

大気微量成分のデータ同化 ～観測とモデルの融合～

関山剛（気象庁気象研究所）

1. はじめに

気象学とは大気の科学であり、すなわち大気の現象を探る学問である。したがって、その現象を観測しなければまずは話が始まらない。しかし、地球大気は唯一無二、それでいて人類を圧倒する巨大で複雑なシステムであり、そこで起こるすべての現象を観測して把握するなど不可能であろう。さらに言えば気象学は科学でありながら他の分野（物理学や化学や生物学）ならば当たり前の作業である実験や追試というものがほとんど無理なのである。

だから先人は知恵を尽くして複雑な現象の中からエッセンスを抜き出し、技術の粋を集めてそれを観測し、帰納的に分析することによって大気現象を調べてきた。そこから『見ることが出来なかった現象』あるいは『これから起こるかもしれない現象』も類推してきた。しかしながら、この方法ではあまりにも見えないものや予測できないものが多すぎた。

そこで登場したのが大気シミュレーションモデル（以下、モデル）である。モデルは人類にとって既知の物理法則や化学反応をプログラムに組み込んで、高性能なコンピュータを用いてそれを計算することにより大気現象をシミュレートする。コンピュータの計算性能が許せば、地球上のどの場所でもシミュレートすることができるし、未来の現象の予測も可能である。

また、ある大気現象の再現をモデルが失敗した場合でも（例えば極渦崩壊とか局地豪雨とか黄砂発生の予測は難しい）、その原因をモデルの中で探ることによって現象のメカニズムを調べることができる。さらに現実では起こりえない設定（コリオリ因子や太陽放射を変化させるとか、地形を大規模に改変するとか）を行うことによって、なんと仮想実験も可能になる。大気中の温室効果ガス濃度も自由に設定できるから、気候変動予測もできてしまうのだ。

しかし、モデルの中の世界はバーチャルリアリティであり、所詮は作り物の世界だ。モデルが再現する大気は決して現実大気と同一の挙動を示すことはない。モデルによるシミュレーション結果が現実とずれる原因は大きく分けて二つある。一つはモデル自体の精度の問題、もう一つは大気が複雑系であ

りカオスであることの問題である。どちらにしても、モデルを動かしているだけでは現実の現象をままねく『見る』ことなど不可能である。

ちなみに天気予報の精度を上げようとした場合、最近の予報モデルの性能はそこそこ良いので、大気のカオス性すなわち鋭敏な初期値依存性が大きな問題となるケースが多い。したがって、精度の高い初期値を入手すること、すなわち過去のある瞬間における大気全体の状態を極めて正確に把握することが重要となる。ところが観測だけでそれを果たそうとするのは、我々人類の技術力と経済力ではほとんど無理だろう。

ではどうしたらよいのか。我々の手元には2種類の不完全な情報がある。一つは観測、もう一つはモデル結果である。前者は紛れもない真実を教えてくれるが、欲しい情報を欲しいだけ手に入れることは無理である。後者はどんな場所／どんな時間／どんな物理量も再現し情報を提供してくれるが、真実の世界ではない。この相補的な2種類の情報を融合することができたら、それぞれの欠点を打ち消しあってくれないのではないだろうか。そう、それを可能にするのがデータ同化である。

2. データ同化とは

融合といっても、足して2で割るような単純な方法では無理である。数学的な理論に裏打ちされた統計的に意味のある方法を用いるべきであろう。ベイズ統計学によれば、あるモデル結果（これを統計学用語で第一推定値と呼ぶ）と観測値、およびそれらの確からしさの情報が与えられれば、第一推定値（＝モデル結果）と観測値のどちらよりも確からしい解析値を得ることができる。すなわち情報の相補的融合が可能であることは数学的には明らかなのである。

このとき高精度の第一推定値あるいは観測値があればあるほど、精度の高い解析値を計算することができるのは当然である。またそれだけでなく、理論的な考察によると、個々の第一推定値同士の相関（＝関連性の強さ；例えばある地点の風向風速と別の地点の風向風速の関連性、あるいはある地点の異なる時刻の気温の関連性）を見積もることができ

ば、複数の観測値とそこから遠く離れた場所・時間の第一推定値とを同時に融合することも可能であることが示されている。すなわち、第一推定値と観測値は同数存在する必要はなく、観測値が第一推定値に比べて圧倒的に少なかったとしても、例え観測値が得られない場所や時間があつたとしても、第一推定値（モデル結果）と観測値の融合は可能なのである。

このような情報融合の計算を実現するアルゴリズムにはこれまで様々なものが提案されてきた。もちろんその中には精度の高い手法もあれば低い手法もある。一般的には精度の高い手法ほどコンピュータで計算する際の計算負荷が莫大になるので、計算機資源をどれほど用意できるかは重要である。計算アルゴリズムの中には理論的には極めて高精度な解析値を得られることが予言されているが、現代の計算機性能のレベルでは（例え世界最高速のスパコンを使おうとも）実際的には計算不可能なものもある。気象学の分野で実用化されているものの中で最新かつ最高精度の計算アルゴリズムには、4次元変分法と呼ばれるものとアンサンブル・カルマン・フィルタと呼ばれるものがある。それらの仕組みは次章で概説する。

これらモデル結果と観測値を融合して解析値を得るための統計数理的手法はデータ同化と呼ばれる。「同化」という用語は、例えば生物学においては、「炭酸同化作用」というように使われている。ちなみに炭酸同化とは光合成のことである。炭酸（＝二酸化炭素）を体内に吸収して、それを原料に有機物を合成し、その有機物を自らの体の一部に取り込むことから「同化」と呼ばれている。それと同じように、観測値（データ）を取り込んでモデルの状態変数（＝モデルの中の仮想現実の状態）と一体化させるから「データ同化」なのだ。

データ同化は観測とモデル結果の双方の情報を補完し、現実とよく一致する（と確からしく推定される）大気状態を再現する。言い換えればデータ同化とは観測とモデルという気象学における伝統的な二つの方法論の橋渡し役であり、それら伝統的なアプローチ単独では得られない視点を人類に与える。データ同化が情報のイノベーションと呼ばれる所以である。

3. データ同化のしくみ

ところでデータ同化という解析手法は気象学だけで使われているわけではない。海洋学や地震学といった気象学に近い分野でも近年盛んに利用されている。また、通信工学や制御工学の分野では、ノ

イズに乱された観測データから意味のある信号を取り出すためのフィルタとしてデータ同化と同じ統計的手法が利用されている。身近な例を挙げると、カーナビが極めて高精度に自動車の位置を表示してくれるのはカーナビの中でデータ同化の計算が行われているからである。人工衛星の制御にもデータ同化は利用されており、さらには経済学や金融工学にも応用されている。したがって以下の章節ではデータ同化の歴史とアルゴリズムについて事細かに述べることは避け、気象学に関連するトピックスに話題を絞って話を進めることをご承知おき願いたい。

3-1. 基本的な方法

データ同化と聞くと難しく考えてしまうかもしれないが、その基本的な仕組みは重み付き平均である。話を簡単にするために1変数モデルを考えてみる。その変数 x はある場所・ある時刻の気温あるいは気圧、あるいは黄砂濃度でも何でも構わない。そして、その変数 x の推定値が2つ得られたとしよう。それらを次式のように x_1 と x_2 とする。この2つの値はモデル結果および観測値と考えればよい。現実大気においては変数 x の本当の値（＝真値） x^t を人類は決して知ることができない。たとえ測定誤差0の観測値を得られたとしても、空間代表性誤差の排除はできない。したがって x_1 と x_2 の誤差 ε_1 と ε_2 の正確な値は知ることができないのだ。

$$\begin{aligned} x_1 &= x^t + \varepsilon_1 \\ x_2 &= x^t + \varepsilon_2 \end{aligned} \quad (1)$$

しかし、我々は真値 x^t になるべく近い解析値 x_a を推定したい。そこで(1)式に対して次の2つの条件を仮定する。この仮定が現実世界にそのまま当てはまるとは言い難いが、それほど現実離れしているわけでもない。したがって、とりあえずは受け入れてみよう。

その1) x_1 と x_2 の平均値はどちらも x^t に等しい。平均値（期待値）を表す記号を $\langle \rangle$ とすると、

$$\langle x_1 \rangle = \langle x_2 \rangle = x^t \quad (2)$$

となる。したがって、 ε_1 と ε_2 の平均値は

$$\langle \varepsilon_1 \rangle = \langle \varepsilon_2 \rangle = 0 \quad (3)$$

である。

その2) x_1 と x_2 の誤差 ε_1 と ε_2 は無相関であり、それぞれ完全に無関係に変動する。つまり、

$$\langle \varepsilon_1 \varepsilon_2 \rangle = 0 \quad (4)$$

となる。

以上の仮定のもとで、解析値 χ_a を (5) 式のような形 (重み付き平均) で表すと、重み α の最適解は解析値 χ_a の誤差分散 σ_a^2 を最小化させることによって得られる。これを線形最小分散推定と呼ぶ。

$$\chi_a = \alpha \chi_1 + (1 - \alpha) \chi_2 \quad (5)$$

χ_a の誤差分散 σ_a^2 は χ_a と χ^t の差の二乗の平均であるから、

$$\begin{aligned} \sigma_a^2 &= \langle (\chi_a - \chi^t)^2 \rangle \\ &= \alpha^2 \langle \varepsilon_1^2 \rangle + 2\alpha \langle \varepsilon_1 \varepsilon_2 \rangle \\ &\quad + (1 - \alpha)^2 \langle \varepsilon_2^2 \rangle \\ &= \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2 \quad (6) \\ &\quad (\because \langle \varepsilon_1 \varepsilon_2 \rangle = 0) \end{aligned}$$

となる。ここで σ_1^2 と σ_2^2 はそれぞれ χ_1 と χ_2 の誤差分散である。(6) 式の右辺を α で微分し、それを 0 と置いて方程式を解けば σ_a^2 が最小となる時の重み α の値が以下のように求められる。

$$\alpha = \sigma_2^2 / (\sigma_1^2 + \sigma_2^2) \quad (7)$$

この α の値を (5) 式に代入して計算した解析値 χ_a は最適推定値とも呼ばれる。また、 σ_1^2 と σ_2^2 はモデル結果および観測値の誤差分散すなわち確かからしきであり、何らかの方法で見積もってやる必要がある。これら一連の線形最小分散推定は図 1 のように解釈することが可能である。

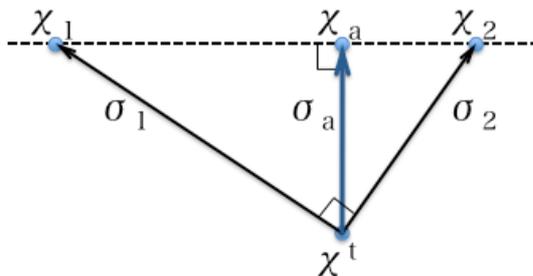


図 1. 線形最小分散推定における真値、推定値、および解析値の幾何学的関係。

2つの推定値 χ_1 と χ_2 は真値 χ^t を原点として長さがそれぞれ σ_1 と σ_2 のベクトルと考えられる。また、 χ_1 と χ_2 の誤差は無相関であるので、この2つのベクトルは直交している (直角の関係)。このとき最適に推定された解析値は χ_1 と χ_2 を結ぶ直線上に位置し、 χ^t からの距離 (= 誤差) が最短 (= 最小) となる点すなわち χ^t から下ろした垂線上にある。この図から明らかなように、ベイズ統計学の教えるとおり、解析値の誤差 σ_a は元の2つの推定値の誤差 σ_1 と σ_2 よりも小さくなっている。

実際のデータ同化では、変数 χ は空間的に横にも縦にも数多く分布し、観測値も複数存在している。複数の観測値を複数のモデル結果と融合するためには一度に複数の重み α を使わなければならない。そのためには個々の χ の誤差分散だけでなく、異なる地点の χ の間の関連性すなわち共分散の大きさも見積もらなければならない。例えば、ある地点の気温はその近くの地点の気温と良い相関を持つが (= 共分散が大きい)、遠く離れた地点の気温とは相関が低い (= 共分散が小さい)。このため重み α を算出する作業は複雑になり、ベクトルや行列の計算が必要になってくるが、基本的な考え方は上記のような1変数の場合と同じである。

では、実際問題として離れた場所の観測値に対する重み α を知るにはどうすればよいのだろうか。最も簡単な方法は、細かなことは気にしないで経験的に適当な α を与えてしまうことである。今を去ること数十年前、数値天気予報の揺籃期にはそんな方法も使われた。しかし、もうちょっと精度の高い α の値が欲しい。そこで、重みを適当に与えるのではなく、長期間のモデル結果と観測値の統計平均をもとに共分散の分布を見積もり、重み α の値を計算する方法が使われるようになった。統計的に尤もらしい α の値を事前に準備しておくのである。この方法を最適内挿法 (optimal interpolation, OI) と呼ぶ。統計的に最適なデータ内挿を行うという意味である。OI におけるモデル結果 (格子点値) と観測値の関係を模式的に示したのが図 2 である。ここでは1つの格子点値に対して、複数の観測値を重み付け平均して解析値を作成している。

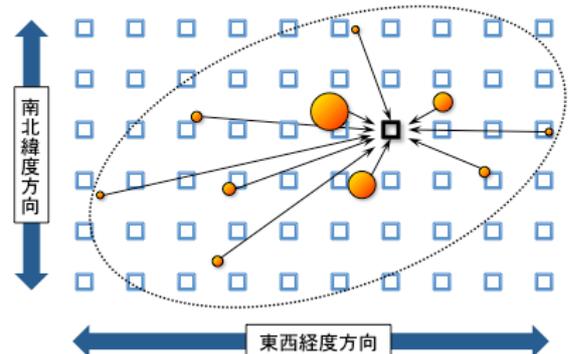


図 2. 最適内挿法における観測値の選別。モデル結果 (格子点値) を□、観測値を●で示している。●の大きさは重み α の相対的な大きさを表している。楕円の外の観測は使用しない。

OI は比較的単純な手法で計算機負荷が低いため、1970年代には世界中の気象局で主流のデータ同化法となった。OI を採用した気象局の中には日本の気象庁も含まれる。そして、気象庁では 2001 年ま

でOIが使われ続けた。しかし、OIには致命的な欠点が幾つかあった。

一つは、モデルの予報する物理量と観測する物理量は同じものでなければならないことである。つまりモデルが予報する気温・気圧・風向風速・湿度など以外の物理量を観測値として得ても（例えば人工衛星による放射輝度観測など）、データ同化ができないのだ。また異なる物理量、例えば気温の観測値を風向風速のモデル結果にデータ同化するというのも無理である。これは実に不便である。

もう一つの問題点は、本当ならば大気の流れに依存して時々刻々変化するはずの共分散の分布すなわち重み α の分布が固定されてしまっていることである。台風や黄砂といった頻度は低いが一度起きれば大きな変化を伴う現象をデータ同化しようとした時、こんなことでは精度の良い解析は不可能である。

さらに、本来モデルの中では、仮想現実の世界とはいえ、すべての状態変数が物理学的なバランスのとれた状態になっている。しかしOIはそのバランスを無視して観測値をねじ込むため、得られた解析値は（個々の値の誤差は小さくとも）全体的な均衡の崩れた状態になってしまう。これを数値天気予報の初期値として用いると、モデルがその不均衡状態を解消するように初動するため、多くのノイズが発生して予測精度が著しく低下してしまう。

これらの欠点を克服するため、OIを超える様々なデータ同化手法が開発され実用化されてきた。その第一陣が変分法である。

3-2. 変分法

実は前節で解説した線形最小分散推定のほかに、統計数理的に解析値を推定する方法がもう一つある。それは最尤推定と呼ばれ、変分法と呼ばれるデータ同化手法の基礎となっている。最尤とは最ももっともらしいという意味であり、最小分散推定がなるべく誤差分散を小さくしようと努力する手法であるのに対し、最尤推定は解析値が真値に近づく確率を最大化させようと努力する手法である。両者は理想条件下では同じ解を与えることが数学的に証明されている（実際のデータ同化は理想条件とはほど遠い条件で行わなければならないが…）。

具体的には、次式のような評価関数と呼ばれるものを定義し、その評価関数 $J(x)$ の値を最小化させるような変数 x の値を解析値 x_a とする。

$$J(x) = (x_1 - x) / 2 \sigma_1^2 + (x_2 - x) / 2 \sigma_2^2 \quad (8)$$

実際のデータ同化では変数 x は1つの値ではなく、大気全体の状態あるいは観測全体の値を表す。したがって、最適な x を探すためにはモデルの大気の状態をすべて一括して変化させ評価関数 J の最小値を見つけ出す必要がある。その手法を変分法と呼ぶ。そもそも変分法とは最速降下線問題などを解く際に用いられる微分積分法に似た数学上の技巧であり、通常の変分法が数の関数を扱うのに対して、関数の関数を取り扱うのが変分法である。

ところが、変分法を気象学のような大規模問題でそのまま解くことはほぼ不可能である。変数 x は大気の任意の場所の任意の物理量を表す必要があり、例えば現代の全球大気モデルでは水平方向の格子点数が東西と南北それぞれで百個～千個程度、鉛直層の数が数十層、その各格子点上に気温・気圧・風向風速・湿度・大気微量成分濃度などの物理量が設定されているので、ざっと計算してモデルの中の変数の数は1000万個のオーダーとなる。そして、それらすべての値が物理法則や化学反応によって相互に影響し合っているのだ。

そこで、アジョイント法と呼ばれる近似解法の研究が1970年代から1980年代にかけて始まった。アジョイントとは代数学の用語で随伴行列のことである。随伴行列を用いて評価関数の勾配を見積もり、その勾配の値を用いて反復計算を大量に繰り返すことによって評価関数の最小値を探し出す仕組みである。

そして最初に実用化された変分法データ同化システムは米国気象局による3次元変分法(3-Dimensional Variational Method, 3D-Var)であった(1991年)。3D-Varではある瞬間における東西・南北・上下の3次元空間全体のモデル結果とその瞬間における観測値のすべてを一括して取り扱い、異なる物理量もすべて同時に解析する。さらに、評価関数 $J(x)$ には既知の物理法則に基づいた拘束条件を付け足して計算することも可能である。その結果、3D-Varによって得られた解析値全体の物理的バランスはOIに比べて著しく向上した。

また、3D-Varでは評価関数 $J(x)$ を計算する際、解析値と観測値の差を計算しなければならないが、解析値と観測値はそのまま直接比較する必要はなく、もし観測値が解析値(気温・気圧・湿度など)から導出することのできる物理量であれば(例えば放射輝度など)、その導出した値と観測値を比較して差を計算し、評価関数の最小値を探索することができる。ということは、モデルで扱う物理量と観測で扱う物理量は相互に変換可能でさえあれば異なる

っていても構わないのである。これは極めて便利である。ちなみにモデル物理量を観測物理量へ変換する計算プログラムを観測演算子と呼ぶ。

しかし、3D-Var にはまだ改良すべき点があった。3D-Var では本来ならば大気の流れに依存して時々刻々変化するはずの誤差共分散の分布がOIと同様に固定されたままなのである。これを改良するには、変分法の計算を3次元に限定せず、時間軸を含めた4次元に拡張すればよい。これを4次元変分法(4-Dimensional Variational Method, 4D-Var)と呼ぶ。

ところが4D-Varはモデルシミュレーションの反復計算を大量に必要とするため、OIや3D-Varに比べて計算機負荷が極めて高い。そのため4D-Varの実用化は、計算負荷を下げるスキームの研究が進み、かつ大型計算機の性能が飛躍的に高まるまで待たなければならなかった。そして1997年、ついにヨーロッパ中期予報センター(ECMWF)が4D-Varを現業化した。それを追って、日本の気象庁も2002年に4D-Varを導入した。現在では世界を代表する気象局のほとんどは4D-Varを採用し、データ同化アルゴリズムの主流となっている。

4D-Varの基本的な仕組みは3D-Varと同じであるが、解析範囲(=評価関数 $J(\chi)$ の最小値を探索する範囲)が4次元時空間に拡張されている。時間方向の解析範囲を同化ウィンドウと呼ぶ。4D-Varによる解析の概念を図3に示す。

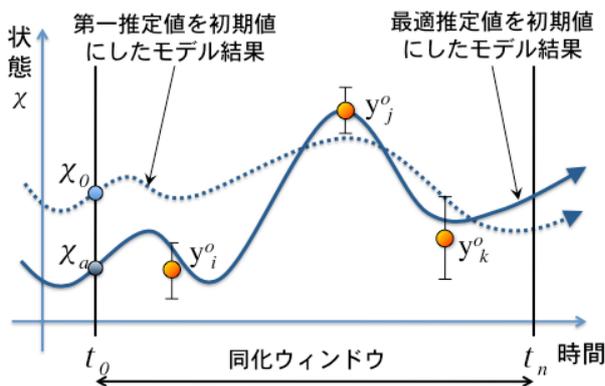


図3. 4次元変分法の概念図。 χ は状態変数、 y^o は観測値(縦線は誤差の大きさ)。ここでは同化ウィンドウ(t_0 から t_n)の中で3つの観測値が得られている。2本の曲線はそれぞれ第一推定値と最適推定値(=解析値)を初期値として時間積分を行ったモデル結果の軌跡。

4D-Varは3D-Varの利点に加えて、以下の重要な特徴を持つ。一つは観測値を観測した時刻に合わせてデータ同化することが可能である。当たり前のようだが、実はOIも3D-Varもこれが出来ない。少しずつ異なった時刻に観測した観測値が多数得

られた場合、それらすべてを同時刻の観測として扱わなければならなかったのである。4D-Varでは同化ウィンドウの中の様々な時刻に観測された様々な物理量との間で『最も整合性が高くなるようなモデルシミュレーション』が実行される初期値が解析値として得られる。

もう一つの4D-Varの特徴はさらに重要である。評価関数の最小値探索のための反復計算は時間軸方向にも行われるため、その間に誤差共分散の分布が修正され、大気の流れに合致した誤差共分散が暗黙のうちに推定されるのである。OIや3D-Varが静的なデータ同化だとすれば、4D-Varは人類が初めて手に入れた動的なデータ同化である。

以上のように良いことづくめのような4D-Varであるが、痛恨の欠点の一つがある。開発コストが非常に高いのである。評価関数の最小値探索の際に観測演算子だけでなくモデル本体の反復計算も行わなければならないため、モデルのアジョイントを用意する必要がある。巨大で複雑な大気モデルのアジョイントを開発するためには多大な人的資源と時間が必要とされる。アジョイントはモデルの構造に依存するため、モデルが改変されればアジョイントも改変しなければならない。この開発コストの高さが一般研究者への4D-Varの普及を拒んできた。

このような状況の中、2000年代になって彗星のごとく現れたのがアンサンブル・カルマン・フィルタ(カルマン・フィルタの一種)である。

3-3. カルマン・フィルタ

カルマン・フィルタは1960年にKalmanによって提案された最適制御のためのフィルタリング理論であるが、その名を一躍有名にしたのはアポロ計画で月探査船の軌道制御に用いられたことによる。カルマン・フィルタの基本的な原理はOIと同じ線形最小分散推定であり、時々刻々の重み α を動的に推定しながら重み付き平均の計算を行うことで精度の高いデータ同化を実現する。

カルマン・フィルタを気象学で利用する試みは古くから行われていたが、実用化は困難であった。カルマン・フィルタは4D-Varと異なり、時間の流れ(=大気の流れ)に依存する誤差共分散の分布を陽に(=明示的に)計算する必要があり、気象学に應用しようとすればモデルの中の変数の個数と同じ次元数の行列計算(すなわち1000万×1000万の配列要素を持つ行列の計算)を必要とする。これは現在の最新の計算機環境でも実現不可能である。

そのため、計算量が少なくすむ近似的なカルマン・フィルタのアルゴリズムが1990年代以降に幾

つか提唱された。その中で特に注目されたのがアンサンブル・カルマン・フィルタ (EnKF) である。EnKF は誤差共分散の時間変化をアンサンブル予報の結果によって近似し、計算量を劇的に減らしている。気象学におけるアンサンブルとは予報の初期値に微小の差を加えて少しずつ状態の異なった数多くの初期値を用意し、そこから複数のモデル計算を行う手法である。大気予報計算を一回だけのモデル計算で行おうとすると、特に中長期予報の場合、大気の鋭敏な初期値依存性が原因で予報を大きく外す場合がある。アンサンブルを使えば、断言的な予報を行うのは無理な場合でも、将来の大気の状態について幾つかの可能性を提示することができる。そのため今では世界の主な気象局はアンサンブル予報を導入している。

カルマン・フィルタのアルゴリズムは2つのステップに大別できる。一方は誤差共分散の時間変化を計算するプロセス、もう一方は線形最小分散推定に基づいて解析値を計算するプロセスであるが、計算量が膨大なのは前者である。ところがそのプロセスで必要な情報は結局のところ将来のモデル誤差の大きさなのであるから、アンサンブル予報で得られる情報と同質である。それがアンサンブル予報によって誤差共分散の時間変化を近似できる所以である。

その上、EnKF で特筆すべき特徴の一つは、データ同化に関する計算とモデル計算が完全に独立しており、計算機上での実装が4D-Var に比べて極めて容易であることだ。データ同化に必要な情報はアンサンブル予報結果だけであるから、4D-Var のようにモデルの構造に依存するアジョイントを用意する必要がない。反復計算を行うことはないから、観測演算子のアジョイントすら必要ない。さらに、次の時間ステップにおいて必要となるアンサンブル計算の初期値の微小差違 (これを初期摂動という) は EnKF が自動的に生成してくれる。その初期摂動はランダムに決定されるのではなく、大気のカオス性に基づいて最も適切な場所/適切な物理量/適切な大きさを EnKF が判断して、物理的な整合性を保った状態で作成される。

ではいったい幾つのアンサンブルメンバーを用意すれば実用的な近似が可能なのだろうか。何万・何十万のメンバーが必要となれば実用化は無理である。ちなみに一般的なアンサンブル予報では数十個程度のメンバーが用意される。そして、幸いなことにこれまでの研究ではそれと同程度のオーダー (数十から百個程度) のメンバーを計算すれば EnKF でも 4D-Var に匹敵する精度でデータ同化が

可能であることが示されている。数百万個の状態変数を持つ大気モデルの予報誤差が僅か数十個のアンサンブルメンバーで近似できることは驚異である。その結果、EnKF は 4D-Var と同程度の計算機負荷で実用化可能となったのである。EnKF による解析の概念を図4に示す。

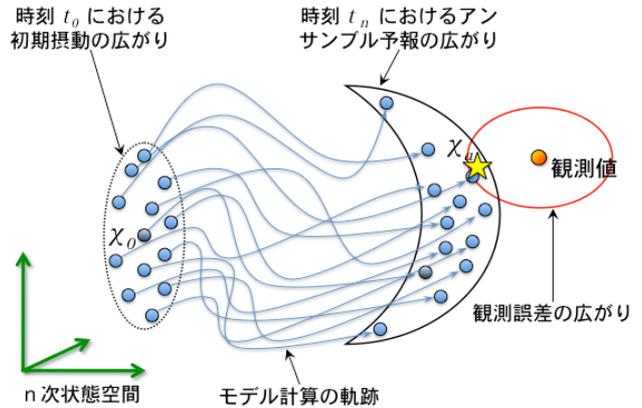


図4. アンサンブル・カルマン・フィルタの概念図。 x_0 は初期値の平均、赤●印は時刻 t_n における観測値、 x_n (黄色の☆印) は最適推定値 (= 解析値)。1本1本の曲線矢印は個々のアンサンブルメンバーについて計算したモデルの状態の軌跡。

EnKF と 4D-Var との優劣は現在大きな議論的となっている。今のところ現業で天気予報に使われている 4D-Var が完全に EnKF と入れ替わった例はないが、ECMWF や日本の気象庁などでは精力的に EnKF の試験が行われている。また、EnKF の実装が 4D-Var に比べて極めて容易であることから、大学など小規模なグループでの研究では EnKF が盛んに利用され始めている。

EnKF の利点は実装の容易さだけではない。実用の立場から見ると、EnKF の導入にはアンサンブル予報の初期摂動を効率よく作成してくれるという副次効果がある (そもそも普段からアンサンブル予報を現業化していれば、少ない追加コストでデータ同化も行える)。研究的な立場から見ても、誤差共分散情報が追加コストをかけずに手に入ることで様々な実験が可能である。また、オリジナルのカルマン・フィルタでは OI や 3D-Var のように観測値が手に入る毎に解析計算を実行する必要があったが、4次元アンサンブル・カルマン・フィルタという手法が開発されてからは複数の時刻の観測値を一括してデータ同化することが可能となり、この点でも 4D-Var と比べて遜色ないと言ってよい。

ところで実装の容易さという点では、天気予報の分野以上に大気化学の分野で EnKF は福音であった。4D-Var で必要なアジョイントの作成にはモデルの線形化 (= 単純化) の作業が必要なため、非線

形性の極めて強い(=方程式が極めて複雑な)化学反応や拡散沈着過程を数多く含む大気化学モデル(オゾンやエアロゾルなど大気微量成分を取り扱うモデル)ではアジョイントの作成がほぼ不可能であった。しかし現在ではアジョイントの要らないEnKFの出現によって大気微量成分のデータ同化が気象学における最先端の研究の一つとなっている。

4. データ同化の実際

4次元 EnKF による大気微量成分のデータ同化の一例として、ライダー観測を用いた黄砂エアロゾルの解析を紹介しよう。このデータ同化には黄砂/硫酸塩/海塩/黒色炭素/有機炭素のエアロゾルをそれぞれ取り扱うことのできる全球大気モデルと、人工衛星あるいは地上に設置されたライダー装置によって得られたエアロゾル減衰後方散乱係数の観測値を用いた。

モデルでは減衰後方散乱係数を直接には予報しないので、観測演算子プログラムによってモデルの状態変数から変換した。また、この解析では大気中の黄砂濃度だけでなく黄砂発生源(ゴビ砂漠、タクラマカン砂漠など)における黄砂舞い上がり強度を推定する係数(これをモデルパラメータという)の最適値も EnKF によって動的に推定している。このように大気の状態量だけでなくモデルパラメータの値も動的推定が可能なのは EnKF や変分法など高度なデータ同化の特徴である。

まず、人工衛星 CALIPSO (米国) に搭載されたライダーによって得られた観測値をデータ同化した結果を図5に示す。ライダーは鉛直解像度と時間解像度は非常に高いが、直上もしくは直下の一点しか観測できない。そのようなタイプの観測値をデータ同化するにはOIや3D-Varでは非力過ぎ、4次元 EnKF のような強力なデータ同化スキームが必要となる。図5から明らかのように、CALIPSOデータ同化の結果は気象官署での黄砂観測や別の人工衛星観測と極めて良い一致を示す。

次に、地上ライダーによって得られた観測値を先ほどと同じ解析システムによってデータ同化した結果を示す(図6)。用いた地上ライダーは国立環境研究所によって運営されている東アジアネットワークの8地点である。比較のため、CALIPSO観測値のデータ同化結果、データ同化をしなかった場合の推定結果、他の人工衛星による観測結果を示す。データ同化によって再現性が改善していることが見て取れるが(特に朝鮮半島付近)、地上ライダーのデータ同化結果にバイアスが見られるため、その

改善が今後の目標である。これらのデータ同化結果は黄砂予測の初期値として利用予定である。

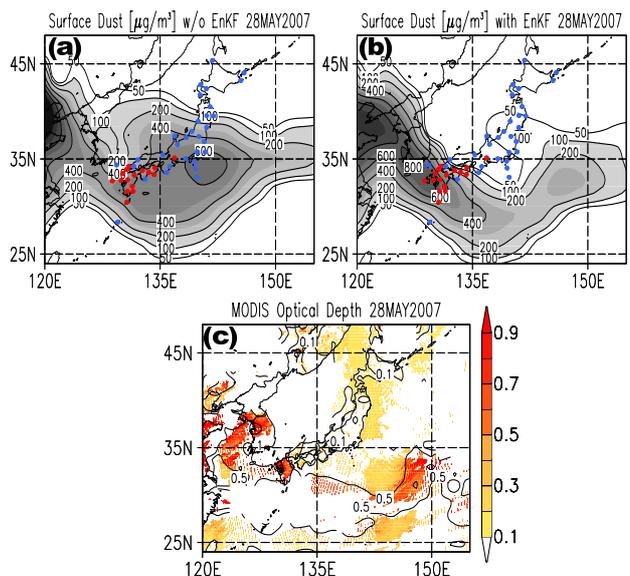


図5. 2007年5月28日に黄砂を実際に観測した(しなかった)日本の気象官署を赤○印(青○印)で表し、そのとき推定された地表での黄砂濃度を等値線で表している。灰色の領域は特に黄砂が濃い場所。等値線は(a)データ同化無しでモデルが推定計算した結果、および(b)CALIPSOデータ同化によって解析された推定結果。(c)は同じ日に別の人工衛星によって観測された光学的厚さ(=空気の濁り具合の鉛直積算値)の観測値。

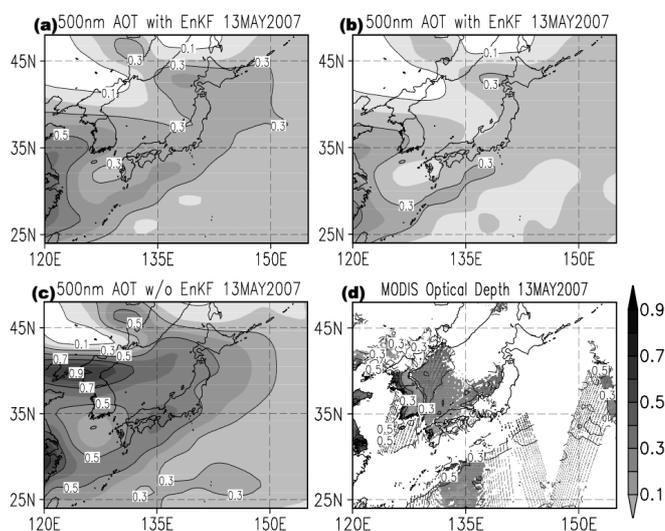


図6. 等値線は2007年5月13日における光学的厚さの値。主に黄砂の影響で空気が濁っている。(a)CALIPSOの観測値をデータ同化した解析結果、(b)国立環境研ライダーネットワークの観測値をデータ同化した解析結果、(c)データ同化無しでモデルが推定計算した結果、(d)別の人工衛星によって独立に観測された結果。

参考文献

さらにデータ同化について勉強をしたいという諸氏には以下の教科書(和書)を推薦する。
淡路敏之, 蒲地政文, 池田元美, 石川洋一, 2009: データ同化. 京都大学学術出版会, 284pp.
露木義, 川畑拓矢, 2008: 気象研究ノート第217号 気象学におけるデータ同化. 日本気象学会, 277pp.