時間
高崎市 (D)	3/15 13-14	-2 時間	-1 時間	-1 時間	-2 時間
茅ヶ崎市 (E)	3/15 13-14	-1 時間	-1 時間	-1 時間	-2 時間
福島市 (F)	3/15 17-18	+2 時間	+1 時間	+1 時間	+2 時間
絶対値の合計		8 時間	7 時間	5 時間	11 時間

TEST2 では変数局所化で風速場の情報も取り入れながらデータ同化しているの、期待どおりの結果である。

一方、TEST3 では STD よりも絶対値誤差が大きくなってしまっている。これは濃度場の誤差やサンプリングエラーを気温場や気圧場が拾い、それが力学を通じて風速場の解析精度を悪化させ、それにより移流拡散計算の精度が低下して濃度場を悪化させたと考えれば説明がつく。図 4 で見たとおり TEST3 の地上風解析精度の悪化はわずか 0.05 m/s であるが、極めて小さな風速場の誤差が大きな濃度場の誤差を作り出す可能性があることは Nachamkin et al. (2007) や Sekiyama et al. (2017) も報告している。濃度場の誤差は大気微量成分の流路に沿った領域での風速場誤差を蓄積しているからである。また、相関の対象となる変数が増えるため、濃度場が余分なサンプリングエラーを拾っている可能性も大いに考えられる。これはアンサンブルメンバー数がわずか 20 個であることの弊害である。

本研究では変数局所化を使っても 4D-Var による先行研究と同じく風速場の解析精度を明瞭に向上させることができなかった。また、濃度観測データの扱いにおいて、常用対数を使うだけでなく自然対数を使う／濃度値をそのまま使うなど本稿で紹介した設定以外も試みた。濃度観測誤差の設定でも、本稿で使った値(1 桁)以外の値も大小試みた。だが、いずれの場合も本稿で紹介した解析精度を超えるものはなかった。結局のところ、誤差共分散推定の不確かさの改善が重要であろう。まず、本研究の問題点としてアンサンブルメンバー数を十分に大きく取

れなかったことによる誤差共分散の不確実性が上げられる。さらに本研究では特に移流拡散過程にかかわるモデルの不完全性が誤差共分散の推定に悪影響を及ぼした可能性は高い。したがってデータ同化手法の改良とともに移流拡散過程の不完全性を減らす努力が必要である。

例えば、Sekiyama and Kajino (2020)によると福島原発事故大気汚染シミュレーションでは予報モデルの水平解像度が3 km から1 km へ向上すると、複雑地形の再現性が上がり、放射性セシウム濃度の分布再現性が向上することが報告されている。しかし水平解像度が3 km から1 km になると、計算量の増加は流体力学の CFL 条件を考慮する必要があるため、水平解像度増加比の3乗倍すなわち $3^3 = 27$ 倍の計算機資源が必要になる。本研究でも気象研究所のスーパーコンピュータ資源を多く使わせて貰ったが、さらに27倍の計算機資源をすぐに確保することは難しい。しかしながら計算機性能の向上を待って本研究を高解像度化した実験を実施することは大いに期待が持てる。高解像度化だけでなく、アンサンブルメンバー数の拡大もサンプリングエラーの直接的な低減に期待が持てるが、ノイズの大きさはサンプル数の比の平方根でしか減少しないため、やはり計算機性能の大幅な向上を待つ必要がある。

一方、風速場の解析とは異なり、濃度場の解析精度は変数局所化を使った同時データ同化によって大きく改善した。このことを現実大気において実証した本研究の成果は大きい。もし大気汚染物質濃度の予報サービス向上が本研究の目的ならば、この成果はすぐに役立つものである。その場合、EnKFによる解析値推定計算は1回のステップで2度行い(それぞれ本研究でのSTDとTEST2に相当)、気象場の解析値とアンサンブル摂動はSTDで作成し続けて、濃度場の解析値とアンサンブル摂動はTEST2で作成しつつTEST2で作成した気象場は捨てる、という方法が良いだろう。EnKFの解析計算はアンサンブルシミュレーション計算に比べるとかなり計算機負荷が小さいのでそのような方法も可能である。

5. おわりに

現実大気による変数局所化の効果を検証したが、先行研究と同じく風速場の解析精度を明瞭に向上させることはできなかった。しかし濃度観測値を用いた風速場の解析精度向上は気象予報の技術開発に革新をもたらすものであり、今後も様々な試みが続くだろう。そこでは対象にするトレーサーの種類や使用する数値モデルの不完全性、アンサンブルメンバー数の検討が重要であると考えられる。もし本研究の当初目標が成功していれば、放射性同位体の恣意的放出は当然無理であるが、例えば非放射性の同位体のような無害なトレーサーを人為的かつ定期的環境中へ放出し、その濃度を各地で精密測定することによって、天気予報の精度向上も期待できたはずである。

いかに変数局所化によって重要なシグナルを選択して取りだそうとしても、サンプリングエラーがそのシグナルを上回るような設定では望んだ計算結果を得られないことは心得ておくべきである。ただし実験設定によっては、本研究で濃度場の解析精度だけが向上したように、片利的な結果が得られる可能性はある。そのような成果も目的によっては十分に役立つ可能性がある。変数局所化の実用化においても通常のEnKFと同様にサンプリングエラーを減らすことが重要であり、誤差共分散推定の不確実性の要因特定と改善が求められる。

謝 辞

本研究はJSPS科研費17K00533および(独)環境再生保全機構環境研究総合推進費1-1802の助成を受けたものです。局所アンサンブル変換カルマンフィルタのソースコード利用に関しては気象庁の國井勝氏および横田祥氏の協力を得ました。

参 考 文 献

- Allen, D. R., Hoppel, K. W. and Kuhl, D. D. (2018). Extraction of wind and temperature information from hybrid 4D-Var assimilation of stratospheric ozone using NAVGEM, *Atmospheric Chemistry and Physics*, **18**, 2999–3026.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**, 10143–10162.
- Han, W. and McNally, A. P. (2010). The 4D-Var assimilation of ozone-sensitive infrared radiances measured by IASI, *Quarterly Journal of the Royal Meteorological Society*, **136**, 2025–2037.
- Hunt, B. R., Kostelich, E. and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D: Nonlinear Phenomena*, **230**, 112–126.
- Kajino, M., Deushi, M., Sekiyama, T. T., Oshima, N., Yumimoto, K., Tanaka, T., Ching, J., Hashimoto, A., Yamamoto, T., Ikegami, M., Kamada, A., Miyashita, M., Inomata, Y., Shima, S., Takami, A., Shimizu, A., Hatakeyama, S., Sadanaga, Y., Irie, H., Adachi, K., Zaizen, Y., Igarashi, Y., Ueda, H., Maki, T. and Mikami, M. (2019). NHM-Chem, the Japan Meteorological Agency’s regional meteorology-chemistry model: Model evaluations toward the consistent predictions of the chemical, physical and optical properties of aerosols, *Journal of the Meteorological Society of Japan*, **97**, 337–374.
- Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. and Ballabrera-Poy, J. (2007). 4-D-Var or ensemble Kalman filter?, *Tellus A*, **59**, 758–773.
- Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T. and Ide, K. (2011). “Variable localization” in an ensemble Kalman filter: Application to the carbon cycle data assimilation, *Journal of Geophysical Research*, **116**, D09110.
- Katata, G., Chino, M., Kobayashi, T., Terada, H., Ota, M., Nagai, H., Kajino, M., Draxler, R., Hort, M. C., Malo, A., Torii, T. and Sanada, Y. (2015). Detailed source term estimation of the atmospheric release for the Fukushima Daiichi Nuclear Power Station accident by coupling simulations of an atmospheric dispersion model with an improved deposition scheme and oceanic dispersion model, *Atmospheric Chemistry and Physics*, **15**, 1029–1070.
- Kondo, K. and Miyoshi, T. (2016). Impact of removing covariance localization in an ensemble Kalman filter: Experiments with 10240 members using an intermediate AGCM, *Monthly Weather Review*, **144**, 4849–4865.
- Kunii, M. (2014). Mesoscale data assimilation for a local severe rainfall event with the NHM-LETKF system, *Weather and Forecasting*, **29**, 1093–1105.
- 三好建正 (2008). カルマンフィルタ, 『気象研究ノート第 217 号「気象学におけるデータ同化」』(露木義, 川畑拓矢 編), 69–95, 日本気象学会, 東京.
- Nachamkin, J. E., Cook, J., Frost, M., Martinez, D. and Sprung, G. (2007). Evaluation of dispersion forecasts driven by atmospheric model output at coarse and fine resolution, *Journal of Applied Meteorology and Climatology*, **46**, 1967–1980.
- Nakanishi, M. and Niino, H. (2006). An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog, *Boundary-Layer Meteorology*, **19**, 397–407.
- Nakanishi, M. and Niino, H. (2009). Development of an improved turbulence closure model for the atmospheric boundary layer, *Journal of the Meteorological Society of Japan*, **87**, 895–912.
- Oura, Y., Ebihara, M., Tsuruta, H., Nakajima, T., Ohara, T., Ishimoto, M., Sawahata, H., Katsumura, Y. and Nitta, W. (2015). A database of hourly atmospheric concentrations of radiocesium (^{134}Cs and ^{137}Cs) in suspended particulate matter collected in March 2011 at 99 air pollution monitoring stations in Eastern Japan, *Journal of Nuclear and Radiochemical Sciences*, **15**, 15–26.

- Sato, Y., Takigawa, M., Sekiyama, T. T., Kajino, M., Terada, H., Nagai, H., Kondo, H., Uchida, J., Goto, D., Quélo, D., Mathieu, A., Quérel, A., Fang, S., Morino, Y., von Schoenberg, P., Grahn, H., Brännström, N., Hirao, S., Tsuruta, H., Yamazawa, H. and Nakajima, T. (2018). Model intercomparison of atmospheric ^{137}Cs from the Fukushima Daiichi Nuclear Power Plant accident: Simulations based on identical input data, *Journal of Geophysical Research: Atmospheres*, **123**, 11748–11765.
- Sekiyama, T. T. and Kajino, M. (2020). Reproducibility of surface wind and tracer transport simulations over complex terrain using 5-, 3-, and 1-km-grid models, *Journal of Applied Meteorology and Climatology*, **59**, 937–952.
- Sekiyama, T. T., Kunii, M. and Kajino, M. (2017). The impact of surface wind data assimilation on the predictability of near-surface plume advection in the case of the Fukushima nuclear accident, *Journal of the Meteorological Society of Japan*, **95**, 447–454.
- 関山剛, 寺田宏明, 鶴田治雄, 大浦泰嗣, 森泉純 (2022). モデル検証のための福島原発事故データ, 『気象研究ノート「点発生源からのメソスケール拡散シミュレーション」』(近藤裕昭, 山澤弘実 編), 日本気象学会, 東京 (in print).
- Semane, N., Peuch, V.-H., Pradier, S., Desroziers, G., El Amraoui, L., Brousseau, P., Massart, S., Chapnik, B. and Peuch, A. (2009). On the extraction of wind information from the assimilation of ozone profiles in Météo-France 4-D-Var operational NWP suite, *Atmospheric Chemistry and Physics*, **9**, 4855–4867.
- Tsuruta, H., Oura, Y., Ebihara, M., Ohara, T. and Nakajima, T. (2014). First retrieval of hourly atmospheric radionuclides just after the Fukushima accident by analyzing filter-tapes of operational air pollution monitoring stations, *Scientific Reports*, **4**, 6717.
- Walcek, C. L. and Aleksic, N. (1998). A simple but accurate mass conservative peak-preserving mixing ratio bounded advection algorithm with Fortran code, *Atmospheric Environment*, **32**, 3863–3880.

Simultaneous Data Assimilation of Meteorological Fields and Atmospheric Concentration Fields Using Variable Localization in the Ensemble Kalman Filter

Tsuyoshi Thomas Sekiyama and Mizuo Kajino

Meteorological Research Institute, Japan Meteorological Agency

The ensemble Kalman filter (EnKF) explicitly derives background error covariance matrixes, which are subsequently used to calculate the Kalman gains. The derived background error covariance can be modified during the data assimilation process. For example, the covariance can be arbitrarily increased to spread the ensemble perturbation (i.e., covariance inflation) or decreased according to the physical distance between state variables (i.e., covariance localization). Covariance localization is a key reason why the EnKF works practically with a much smaller number of samples than the degrees of freedom of the system in meteorology. Covariance localization is applicable not only to physical distances but also between state variables with small correlations (i.e., variable localization). In this study, while simultaneously assimilating meteorological data (wind, temperature, pressure, etc.) and atmospheric concentration data using the EnKF, we attempted to zero the covariances for combinations of variables with small correlations by variable localization. This improved the analysis accuracy of the concentration distribution using wind observation information to minimize the effects of sampling errors. In contrast, under the conditions of this study, the analysis accuracy could not be improved for the wind distribution using the information of the concentration observations.