

データ同化

——観測データをシミュレーションの世界に融合させる

Data Assimilation: Incorporating Observations into Simulated Variables

伊藤耕介

Kosuke Ito

琉球大学理学部助教、日本気象学会台風研究連絡会代表／1981年生まれ。京都大学理学部卒業。
同大学院理学研究科修了。博士(理学)。気象学。共著に『台風の正体』。日本地震学会論文賞受賞(共著)。

天気図には地上気圧の分布が書いてある。低気圧の近くでは天気が悪くなり、高気圧に覆われているところではおおむね天気がよいので、概況を知るには便利な図である。しかし、よく見てみると、観測がほとんど得られないはずの日本のはるか南の海上にもきっちりと数字が入っていることに気付く。実は、天気図上の地上気圧の分布は観測値ではなく、観測値を数値シミュレーションモデルの世界に融合させ、最適な状態として推定した結果がもとになっている。このような融合を「データ同化」という。

そもそも、気象学における数値シミュレーションとは、コンピュータ上の格子で区切られた仮定の地球上で風速・気圧・湿度・温度・水蒸気量・液体の水の量といった大気の状態を定義し、ニュートンの運動法則や質量保存の法則といった物理法則に従って将来に向かって計算していくことを指す。現在使われる主要なデータ同化手法は、これらの物理法則と矛盾がないことを条件として課したうえで、観測値とも数値シミュレーション結果とも整合的になる状態を探索していく。

データ同化にはさまざまな用途がある。そのひとつは、ばらばらに点在している観測情報を、規則正しい格子点上に配置させることである。天気図の作成にあたっては、この利点が生かされている。また、天気予報の精度向上もデータ同化の大きな利点のひとつである。天気予報を行うにあたっては、現在の大気の状態の推定値を初期状態として数値シミュレーションシステムに入力する必要がある。そこで、天気予報を行う気象庁などの機関は、直近の観測データを同化して最適推定を行ったのち、それを入力値として将来予測を実施している。つまり、データ同化システムの改善は、近年の天気予報の精度向上にも大きく貢献してきたのである。さらに、データ同化を通じて、過去の大気循環や成層圏の全体の構造を復元する試みも行われている。過去の限られた観測データを最大限有効活用し、大気や

海洋全体の状態を推定しようというのである。このような試みは、地球温暖化に代表される気候変動の研究に役立てられている。

データ同化にはさまざまな手法があるので、誌面の都合上、すべてについて詳しく述べることはできないが(淡路ほか『データ同化』[京都大学出版会]参照)、誤解を恐れずに言えば、データ同化手法は「観測値やシミュレーション結果に見られる物理量の誤差の分散や共分散を利用し、最適推定を行う数学的な仕組み」ということができる。このことを、①あるひとつの物理量に関してシミュレーション結果と観測値が同時に得られた場合、②シミュレーションで複数の物理量が計算されたのに対し、観測値がひとつだけ得られた場合、に分けて説明しよう。

まず、①の場合だが、シミュレーション結果を x_s 、その誤差標準偏差を σ_s 、観測値を x_o 、その誤差標準偏差を σ_o とする。このとき、

$$x_d = x_s + \frac{\sigma_s^2}{\sigma_o^2 + \sigma_s^2} (x_o - x_s)$$

という変数 x_d をつくると、その誤差標準偏差 σ_d は、 σ_s および σ_o のいずれよりも小さくなることが証明できる(証明は「データ同化」第1章を参照。厳密にはいくつかの仮定が必要)。ここでは、 x_d がデータ同化結果に相当する。すなわち、観測値と数値シミュレーション結果という二つの情報が手に入るとき、両

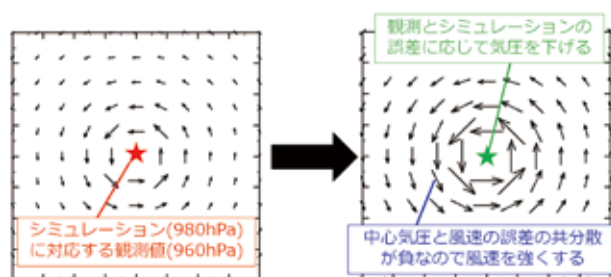


図1 台風観測におけるデータ同化の応用例

者を融合させたデータ同化結果 x_s の誤差の方が、観測値の誤差よりも小さいと期待されるのである。

②の例として、台風における観測を考えてみる。数値シミュレーションによって台風が再現され、シミュレーション上の中心気圧が p_s 、中心から50kmの地点で風速が v_s であったとしよう¹⁾。ここで、台風の中心が洋上の観測ブイ上を通過し、中心気圧が p_o ($< p_s$) であることがわかったとする。データ同化において、中心気圧を修正すべきなのは当然だが、風速になんの修正も加えないというのはもったいない。というのも、台風近傍では中心気圧が低いときに風速が強いからである。そこで、共分散の情報を取り込み、中心気圧を下げるだけでなく、風速も強くなるように修正する。これを式で書くと、

$$p_a = p_s + \frac{\sigma_{p,s}^2}{\sigma_{p,o}^2 + \sigma_{p,s}^2} (p_o - p_s)$$

$$v_a = v_s + \frac{r_{pv} \sigma_{p,s} \sigma_{v,s}}{\sigma_{p,o}^2 + \sigma_{p,s}^2} (p_o - p_s)$$

となる。下付き文字の p, v は中心気圧と風速の誤差の標準偏差を表し、 r_{pv} は中心気圧と風速の誤差の相関係数を表す。このように、中心気圧の観測値に共分散の情報を組み合わせて風速も修正することで、全体的に誤差をさらに小さくすることができる。

気象学においては、②のように、観測データを有効活用して異なる物理量の補正をかけるという点は重要である。なぜなら、興味のあるすべての時間・地点・高度で、すべての物理量が観測されることは普通ありえないからである。気象官署やアメダス地点、航空機、船舶、衛星、レーダーなどによって、さまざまな観測データが日々得られているが、すべての情報が完備されているわけではない。特に、顕著現象の周辺や海上・高層の状態を正確にかつ密に観測することは非常に難しいのでこの点が重要となる。

つまり、よりよいデータ同化システムは、物理量の誤差の共分散を正確に推定し、観測データから最大限情報を取り出すことのできるシステムと言い換えられる。しかし、変数同士の共分散の組合せは数値シミュレーションで扱う変数のおよそ2乗÷2個あるため、直接計算には膨大な計算機資源を要する。そのため、かつては、最適内挿法や3次元変分法(3D-Var)と呼ばれる簡便な手法が採用されていた。これらの手法では、過去の統計に基づいて、物理量間の誤差の共分散を特徴的な構造で固定化しておく。これにより、予報のたびに共分散を計算する必要がなくなり、計算時間を大幅に短縮することができる。現在でも、気象庁は

日本域の局地モデル予報のデータ同化には、限られた時間しか割くことができないため3D-Varを用いている。

近年では、計算機性能の向上により、数時間ごとに誤差の共分散の時間変化を計算するのが一般的になってきた。このようなシステムは、空間と時間の4次元を扱うことから、4次元データ同化システムと呼ばれている。とはいえ、超多自由度システムの誤差共分散のすべての要素を時間発展させるのは計算量から言って非現実的なので、現実機関は近似的な計算を行うアルゴリズムを採用している。その例としては、気象庁が多くのシステムに採用している4次元変分法(4D-Var)、アンサンブルカルマンフィルタ(EnKF)、両者を組み合わせたハイブリッドシステムなどが挙げられる。

4D-Varでは、まず、シミュレーション結果と観測データとの差異の総和などを用いて評価関数を定義する。そして、数値シミュレーションモデルをある固定期間内で走らせて、物理的な制約条件を与えながら、評価関数を最小化するような初期条件を探索する。このとき、評価関数と初期条件との関係を時間発展モデルで計算し、効果的な修正を実施する。この手法は直接的には誤差の共分散を計算しないが、数学的には、誤差の共分散の時間変化を計算したのとはほぼ等価な結果が得られることになる。EnKFは初期条件を少しずつ変えて複数の予報を開始し、そのバラつき具合と相関から、分散・共分散の時間変化を推定する手法である。例えば、少しだけ初期条件の異なる10回の計算を行い、東京と南海上のA地点の地上気温に正の共分散があることがわかったら、東京の地上気温観測データを使って、A地点の地上気温を補正できる。4D-VarとEnKFには一長一短があるため、両者の欠点を補いあうハイブリッドデータ同化システムの開発も進んでいる。筆者らのグループでもハイブリッド化によって顕著現象の予測精度がよくなることを示してきた。ハイブリッドデータ同化システムは気象庁でも次世代計算機の導入に合わせて採用される予定となっている。

近年では海洋の解析にも4次元データ同化手法を取り入れるのが一般的となってきたほか、東大地震研を中心としたグループは、昨年、GPSに基づいた観測を4D-Varで同化すると、すべり量の予測精度が改善することを世界で初めて示した。これは、余震などの確率予測にもつながる画期的な成果である。その他、航空工学・土木工学・材料工学などの分野でもデータ同化研究の成果が出始めている。データ同化は、観測データを数値シミュレーションの世界で扱う統合的なプラットフォームとみなすことができるため、気象学にとどまらず、今後も他分野へと広がりを見せていくことであろう。