

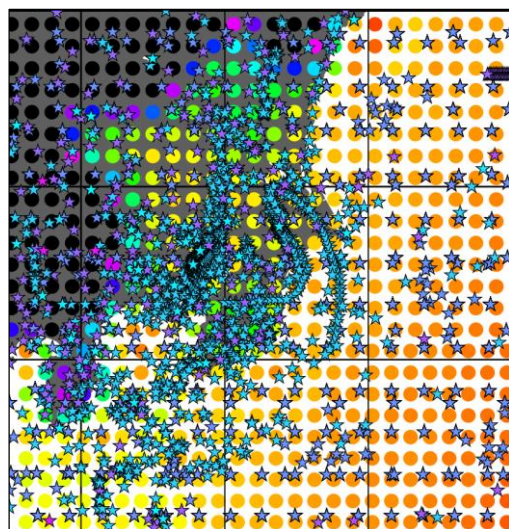


第1章 基礎編

1.3 データ同化

データ同化とは

第一推定値(前回の数値予報の値)と観測値及びそれらの誤差情報をもとに、第一推定値を観測値で修正して解析値(最も確からしい値)を作成する処理



数値予報モデルの格子点値(○)と
入電した観測データ(☆)

観測データと第一推定値(数値予報結果)の特徴

| | 観測値 | 第一推定値 |
|---|----------------|-------------------------|
| ◎ | 基本的には現実を反映している | 全格子・全要素のデータが利用可能 |
| △ | 観測誤差を内包 | 予報誤差を内包 |
| × | 全格子・全要素のデータがない | 観測値と比べると、現実に沿っているか定かでない |

- 第一推定値は全格子点・全要素の情報を持っている。
⇒「たたき台」として利用する。
- 観測値は現実を反映している。
⇒観測値を利用してたたき台を「修正」する。

数値予報モデルを実行するためには、初めに3次元空間の全ての格子点で初期時刻の気温や風、水蒸気量などの大気状態を表す物理量(初期値)を与えなければならない。初期値の精度は数値予報の精度に決定的な意味を持つため、精度の高い初期値を作成することは非常に重要となる(本田ほか 2018)。ここではデータ同化による初期値の作成について簡単に述べる。

表に観測データと第一推定値(前回の数値予報モデルの予報値)の特徴をまとめている。観測データは基本的に現実を反映しているが、利用可能な地点や高度、あるいはデータの種類が限られたものになっている。第一推定値は全ての格子点上で全てのデータを利用できるが、現実に沿ったものとは限らない。このように観測データと第一推定値のメリットとデメリットは互いに相補的になっていて、また観測データと第一推定値はそれぞれ観測誤差、予報誤差を内包している。図は数値予報モデルの格子点と実際に入電した観測データの分布を示している。数値予報モデルの格子点は3次元空間内に規則正しく並び、これら全ての格子点上で全ての要素のデータが利用できる。一方で観測データの分布はその多くが陸上に集中していて海上では少なく、まばらな分布になっていることが図から読み取れる。

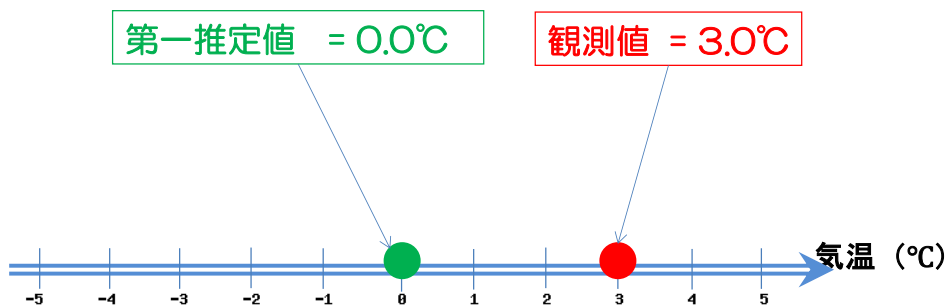
このようにそれぞれメリットとデメリットを持つ観測データと第一推定値をうまく使って、数値予報の初期値を作成する処理をデータ同化、または客観解析と呼ぶ。データ同化では、全格子点上で全要素のデータが利用できる第一推定値を初期値のたたき台として利用し、現実を反映している観測データによってたたき台(第一推定値)を修正することで、精度の良い初期値を作成する。データ同化によって作成された初期値は解析値とも呼ばれ、観測データと第一推定値の双方の利点を活用しつつ、それぞれに内包される観測誤差と予報誤差(背景誤差)も考慮して作成されることから、解析値は最も確からしい値として得られる。

データ同化手法には4次元変分法や3次元変分法、最適内挿法、アンサンブルカルマンフィルタなどがある。

データ同化の例

例) ある場所の気温が 3.0°C と観測された。一方、同じ場所・同じ時刻での第一推定値の気温の予想は、 0.0°C であった。

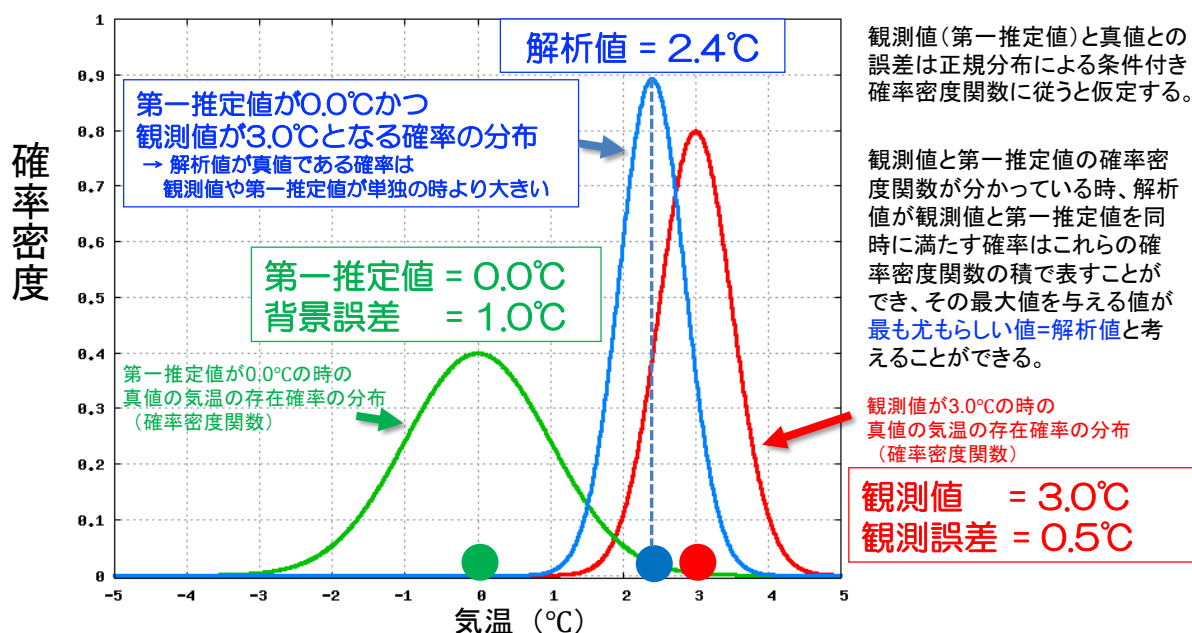
解析値をどのようにして求めるか？



データ同化では、観測値を単純に解析値とするのではなく、観測値と第一推定値の双方に重みをつけて内挿して解析値を求める。解析値が観測値側に寄るかそれとも第一推定値側に寄るかは、データ同化を実行する際にそれぞれの値の持つ誤差の統計的性質として与えられる「観測誤差」と「背景誤差」(第一推定値の誤差)の設定によって決まることになる。

簡単な例として、図のように、ある場所の気温が 3.0°C と観測され、同じ場所・同じ時刻での第一推定値の気温の予想は 0.0°C であった場合、解析値をどのように求めるかについて考える。

データ同化の例 (最尤推定法)



観測値(第一推定値)と真値との誤差は正規分布による条件付き確率密度関数に従うと仮定する。

観測値と第一推定値の確率密度関数が分かっている時、解析値が観測値と第一推定値を同時に満たす確率はこれらの確率密度関数の積で表すことができ、その最大値を与える値が最も尤もらしい値=解析値と考えることができる。

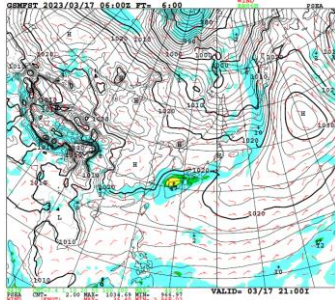
観測値が3.0°Cの時の真値の気温の存在確率の分布(確率密度関数)

観測値 = 3.0°C
観測誤差 = 0.5°C

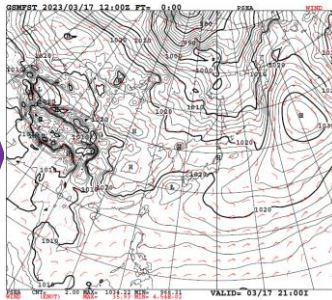
図は、最尤推定法という考え方で解析値を求める場合の例を示す。観測値、第一推定値には、それぞれ統計的な性質として与えられる「観測誤差」、「背景誤差」が存在する。観測値だけが分かっているときには、実際(真値)の気温は、観測値の周りに存在すると考えられ、第一推定値についても同様である。観測値、第一推定値と真値との誤差はそれぞれ正規分布による条件付き確率密度関数に従うと仮定する。解析値の確率の分布は観測誤差、背景誤差の大きさに依存し、観測値と第一推定値の確率密度関数が分かっている場合、解析値が観測値と第一推定値を同時に満たす確率はこれらの確率密度関数の積で表せる。その最大値を与える値が最も尤もらしい値=解析値と考えることができる。例では、青色の確率密度が最大となる場合の気温は2.4°Cで、この値が解析値となる。このように、観測誤差が背景誤差より小さい場合は解析値は観測データ側に寄り、逆に観測誤差が大きい場合は解析値は第一推定値側に寄ることになる。

データ同化による初期値の作成

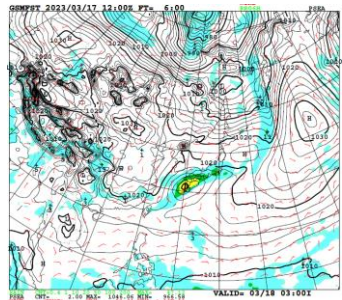
第一推定値



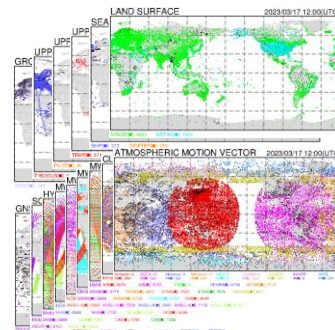
初期値



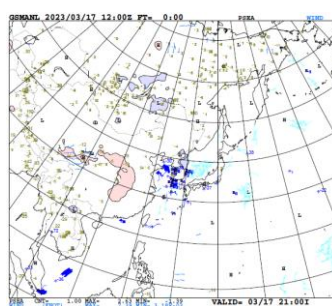
予報値



観測データ



修正量(インクリメント)



データ同化(客観解析)

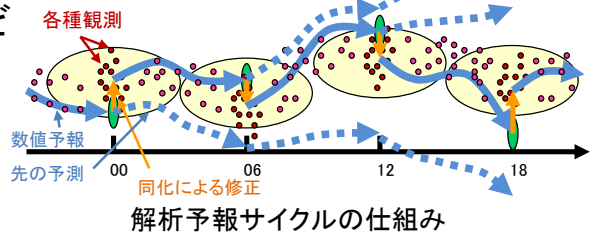
予測計算
(数値予報モデルの実行)

データ同化では、前回の数値予報モデルの予報値である第一推定値を、観測データで修正することによって初期値を作成する。作成された初期値からは数値予報モデルを実行することによって予報値が得られ、次のデータ同化の第一推定値として利用される。データ同化には地上観測や高層観測などによる直接観測のデータや、地上リモートセンシングや衛星観測といった遠隔観測のデータなど、様々な観測データが用いられる。精度の良い初期値を作成するためにはデータ同化に用いる観測データの品質が重要で、異常データが用いられた場合は初期値の精度は著しく悪化する。このためデータ同化には品質管理によって一定の品質を保った観測データのみが利用される。データ同化によって作成された初期値と、たたき台として利用された第一推定値との差は修正量(インクリメント)と呼ばれ、第一推定値が観測データによってどれだけ修正されたかを表している。異常に大きなインクリメントが入っている場合や、逆にインクリメントが全く入っていない場合などは、それぞれ異常データが利用されていることや観測データが利用されていないといったことが考えられ、作成された初期値の精度に注意する必要がある。

解析予報サイクル

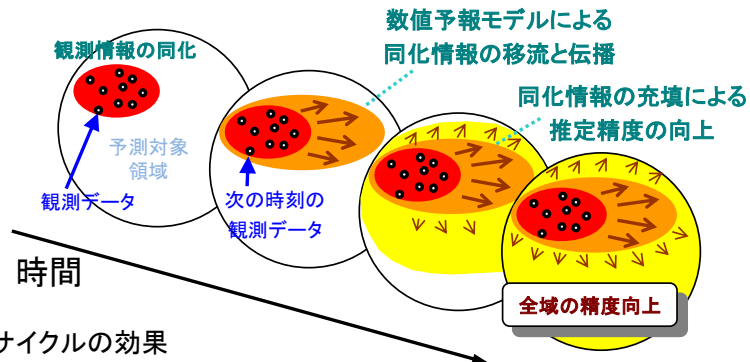
- 数値予報では予測時間が長いほど誤差が増大する。

- 短時間の予測とデータ同化を繰り返して予測精度を保つ。
- 一度で修正できない大きな誤差は次のデータ同化で修正。



- 解析と予測の繰り返しを「解析予報サイクル」と呼ぶ。

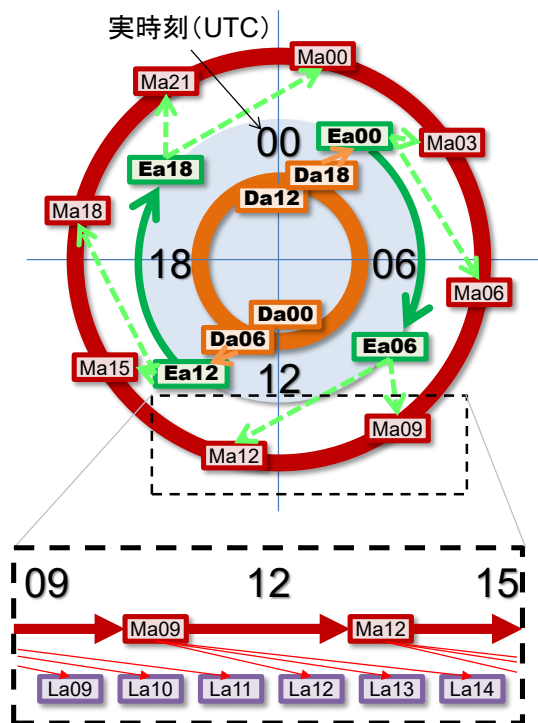
- サイクルを繰り返すことで、海上などの観測が少ない領域でも解析精度の向上が期待できる。



数値予報では、予測時間が長くなるほど誤差が増大するため、予報値は次第に真値（現実の気象状態）から離れていく。そこで短い予測時間の予報値を第一推定値に利用してデータ同化を行うことで真値に近いと期待される初期値を作成し、数値予報モデルを実行して次のデータ同化の第一推定値として利用することを繰り返し、初期値や予報値の精度の保持を図っている。このデータ同化と数値予報の繰り返しを解析予報サイクルと呼ぶ。例えば第一推定値が大きな誤差を持っていて一度のデータ同化で誤差を修正しきれない場合でも、解析予報サイクルでは次のサイクルのデータ同化によって修正され、初期値や予報値の精度が保たれる。特に短期予報や週間予報の範囲では、初期値の精度が数値予報の精度に大きな影響を与えるため、解析予報サイクルによって常に高い精度の初期値を作成することが必要となる。

1つの観測種別のデータは局所的に存在するため、一度のデータ同化では観測が存在する領域の周辺以外では第一推定値が修正されず、データ同化による初期値の改善の影響は限定的である。しかし、データ同化により観測情報が同化された初期値から数値予報モデルを実行し、その予報値を次のデータ同化の第一推定値として利用することで、観測データの影響は数値予報モデルを介して周辺に広がっていく。解析予報サイクルによって予測対象領域全域での精度向上が期待され、このことは初期値や予報値の精度向上において大きな意味を持つ。

解析予報サイクルスケジュール



全球サイクル解析 (Da)

- 一番内側で自己完結したサイクルで精度維持のための解析
- そのため観測データの待ち時間が最も長い

全球速報解析 (Ea)

- 全球予報のための解析
- Ea00,12はDa18,06解析値から第一推定値を作ることで、より高い精度の維持を図っている

メソ解析 (Ma)

- メソ予報のための解析
- 全球解析からはほぼ独立しているものの、側面境界値を全球予報からもらっている

局地解析 (La)

- 局地予報のための解析
- メソ解析・予測値を最初の第一推定値として使っており、サイクルとはなっていない

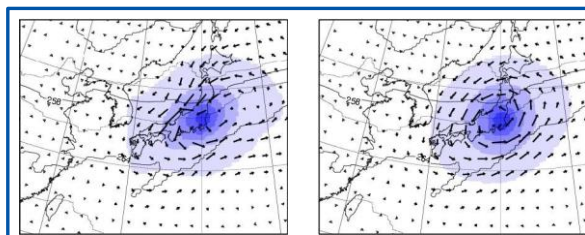
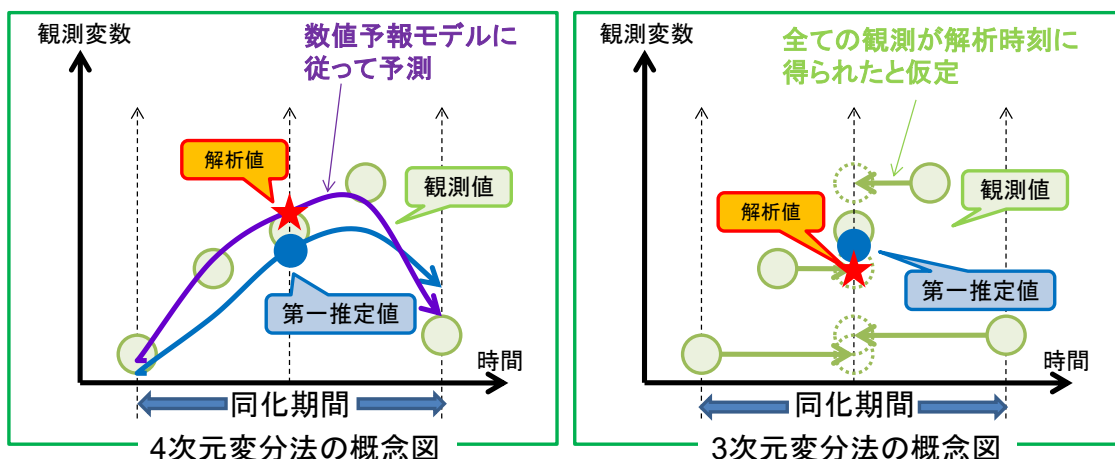
解析予報サイクルと数値予報システムの関係を示す。Daは全球サイクル解析 (Delayed analysis)、Eaは全球速報解析 (Early analysis)、Maはメソ解析 (Meso analysis)、Laは局地解析 (Local analysis) を表し、2桁の数字は解析対象時刻 (UTC) を表す。DaおよびMaの輪は、これらが解析予報サイクルであることを意味し、矢印は第一推定値や側面境界値の提供に関する関係を示している。

DaとEaはともに全球解析であり、それぞれ異なる目的で実行されている。Daは精度を維持するための解析として自己完結したサイクルで、多くの観測データを利用するためにデータ取得打ち切り時間が最も長く、00,12UTCでは11時間50分、06,18UTCでは7時間50分に設定されている。Eaは全球モデル (GSM) を実行するための解析で、数値予報プロダクトを迅速に提供するためにデータ取得打ち切り時間は2時間20分に設定されている。なおEa00,12では、精度を維持するためにそれぞれDa18,06の解析値から第一推定値を作成している。

Maはメソモデル (MSM) を実行するための解析で、データ取得打ち切り時間は50分に設定されている。DaやEaからほぼ独立したサイクルとなっているが、解析を実行するための側面境界値はGSMから提供されている。

Laは局地モデル (LFM) を実行するための解析で、データ取得打ち切り時間は30分に設定されている。解析を実行するための最初の第一推定値はMaやMSMから提供されており、サイクルとはなっていない。

変分法によるデータ同化



4次元変分法(左)と3次元変分法(右)による1点同化の例

ここではデータ同化手法の1つである変分法(4次元変分法と3次元変分法)について簡単に紹介する。4次元変分法は高精度な初期値が必要となる全球解析(門脇 2005; 西嶋・室井 2006)やメソ解析(本田・澤田 2008,2009)で、3次元変分法はより迅速な処理が求められる局地解析(永戸ほか 2012;幾田 2015)で用いられるデータ同化手法である。スライドに4次元変分法と3次元変分法の概念図を示している。変分法では解析時刻の前後に観測されたデータを利用したデータ同化を行う。この同化期間を同化窓(または同化ウィンドウ)と呼ぶ。

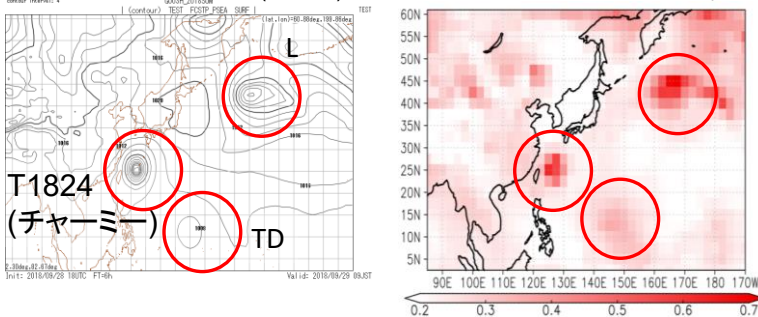
4次元変分法では数値予報モデルを実行することで、同化窓内の各時刻に観測されたデータを大気状態の時間変化を考慮して利用できる(露木 1997,2002)。一方、3次元変分法では数値予報モデルは実行せず、同化窓内の観測データは全て解析時刻に観測されたとして同化を行う。また4次元変分法では数値予報モデルを繰り返し実行して少しずつ第一推定値を修正することで、得られる解析値は力学的にバランスのとれたものになっているが、数値予報モデルを実行しない3次元変分法では必ずしも力学的にバランスのとれた解析値が得られるとは限らない。下の2枚の図は気温の一点観測を4次元変分法と3次元変分法を用いて同化した場合のインクリメント(気温:塗りつぶし、風:矢印)と気温の等値線を示している。4次元変分法では気温の等値線に沿って変形したインクリメントとなっているが、3次元変分法ではインクリメントが等方的になっていることが分かる。

このように4次元変分法は3次元変分法よりも高精度な解析値が得られるデータ同化手法となっている。データ同化手法として変分法他には、地上解析や積雪解析、波浪データ同化などに用いられている最適内挿法や、アンサンブル予報の初期摂動の作成に利用されるアンサンブルカルマンフィルタなどがある。

ハイブリッド同化

- 基本的な変分法データ同化では、背景誤差として気候学的背景誤差を使用
 - 第一推定値の誤差が流れ(気象場)に依存しない
⇒実際には場に応じて背景誤差(予報の当たり/外れやすさ)は違うはず
- ハイブリッド同化では、気候学的背景誤差に加えてアンサンブル予報から見積もられる予報誤差を利用することで解析精度を向上
 - アンサンブル由来の背景誤差として以下を利用
 - 全球解析: LETKF(局所アンサンブル変換カルマンフィルタ)による予報摂動
 - 局地解析: メソアンサンブル予報システムによる予報摂動

第一推定値の地上気圧(FT=6h) LETKFによる地上気圧のばらつき(FT=6h)



全球解析の例

・擾乱周辺ではばらつき大
(流れに依存する誤差)

・この情報を4次元変分法の中
で利用

データ同化では、観測値を単純に解析値とするのではなく、観測値と第一推定値の双方に重みをつけて内挿して解析値を求めることになる。解析値が観測値側に寄るかそれとも第一推定値側に寄るかは、データ同化を実行する際にそれぞれの値の持つ誤差の統計的性質として与えられる「観測誤差」と「背景誤差」(第一推定値の誤差)の設定によって決まることになる。観測誤差が背景誤差より小さい場合は解析値は観測データ側に寄り、逆に観測誤差が大きい場合は解析値は第一推定値側に寄ることになる。基本的な変分法データ同化では、背景誤差に過去の一定期間のデータから統計的に計算した予測誤差(「気候学的背景誤差」)を用いているため、日々の気象条件に応じた不確実性の違いを正確に表現することができない。そのため、全球解析や局地解析では、変分法で用いる背景誤差について、気候学的に事前に見積もった値に加えて図に示すようにアンサンブル予報から見積もられる予報摂動(ばらつき)を利用することで、実際の大気の状態に応じた誤差情報も考慮する手法(ハイブリッド同化)を用いている(数値予報開発センター 2021; 数値予報開発センター 2022)。これにより、観測データの情報をより適切に初期値に反映させることを可能としている。

参考文献

- 幾田泰醇, 2015: 局地解析の更新と改良. 平成27年度 数値予報研修テキスト, 気象庁予報部, 2-8.
- 永戸久喜, 藤田匡, 原旅人, 2012: 局地モデルの本運用. 平成24年度数値予報研修テキスト, 気象庁予報部, 72-86.
- 門脇隆志, 2005: 全球4次元変分法. 数値予報課報告・別冊第51号, 気象庁予報部, 100-105.
- 数値予報開発センター, 2021: ハイブリッド同化の全球解析への導入. 令和2年度数値予報開発センター年報, 気象庁数値予報開発センター, 18-22
- 数値予報開発センター, 2022: 局地解析へのハイブリッド同化手法の導入. 令和3年度数値予報開発センター年報, 気象庁数値予報開発センター, 102-105
- 露木義, 1997: 変分法によるデータ同化. 数値予報課報告・別冊第43号, 気象庁予報部, 102-165.
- 露木義, 2002: 変分法によるデータ同化の基礎. 数値予報課報告・別冊第48号, 気象庁予報部, 1-16.
- 西嶋信, 室井ちあし, 2006: データ同化システムの概要. 平成18年度数値予報研修テキスト, 気象庁予報部, 11-13.
- 本田有機, 澤田謙, 2008: 非静力学メソ4次元変分法. 平成20年度数値予報研修テキスト, 気象庁予報部, 48-52.
- 本田有機, 澤田謙, 2009: 非静力学メソ4次元変分法の現業化. 平成21年度数値予報研修テキスト, 気象庁予報部, 65-71.
- 本田有機, 藤田匡, 計盛正博, 室井ちあし, 佐藤芳昭, 2018: データ同化. 平成30年度数値予報研修テキスト, 気象庁予報部, 84-89.